Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2446859 A1 2002/11/21

(21) 2 446 859

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2002/05/13
(87) Date publication PCT/PCT Publication Date: 2002/11/21
(85) Entrée phase nationale/National Entry: 2003/11/07
(86) N° demande PCT/PCT Application No.: US 2002/014876
(87) N° publication PCT/PCT Publication No.: 2002/092764
(30) Priorités/Priorities: 2001/05/11 (60/290,071) US; 2001/05/17 (60/291,311) US; 2002/02/01 (60/353,058) US; 2002/03/04 (60/361,293) US

(51) Cl.Int.^{7}/Int.Cl.^{7} C12N 15/12, A61K 45/00, A61P 9/00, A61K 48/00, A61K 31/7088, A61K 38/22, A61K 38/18, A61K 38/17, A01K 67/027, C12Q 1/00, C12N 5/06, C12N 15/18, C12N 15/89, C12N 15/85

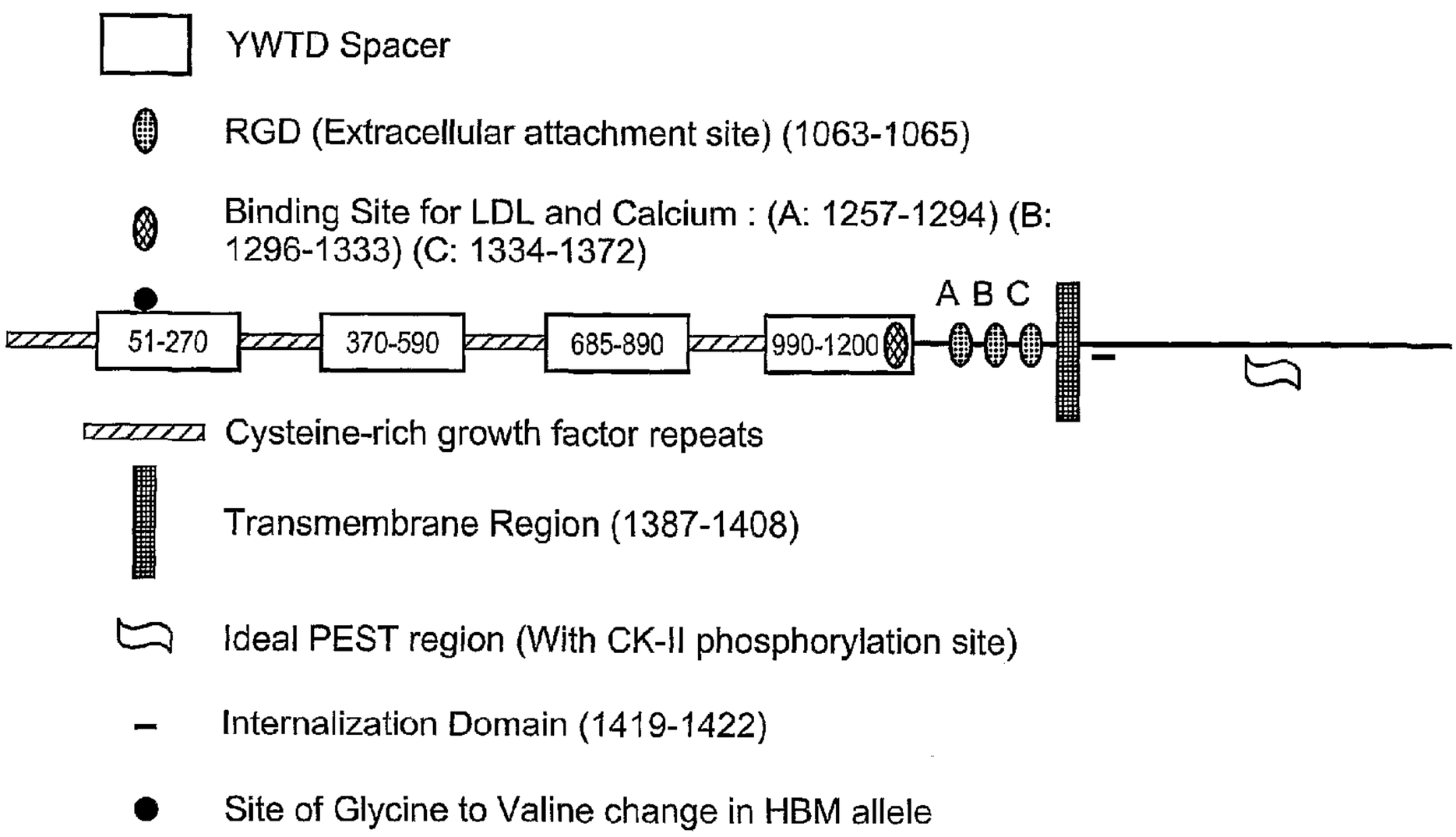
(71) Demandeur/Applicant:
WYETH, US

(72) Inventeurs/Inventors:
BABIJ, PHILIP, US;
BEX, FREDERICK JAMES III, US;
YAWORSKY, PAUL J., US;
BODINE, PETER VAN NEST, US

(74) Agent: BARRIGAR INTELLECTUAL PROPERTY LAW

(54) Titre : MODELE DE MODULATION DE LA MASSE OSSEUSE D'UN'ANIMAL TRANSGENIQUE
(54) Title: TRANSGENIC ANIMAL MODEL OF BONE MASS MODULATION

Model for a LDL Receptor-Related protein, Zmax1



(57) Abrégé/Abstract:
The present invention relates to methods and materials used to express the HBM protein in animal cells and transgenic animals. The present invention also relates to transgenic animals expressing the high bone mass gene, the corresponding wild-type gene, and mutants thereof. The invention provides nucleic acids, including coding sequences, oligonucleotide primers and probes, proteins, cloning vectors, expression vectors, transformed hosts, methods of developing pharmaceutical compositions, methods of identifying molecules involved in bone development, and methods of diagnosing and treating diseases involved in bone development. In preferred embodiments, the present invention is directed to methods for treating, diagnosing and preventing osteoporosis.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 November 2002 (21.11.2002)

PCT

(10) International Publication Number
WO 02/092764 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/14876

(22) International Filing Date: 13 May 2002 (13.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/290,071	11 May 2001 (11.05.2001)	US
60/291,311	17 May 2001 (17.05.2001)	US
60/353,058	1 February 2002 (01.02.2002)	US
60/361,293	4 March 2002 (04.03.2002)	US

(71) Applicants *(for all designated States except US)*: **GENOME THERAPEUTICS CORPORATION** [US/US]; 100 Beaver Street, Walthham, MA 02453-8443 (US). **WYETH** [US/US]; Five Giralda Farms, Madison, NJ 07928 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BABIJ, Philip** [GB/US]; 75 Adams Street, Dunstable, MA 01827 (US). **BEX, Frederick, James, III** [US/US]; 4 Lakeshore Drive, Newtown Square, PA 19073 (US). **YAWORSKY, Paul, J.** [US/US]; 13 Hobart Lane, Rockland, MA 02370 (US). **BODINE, Peter, Van Nest** [US/US]; 1742 Academy Lane, Havertown, PA 19083 (US).

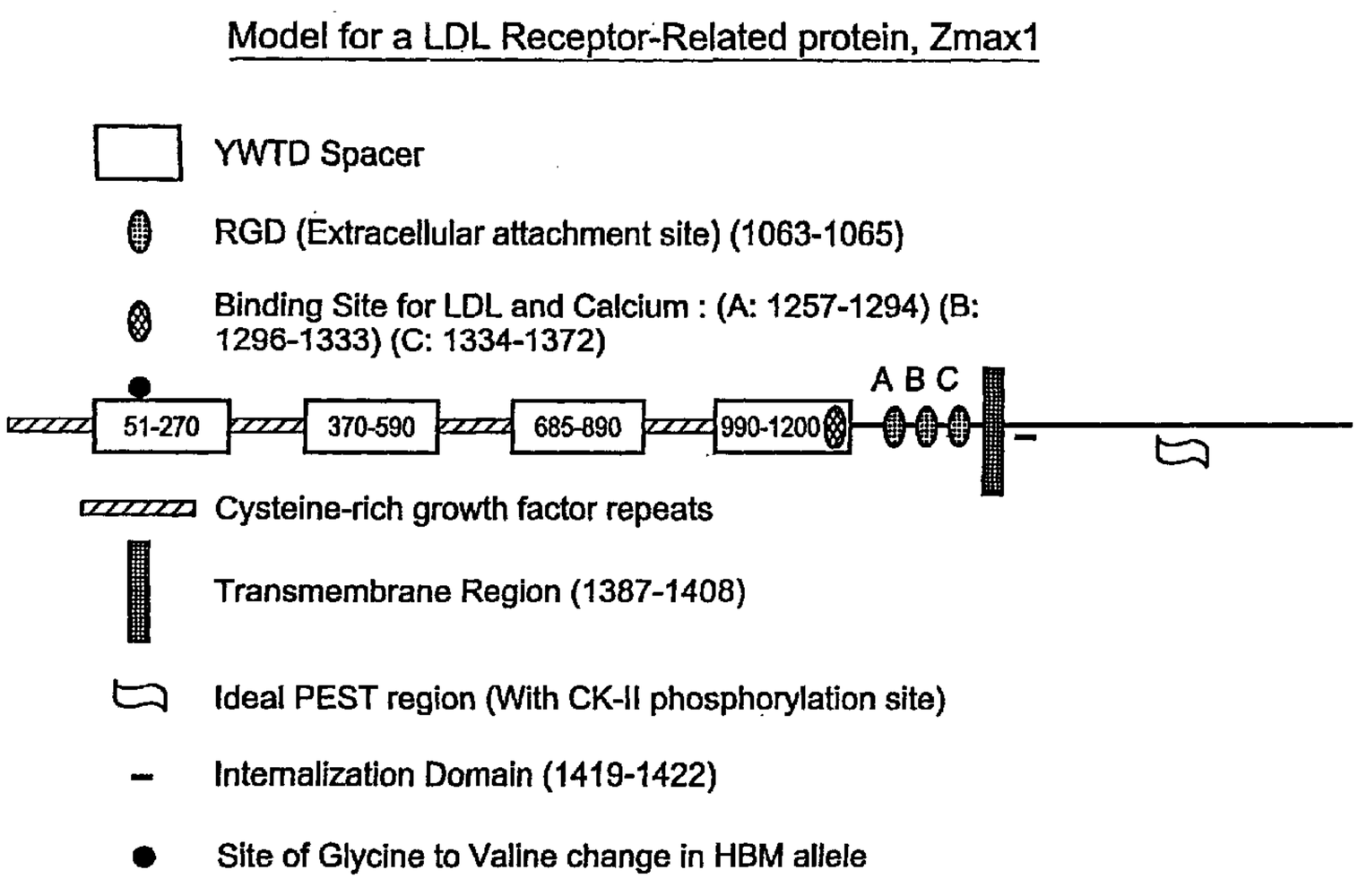
(74) Agents: **REA, Teresa, Stanek** et al.; Burns, Doane, Swecker & Mathis, L.L.P., P.O. Box 1404, Alexandria, VA 22313-1404 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,

[Continued on next page]

(54) Title: TRANSGENIC ANIMAL MODEL OF BONE MASS MODULATION



(57) Abstract: The present invention relates to methods and materials used to express the HBM protein in animal cells and transgenic animals. The present invention also relates to transgenic animals expressing the high bone mass gene, the corresponding wild-type gene, and mutants thereof. The invention provides nucleic acids, including coding sequences, oligonucleotide primers and probes, proteins, cloning vectors, expression vectors, transformed hosts, methods of developing pharmaceutical compositions, methods of identifying molecules involved in bone development, and methods of diagnosing and treating diseases involved in bone development. In preferred embodiments, the present invention is directed to methods for treating, diagnosing and preventing osteoporosis.

WO 02/092764 A2

WO 02/092764 A2

GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 260

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 260

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

TRANSGENIC ANIMAL MODEL OF BONE MASS MODULATION

FIELD OF THE INVENTION

The present invention relates generally to the field of genetics, genomics and molecular biology. The invention relates to methods and materials used to isolate, detect and sequence a high bone mass gene and corresponding wild-type gene, and mutants thereof. The present invention also relates to the high bone mass (HBM) gene, the corresponding wild-type gene, and mutants thereof. The genes identified in the present invention are implicated in the ontology and physiology of bone development. The invention also provides nucleic acids, proteins, cloning vectors, expression vectors, transformed hosts, methods of developing pharmaceutical compositions, methods of identifying molecules involved in bone development, and methods of diagnosing and treating diseases involved in bone development. The invention further relates to transgenic animals for studying the HBM phenotype, the mechanism of action of the *HBM* gene, and factors and treatments affecting normal and abnormal bone conditions. In preferred embodiments, the present invention is directed to methods for treating, diagnosing, preventing and screening for normal and abnormal conditions of bone, including metabolic bone diseases such as osteoporosis.

BACKGROUND OF THE INVENTION

Two of the most common types of osteoporosis are postmenopausal and senile osteoporosis. Osteoporosis affects men as well as women, and, taken with other abnormalities of bone, presents an ever-increasing health risk for an aging population. The most common type of osteoporosis is that associated with menopause. Most women lose between 20-60% of the bone mass in the trabecular compartment of the bone within 3-6 years after the cessation of menses. This rapid loss is generally associated with an increase of bone resorption and formation. However, the resorptive cycle is more dominant and the result is a net loss of bone

mass. Osteoporosis is a common and serious disease among postmenopausal women. There are an estimated 25 million women in the United States alone who are afflicted with this disease. The results of osteoporosis are both personally harmful, and also account for a large economic loss due to its chronicity and the need for extensive and long-term support (hospitalization and nursing home care) from the disease sequelae. This is especially true in more elderly patients. Additionally, while osteoporosis is generally not thought of as a life-threatening condition, a 20-30% mortality rate is related to hip fractures in elderly women. A large percentage of this mortality rate can be directly associated with postmenopausal osteoporosis.

The most vulnerable tissue in the bone to the effects of postmenopausal osteoporosis is the trabecular bone. This tissue is often referred to as spongy bone and is particularly concentrated near the ends of the bone near the joints and in the vertebrae of the spine. The trabecular tissue is characterized by small structures which inter-connect with each other as well as the more solid and dense cortical tissue which makes up the outer surface and central shaft of the bone. This criss-cross network of trabeculae gives lateral support to the outer cortical structure and is critical to the biomechanical strength of the overall structure. In postmenopausal osteoporosis, it is primarily the net resorption and loss of the trabeculae which lead to the failure and fracture of the bone. In light of the loss of the trabeculae in postmenopausal women, it is not surprising that the most common fractures are those associated with bones which are highly dependent on trabecular support, e.g., the vertebrae, the neck of the femur, and the forearm. Indeed, hip fracture, Colle's fractures, and vertebral crush fractures are indicative of postmenopausal osteoporosis.

One of the earliest generally accepted methods for treatment of postmenopausal osteoporosis was estrogen replacement therapy. Although this therapy frequently is successful, patient compliance is low, primarily due to the undesirable side-effects of chronic estrogen treatment. Frequently cited side-effects

of estrogen replacement therapy include reinitiation of menses, bloating, depression, and fear of breast or uterine cancer. In order to limit the known threat of uterine cancer in those women who have not undergone a hysterectomy, a protocol of estrogen and progestin cyclic therapy is often employed. This protocol is similar to that which is used in birth control regimens, and often is not tolerated by women because of the side-effects characteristic of progestin. More recently, certain antiestrogens, originally developed for the treatment of breast cancer, have been shown in experimental models of postmenopausal osteoporosis to be efficacious. Among these agents is raloxifene (See, U.S. Patent No. 5,393,763, and Black et al., *J. Clin. Invest.*, 93:63-69 (1994)). In addition, tamoxifene, a widely used clinical agent for the treatment of breast cancer, has been shown to increase bone mineral density in post menopausal women suffering from breast cancer (Love et al., *N. Engl. J. Med.*, 326:852-856 (1992)).

Another therapy for the treatment of postmenopausal osteoporosis is the use of calcitonin. Calcitonin is a naturally occurring peptide which inhibits bone resorption and has been approved for this use in many countries (Overgaard et al., *Br. Med. J.*, 305:556-561 (1992)). The use of calcitonin has been somewhat limited, however. Its effects are very modest in increasing bone mineral density and the treatment is very expensive. Another therapy for the treatment of postmenopausal osteoporosis is the use of bis-phosphonates. These compounds were originally developed for use in Paget's disease and malignant hypercalcemia. They have been shown to inhibit bone resorption. Alendronate, one compound of this class, has been approved for the treatment of postmenopausal osteoporosis. These agents may be helpful in the treatment of osteoporosis, but these agents also have potential liabilities which include osteomalacia, extremely long half-life in bone (greater than 2 years), and possible "frozen bone syndrome," e.g., the cessation of normal bone remodeling.

Senile osteoporosis is similar to postmenopausal osteoporosis in that it is marked by the loss of bone mineral density and resulting increase in fracture rate,

morbidity, and associated mortality. Generally, it occurs in later life, i.e., after 70 years of age. Historically, senile osteoporosis has been more common in females, but with the advent of a more elderly male population, this disease is becoming a major factor in the health of both sexes. It is not clear what, if any, role hormones such as testosterone or estrogen have in this disease, and its etiology remains obscure. Treatment of this disease has not been very satisfactory. Hormone therapy, estrogen in women and testosterone in men, has shown equivocal results; calcitonin and bis-phosphonates may be of some utility.

The peak mass of the skeleton at maturity is largely under genetic control. Twin studies have shown that the variance in bone mass between adult monozygotic twins is smaller than between dizygotic twins (Slemenda et al., *J. Bone Miner. Res.*, 6:561-567 (1991); Young et al., *J. Bone Miner. Res.*, 6:561-567 (1995); Pocock et al., *J. Clin. Invest.*, 80:706-710 (1987); Kelly et al., *J. Bone Miner. Res.*, 8:11-17 (1993)), and it has been estimated that up to 60% or more of the variance in skeletal mass is inherited (Krall et al., *J. Bone Miner. Res.*, 10:S367 (1993)). Peak skeletal mass is the most powerful determinant of bone mass in elderly years (Hui et al., *Ann. Int. Med.*, 111:355-361 (1989)), even though the rate of age-related bone loss in adult and later life is also a strong determinant (Hui et al., *Osteoporosis Int.*, 1:30-34 (1995)). Since bone mass is the principal measurable determinant of fracture risk, the inherited peak skeletal mass achieved at maturity is an important determinant of an individual's risk of fracture later in life. Thus, study of the genetic basis of bone mass is of considerable interest in the etiology of fractures due to osteoporosis.

Recently, a strong interest in the genetic control of peak bone mass has developed in the field of osteoporosis. The interest has focused mainly on candidate genes with suitable polymorphisms to test for association with variation in bone mass within the normal range, or has focused on examination of genes and gene loci associated with low bone mass in the range found in patients with osteoporosis. The vitamin D receptor locus (VDR) (Morrison et al., *Nature*,

367:284-287 (1994)), PTH gene (Howard et al., *J. Clin. Endocrinol. Metab.*, 80:2800-2805 (1995); Johnson et al., *J. Bone Miner. Res.*, 8:11-17 (1995); Gong et al., *J. Bone Miner. Res.*, 10:S462 (1995)) and the estrogen receptor gene (Hosoi et al., *J. Bone Miner. Res.*, 10:S170 (1995); Morrison et al., *Nature*, 367:284-287 (1994)) have figured most prominently in this work. These studies are difficult because bone mass (the phenotype) is a continuous, quantitative, polygenic trait, and is confounded by environmental factors such as nutrition, co-morbid disease, age, physical activity, and other factors. Also, this type of study design requires large numbers of subjects. In particular, the results of VDR studies to date have been confusing and contradictory (Garnero et al., *J. Bone Miner. Res.*, 10:1283-1288 (1995); Eisman et al., *J. Bone. Miner. Res.*, 10:1289-1293 (1995); Peacock, *J. Bone Miner. Res.*, 10:1294-1297 (1995)). Furthermore, the work thus far has not shed much light on the mechanism(s) whereby the genetic influences might exert their effect on bone mass.

While it is well known that peak bone mass is largely determined by genetic rather than environmental factors, studies to determine the gene loci (and ultimately the genes) linked to variation in bone mass are difficult and expensive. Study designs which utilize the power of linkage analysis, e.g., sib-pair or extended family, are generally more informative than simple association studies, although the latter do have value. However, genetic linkage studies involving bone mass are hampered by two major problems. The first problem is the phenotype, as discussed briefly above. Bone mass is a continuous, quantitative trait, and establishing a discrete phenotype is difficult. Each anatomical site for measurement may be influenced by several genes, many of which may be different from site to site. The second problem is the age component of the phenotype. By the time an individual can be identified as having low bone mass, there is a high probability that their parents or other members of prior generations will be deceased and therefore unavailable for study, and younger generations may not have even reached peak bone mass, making their phenotyping uncertain for genetic analysis.

Regardless, linkage analysis can be used to find the location of a gene causing a hereditary "disorder" and does not require any knowledge of the biochemical nature of the disorder, i.e., a mutated protein that is believed to cause the disorder does not need to be known. Traditional approaches depend on assumptions concerning the disease process that might implicate a known protein as a candidate to be evaluated. The genetic localization approach using linkage analysis can be used to first find the general chromosomal region in which the defective gene is located and then to gradually reduce the size of the region in order to determine the location of the specific mutated gene as precisely as possible. After the gene itself is discovered within the candidate region, the messenger RNA and the protein are identified and, along with the DNA, are checked for mutations.

The genetic localization approach has practical implications since the location of the disease can be used for prenatal diagnosis even before the altered gene that causes the disease is found. Linkage analysis can enable families, even many of those that do not have a sick child, to know whether they are carriers of a disease gene and to evaluate the condition of an unborn child through molecular diagnosis. The transmission of a disease within families, then, can be used to find the defective gene. As used herein, reference to "high bone mass" (HBM) is analogous to reference to a disease state, although from a practical standpoint high bone mass can actually help a subject avoid the disease known as osteoporosis.

Linkage analysis is possible because of the nature of inheritance of chromosomes from parents to offspring. During meiosis, the two parental homologues pair to guide their proper separation to daughter cells. While they are lined up and paired, the two homologues exchange pieces of the chromosomes, in an event called "crossing over" or "recombination." The resulting chromosomes are chimeric, that is, they contain parts that originate from both parental homologues. The closer together two sequences are on the chromosome, the less likely that a recombination event will occur between them, and the more closely linked they are. In a linkage analysis experiment, two positions on the

chromosomes are followed from one generation to the next to determine the frequency of recombination between them. In a study of an inherited disease, one of the chromosomal positions is marked by the disease gene or its normal counterpart, i.e., the inheritance of the chromosomal region can be determined by examining whether the individual displays symptoms of the disorder or not. The other position is marked by a DNA sequence that shows natural variation in the population such that the two homologues can be distinguished based on the copy of the "marker" sequence that they possess. In every family, the inheritance of the genetic marker sequence is compared to the inheritance of the disease state. If, within a family carrying an autosomal dominant disorder such as high bone mass, every affected individual carries the same form of the marker and all the unaffected individuals carry at least one different form of the marker, there is a great probability that the disease gene and the marker are located close to each other. In this way, chromosomes may be systematically checked with known markers and compared to the disease state. The data obtained from the different families is combined, and analyzed together by a computer using statistical methods. The result is information indicating the probability of linkage between the genetic marker and the disease allowing different distances between them. A positive result can mean that the disease is very close to the marker, while a negative result indicates that it is far away on that chromosome, or on an entirely different chromosome.

Linkage analysis is performed by typing all members of the affected family at a given marker locus and evaluating the co-inheritance of a particular disease state with the marker probe, thereby determining how often the two of them are co-inherited. The recombination frequency can be used as a measure of the genetic distance between two gene loci. A recombination frequency of 1% is equivalent to 1 map unit, or 1 centiMorgan (cM), which is roughly equivalent to 1,000 kb of DNA. This relationship holds up to frequencies of about 20% or 20 cM.

The entire human genome is 3,300 cM long. In order to find an unknown disease gene within 5-10 cM of a marker locus, the whole human genome can be

searched with roughly 330 informative marker loci spaced at approximately 10 cM intervals (Botstein et al., *Am. J. Hum. Genet.*, 32:314-331 (1980)). The reliability of linkage results is established by using a number of statistical methods. The method most commonly used for the analysis of linkage in humans is the LOD score method (Morton, *Prog. Clin. Biol. Res.*, 147:245-265 (1984), Morton et al., *Am. J. Hum. Genet.*, 38:868-883 (1986)) which was incorporated into the computer program, LIPED, by Ott, *Am. J. Hum. Genet.*, 28:528-529 (1976). LOD scores are the logarithm of the ratio of the likelihood that two loci are linked at a given distance to that they are not linked (>50 cM apart). The advantage of using logarithmic values is that they can be summed among families with the same disease. This becomes necessary given the relatively small size of human families.

By convention, a total LOD score greater than + 3.0 (that is, odds of linkage at the specified recombination frequency being 1000 times greater than odds of no linkage) is considered to be significant evidence for linkage at that particular recombination frequency. A total LOD score of less than - 2.0 (that is, odds of no linkage being 100 times greater than odds of linkage at the specified frequency) is considered to be strong evidence that the two loci under consideration are not linked at that particular recombination frequency. Until recently, most linkage analyses have been performed on the basis of two-point data, which is the relationship between the disorder under consideration and a particular genetic marker. However, as a result of the rapid advances in mapping the human genome over the last few years, and concomitant improvements in computer methodology, it has become feasible to carry out linkage analyses using multi-point data. Multi-point analysis provide a simultaneous analysis of linkage between the disease and several linked genetic markers, when the recombination distance among the markers is known.

Multi-point analysis is advantageous for two reasons. First, the informativeness of the pedigree is usually increased. Each pedigree has a certain amount of potential information, dependent on the number of parents heterozygous

for the marker loci and the number of affected individuals in the family. However, few markers are sufficiently polymorphic as to be informative in all those individuals. If multiple markers are considered simultaneously, then the probability of an individual being heterozygous for at least one of the markers is greatly increased. Second, an indication of the position of the disease gene among the markers may be determined. This allows identification of flanking markers, and thus eventually allows isolation of a small region in which the disease gene resides. Lathrop et al., *Proc. Natl. Acad. Sci. USA*, 81:3443-3446 (1984) have written the most widely used computer package, LINKAGE, for multi-point analysis.

There is a need in the art for identifying the gene associated with a high bone mass phenotype. There is also a need for tools for the study of the high bone mass gene and phenotype. More generally there is need for the development of diagnostic tools and treatments. The present invention is directed to these, as well as other, important ends.

SUMMARY OF THE INVENTION

The present invention describes the identification of the *LRP5* gene and the *HBM* allele of the *LRP5* gene on chromosome 11q13.3 by genetic linkage and mutation analysis. The *LRP5* gene and the LRP5 protein which it encodes have previously been referred to as *Zmax1* and Zmax1 (also Zmax) by the inventors and coworkers. The gene and its product have also been referred to by others using the designation LR3. It is understood that Zmax, Zmax1, LRP5, and LR3 are synonymous terms. The use of genetic markers linked to the genes has aided their discovery. By using linkage analysis and mutation analysis, persons predisposed to HBM may be readily identified. Cloning methods using Bacterial Artificial Chromosomes have enabled the inventors to focus on the chromosome region of 11q13.3 and to accelerate the sequencing of the autosomal dominant gene. In addition, the invention identifies the *LRP5* gene and the *HBM* gene, and identifies

the guanine-to-thymine polymorphism mutation at position 582 in the *LRP5* gene that produces the *HBM* gene and the HBM phenotype.

The present invention identifies the *LRP5* gene and the *HBM* gene, which can be used to determine if people are predisposed to HBM and, therefore, not susceptible to diseases characterized by reduced bone density, including, for example, osteoporosis, or are predisposed and susceptible to diseases characterized by abnormally high bone density, such as, for example, osteopetrosis. Older individuals carrying the *HBM* gene express the HBM protein, and, therefore, do not develop osteoporosis. In other words, the *HBM* gene is a suppressor of osteoporosis. This *in vivo* observation is a strong evidence that treatment of normal individuals with the *HBM* gene or protein, or fragments thereof, will ameliorate osteoporosis.

The present invention provides expression vectors for LRP5 and HBM which are useful for the study of bone density modulation in animal models. The expression vectors comprise promoters which drive the expression of LRP5 and HBM ubiquitously in animal tissues and specifically in bone tissues. The expression vectors also serve to provide linear nucleic acid sequences for the creation of transgenic and other genetically modified animals.

One embodiment provides vectors for gene targeting in mice and other animals for the purpose of creating knock-out mice which do not express LRP5 and knock-in mice which express the homologous mouse HBM protein. A conditional knock-out/knock-in vector is provided which allows *in-vitro* deletion of a knock-out cassette in pre-fusion zygotes. The present invention provides animal embryonic stem cells which comprise recombinant DNA of the gene targeting vectors.

In another embodiment, animal cells expressing LRP5 and HBM are provided for use in investigating modulators of bone density.

Yet another embodiment provides transgenic animals expressing the *LRP5* gene and the *HBM* gene or other related variants under the control of general promoters and bone specific promoters. Transgenic animals are also provided

wherein either the endogenous LRP5 gene or a heterologous LRP5 or HBM gene is under the control of inducible or conditional promoters such as for example the GENESWITCH® System. The present invention provide methods using these animals for the study of the HBM phenotype and its molecular mechanism, for the development of diagnostic and screening tools, and for the testing and development of treatments and therapeutic compounds.

Moreover, such treatment will be indicated in the treatment of bone lesions, particularly bone fractures, for bone remodeling in the healing of such lesions. For example, persons predisposed to or suffering from stress fractures (i.e., the accumulation of stress-induced microfractures, eventually resulting in a true fracture through the bone cortex) may be identified and/or treated by means of the invention. Moreover, the methods and compositions of the invention will be of use in the treatment of secondary osteoporosis, where the course of therapy involves bone remodeling, such as endocrine conditions accompanying corticosteroid administration, hyperthyroidism, hypogonadism, hematologic malignancies, malabsorption and alcoholism, as well as disorders associated with vitamin D and/or phosphate metabolism, such as osteomalacia and rickets, and diseases characterized by abnormal or disordered bone remodeling, such as Paget's disease, and in neoplasms of bone, which may be benign or malignant.

In various embodiments, the present invention is directed to nucleic acids, proteins, vectors, transformed hosts expressing HBM and LRP5, and transgenic animals carrying the human *HBM* and *LRP5* genes and related variants, knock-in animals for HBM homologues or knock-out animals for these genes.

Additionally, the present invention is directed to applications of the above embodiments of the invention including, for example, gene therapy, pharmaceutical development, and diagnostic assays for bone development disorders. In preferred embodiments, the present invention is directed to methods for treating, diagnosing, preventing and screening for osteoporosis.

Another aspect of the invention is to provide transgenic animals having somatic and/or germ cells comprising a nucleic acid which comprises a promoter region that directs protein expression in animal and/or human cells operably linked to a sequence comprising at least 15 contiguous nucleotides of SEQ ID NO: 2 including at least the thymine at position 582 of SEQ ID NO: 2.

Other embodiments contemplated includes a transgenic animal having somatic and/or germ cells comprising a nucleic acid which comprises a sequence which encodes SEQ ID NO: 4 and which includes at least a codon for the valine corresponding to the valine at position 171 of SEQ ID NO: 4, and wherein the nucleic acid further comprises an operably linked promoter region that directs protein expression in animal and/or human cells; a transgenic animal for the study of bone density modulation having somatic and/or germ cells comprising a nucleic acid which comprises a promoter region that directs protein expression in animal and/or human cells operably linked to a sequence comprising at least 15 contiguous nucleotides of SEQ ID NO: 1. Also contemplated are the progeny of such animals. The animals are preferably mice, but can include any non-human animal (e.g., primates, canines, felines, rodents, ovines, bovines, and the like).

Such animals are useful for the study of bone density or bone mass modulation and the development of methods and treatments for affecting bone density or bone mass modulation. Modulation of bone density and/or bone mass can be assessed by changes in one or more parameters such as bone mineral density, bone strength, trabecular number, bone size, and bone tissue connectivity.

Another object of the invention is to provide an animal embryo comprising a nucleic acid which comprises a promoter region that directs protein expression in animal and/or human cells operably linked to a sequence comprising at least 15 contiguous nucleotides of SEQ ID NO: 2 including at least the thymine at position 582 of SEQ ID NO: 2.

It is a further object of the invention to provide a nucleic acid for gene targeting by homologous recombination comprising a first section homologous to

mouse *LRP5* gene and a second section homologous to another section of mouse *LRP5* gene, and between the first and the second section a middle section comprising an engineered deletion of a portion of the *LRP5* gene, a nucleic acid sequence change, or a nucleic acid insertion, and wherein the nucleic acid is capable of homologous recombination with the endogenous gene.

Another object of the invention is to provide a method of producing a transgenic animal, and preferably a transgenic mouse, whose genome comprises an alteration of the gene encoding LRP5. This method can comprise:

(a) providing the a nucleic acid of which encodes LRP5 or HBM;

(b) introducing the nucleic acid into mouse embryonic stem cells;

(c) selecting those embryonic stem cells that comprise the nucleic acid;

(d) introducing an embryonic stem cells of step (c) into a mouse blastocyst;

(e) transferring the blastocyst of step (d) to a pseudopregnant mouse; and

(f) allowing the transferred blastocyst to develop into a mouse chimeric for the nucleic acid.

In another aspect of the invention, the animals obtained as described can then be further bred by for example, breeding the chimeric mouse to a wild-type mouse to obtain mice heterozygous for the alteration; and breeding the heterozygous mice to generate mice homozygous for the alteration.

Another aspect of the invention is to provide a method for identifying agents which modulate HBM expression comprising the steps of: (a) providing cells according to claim 13; (b) exposing the cells to a test compound; and (c) measuring the expression of HBM.

It is another object of the invention to study bone mass modulators by (a) providing a first group of transgenic animals as described above; (b) administering a test compound; and (c) measuring at least one parameter of development in the animals administered a test compound. Test compounds can include but are not limited to a hormone, a growth factor, a peptide, RNA, DNA, a mineral, a vitamin, a natural product, or a synthetic organic compound.

In another aspect, bone mass modulation and bone development can be studied by a method utilizing a group of transgenic animals, as described above, administering an experimental procedure to the animals, and measuring a parameter of development. Experimental procedures include, for example, ovariectomy, restricted bone loading, and increased bone loading.

Another aspect of the invention provides a reagent set for quantifying human LRP5 mRNA or HBM mRNA comprising the isolated nucleic acid sequences (SEQ ID NOS: 689-697):

(1) 5'-GTCAGCCTGGAGGAGTTCTCA-3';

5'-TCACCCTTGGCAATACAGATGT-3'; and,

6-FAM-5'-CCCACCCATGTGCCCGTGACA-3'; or

(2) 5'-CGTGATTGCCGACGATCTC-3';

5'-TTCCGGCCGCTAGTCTTGT-3'; and,

6-FAM-5'-CGCACCCGTTCGGTCTGACGCAGTAC-3'.

Another reagent set includes

5'-CTTTCCCCACGAGTATGTTGGT-3'; and,

5'-AAGGGACCGTGCTGTGAGC-3'; and,

6-FAM-5'-AGCCCCTCATGTGCCTCTCAACTTCATAG-3'.

Another aspect of the invention provides for variants of SEQ ID NO:3 which contain one or more of the following amino acid substitutions: G171V, A214V, E128V, A65V, G199V, M282V, G479V, G781V, Q1087V, G171K, G171F, G171I, G171Q.

It is another object of the invention to provide for corresponding variants in LRP6. Preferred variants in LRP6 include G158V.

It is yet a further embodiment of the invention to provide a method for studying the effect of HBM on other bone disorders comprising the steps of: (a) providing embryos of animals with a bone disorder phenotype; (b) introducing the nucleic acid of encoding LRP5, HBM, a variant thereof into a first group of the embryos; (c) transferring the embryos to pseudopregnant mice; and, (d) measuring

at least one parameter of development in the resultant mice. The nucleic acid can originate from any animal and is not limited to the human LRP5 or human HBM.

Another aspect of the invention provides a method for studying cardiac disorders related to LRP5 or HBM comprising the steps of: (a) providing a first group of transgenic animals as described above and (b) measuring at least one parameter of cardiac health in the animals administered a test compound. In a further method, these animals can be used in a screen of putative cardiac drugs for efficacy.

It is yet another embodiment of the invention to provide methods of evaluating cardio-protective treatments for bone mass modulation effects comprising providing a first group of animals according to claim 9; administering a cardio-protective treatment to a subgroup of the first group the first group of animals; and measuring at least one parameter of bone modulation in at least the treated mice.

Another aspect of the invention is to provide a method for studying modulators of bone mass comprising the steps of: (a) providing a first group of animals having somatic and/or germ cells comprising a nucleic acid which comprises a promoter region that directs protein expression in animal and/or human cells operably linked to a sequence comprising at least 15 contiguous nucleotides of SEQ ID NO: 1; (b) administering a test compound; and (c) measuring at least one parameter of bone development in the animals administered a test compound.

Another aspect is to provide a method for studying the effect of an experimental procedure on bone mass comprising the steps of: (a) providing a first group of animals having somatic and/or germ cells comprising a nucleic acid which comprises a promoter region that directs protein expression in animal and/or human cells operably linked to a sequence comprising at least 15 contiguous nucleotides of SEQ ID NO: 1; (b) administering an experimental procedure; and (c) measuring at least one parameter of bone mass to assess bone modulation in the animals administered an experimental procedure.

These and other aspects of the present invention are described in more detail below.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 shows the pedigree of the individuals used in the genetic linkage studies. Under each individual is an ID number, the z-score for spinal BMD, and the allele calls for the critical markers on chromosome 11. Solid symbols represent "affected" individuals. Symbols containing "N" are "unaffected" individuals. DNA from 37 individuals was genotyped. Question marks denote unknown genotypes or individuals who were not genotyped.

Fig. 2 depicts the BAC/STS content physical map of the HBM region in 11q13.3. STS markers derived from genes, ESTs, microsatellites, random sequences, and BAC endsequences are denoted above the long horizontal line. For markers that are present in GDB the same nomenclature has been used. Locus names (D11S####) are listed in parentheses after the primary name if available. STSs derived from BAC endsequences are listed with the BAC name first followed by L or R for the left and right end of the clone, respectively. The two large arrows indicate the genetic markers that define the HBM critical region. The horizontal lines below the STSs indicate BAC clones identified by PCR-based screening of a nine-fold coverage BAC library. Open circles indicate that the marker did not amplify the corresponding BAC library address during library screening. Clone names use the following convention: B for BAC, the plate, row and column address, followed by -H indicating the HBM project (i.e., B36F16-H).

Figs. 3A-3F show the genomic structure of LRP5 with flanking intron sequences. Translation is initiated by the underlined "ATG" in exon 1. The site of the polymorphism in the *HBM* gene is in exon 3 and is represented by the underlined "G," whereby this nucleotide is a "T" in the *HBM* gene. The 3' untranslated region of the mRNA is underlined within exon 23 (exon 1, SEQ ID NO:40; exon 2, SEQ ID NO:41; exon 3, SEQ ID NO:42; exon 4, SEQ ID NO:43;

exon 5, SEQ ID NO:44; exon 6, SEQ ID NO:45; exon 7, SEQ ID NO:46; exon 8, SEQ ID NO:47; exon 9, SEQ ID NO:48; exon 10, SEQ ID NO:49; exon 11, SEQ ID NO:50; exon 12, SEQ ID NO:51; exon 13, SEQ ID NO:52; exon 14, SEQ ID NO:53; exon 15, SEQ ID NO:54; exon 16, SEQ ID NO:55; exon 17, SEQ ID NO:56; exon 18, SEQ ID NO:57; exon 19, SEQ ID NO:58; exon 20, SEQ ID NO:59; exon 21, SEQ ID NO:60; exon 22, SEQ ID NO:61; and exon 23; SEQ ID NO:62).

Fig. 4 shows the domain organization of LRP5, including the YWTD spacers, the extracellular attachment site, the binding site for LDL and calcium, the cysteine-rich growth factor repeats, the transmembrane region, the ideal PEST region with the CK-II phosphorylation site and the internalization domain. Fig. 4 also shows the site of the glycine to valine change that occurs in the HBM protein. The signal peptide is located at amino acids 1-31, the extracellular domain is located at amino acids 32-1385, the transmembrane segment is located at amino acids 1386-1413, and the cytoplasmic domain is located at amino acids 1414-1615.

Fig. 5 is a schematic illustration of the BAC contigs B527D12 and B200E21 in relation to the *HBM* gene.

Figs. 6A-6J show the nucleotide (SEQ ID NO: 1) and amino acid (SEQ ID NO: 3) sequences of the wild-type gene, LRP5. The location for the base pair substitution at nucleotide 582, a guanine to thymine, (SEQ ID NOS: 2, 4) is underlined. This allelic variant is the *HBM* gene. The *HBM* gene encodes for a protein with an amino acid substitution of glycine to valine at position 171. The 5' untranslated region (UTR) boundaries bases 1 to 70, and the 3' UTR boundaries bases 4916-5120.

Figs. 7A and 7B are northern blot analysis showing the expression of LRP5 in various tissues.

Fig. 8 is a PCR product analysis.

Fig. 9 is allele specific oligonucleotide detection of the LRP5 exon 3 mutation.

Fig. 10 is the cellular localization of mouse LRP5 by *in situ* hybridization at 100X magnification using sense and antisense probes.

Fig. 11 is the cellular localization of mouse LRP5 by *in situ* hybridization at 400X magnification using sense and antisense probes.

Fig. 12 is the cellular localization of mouse LRP5 by *in situ* hybridization of osteoblasts in the endosteum at 400X magnification using sense and antisense probes.

Fig. 13 shows antisense inhibition of LRP5 expression in MC-3T3 cells.

Fig. 14 shows a LRP5 Exon3 Allele Specific Oligonucleotide (ASO) assay which illustrates the rarity of the HBM allele (right panels; T-specific oligo; 58°C Wash) as compared to the wild-type LRP5 allele (left panels, G-specific oligo; 55°C Wash). The positive spots appearing in the right panels were positive controls.

Fig. 15 depicts a model representing the potential role of LRP5 (Zmax1) in focal adhesion signaling.

Fig. 16 depicts a schematic of two LRP5 gene targeting vectors for the knock-out of endogenous mouse LRP5 or conditional knock-in of the HBM polymorphism. B, X, and R indicate *BamH*I, *Xba*I, and *Eco*RI sites in DNA BAC 4735P5 respectively. Exons 3, 4, and 5 are indicated by black rectangles. A G -> T base change is engineered at base 24 of exon 3 to produce the HBM polymorphism. The location of a LoxP flanked cassette containing a neomycin resistance gene and a synthetic pause sequence and probes used for screening and characterizing of ES cell clones are also indicated.

Fig. 17 confirms expression by the transgenic (*i.e.*, HBMMCBA and HBMMTIC) and wild-type (*i.e.*, ZmaxWTCBA and ZmaxWTTIC) plasmid constructs. These constructs were transiently transfected into HOB-02-02 cells and the mRNA levels determined using TaqMan® quantitative PCR. HBMMCBA and ZmaxWTCBA are shown in the left column (*i.e.* CMVβActin) and HBMMTIC and ZmaxWTTIC are shown in the right column (*i.e.* Type I collagen) of the Table.

Fig. 18 depicts a comparison between the human and mouse TaqMan® Primer/Probe sets. HOB (HOB-03-C5) and mouse (MC-3T3-E1) osteoblastic cell mRNA was analyzed using the probes and primers.

Fig. 19 depicts the quantification of human Zmax-1 mRNA expressed in a mixed human and mouse RNA background using the TaqMan® Primer/Probe sets. Results are presented in Human LRP5 mRNA added (ng) versus Human LRP5 mRNA measured (ng).

Fig. 20 depicts expression of HBM in transgenic mice based on mRNA expression analyzed by TaqMan®.

Fig. 21A-C depicts the analysis of various transgenic mouse lines that express the HBMMCBA construct in spine (Fig. 21A), femur (Fig. 21B) and total body (Fig. 21C).

Fig. 21D-F depicts the analysis of various transgenic mouse lines that express the HBMMTIC construct in spine (Fig. 21D), femur (Fig. 21E) and total body (Fig. 21F).

Fig. 21G-L depict the analysis of transgenic mouse lines that express the HBMMTIC construct (Lines 19 and 35) in spine, femur and total body through 17 weeks.

Fig. 22 depicts changes in BMD in HBM transgenic mice (*i.e.*, HBMMCBA and HBMMTIC constructs) at 5 weeks using *in vivo* pDXA* analysis. The BMD changes are presented as compared to wild-type animals which were also only 5 weeks old.

Fig. 23 depicts changes in BMD in HBM transgenic mice (*i.e.*, HBMMCBA and HBMMTIC constructs) at 9 weeks using *in vivo* pDXA* analysis. The BMD changes are presented as compared to wild-type animals which were also only 9 weeks old.

Fig. 24 (A-D) presents the sequence of the HBMGI_2AS vector insert (SEQ ID NO: 759).

Fig. 25 (A-D) presents the sequence of the ZMAXGI_3AS vector insert (SEQ ID NO: 760).

Fig. 26 (A-C) presents an alignment of human (SEQ ID NO: 761) and mouse (SEQ ID NO: 762) LRP5 amino acid sequences.

Fig. 27 (A-C) presents an alignment of human LRP5 (SEQ ID NO: 763) and LRP6 (SEQ ID NO: 764) amino acid sequences.

Fig. 28 illustrates an apparatus for testing the effects of loading on bone growth in a mouse.

Fig. 29 presents a histomorphological illustration of the effects of bone loading on bone growth in HBM transgenic and non-transgenic mice.

DETAILED DESCRIPTION OF THE INVENTION

To aid in the understanding of the specification and claims, the following definitions are provided.

"Gene" refers to a DNA sequence that encodes through its template or messenger RNA a sequence of amino acids characteristic of a specific peptide. The term "gene" includes intervening, non-coding regions, as well as regulatory regions, and can include 5' and 3' ends.

By "nucleic acid" is meant to include single stranded and double stranded nucleic acids, DNAs, RNAs (e.g., mRNA, tRNAs), cDNAs, recombinant DNA (rDNA), rRNAs, antisense nucleic acids, oligonucleotides, and oligomers, and polynucleotides. May also include hybrids such as triple stranded regions of RNA and/or DNA or double stranded RNA:DNA hybrids. And may include modified bases such as biotinylated, tritylated, fluorophor, inosine, and etc.

"Gene sequence" refers to a nucleic acid molecule, including DNA which contains a non-transcribed or non-translated sequence, which comprises a gene. The term is also intended to include any combination of gene(s), gene fragment(s),

non-transcribed sequence(s) or non-translated sequence(s) which are present on the same DNA molecule.

The nucleic acid sequences of the present invention may be derived from a variety of sources including DNA, cDNA, synthetic DNA, synthetic RNA or combinations thereof. Such sequences may comprise genomic DNA which may or may not include naturally occurring introns. Moreover, such genomic DNA may be obtained in association with promoter regions and/or poly (A) sequences. The sequences, genomic DNA or cDNA may be obtained in any of several ways. Genomic DNA can be extracted and purified from suitable cells by means well known in the art. Alternatively, mRNA can be isolated from a cell and used to produce cDNA by reverse transcription or other means.

"cDNA" refers to complementary or copy DNA produced from an RNA template by the action of RNA-dependent DNA polymerase (reverse transcriptase). Thus, a "cDNA clone" means a duplex DNA sequence for which one strand is complementary to an RNA molecule of interest, carried in a cloning vector or PCR amplified. cDNA can also be single stranded after first strand synthesis by reverse transcriptase. In this form it is a useful PCR template and does not need to be carried in a cloning vector. This term includes genes from which the intervening sequences have been removed. Thus, the term "gene", as sometimes used generically, can also include nucleic acid molecules comprising cDNA and cDNA clones.

"Recombinant DNA" means a molecule that has been engineered by splicing *in vitro* a cDNA or genomic DNA sequence or altering a sequence by methods such as PCR mutagenesis.

"Cloning" refers to the use of *in vitro* recombination techniques to insert a particular gene or other DNA sequence into a vector molecule. In order to successfully clone a desired gene, it is necessary to use methods for generating DNA fragments, for joining the fragments to vector molecules, for introducing the

composite DNA molecule into a host cell in which it can replicate, and for selecting the clone having the target gene from amongst the recipient host cells.

"cDNA library" refers to a collection of recombinant DNA molecules containing cDNA inserts which together comprise the entire or a partial repertoire of genes expressed in a particular tissue or cell source. Such a cDNA library can be prepared by methods known to one skilled in the art and described by, for example, Cowell and Austin, "cDNA Library Protocols," Methods in Molecular Biology (1997).

"Cloning vehicle" refers to a plasmid or phage DNA or other DNA sequence which is able to replicate in a host cell. This term can also include artificial chromosomes such as BACs and YACs. The cloning vehicle is characterized by one or more endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the DNA, which may contain a marker suitable for use in the identification of transformed cells.

"Expression" refers to the process comprising transcription of a gene sequence and subsequent processing steps, such as translation of a resultant mRNA to produce the final end product of a gene. The end product may be a protein (such as an enzyme or receptor) or a nucleic acid (such as a tRNA, antisense RNA, or other regulatory factor). The term "expression control sequence" refers to a sequence of nucleotides that control or regulate expression of structural genes when operably linked to those genes. These include, for example, the lac systems, the trp system, major operator and promoter regions of the phage lambda, the control region of fd coat protein and other sequences known to control the expression of genes in prokaryotic or eukaryotic cells. Expression control sequences will vary depending on whether the vector is designed to express the operably linked gene in a prokaryotic or eukaryotic host, and may contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements and/or translational initiation and termination sites.

"Expression vehicle" refers to a vehicle or vector similar to a cloning vehicle but which is capable of expressing a gene which has been cloned into it, after transformation into a host. The cloned gene is usually placed under the control of (i.e., operably linked to) an expression control sequence.

"Operator" refers to a DNA sequence capable of interacting with the specific repressor, thereby controlling the transcription of adjacent gene(s).

"Promoter" refers to a DNA sequence that can be recognized by an RNA polymerase. The presence of such a sequence permits the RNA polymerase to bind and initiate transcription of operably linked gene sequences.

"Promoter region" is intended to include the promoter as well as other gene sequences which may be necessary for the initiation of transcription. The presence of a promoter region is sufficient to cause the expression of an operably linked gene sequence. The term "promoter" is sometimes used in the art to generically indicate a promoter region. Many different promoters are known in the art which direct expression of a gene in a certain cell types. Tissue-specific promoters can comprise nucleic acid sequences which cause a greater (or decreased) level of expression in cells of a certain tissue type.

"Operably linked" means that the promoter controls the initiation of expression of the gene. A promoter is operably linked to a sequence of proximal DNA if upon introduction into a host cell the promoter determines the transcription of the proximal DNA sequence(s) into one or more species of RNA. A promoter is operably linked to a DNA sequence if the promoter is capable of initiating transcription of that DNA sequence.

"Prokaryote" refers to all organisms without a true nucleus, including bacteria.

"Eukaryote" refers to organisms and cells that have a true nucleus, including mammalian cells.

"Host" includes prokaryotes and eukaryotes, such as yeast and filamentous fungi, as well as plant and animal cells. The term includes an organism or cell that is the recipient of a replicable expression vehicle.

The term "animal" is used herein to include all vertebrate animals, except humans. It also includes an individual animal in all stages of development, including embryonic and fetal stages.

A "transgenic animal" is an animal containing one or more cells bearing genetic information received, directly or indirectly, by deliberate genetic manipulation or by inheritance from a manipulated progenitor at a subcellular level, such as by microinjection or infection with a recombinant viral vector (e.g., adenovirus, retrovirus, herpes virus, adeno-associated virus, lentivirus). This introduced DNA molecule may be integrated within a chromosome, or it may be extra-chromosomally replicating DNA.

"Embryonic stem cells" or "ES cells" as used herein are cells or cell lines usually derived from embryos which are pluripotent meaning that they are un-differentiated cells. These cells are also capable of incorporating exogenous DNA by homologous recombination and subsequently developing into any tissue in the body when incorporated into a host embryo. It is possible to isolate pluripotent cells from sources other than embryonic tissue by methods which are well understood in the art.

Embryonic stem cells in mice have enabled researchers to select for transgenic cells and perform gene targeting. This allows more genetic engineering than is possible with other transgenic techniques. For example, mouse ES cells are relatively easy to grow as colonies *in vitro*. The cells can be transfected by standard procedures and transgenic cells clonally selected by antibiotic resistance. See, for example, Doetschman et al., 1994, *Gene transfer in embryonic stem cells*. In Pinkert (Ed.) Transgenic Animal Technology: A Laboratory Handbook. Academic Press Inc., New York, pp.115-146. Furthermore, the efficiency of this process is such that sufficient transgenic colonies (hundreds to thousands) can be

produced to allow a second selection for homologous recombinants. Mouse ES cells can then be combined with a normal host embryo and, because they retain their potency, can develop into all the tissues in the resulting chimeric animal, including the germ cells. The transgenic modification can then be transmitted to subsequent generations.

Methods for deriving embryonic stem (ES) cell lines *in vitro* from early preimplantation mouse embryos are well known. See for example, Evans et al., 1981 Nature 29:154-156 and Martin, 1981, Proc. Nat. Aca. Sci. USA, 78:7634-7638. ES cells can be passaged in an undifferentiated state, provided that a feeder layer of fibroblast cells or a differentiation inhibiting source is present.

The term "somatic cell" indicates any animal or human cell which is not a sperm or egg cell or is capable of becoming a sperm or egg cell. The term "germ cell" or "germ-line cell" refers to any cell which is either a sperm or egg cell or is capable of developing into a sperm or egg cell and can therefore pass its genetic information to offspring. The term "germ cell-line transgenic animal" refers to a transgenic animal in which the genetic information was incorporated in a germ line cell, thereby conferring the ability to transfer the information to offspring. If such offspring in fact possess some or all of that information, then they, too, are transgenic animals.

The genetic alteration of genetic information may be foreign to the species of animal to which the recipient belongs, or foreign only to the particular individual recipient. In the last case, the altered or introduced gene may be expressed differently than the native gene.

"Fragment" of a gene refers to any portion of a gene sequence. A biologically active fragment refers to any portion of the gene that retains at least one biological activity of that gene.

"Biologically active" refers to those forms of proteins and polypeptides, including conservatively substituted variants, alleles of genes encoding a protein or polypeptide fragments of proteins which retain a biological and/or immunological

activity of the wild-type protein or polypeptide. Preferably the activity is one which induces a change in bone mass development or phenotype. Biologically active also refers the capability to modulate a signaling pathway associated with LRP5 (Zmax1), LPR6, and HBM such as the Wnt pathway whether directly or indirectly and whether *in vivo* or in and *in vitro* assay.

By “effective amount” or “dose effective amount” or "therapeutically effective amount" is meant an amount of an agent which modulates a biological activity of the polypeptide of the invention.

"Variant" refers to a gene that is substantially similar in structure and biological activity or immunological characteristics to either the entire gene or to a fragment of the gene. Provided that the two genes possess a similar activity, they are considered variant as that term is used herein even if the sequence of encoded amino acid residues is not identical.

"Amplification of nucleic acids" refers to methods such as polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known in the art and described, for example, in U.S. Patent Nos. 4,683,195 and 4,683,202. Reagents and hardware for conducting PCR are commercially available. Primers useful for amplifying sequences from the HBM region are preferably complementary to, and hybridize specifically to sequences in the HBM region or in regions that flank a target region therein. HBM sequences generated by amplification may be sequenced directly. Alternatively, the amplified sequence(s) may be cloned prior to sequence analysis.

"Antibody" is meant to include but not limited to polyclonal, monoclonal, chimeric, human, humanized, bispecific, multispecific, primatized™ antibodies. The term "antibodies" preferably refers to polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, that can bind to the HBM proteins and fragments thereof or to nucleic acid sequences from the HBM region, particularly from the HBM locus or a portion thereof. The term

antibody is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Proteins may be prepared synthetically in a protein synthesizer and coupled to a carrier molecule and injected over several months into rabbits. Rabbit sera is tested for immunoreactivity to the HBM protein or fragment. Monoclonal antibodies may be made by injecting mice with the proteins, or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with HBM protein or fragments thereof. Harlow et al., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1988). These antibodies will be useful in assays as well as pharmaceuticals.

The *LRP5* gene and the LRP5 protein which it encodes have previously been referred to as *Zmax1* and Zmax1 by the inventors. The gene and its product have also been referred to in the art with the designation LR3. It is understood that Zmax, Zmax1, LRP5, and LR3 are synonymous terms. "HBM protein" refers to a protein that is identical to a LRP5 protein except that it contains an alteration of glycine 171 to valine. An HBM protein is defined for any organism that encodes a LRP5 true homolog. For example, a mouse HBM protein refers to the mouse LRP5 protein having the glycine 170 to valine substitution.

In one embodiment of the present invention, "*HBM* gene" refers to the genomic DNA sequence found in individuals showing the HBM characteristic or phenotype, where the sequence encodes the protein indicated by SEQ ID NO: 4. The *HBM* gene and the *LRP5* gene are allelic. The protein encoded by the *HBM* gene has the property of causing elevated bone mass, while the protein encoded by the *LRP5* gene does not. The *HBM* gene and the *LRP5* gene differ in that the *HBM* gene has a thymine at position 582, while the *LRP5* gene has a guanine at position 582. The *HBM* gene comprises the nucleic acid sequence shown as SEQ ID NO: 2. The *HBM* gene may also be referred to as an "HBM polymorphism."

In alternative embodiments of the present invention, "HBM gene" may also refer to any allelic variant of LRP5 (Zmax1) or LRP6 which results in the HBM

phenotype. Such variants may include alteration from the wild-type protein coding sequence as described herein and/or alteration in expression control sequences of LRP5. A preferred example of such a variant is an alteration of the endogenous LRP5 promoter region resulting in increased expression of the LRP5 protein.

"Normal," "wild-type," "unaffected" and "LRP5" all refer to the genomic DNA sequence that encodes the protein indicated by SEQ ID NO: 3. The *LRP5* gene has a guanine at position 582. The *LRP5* (*Zmax1*) gene comprises the nucleic acid sequence shown as SEQ ID NO: 1. "Normal," "wild-type," "unaffected" and "LRP5" also refer to allelic variants of the genomic sequence that encodes proteins that do not contribute to elevated bone mass. The *LRP5* gene is common in the human population, while the *HBM* gene is rare.

"5YWTD+EGF" refers to a repeat unit found in the LRP5 protein, consisting of five YWTD repeats followed by an EGF repeat.

"Bone development" generally refers to any process involved in the change of bone over time, including, for example, normal development, changes that occur during disease states, and changes that occur during aging. This may refer to structural changes in and dynamic rate changes such as growth rates, resorption rates, bone repair rates, and etc. "Bone development disorder" particularly refers to any disorders in bone development including, for example, changes that occur during disease states and changes that occur during aging. Bone development may be progressive or cyclical in nature. Aspects of bone that may change during development include, for example, mineralization, formation of specific anatomical features, and relative or absolute numbers of various cell types.

"Bone modulation" or "modulation of bone formation" refers to the ability to affect any of the physiological processes involved in bone remodeling, as will be appreciated by one skilled in the art, including, for example, bone resorption and appositional bone growth, by, inter alia, osteoclastic and osteoblastic activity, and may comprise some or all of bone formation and development as used herein.

Bone is a dynamic tissue that is continually adapting and renewing itself through the removal of old or unnecessary bone by osteoclasts and the rebuilding of new bone by osteoblasts. The nature of the coupling between these processes is responsible both for the modeling of bone during growth as well as the maintenance of adult skeletal integrity through remodeling and repair to meet the everyday needs of mechanical usage. There are a number of diseases of bone that result from an uncoupling of the balance between bone resorption and formation. With aging there is a gradual "physiologic" imbalance in bone turnover, which is particularly exacerbated in women due to menopausal loss of estrogen support, that leads to a progressive loss of bone. The reduction in bone mass and deterioration in bone architecture results in an increase in bone fragility and susceptibility to spontaneous fractures. For every 10 percent of bone that is lost the risk of fracture doubles. Individuals with bone mineral density (BMD) in the spine or proximal femur 2.5 or more standard deviations below normal peak bone mass are classified as osteoporotic. However, osteopenic individuals with BMD between 1 and 2.5 standard deviations below the norm are clearly at risk of suffering bone loss related disorders.

Bone modulation may be assessed by measuring parameters such as bone mineral density (BMD) and bone mineral content (BMC) by pDXA X-ray methods, bone size, thickness or volume as measured by X-ray, bone formation rates as measured for example by calcien labeling, total, trabecular, and mid-shaft density as measured by pQCT and/or μCT methods, connectivity and other histological parameters as measured by μCT methods, mechanical bending and compressive strengths as preferably measured in femur and vertebrae respectively. Due to the nature of these measurements, each may be more or less appropriate for a given situation as the skilled practitioner will appreciate. Furthermore, parameters and methodologies such as a clinical history of freedom from fracture, bone shape, bone morphology, connectivity, normal histology, fracture repair rates, and other bone quality parameters are known and used in the art. Most preferably, bone quality

may be assessed by the compressive strength of vertebra when such a measurement is appropriate. Bone modulation may also be assessed by rates of change in the various parameters. Most preferably, bone modulation is assessed at more than one age.

"Normal bone density" refers to a bone density within two standard deviations of a Z score of 0 in the context of the HBM linkage study. In a general context, the range of normal bone density parameters is determined by routine statistical methods. A normal parameter is within about 1 or 2 standard deviations of the age and sex normalized parameter, preferably about 2 standard deviations. A statistical measure of meaningfulness is the P value which can represent the likelihood that the associated measurement is significantly different from the mean. Significant P values are $P < 0.05$, 0.01, 0.005, and 0.001, preferably at least $P < 0.01$.

"HBM" refers to high bone mass although this term may also be expressed in terms of bone density, mineral content, and size.

The "HBM phenotype" may be characterized by an increase of about 2 or more standard deviations, preferably 2, 2.5, 3, or more standard deviations in 1, 2, 3, 4, 5, or more quantitative parameters of bone modulation, preferably bone density and mineral content and bone strength parameters, above the age and sex norm for that parameter. The HBM phenotype is characterized by statistically significant increases in at least one parameter, preferably at least 2 parameters, and more preferably at least 3 or more parameters. The HBM phenotype may also be characterized by an increase in one or more bone quality parameters and most preferably increasing parameters are not accompanied by a decrease in any bone quality parameters. Most preferably, an increase in bone modulation parameters and/or bone quality measurements is observed at more than one age.

A "LRP5 system" refers to a purified protein, cell extract, cell, animal, human or any other composition of matter in which LRP5 is present in a normal or mutant form.

The term "isolated" refers to a substance altered by hand of man from the natural environment. An isolated peptide may be for example in a substantially pure form or otherwise displaced from its native environment such as by expression in an isolated cell line or transgenic animal. An isolated sequence may for example be a molecule in substantially pure form or displaced from its native environment such that at least one end of said isolated sequence is not contiguous with the sequence it would be contiguous with in nature.

A "surrogate marker" refers to a diagnostic indication, symptom, sign or other feature that can be observed in a cell, tissue, human or animal that is correlated with the *HBM* gene or elevated bone mass or both, but that is easier to measure than bone density. The general concept of a surrogate marker is well accepted in diagnostic medicine.

The present invention encompasses the *LRP5* gene and LRP5 protein in the forms indicated by SEQ ID NOS: 1 and 3, respectively, and other closely related variants, as well as the adjacent chromosomal regions of *LRP5* necessary for its accurate expression. In a preferred embodiment, the present invention is directed to at least 15 contiguous nucleotides of the nucleic acid sequence of SEQ ID NO: 1.

I. Introduction

The present invention also encompasses the *HBM* gene and HBM protein in the forms indicated by SEQ ID NO: 2 and 4, respectively, and other closely related variants, as well as the adjacent chromosomal regions of the HBM gene necessary for its accurate expression. In a preferred embodiment, the present invention is directed to at least 15 contiguous nucleotides of the nucleic acid sequence of SEQ ID NO: 2. More preferably, the present invention is directed to at least 15 contiguous nucleotides of the nucleic acid sequence of SEQ ID NO: 2, wherein one of the 15 contiguous nucleotides is the thymine at nucleotide 582.

The invention also relates to the nucleotide sequence of the *LRP5* gene region, as well as the nucleotide sequence of the *HBM* region. More particularly, a

preferred embodiment are the BAC clones containing segments of the *LRP5* gene region B200E21-H and B527D12-H. A preferred embodiment is the nucleotide sequence of the BAC clones consisting of SEQ ID NOS: 5-12.

The invention also concerns the use of the nucleotide sequence to identify DNA probes for the *LRP5* gene and the HBM gene, PCR primers to amplify the *LRP5* gene and the *HBM* gene, nucleotide polymorphisms in the *LRP5* gene and the *HBM* gene, and regulatory elements of the *LRP5* gene and the *HBM* gene.

This invention describes the further localization of the chromosomal location of the *LRP5* gene and *HBM* gene on chromosome 11q13.3 between genetic markers D11S987 and SNP_CONTIG033-6, as well as the DNA sequences of the *LRP5* gene and the *HBM* gene. The chromosomal location was refined by the addition of more genetic markers to the mapping panel used to map the gene, and by the extension of the pedigree to include more individuals. The pedigree extension was critical because the new individuals that have been genotyped harbor critical recombination events that narrow the region. To identify genes in the region on 11q13.3, a set of BAC clones containing this chromosomal region was identified. The BAC clones served as a template for genomic DNA sequencing, and also as a reagent for identifying coding sequences by direct cDNA selection. Genomic sequencing and direct cDNA selection were used to characterize more than 1.5 million base pairs of DNA from 11q13.3. The *LRP5* gene was identified within this region and the *HBM* gene was then discovered after mutational analysis of affected and unaffected individuals.

When a gene has been genetically localized to a specific chromosomal region, the genes in this region can be characterized at the molecular level by a series of steps that include: cloning of the entire region of DNA in a set of overlapping clones (physical mapping), characterization of genes encoded by these clones by a combination of direct cDNA selection, exon trapping and DNA sequencing (gene identification), and identification of mutations in these genes by

comparative DNA sequencing of affected and unaffected members of the HBM kindred (mutation analysis).

Physical mapping is accomplished by screening libraries of human DNA cloned in vectors that are propagated in *E. coli* or *S. cereviseae* using PCR assays designed to amplify unique molecular landmarks in the chromosomal region of interest. To generate a physical map of the HBM candidate region, a library of human DNA cloned in Bacterial Artificial Chromosomes (BACs) was screened with a set of Sequence Tagged Site (STS) markers that had been previously mapped to chromosome 11q12-q13 by the efforts of the Human Genome Project.

STSs are unique molecular landmarks in the human genome that can be assayed by PCR. Through the combined efforts of the Human Genome Project, the location of thousands of STSs on the twenty-two autosomes and two sex chromosomes has been determined. For a positional cloning effort, the physical map is tied to the genetic map because the markers used for genetic mapping can also be used as STSs for physical mapping. By screening a BAC library with a combination of STSs derived from genetic markers, genes, and random DNA fragments, a physical map comprised of overlapping clones representing all of the DNA in a chromosomal region of interest can be assembled.

BACs are cloning vectors for large (80 kilobase to 200 kilobase) segments of human or other DNA that are propagated in *E. coli*. To construct a physical map using BACs, a library of BAC clones is screened so that individual clones harboring the DNA sequence corresponding to a given STS or set of STSs are identified. Throughout most of the human genome, the STS markers are spaced approximately 20 to 50 kilobases apart, so that an individual BAC clone typically contains at least two STS markers. In addition, the BAC libraries that were screened contain enough cloned DNA to cover the human genome six times over. Therefore, an individual STS typically identifies more than one BAC clone. By screening a six-fold coverage BAC library with a series of STS markers spaced approximately 50 kilobases apart, a physical map consisting of a series of overlapping BAC clones,

i.e. BAC contigs, can be assembled for any region of the human genome. This map is closely tied to the genetic map because many of the STS markers used to prepare the physical map are also genetic markers.

When constructing a physical map, it often happens that there are gaps in the STS map of the genome that result in the inability to identify BAC clones that are overlapping in a given location. Typically, the physical map is first constructed from a set of STSs that have been identified through the publicly available literature and World Wide Web resources. The initial map consists of several separate BAC contigs that are separated by gaps of unknown molecular distance. To identify BAC clones that fill these gaps, it is necessary to develop new STS markers from the ends of the clones on either side of the gap. This is done by sequencing the terminal 200 to 300 base pairs of the BACs flanking the gap, and developing a PCR assay to amplify a sequence of 100 or more base pairs. If the terminal sequences are demonstrated to be unique within the human genome, then the new STS can be used to screen the BAC library to identify additional BACs that contain the DNA from the gap in the physical map. To assemble a BAC contig that covers a region the size of the HBM candidate region (2,000,000 or more base pairs), it is often necessary to develop new STS markers from the ends of several clones.

After building a BAC contig, this set of overlapping clones serves as a template for identifying the genes encoded in the chromosomal region. Gene identification can be accomplished by many methods. Three methods are commonly used: (1) a set of BACs selected from the BAC contig to represent the entire chromosomal region can be sequenced, and computational methods can be used to identify all of the genes, (2) the BACs from the BAC contig can be used as a reagent to clone cDNAs corresponding to the genes encoded in the region by a method termed direct cDNA selection, or (3) the BACs from the BAC contig can be used to identify coding sequences by selecting for specific DNA sequence motifs in a procedure called exon trapping. The present invention includes genes identified by the first two methods.

To sequence the entire BAC contig representing the HBM candidate region, a set of BACs was chosen for subcloning into plasmid vectors and subsequent DNA sequencing of these subclones. Since the DNA cloned in the BACs represents genomic DNA, this sequencing is referred to as genomic sequencing to distinguish it from cDNA sequencing. To initiate the genomic sequencing for a chromosomal region of interest, several non-overlapping BAC clones are chosen. DNA for each BAC clone is prepared, and the clones are sheared into random small fragments which are subsequently cloned into standard plasmid vectors such as pUC18. The plasmid clones are then grown to propagate the smaller fragments, and these are the templates for sequencing. To ensure adequate coverage and sequence quality for the BAC DNA sequence, sufficient plasmid clones are sequenced to yield six-fold coverage of the BAC clone. For example, if the BAC is 100 kilobases long, then phagemids are sequenced to yield 600 kilobases of sequence. Since the BAC DNA was randomly sheared prior to cloning in the phagemid vector, the 600 kilobases of raw DNA sequence can be assembled by computational methods into overlapping DNA sequences termed sequence contigs. For the purposes of initial gene identification by computational methods, six-fold coverage of each BAC is sufficient to yield ten to twenty sequence contigs of 1000 base pairs to 20,000 base pairs.

The sequencing strategy employed in this invention was to initially sequence "seed" BACs from the BAC contig in the HBM candidate region. The sequence of the "seed" BACs was then used to identify minimally overlapping BACs from the contig, and these were subsequently sequenced. In this manner, the entire candidate region was sequenced, with several small sequence gaps left in each BAC. This sequence served as the template for computational gene identification. One method for computational gene identification is to compare the sequence of BAC contig to publicly available databases of cDNA and genomic sequences, e.g. unigene, dbEST, genbank. These comparisons are typically done using the BLAST family of computer algorithms and programs (Altschul et al., *J. Mol. Biol.*, 215:403-410

(1990)). The BAC sequence can also be translated into protein sequence, and the protein sequence can be used to search publicly available protein databases, using a version of BLAST designed to analyze protein sequences (Altschul et al., *Nucl. Acids Res.*, 25:3389-3402 (1997)). Another method is to use computer algorithms such as MZEF (Zhang, *Proc. Natl. Acad. Sci.*, 94:565-568 (1997)) and GRAIL (Uberbacher et al., *Methods Enzymol.*, 266:259-281 (1996)), which predict the location of exons in the sequence based on the presence of specific DNA sequence motifs that are common to all exons, as well as the presence of codon usage typical of human protein encoding sequences.

In addition to identifying genes by computational methods, genes were also identified by direct cDNA selection (Del Mastro et al., *Genome Res.* 5(2):185-194 (1995)). In direct cDNA selection, cDNA pools from tissues of interest are prepared, and the BACs from the candidate region are used in a liquid hybridization assay to capture cDNA which base-pairs to coding regions in the BAC. In the methods described herein, the cDNA pools were created from several different tissues by random priming the first strand cDNA from poly-A RNA, synthesizing the second strand cDNA by standard methods, and adding linkers to the ends of the cDNA fragments. The linkers are used to amplify the cDNA pools. The BAC clones are used as a template for *in vitro* DNA synthesis to create a biotin labeled copy of the BAC DNA. The biotin labeled copy of the BAC DNA is then denatured and incubated with an excess of the PCR amplified, linkered cDNA pools which have also been denatured. The BAC DNA and cDNA are allowed to anneal in solution, and heteroduplexes between the BAC and the cDNA are isolated using streptavidin coated magnetic beads. The cDNA which is captured by the BAC is then amplified using primers complimentary to the linker sequences, and the hybridization/selection process is repeated for a second round. After two rounds of direct cDNA selection, the cDNA fragments are cloned, and a library of these direct selected fragments is created.

The cDNA clones isolated by direct selection are analyzed by two methods. Since a pool of BACs from the HBM candidate region is used to provide the genomic DNA sequence, the cDNAs must be mapped to individual BACs. This is accomplished by arraying the BACs in microtiter dishes, and replicating their DNA in high density grids. Individual cDNA clones are then hybridized to the grid to confirm that they have sequence identity to an individual BAC from the set used for direct selection, and to determine the specific identity of that BAC. cDNA clones that are confirmed to correspond to individual BACs are sequenced. To determine whether the cDNA clones isolated by direct selection share sequence identity or similarity to previously identified genes, the DNA and protein coding sequences are compared to publicly available databases using the BLAST family of programs.

The combination of genomic DNA sequence and cDNA sequence provided by BAC sequencing and by direct cDNA selection yields an initial list of putative genes in the region. The genes in the region were all candidates for the HBM locus. To further characterize each gene, Northern blots were performed to determine the size of the transcript corresponding to each gene, and to determine which putative exons were transcribed together to make an individual gene. For Northern blot analysis of each gene, probes were prepared from direct selected cDNA clones or by PCR amplifying specific fragments from genomic DNA or from the BAC encoding the putative gene of interest. The Northern blots gave information on the size of the transcript and the tissues in which it was expressed. For transcripts which were not highly expressed, it was sometimes necessary to perform a reverse transcription PCR assay using RNA from the tissues of interest as a template for the reaction.

Gene identification by computational methods and by direct cDNA selection provides unique information about the genes in a region of a chromosome. When genes are identified, then it is possible to examine different individuals for mutations in each gene.

The present invention also encompasses the *HBM* gene and HBM protein in the forms indicated by SEQ ID NO: 2 and 4, respectively, and other closely related variants, as well as the adjacent chromosomal regions of the *HBM* gene necessary for its accurate expression. In a preferred embodiment, the present invention is directed to an isolated nucleic acid sequence of SEQ ID NO: 2, as well as variants thereof. Variants of SEQ ID NO: 2 include polynucleotides having at least about 90%, preferably 95%, or more preferably 98% similarity or identity to the nucleic acid sequence of SEQ ID NO: 2 or fragments thereof. Therefore, sequences which are 96%, 97%, and 99% similar to SEQ ID NO: 2 or fragments thereof are also contemplated herein.

Determination of the degree of variation between a high bone mass (HBM) variant can be performed using BLAST or FASTA or other suitable algorithm using standard default parameters. Preferably, identity will be determined for coding regions of SEQ ID NO: 2, but can also include non-coding domains. Additionally, alignment programs can be used to identify conserved sequences or potential motifs across different animal species. Alignment programs can also be used to align the nucleic acid and/or protein sequences of related genes and the proteins that they encode. Preferred alignment programs include CLUSTALW, PILEUP and GAP, and would preferably be used with default parameters. For example, such programs can be used to align the sequences of LRP5 (also known as Zmax1), HBM, LDL receptor-related protein 6 (LRP6) and related sequences.

By a polynucleotide having a nucleotide sequence at least, for example, 90% "similar" to a reference nucleotide sequence encoding a polypeptide, is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to ten point mutations per each 100 nucleotides of the reference nucleotide sequence. These mutations of the reference sequence may occur at any location in SEQ ID NO: 2 and may be silent, or may or may not encode an amino acid substitution.

Another embodiment contemplates that such polynucleotide variants of SEQ ID NO: 2 comprise nucleic acid sequences which are at least 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, or 500 contiguous nucleotides of SEQ ID NO: 2. More preferably, such polynucleotide variants have a contiguous nucleic acid sequence corresponding with the polymorphism at nucleotide 582 (G→T substitution) of SEQ ID NO: 2 or other variants of SEQ ID NO: 2, which comprise a mutation which modulates bone mass when the polypeptide encoded thereby is administered to a subject. All variants of SEQ ID NO: 2 contemplated possess the characteristic of encoding a protein or polypeptide which when administered to a subject induces bone modulation. Additional variants which may be responsible for modulating bone mass when administered to a subject may lie within the domain known to contain the HBM polymorphism and which encodes the beta propeller domain (YWTD motifs). Alternatively, other variants of LRP5 which modulate bone mass and/or result in an HBM phenotype in a subject may be due to mutations in the nucleic acid sequences encoding any of the other conserved domains of LRP5, such as those set forth in Figure 4 (e.g., the RGD extracellular attachment site, the binding site for LDL and calcium, the cysteine rich growth factor repeats, the ideal PEST region, and the internalization domain)

HBM polynucleotides contemplated include those which hybridize under stringent conditions to SEQ ID NO: 2. Hybridization methods are known in the art and include, but are not limited to: (a) washing with 0.1X SSPE (0.62 M NaCl, 0.06 M $NaH_2PO_4 \cdot H_2O$, 0.075 M EDTA, pH 7.4) and 0.1% sodium dodecyl sulfate (SDS) at 50°C; (b) washing with 50% formamide, 5X SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6-8), 0.1% sodium pyrophosphate, 5X Denhardt's solution, sonicated salmon sperm DNA (50 μg/ml), 0.1% SDS and 10% dextran sulfate at 42°C, followed by washing at 42°C in 0.2X SSC and 0.1% SDS; and (c) washing with of 0.5 M $NaPO_4$, 7% SDS at 65°C followed by washing at 60°C in 0.5X SSC and 0.1% SDS. Additional conditions under which HBM variants can be isolated by hybridization to SEQ ID NO: 2 or nucleic acid

fragments thereof can be performed by varying the hybridization temperature. High stringency hybridization conditions are those performed at about 20°C below the melting temperature (T_m) of SEQ ID NO: 2 or fragments thereof. Preferred stringency is performed at about 5-10°C below the T_m of SEQ ID NO: 2 or fragments thereof. Additional hybridization conditions can be prepared as described in Chapter 11 of Sambrook et al., Molecular Cloning: A Laboratory Manual (1989), or as would be known to the artisan of ordinary skill.

Alternatively, mammalian libraries (e.g., equine, primate, caprine, bovine, ovine, feline, porcine, and canine) can be probed using degenerate primers and polymerase chain reaction (PCR) techniques to identify variants of SEQ ID NO: 2 or fragments thereof. Preferably primers are utilized which hybridize under stringent conditions to the open reading frame of SEQ ID NO: 2, or to non-coding portions of the sequence. More preferably, such primers hybridize to conserved domains within SEQ ID NO: 2. For example, conserved domains include those coding for the YWTD beta-propeller domains or other domains, such as those listed in Figure 4. Preferred primers are typically 15 nucleotides in length, but can vary to be at least, about 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35 or 40 nucleotides in length. Heterologous hybridization is to amplify the target gene or nucleic acid sequence using degenerate PCR primers. Probes for variants of SEQ ID NO: 2 and the polypeptide encoded thereby can be obtained by preparing mixed oligonucleotides of greater than 10, preferably of 15 or more, nucleotides in length representing all possible nucleotide sequences which could encode the corresponding amino acid sequences (e.g., SEQ ID NO: 4 fragments thereof). This method is clearly documented by Gould et al., 1989, Proc. Natl. Acad. Sci. USA 86(6): 1934-8.

Another embodiment includes nucleic acids which encode an HBM polypeptide which is at least about 90% similar to SEQ ID NO: 4 and fragments thereof, and which when administered to a subject modulate bone mass in that subject. Such HBM polypeptides include variants which have a valine

corresponding to position 171 of SEQ ID NO: 4 (Gly to Val substitution) or 170 of the mouse homolog. Other preferred embodiments include high bone mass polypeptides which have at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 500 or more contiguous amino acids of SEQ ID NO: 4. Such contemplated contiguous sequences preferably overlap with a polymorphism corresponding to high bone mass, such as valine-171 of SEQ ID NO: 4. Also contemplated are the polynucleotides encoding polypeptides which are at least about 95%, 96%, 97%, 98% and 99% similar to SEQ ID NO: 4 and fragments thereof.

In another embodiment, a synthetic nucleic acid encoding SEQ ID NO: 4 is contemplated wherein the nucleic acid sequence has been conservatively substituted based on the degeneracy of the code such that no amino acids are altered in SEQ ID NO: 4, but perhaps wherein the resulting synthetic polynucleotide encoding said SEQ ID NO: 4 is one that is at least, about 50% similar to SEQ ID NO: 2.

By a polypeptide having an amino acid sequence of at least, for example, 95% "identity" to a reference amino acid sequence of SEQ ID NO: 4 or fragment thereof is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid sequence of SEQ ID NO: 4. In other words, to obtain a polypeptide having an amino acid sequence 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

Additional HBM polypeptides and nucleic acids which encode said HBM polypeptides are contemplated wherein amino acid residues are conservatively substituted. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie et al., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," Science 247: 1306-10 (1990), wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selections or screens to identify sequences that maintain functionality. These studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at a certain position of the protein. Numerous phenotypic substitutions are described in Bowie et al., *supra*, and the references cited therein, which is herein incorporated by reference in its entirety. Preferred substitutions would be in domains which are less conserved across species, and which do not correspond to a structurally or functionally important domain (e.g., a binding site, catalytic site, or beta propeller or other domain described in Figure 4).

Variants of LRP5/HBM

A structural model of the LRP5 first beta-propeller module was generated based on model predictions from two YWTD- propeller containing molecules (chicken LRP1 and human nidogen, with Swiss-PROT references LRP1_CH7 and NIDO_HU1 respectively) as described in Springer et al., (1998) *J. Molecular Biology,* 283:837-862. Based on the model, certain amino acid residues were identified as important variants of HBM/LRP5. The following three categories provide examples of such variants:

The shape of the beta-propeller resembles a disk with inward-sloping sides and a hole down the middle. Residue 171 is in a loop on the outer or top surface of

the domain in blade 4 of propeller module 1. Thus, variants comprising changed residues in structurally equivalent positions in other blades; as well as residues that are slightly more interior to the binding pocket, but still accessible to the surface, are important embodiments of the present invention for the study of bone mass modulation by HBM, for the development of pharmaceuticals and treatments of bone mass disorders, and for other objectives of the present invention. The following are examples of such variants:

A214V (a position equivalent to 171 in blade 5; alanine is not conserved in other propellers),

E128V (a position equivalent to 171 in blade 3; glutamate is not conserved in other propellers),

A65V (a position equivalent to 171 in blade 2; alanine is conserved in propellers 1-3 but not 4),

G199V (an accessible interior position in blade 5; glycine is conserved in propellers 1-3 but not 4), and

M282V (accessible interior position in blade 1; methionine is conserved in propellers 1-3 but not 4).

LRP5 has four beta-propeller structures; the first three beta-propeller modules conserve a glycine in the position corresponding to residue 171 in human LRP5. Therefore, variants bearing a valine in the equivalent positions in the other propellers are important embodiments of the present invention. The following variants are useful for the study of bone mass modulation by HBM, for the development of pharmaceuticals and treatments of bone mass disorders, and for other objectives of the present invention: G479V, G781V, and Q1087V.

The G171V HBM polymorphism results in “occupied space” of the beta-propeller 1, with the side-chain from the valine residue sticking out into an open binding pocket and potentially altering a ligand/protein interaction. The glycine residue is conserved in LRP5 propellers 1, 2 and 3 but is a glutamine in propeller 4. Therefore, the following variants of HBM are important embodiments of the

present invention for the study of bone mass modulation by HBM, for the development of pharmaceuticals and treatments of bone mass disorders, and for other objectives of the present invention:

G171K (which introduces a charged side-chain),

G171F (which introduces a ringed side-chain),

G171I (which introduces a branched side-chain), and

G171Q (which introduces the propeller 4 residue).

Furthermore, LRP6 is the closest homolog of LRP5. Thus, bone density may also be modulated by LRP6. LRP6 has a beta-propeller structure predicted to be similar, if not identical to LRP5. The position corresponding to glycine 171 in human LRP5 is glycine 158 of human LRP6. Therefore, corresponding variants of LRP6 are an important embodiment of the present invention for the study of the specificity of LRP5 versus its related family member, for the development of pharmaceuticals and treatments of bone mass disorders, and for other objectives of the present invention. Specifically, for example, a glycine to valine substitution at the structurally equivalent position, residue 158, of human LRP6 and similar variants of other species' LRP6 homologs represent important research tools.

One skilled in the art will recognize that these are only a few illustrative examples presented to better describe the present invention and that many other variants may be contemplated within the scope of the present invention.

Methods of determining the bone mass modulating activity of a polypeptide or nucleic acid sequence encoding a polypeptide can be performed using different animal models for studying bone mass. For example, ovariectomized murine models or spontaneously osteoporotic mouse strains can be utilized to determine whether a LRP5 modulating agent correspondingly modulates bone mass in the animal model. For *in vivo* analysis of such mice, see Kalu et al., (1999) *J. Bone Miner. Res.* 14: 593-601 and Shimizu et al., (1999) *Mamm. Genome* 10: 81-7.

Additional *in vivo* assays which can be used are transgenic animals and knockout animals in which expression of LRP5 has been altered or the nucleic acid

encoding HBM introduced. These animals can then be utilized to identify compounds or compositions which modulate bone mass.

The polypeptide of the present invention is preferably provided in an isolated form. By "isolated polypeptide" is intended a polypeptide removed from its native environment. Thus, a polypeptide produced and contained within a recombinant host cell would be considered "isolated" for purposes of the present invention. Also intended as an "isolated polypeptide" are polypeptides that have been purified, partially or substantially, from a recombinant host. Similarly, by "isolated nucleic acid" or "isolated polynucleotide" is meant a nucleic acid sequence which is purified from other nucleic acid and protein contaminants.

The present invention also encompasses the *LRP5* gene and LRP5 protein in the forms indicated by SEQ ID NOS: 1 and 3, respectively, and other closely related variants. The present invention also encompasses the adjacent chromosomal regions of LRP5 necessary for its accurate expression.

In a preferred embodiment, the present invention is directed to at least 15 contiguous nucleotides of the nucleic acid sequence of SEQ ID NO: 1. Variants of the *LRP5* gene and LRP5 protein in the forms indicated by SEQ ID NOS: 1 and 3, respectively may be identified generally as described above for the *HBM* gene and HBM protein without the G171V HBM polymorphism.

The invention also relates to the nucleotide sequence of the *LRP5* gene region, as well as the nucleotide sequence of the *HBM* gene region. More particularly, a preferred embodiment are the BAC clones containing segments of the *LRP5* gene region B200E21-H and B527D12-H. A preferred embodiment is the nucleotide sequence of the BAC clones consisting of SEQ ID NOS: 5-12.

The invention also concerns the use of the nucleotide sequence to identify DNA probes for the *LRP5* gene and the *HBM* gene, PCR primers to amplify the *LRP5* gene and the *HBM* gene, nucleotide polymorphisms in the *LRP5* gene and the *HBM* gene, and regulatory elements of the *LRP5* gene and the *HBM* gene.

II. Phenotyping using DXA Measurements

Spinal bone mineral content (BMC) and bone mineral density (BMD) measurements performed at Creighton University (Omaha, Nebraska) were made by DXA using a Norland Instruments densitometer (Norland XR2600 Densitometer, Dual Energy X-ray Absorptiometry, DXA). Spinal BMC and BMD at other locations used the machinery available. There are estimated to be 800 DXA machines currently operating in the U.S. Most larger cities have offices or imaging centers which have DXA capabilities, usually a Lunar or Hologic machine. Each location that provided spine BMC and BMD data included copies of the printouts from their machines to provide verification that the regions of interest for measurement of BMD have been chosen appropriately. Complete clinical histories and skeletal radiographs were obtained.

The HBM phenotype in human and animal subjects, preferably humans, can be described using criteria such as: very high spinal BMD; a clinical history devoid of any known high bone mass syndrome; and skeletal radiographs showing a normal shape of the appendicular skeleton.

III. Genotyping of Microsatellite Markers

To narrow the genetic interval to a region smaller than that originally reported by Johnson et al., *Am. J. Hum. Genet.*, 60:1326-1332 (1997), additional microsatellite markers on chromosome 11q12-13 were typed. The new markers included: D11S4191, D11S1883, D11S1785, D11S4113, D11S4136, D11S4139, (Dib, et al., *Nature*, 380:152-154 (1996), FGF3 (Polymeropolous, et al., *Nucl. Acid Res.*, 18:7468 (1990)), as well as GTC_HBM_Marker_1, GTC_HBM_Marker_2, GTC_HBM_Marker_3, GTC_HBM_Marker_4, GTC_HBM_Marker_5, GTC_HBM_Marker_6, and GTC_HBM_Marker_7 (*See* Fig. 2).

Blood (20 ml) was drawn into lavender cap (EDTA containing) tubes by a certified phlebotomist. The blood was stored refrigerated until DNA extraction.

DNA has been extracted from blood stored for up to 7 days in the refrigerator without reduction in the quality or quantity of yield. For those subjects that have blood drawn at distant sites, a shipping protocol was successfully used on more than a dozen occasions. Blood samples were shipped by overnight express in a styrofoam container with freezer packs to provide cooling. Lavender cap tubes were placed on individual plastic shipping tubes and then into "zip-lock" biohazard bags. When the samples arrived the next day, they were immediately processed to extract DNA.

The DNA extraction procedure used a kit purchased from Gentra Systems, Inc. (Minneapolis, Minnesota). Briefly, the procedure involved adding 3 volumes of a red blood cell lysis buffer to the whole blood. After incubations for 10 minutes at room temperature, the solution was centrifuged in a Beckman tabletop centrifuge at 2,000 X g for 10 minutes. The white blood cell pellet was resuspended in Cell Lysis Buffer. Once the pellet was completely resuspended and free of cell clumps, the solution was digested with RNase A for 15 minutes at 37°C. Proteins were precipitated by addition of the provided Protein Precipitation Solution and removed by centrifugation. The DNA was precipitated out of the supernatant by addition of isopropanol. This method was simple and fast, requiring only 1-2 hours, and allowed for the processing of dozens of samples simultaneously. The yield of DNA was routinely >8 mg for a 20 ml sample of whole blood and had a MW of >50 kb. DNA was archived by storing coded 50 μg aliquots at -80°C as an ethanol precipitate.

DNA was genotyped using one fluorescently labeled oligonucleotide primer and one unlabeled oligonucleotide primer. Labeled and unlabeled oligonucleotides were obtained from Integrated DNA Technologies, Inc. (Coralville, Iowa). All other reagents for microsatellite genotyping were purchased from Perkin Elmer-Applied Biosystems, Inc. ("PE-ABI") (Norwalk, Connecticut). Individual PCR reactions were performed for each marker, as described by PE-ABI using AmpliTaq™ DNA Polymerase. The reactions were added to 3.5 μl of loading buffer

containing deionized formamide, blue dextran and TAMRA 350 size standards (PE-ABI). After heating at 95°C for 5 minutes to denature the DNA, the samples were loaded and electrophoresed as described in the operator's manual for the Model 377 DNA Sequencer (PE-ABI, Foster City, California). After gel electrophoresis, the data was analyzed using PE-ABI GENESCAN™ and GENOTYPER™ software. First, within the GENESCAN™ software, the lane tracking was manually optimized prior to the first step of analysis. After the gel lane data was extracted, the standard curve profiles of each lane were examined and verified for linearity and size calling. Lanes, which had problems with either of these parameters, were re-tracked and verified. Once all lanes were tracked and the size standards were correctly identified, the data were imported into GENOTYPER™ for allele identification To expedite allele calling (binning), the program Linkage Designer from the Internet web-site of Dr. Guy Van Camp (http://alt.www.uia.ac.be/u/dnalab/ld.html) was used. This program greatly facilitates the importing of data generated by GENOTYPER™ into the pedigree drawing program Cyrillic (Version 2.0, Cherwell Scientific Publishing Limited, Oxford, Great Britain) and subsequent linkage analysis using the program LINKAGE (Lathrop et al., *Am. J. Hum. Genet.*, 37:482-498 (1985)).

IV. Linkage Analysis

Fig. 1 demonstrates the pedigree of the individuals used in the genetic linkage studies for this invention. Specifically, two-point linkage analysis was performed using the MLINK and LINKMAP components of the program LINKAGE (Lathrop et al., *Am. J. Hum. Genet.*, 37:482-498 (1985)). Pedigree/marker data was exported from Cyrillic as a pre-file into the Makeped program and converted into a suitable ped-file for linkage analysis.

The original linkage analysis was performed using three models: (i) an autosomal dominant, fully penetrant model, (ii) an autosomal dominant model with reduced penetrance, and (iii) a quantitative trait model. The HBM locus was

mapped to chromosome 11q12-13 by analyzing DNA for linked markers from 22 members of a large, extended kindred. A highly automated technology was used with a panel of 345 fluorescent markers which spanned the 22 autosomes at a spacing interval ranging from 6-22 cM. Only markers from this region of chromosome 11 showed evidence of linkage (LOD score ~3.0). The highest LOD score (5.74) obtained by two-point and multipoint analysis was D11S987 (map position 55 in Fig. 2). The 95% confidence interval placed the HBM locus between markers D11S905 and D11S937 (map position 41-71 in Fig. 2). Haplotype analysis also places the *LRP5* gene in this same region. Further descriptions of the markers D11S987, D11S905, and D11S937 can be found in Gyapay et al., *Nature Genetics*, Vol. 7, (1994).

In this invention, the inventors report the narrowing of the HBM interval to the region between markers D11S987 and GTC_HBM_Marker_5. These two markers lie between the delimiting markers from the original analysis (D11S905 and D11S937) and are approximately 3 cM from one another. The narrowing of the interval was accomplished using genotypic data from the markers D11S4191, D11S1883, D11S1785, D11S4113, D11S4136, D11S4139, (Dib et al., *Nature*, 380:152-154 (1996)), FGF3 (Polymeropolous et al., *Nucl. Acid Res.*, 18:7468 (1990)) (information about the genetic markers can be found at the internet site of the Genome Database, http://gdbwww.gdb.org/), as well as the markers GTC_HBM_Marker_1, GTC_HBM_Marker_2, GTC_HBM_Marker_3, GTC_HBM_Marker_4, GTC_HBM_Marker_5, GTC_HBM_Marker_6, and GTC_HBM_Marker_7.

As shown in Fig. 1, haplotype analysis with the above genetic markers identifies recombination events (crossovers) in individuals 9019 and 9020 that significantly refine the interval of chromosome 11 to which the *LRP5* gene is localized. Individual 9019 is an HBM-affected individual that inherits a portion of chromosome 11 from the maternal chromosome with the *HBM* gene, and a portion from the chromosome 11 homologue. The portion inherited from the *HBM* gene-

carrying chromosome includes markers D11S935, D11S1313, GTC_HBM_Marker_4, D11S987, D11S1296, GTC_HBM_Marker_6, GTC_HBM_Marker_2, D11S970, GTC_HBM_Marker_3, D11S4113, GTC_HBM_Marker_1, GTC_HBM_Marker_7 and GTC_HBM_Marker_5. The portion from D11S4136 and continuing in the telomeric direction is derived from the non-HBM chromosome. This data places the *LRP5* gene in a location centromeric to the marker GTC_HBM_Marker_5. Individual 9020 is an unaffected individual who also exhibits a critical recombination event. This individual inherits a recombinant paternal chromosome 11 that includes markers D11S935, D11S1313, GTC_HBM_Marker_4, D11S987, D11S1296 and GTC_HBM_Marker_6 from her father's (individual 0115) chromosome 11 homologue that carries the *HBM* gene, and markers GTC_HBM_Marker_2, D11S970, GTC_HBM_Marker_3, GTC_HBM_Marker_1, GTC_HBM_Marker_7, GTC_HBM_Marker_5, D11S4136, D11S4139, D11S1314, and D11S937 from her father's chromosome 11 that does not carry the *HBM* gene. Marker D11S4113 is uninformative due to its homozygous nature in individual 0115. This recombination event places the centromeric boundary of the HBM region between markers D11S1296 and D11S987.

Two-point linkage analysis was also used to confirm the location of the *LRP5* gene on chromosome 11. The linkage results for two point linkage analysis under a model of full penetrance are presented in Table 1 below. This table lists the genetic markers in the first column and the recombination fractions across the top of the table. Each cell of the column shows the LOD score for an individual marker tested for linkage to the *LRP5* gene at the recombination fraction shown in the first row. For example, the peak LOD score of 7.66 occurs at marker D11S970, which is within the interval defined by haplotype analysis.

TABLE 1

Marker	0.0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
D11S935	- infinity	0.39	0.49	0.47	0.41	0.33	0.25	0.17	0.10
D11S1313	- infinity	2.64	2.86	2.80	2.59	2.30	1.93	1.49	1.00
D11S987	- infinity	5.49	5.18	4.70	4.13	3.49	2.79	2.03	1.26
D11S4113	4.35	3.99	3.62	3.24	2.83	2.40	1.94	1.46	0.97
D11S1337	2.29	2.06	1.81	1.55	1.27	0.99	0.70	0.42	0.18
D11S970	7.66	6.99	6.29	5.56	4.79	3.99	3.15	2.30	1.44
D11S4136	6.34	5.79	5.22	4.61	3.98	3.30	2.59	1.85	1.11
D11S4139	6.80	6.28	5.73	5.13	4.50	3.84	3.13	2.38	1.59
FGF3	0.59	3.23	3.15	2.91	2.61	2.25	1.84	1.40	0.92
D11S1314	6.96	6.49	5.94	5.34	4.69	4.01	3.27	2.49	1.67
D11S937	-infinity	4.98	4.86	4.52	4.06	3.51	2.88	2.20	1.47

A single nucleotide polymorphism (SNP) further defines the HBM region. This SNP is termed SNP_Contig033-6 and is located 25 kb centromeric to the genetic marker GTC_HBM_Marker_5. This SNP is telomeric to the genetic marker GTC_HBM_Marker_7. SNP_Contig033-6 is present in HBM-affected individual 0113. However, the HBM-affected individual 9019, who is the son of 0113, does not carry this SNP. Therefore, this indicates that the crossover is centromeric to this SNP. The primer sequence for the genetic markers GTC_HBM_Marker_5 and GTC_HBM_Marker_7 is shown in Table 2 below.

TABLE 2

Marker	Primer (Forward)	Primer (Reverse)
GTC_HBM_ Marker_5	TTTTGGGTACACAATTCAGT CG (SEQ ID NO: 63)	AAAACTGTGGGTGCTTCTGG (SEQ ID NO: 65)
GTC_HBM_ Marker_7	GTGATTGAGCCAATCCTGA GA (SEQ ID NO: 64)	TGAGCCAAATAAACCCCTTCT (SEQ ID NO:66)

The kindred described have several features of great interest, notably that their bones, while very dense, have an absolutely normal shape. The outer

dimensions of the skeletons of the HBM-affected individuals are normal, and, while medullary cavities are present, there is no interference with hematopoiesis. The HBM-affected members seem to be resistant to fracture, and there are no neurologic symptoms, and no symptoms of impairment of any organ or system function in the members examined. HBM-affected members of the kindred live to advanced age without undue illness or disability. Furthermore, the HBM phenotype matches no other bone disorders such as osteoporosis, osteoporosis pseudoglioma, Engelmann's disease, Ribbing's disease, hyperphosphatasemia, Van Buchem's disease, melorheostosis, osteopetrosis, pycnodysostosis, sclerostenosis, osteopoikilosis, acromegaly, Paget's disease, fibrous dysplasia, tubular stenosis, osteogenesis imperfecta, hypoparathyroidism, pseudohypoparathyroidism, pseudopseudohypoparathyroidism, primary and secondary hyperparathyroidism and associated syndromes, hypercalciuria, medullary carcinoma of the thyroid gland, osteomalacia and other diseases. Clearly, the HBM locus in this family has a very powerful and substantial role in regulating bone density, and its identification is an important step in understanding the pathway(s) that regulate bone density and the pathogenesis of diseases such as osteoporosis.

In addition, older individuals carrying the *HBM* gene, and therefore expression of the HBM protein, do not show loss of bone mass characteristic of normal individuals. In other words, the *HBM* gene is a suppressor of osteoporosis. In essence, individuals carrying the *HBM* gene are dosed with the HBM protein, and, as a result, do not develop osteoporosis. This *in vivo* observation is strong evidence that treatment of normal individuals with the *HBM* gene or protein, or a fragment thereof, will ameliorate osteoporosis.

V. Physical Mapping

To provide reagents for the cloning and characterization of the HBM locus, the genetic mapping data described above were used to construct a physical map of the region containing *LRP5* on chromosome 11q13.3. The physical map consists of

an ordered set of molecular landmarks, and a set of BAC clones that contain the *LRP5* gene region from chromosome 11q13.3.

Various publicly available mapping resources were utilized to identify existing STS markers (Olson et al., *Science*, 245:1434-1435 (1989)) in the HBM region. Resources included the GDB, the Whitehead Institute Genome Center, dbSTS and dbEST (NCBI), 11db, the University of Texas Southwestern GESTEC, the Stanford Human Genome Center, and several literature references (Courseaux et al., *Genomics*, 40:13-23 (1997), Courseaux et al., *Genomics*, 37:354-365 (1996), Guru et al., *Genomics*, 42:436-445 (1997), Hosoda et al., *Genes Cells*, 2:345-357 (1997), James et al., *Nat. Genet.*, 8:70-76 (1994), Kitamura et al., *DNA Research*, 4:281-289 (1997), Lemmens et al., *Genomics*, 44:94-100 (1997), Smith et al., *Genome Res.*, 7:835-842 (1997)). Maps were integrated manually to identify markers mapping to the region containing LRP5.

Primers for existing STSs were obtained from the GDB or literature references are listed in Table 3 below. Thus, Table 3 shows the STS markers used to prepare the physical map of the *LRP5* gene region.

TABLE 3: HBM STS Table

STS Name	Locus Nam	Type	GDB Access.	Size (kb	SEQ ID NO: Forward Primer	SEQ ID NO: Reverse Primer	Gene Name
ACTN3		Gene	GDB:197568	0.164	67: CTGGACTACGTGGCCTTCTC	68: TTCAGAAGCACTTGGCTGG	Actinin, alpha 3 - skeletal muscle
PC-B/PC-Y		Gene	GDB:197884	0.125	69: CTCAGTGCCATGAAGATGGA	70: CAAGATCACTCGATCTCCAGG	*Pyruvate Carboxylase*
	D11S2161E	Gene		0.322	71: GTTTCAGGAGACTCAGAGTC	72: TTCTGCAGGTTGCTGTTGAG	Adenosine Receptor (A2) Gene
ADRBK1		Gene	GDB:4590179	0.117	73: TTATTGTGATTTCCCGTGGC	74: GCCCTCTGTCCTGACTTCAGG	Beta-adrenergic receptor kinase
PSANK3		GENE		0.259	75: GAGAAAGAAATAAGGGGACC	76: TGCTTTGTAAAGCACTGAGA	sim. to Human endogenous retrovirus mRNA long terminal repeat
PP1(1/2)/PP1(2/2)		Gene	GDB:197566	0.208	77:GAAGTACGGGCAGTTCAGTGGCCT	78: ATACACCAAGGTCCATGTTCCCCGT	Protein phosphatase 1, catalytic subunit, alpha isoform
GSTP1.PCR1		*Gene*	*GDB:270066*	0.19	79:AGCCTGGGCCACAGCGTGAGACTACGT	80: TCCCGGAGCTTGCACACCCGCTTCACA	Glutathione S-transferase pi
NDUFV1		Gene		0.521	81:CATGTGCCCACCTCATTCAT	82: CAAGATTCTGTAGCTTCTGG	NADH dehydrogenase (ubiquinone) flavoprotein 1 (51kD)
PSANK2		GENE		0.157	83: CAGAGAAGTCAAGGGACTTG	84: ATCCTCTCACATCCCACACT	Aldehyde Dehydrogenase 8 (ALDH8)
PSANK1		EST		0.3	85: CAAGGCTAAAAGACGAAAAA	86: TCAGGAGCATTTCATCTTTT	Human ribosomal protein L37 (PSANK1) pseudogene.
UT5620	D11S1917	MSAT	GDB:314521	0.211	87: AAGTCGAGGCTGCAAGGAG	88: GCCCTGTGTTCCTTTCAGTA	
AFM289ya9	D11S1337	MSAT	GDB:199805	0.287	89: AAGGTGTGAGGATCACTGG	90: AGCTCATGGGGGCTATT	
GALN		Gene		0.322	91: GCTTCTCCGAGTGTATCAAC	92: ATGGCAGAGGACTTAGAACA	*Preprogalanin (GAL1)*
pMS51	D11S97	VNTR	GDB:177850		93: GATCAGCGAACTTCCTCTCGGCTC	94: TCCACATTGAGGACTGTGGGAACG	
BCL1(1)/BCL1(2)		Gene		0.205	95: GCTAATCACAGTCTAACCGA	96: TTGCACTGTCTTGGATGCA	B-cell CLL/lymphoma 1 - Cyclin D1 (PRAD1 gene)
CCND1		Gene	GDB:4590141	0.248	97: GCACAGCTGTAGTGGGGTTCTAGGC	98: CAGGCGCAAAGGACATGCACACGGC	Cyclin D1
FGF4		Gene	GDB:4590113	0.549	99: CACCGATGAGTGCACGTTCAAGGAG	100: CAGACAGAGATGCTCCACGCCATAC	Fibroblast growth factor 4
FGF3.PCR1		Gene	GDB:188627	0.161	101: TTTCTGGGTGTGTCTGAAT	102: ACACAGTTGCTCTAAAGGGT	Fibroblast growth factor 3
AFM164ZF12	D11S913	MSAT	GDB:188151	0.22	103: CATTTGGGAAATCCAGAAGA	104: TAGGTGTCTTATTTTTTGTTGCTTC	
AFMA190YD5		MSAT	GDB:1222329	0.275	105: GACATACCATGAACACTATAAGAGG	106: CAACCCATACCAGGGATAAG	
SHGC-15295	D11S4689	STS	GDB:740600	0.147	107: GAACAAGAGGGGTAAGTTGGC	108: TGAGGACACAGATACTGATGGG	
SHGC-3084	D11S4540	STS	GDB:740102	0.167	109: GAAGTGTTCCCTCTTAAATTCTTTG	110: GAACTATATTGTAGTTAGTGAGGAG	
SHGC-14407	D11S4664	STS	GDB:740516	0.158	111: CCTGTAACCCCCAGTCCC	112: TCTTGCTTCCTAAGTTTCTCGG	
SHGC-10946	D11S4327	Gene	GDB:674522	0.311	113: ACTCCATCCACCTCATCACTG	114: TGCTGTTTGCCTCATCTGAC	*Choline Kinase*
S515	D11S703	STS	GDB:196290	0.166	115: GTGGACAGGCATAGCTGAGG	116: TGTTCACTCTTCTGCCTGCAG	
AFM147XD10	D11S1889	MSAT	GDB:307895	0.183	117: AGCTGGACTCTCACAGAATG	118: CAAGAGGCTGGTAGAAGGTG	
AFMA131YE5	D11S987	MSAT	GDB:195002	0.082	119: GACTCCAGTCTGGGCAATAAAAGC	120: GGTGGCAGCATGACCTCTAAAG	
AFMb358xa9	D11S4178	MSAT	GDB:611922	0.237	121: CAGGCCCAGTCTCTTG	122: CGTGTCCAGATGAAAGTG	
AFMa272yb5	*D11S4113*	MSAT	GDB:608115	0.218	123: ACCTCACGGTGTAATCCC	124: CTTGAAGCCCATCTTTGC	
WI-17803		EST	GDB:4581644	0.15	125: TATTTGCAAAGCTTGAGACTTCT	126: AATCACTGTGCTTTGTTGCC	
SGC31923		EST	GDB:4578606	0.126	127: ACTTTATTGTCAGCGTGGGC	128: ACTCCCTCGATGGCTTCC	
WI-7741	D11S4364	GENE	GDB:677652	0.324	129: GAGCAGGGGAGAGAAGGC	130: CCCAACTGGCTTGTTTTATTG	Transformation-sensistive protein IEF SSP 3521
SGC35223		EST	GDB:4582598	0.13	131: AGCCACTTTATTGTTATTTTGATGC	132: AAGAGTGAACAAAAGCAAACATACC	ZNF162 - splicing factor 1
WI-16754		EST	GDB:4578377	0.15	133: GTGGAGTGTGGGATTGGG	134: TACTGTTCTTGATAAGTATGTCGGC	
WI-6315	D11S4418	EST	GDB:678804	0.224	135: ATGCTTTTGCATGATTCTAATTATT	136: TCCCCCAAAAGAATGTAAAGG	
WI-16915		EST	GDB:4584055	0.125	137: CTGGTCTTCCTTGTGTGCTG	138: ATCACCCAGGCCAGGGAT	Mitogen inducible gene (MIG-2)
SGC30608		EST		0.128	139:TCAGAAGCAGAACTGTTTTTAACA	140: CCTGCTTGAAAGTTCTAGAGCC	
WI-17663		EST	GDB:4583346	0.126	141: CAAGCCGGGTTTTATTGAAA	142: GATGCCAGGACCATGGAC	
WI-6383		Gene	GDB:1222237	0.199	143: GCATATAGAAACAATTTATTGCCG	144: CTCTGAAGCAGGGACCAGAG	Human tat interactive protein (TIP60)
SGC31567		Gene	GDB:4578432	0.207	145: CTACCACACCACACCAGGC	146: CAAGCGAAAGCTGCCTTC	Calcium activated neutral protease large subunit, muCANP, calpain
SGC30658		EST	GDB:4584037	0.15	147: GTTGTCTTGACTTCAGGTCTGTC	148: TTTTCCTTCAACAATCACTACTCC	
SGC34590		EST		0.13	149: GCGTGGGGATATAGAGGTCA	150: TACGTGGCCAAGAAGCTAGG	
SGC33927		EST	GDB:4582382	0.15	151: TAATATATCCCCAGTCTAAGGCAT	152: AGCTTGCAGATGGAGCCC	
WI-8671		EST	GDB:1222235	0.124	153: TGGTTTTAAACCTTTAATGAGAAAA	154: TGTTGATCTATACCCTGTTTCCG	
WI-12334		EST	GDB:1222257	0.127	155: AATTATTTAAAAGAGAGGAAAGGCA	156: TGGCTGTGAACTTCCTCTGA	
WI-18402		EST	GDB:4581874	0.113	157: GGTTACAGAAAAACATTTGAGAGAT	158: TGAGCTTTAGTTCCCTTCTCTG	
WI-18671		EST	GDB:4584947	0.131	159: TTGAAAAACCATTTATTTCACCG	160: TCTGCGGCTGTTGGATTT	Hlark
WI-12856		EST	GDB:4576606	0.209	161: TTGAAAAACCATTTATTTCACCG	162: TGTTCTCTTCTCCCAGCAGG	Hlark
SGC33767		EST	GDB:4581106	0.15	163: CTTTATTGAAAACATTGAGTGCA	164: TTGTCAAATTCCCCCCAAAA	
AFM343YB5		MSAT	GDB:1222332	0.181	165: AAACCACGACCNCCAA	166: CCCTGGAAAGGTAAGATGCT	
SGC33744		EST	GDB:4575826	0.15	167: CTTTTGGTAGAGACAAGGTCTCA	168: TATCTGTCTGTAGTGCTTCAAATGT	
SGC32272		EST	GDB:4581592	0.135	169: GACGAAGGTGATTCAGGGC	170: ACTGAAGAACTCTTGTCCT	
SGC34148		EST	GDB:4583084	0.1	171: CAGATAAAAGAGTCACTATGGCTCA	172: CACTTCTCCCACTTTGTCCC	
WI-18546		EST	GDB:4574598	0.133	173: TTATTGATAAGCATTAGTGAACCCC	174: TGGCAAGTTAGGCACAGTCA	Human 1.1 kb mRNA upregulated in retinoic acid treated HL-60 neutrophilic cells
SGC31103		EST	GDB:4567265	0.1	175: CTATGCCCAGAGATGAACAGG	176: TCCACTAAGGGCTATGTCGC	
SGC30028		Gene	GDB:4580505	0.128	177: GCCAGCTTTATTGAGTAAACTTCC	178: CACTGGAGACTACAAGTGGTGG	Human pyruvate carboxylase precursor
WI-2875	D11S4407	STS	GDB:678546	0.125	179: CATCCCAACCATCACTCAGT	180: GGGGACTAGCTTACAGATTTGA	
SGC36985		Gene	GDB:4577182	0.223	181: AGACTACATTTTGGAACCAGTGG	182: TGAAAGGATATTTATAGCCTGGA	LAR-interacting protein 1b
GCT16B07	D11S4270	STS	GDB:626245	0.137	183: GAAGGTTTTGTCCCTCGATC	184: TGAGGGTTGGGAAGATCATA	

TABLE 3: HBM STS Table

STS Name	Locus Name	Type	GDB Access.	Size (kb)	SEQ ID NO: Forward Primer	SEQ ID NO: Reverse Primer	Gene Name
WI-6504	D11S3974	EST	GDB:588142	0.174	185: CCTTCATAGCCACACCCG	186: CAGCTAACTGTTGACATGCCA	
SGC31049		EST	GDB:4580093	0.15	187: TCTTTACTGTGCTTACAACTTTCCT	188: CAACAGTGCAGTCGGTATCG	
TIGR-A002J17		EST	GDB:1222193	0.199	189: AGATCAGCAAGCAGATAG	190: CATTCCACATGGATAGAC	NDUFV1
WI-5996	D11S2382	EST	GDB:458683	0.1	191: CATACCTATGAGGTGTGCTACAGG	192: GCATTTTCTCATCATCCTTGC	amplaxin (EMS1)
WI-16987		EST	GDB:4575848	0.15	193: TTACAGCCACCAAGGTTTCC	194: AGGTGTGTGTGCCAGGTTGA	Nuclear mitotic apparatus protein 1, NUMA
SGC31912		EST	GDB:4567868	0.101	195: CACTGTTATCTCATTAACTGTGAGG	196: TTTGATTTTGTGTCTCCCAAA	
WI-13500		EST	GDB:4577893	0.15	197: CCCCACTCCCACTTTTATTT	198: CCAGTCACCTTTACTAGTCCTTTG	
CHLC.GAAT1B01.P7933	D11S971	MSAT	GDB:684255	0.103	199: AGGACACAGCCTGCATCTAG	200: ACCAGGCATTGCACTAAAAG	
SGC35519		Gene	GDB:4577180	0.134	201: GATGGGTCACACTAACCTGTCA	202: ACATTTATATTTGGACATGCAACC	LAR-interacting protein 1a mRNA
WI-11974		EST	GDB:1222255	0.108	203: AGCATCTTTAATGTGTCAGGCA	204: ATGTGCTGGGCTGGAAAG	Carnitine palmitoyl transferase I
WI-15244		Gene	GDB:4574740	0.108	205: TCACATTCAAAAATCGGCAA	206: CTGCCTGTGTGGTGTCGC	Beta-adrenergic receptor kinase 1, ADRB1
WI-17496		EST	GDB:4583336	0.131	207: TGTTTTATTTCTCAGTACAAAGCCA	208: GACCTCCTGTGACACCACG	
WI-9159	D11S4381	EST	GDB:678144	0.111	209: CCACCAAATTATTTATAGTTCTGCG	210: GTAAGATTCTCCACTGTTGCACC	FGF4
WI-4232		STS	GDB:1222250	0.175	211: CCTATAATGGGCTGGACCAA	212: ACTCCTCATGTGAAGTCACCG	
SHGC-4167		EST	GDB:4566789	0.161	213: CAGTGTGCACGTTTTCATTT	214: CAGCATCTTCAGCACTTACC	Human DNA helicase gen (SMBP2)
WI-14303		EST	GDB:4576938	0.15	215: CTGCATTTATTATGAGAATCAACAG	216: TGCTGCTGGGAGTCAGAGTC	
WI-16597		EST	GDB:4585666	0.13	217: CAGGGCACTGAGATACACTTACC	218: AAGGATCAAGCAGGCATTTG	
RC29S1CATTFOR/RC29S1CATT	D11S970	MSAT	GDB:191084	0.15	219: ACACATCTCTTCTGTGCCCC	220: TGAACCCTGGAGGCAGAG	
UT979	D11S1296	MSAT	GDB:198525	0.362	221: CATTCCCCAGTTTGCAGAC	222: GTGCTGGGATTACAGGTGT	
1281/1282	D11S1959E	EST	GDB:335216	0.07	223: GCAGAGAAGTCCTGTTAGCC	224: CCATGCTAGAGAAGCACAAC	
D11S468	D11S468	STS		0.096	225: AGTGTGGGGCAGGACCTCTG	226: CAGACAGATAGCCCTGGGTTC	
D11S668	D11S668	STS	GDB:179349	0.143	227: TCCCTCATCCCCTTGTCTGT	228: AGCCCCCCTGGGGATAATC	
RH18048		Gene	GDB:4572853	0.188	229: GATGCTTACCTACCACGGC	230: AGGATTCCTATCTGGGCTATG	Aldehyde dehydrogenase (ALDH8)
IGHMBP2		Gene	GDB:4590087	0.699	231: TGGCAGACCATGCTCCGCCT	232: GAGAAGGCCGGGAGGCTCTG	Human DNA helicase gen (SMBP2)
NUMA		Gene	GDB:4590244	0.277	233: CTCCATCACAACCAGATTTGAGGCT	234: GGGTGTGAGCTGCTGCTGAAGG	Nuclear mitotic apparatus protein 1, NUMA
KRN1		Gene	GDB:4590232	0.228	235: AGTGGGAAACCTCAGGTAGCTCCCGA	236: CAGTTTGGCTCAGACATATGGGGGCA	High sulphur keratin, KRN
Cda1ff06	D11S2302E	EST	GDB:445887	0.091	237: CATTAAGTAGTGGGGGGACAG	238: CAAAGCGACAGTGAGTTAGGG	
RH10753		Gene	GDB:4563588	0.194	239: GGAGTAGACCATGATTACTG	240: CATGGTCTATTTATTCTCG	protein phosphatase 2A , PP2A
EMS1		Gene	GDB:459016	0.64	241: CGCCCTGGATCCTCACACTACA	242: GGGCATCAGGGGATGGGTAGA	Amplaxin
SHGC-11098	DXS9736	Gene	GDB:737674	0.137	243: GCTCCTATCTGTGTTTTGAATGG	244: CCGTGGCATAAGATAAGTAAACG	Androgen Receptor
INPPL1		Gene	GDB:4590093	0.382	245: CTTGGAGCGCTATGAGGAGGGC	246: ATGGCAACTGACCTTCCGTCCTG	51C protein,inositol polyphosphate phosphatase-like 1
RH18051		EST	GDB:4572859	0.195	247: TTGGAGTCACAGGGGC	248: CAGCACTATCCTTGGGG	NOF1
Cda1cc11	D11S2297E	EST	GDB:445869	0.1	249: AACAAAGCTGCTTAGCACCTG	250: GATGAGGACCAACTGGTGAC	
1249/1250	D11S1957E	EST	GDB:335210	0.247	251: TTTTCCAATAATGTGACTTC	252: CAATCCCAACCGTAACAGGC	
NDUFV1	D11S2245E	EST	GDB:445695		253: CTTGATCTCGCCCAGGAAC	254: GCTCGCTGAAGGATGAAGAC	NDUFV1
AFMb032zg5	D11S4136	MSAT	GDB:609546	0.19	255: GAATCGCTTGAACCCAG	256: CCAGGTGGTCTTAACGG	
AFMa059xg9	D11S4196	MSAT	GDB:614025	0.2	257: GAACGTTNTTCATGTAGGCGT	258: TAATGGTCGCTGTCCC	
Cda17c12	D11S2288E	EST	GDB:445842	0.158	259: AGGGAAAATGGTATGTGGGGAG	260: GCAGTGTGTGAAGGCAGG	
SHGC-1364	D11S951E	EST	GDB:456276	0.137	261: AGTGGACAAAATGAGGAAAACAGG	262: CCAACACAGTTTGCTCACATGCC	
RH17410		EST	GDB:4571587	0.126	263: TGACATCTTTGCATTATGGC	264: AGTTATCCCACCTGATACCG	
RH17414		EST	GDB:4571595	0.121	265: AGCTCTTGCTTCTCAGTCCA	266: CAAAAGTTGTTTCTGTGTTTGTTC	
RH17770		EST	GDB:4572301	0.267	267:GCCTCTCAAAGTAGTTGGAACC	268: TGTGTATCCATAGTGCAAAACAG	
SEA		EST	GDB:4590169	0.13	269: CTCAAGGCCAGGCATCACT	270: GGACTCTTCCATGCCAGTG	S13 avian erythroblastosis oncogene homolog
RH10689		EST	GDB:456346C	0.107	271: AATGATGATCTCAACTCTG	272: ACTGAAGAACTCTTGTCCT	
TIGR-A006P20		EST	GDB:4587692	0.236	273: GACATCTGTTAGTCTCATAATTC	274: GGTAACAGTGTCTTGCTT	
TIGR-A007D15		Gene	GDB:4588398	0.24	275: CTATGTACAAAACAGGAAGAG	276: ATCCTAGTTTCCTCTCCTT	Menin gene (MEN1)
TIGR-A008B14		EST	GDB:4588882	0.141	277: GTAAATGAGAAACAGACAAATGA	278: CTATTGGATGTGATATGTTATGG	
TIGR-A008K11		EST	GDB:4589094	0.203	279: AAGTAGAAACAAAATGAGGGAC	280: CCTACCCCAAGGTAACAG	
TIGR-A008P15		EST	GDB:4589662	0.182	281: ACTTCCTATAAATGGAGGTGAG	282: GAGGAGCTTCAAGAGGAA	
TIGR-A008T11		EST	GDB:4589278	0.138	283: CATACTCCTAGACTCAAGGAATC	284: GAATGATGTACATGAATTCTTTG	
TIGR-A008U48		EST	GDB:4589364	0.107	285: GTGTTGAGGAGAAAAGCACT	286: CTCCCAGTAGTCACATTCC	
TIGR-A008X45		EST	GDB:4589838	0.242	287: CAAGTTACAAATAACTTAAGCCG	288: CAAGACCCTATCTCTACAAAAAC	
SHGC-11839	D11S4611	Gene	GDB:740339	0.151	289: TTTATTAGAAGTGACTCTTGGCCC	290: GACTACCTGCCCTCAGCTTG	Folate receptor 2 (FBP2)
NIB1242	D11S4929E	EST	GDB:3888276	0.149	291: TTCTCATGTACAAAGCGGTC	292: CCACTGGCTTCTCTCTTTTT	cGMP-stimulated 3',5'-cyclic nucleotide phosphodiesterase PDE2A3 (PDE2A)
SHGC-13599	D22S1553	Gene	GDB:737558	0.147	293: CACCAGAAGGTTGGGGTG	294: ACTATTACGACATGAAGGCGG	Macrophage Migration Inhibitory factor
SHGC-11867	D11S4331	Gene	GDB:674684	0.14	295: CTCATGCTGGATGACCCC	296: TTGCCTTTCTTGAAACTTAATTCC	P2U Purinoceptor
SHGC-15349	D12S2124	EST	GDB:740819	0.141	297: TCACAGCCTTCAGTCAGGG	298: ACATGCTGTGGCACCATG	
Bda84a05	D11S2235E	EST	GDB:445662	0.095	299: CCTGAGCTACTGCCACAG	300: CCCTGACTTGGACAGTGTCC	
Bda99d07	D11S2238E	EST	GDB:445674	0.09	301: TCAGAGTCACTCCTGCCC	302: CAAATTCAAGCTCATCCAGACC	

TABLE 3: HBM STS Table

STS Name	Locus Name	Type	GDB Access.	Size (kb)	SEQ ID NO: Forward Primer	SEQ ID NO: Reverse Primer	Gene Name
		Gene	GDB:197840	0.3	303: CGGCATTTCATCCAGGAC	304: GGTGTAGGAGGTGCGACAAT	Folate receptor2 (FBP2)
	D11S4284	EST	GDB:626260	0.173	305: TTCCATTTATTGAGCACCTG	306: CTTAAGCCACTGTGTTTTGG	
	D11S4433	Gene	GDB:679143	0.324	307: CCTCCTACACCTGCAAAAGC	308: TGGAAGAACCCCAGAGGAC	Folate receptor3 (FBP3)
		EST	GDB:4578507	0.132	309: AAAGCACAAAAGTAACAGCAACA	310: GTGTGTGGGCCACAATATTG	
		EST	GDB:4575806	0.15	311: AGAGCACCTTTCCTCAGCAC	312: AGAATCTCATCACAGGGGCG	
		EST	GDB:4577492	0.141	313: AAAAAGGACAGTGTCTAAAATTTGA	314: AATTGTTTTTGTTTGTTTTTTGAGT	
		EST	GDB:4567830	0.105	315: GATTTAGGGAGTACAAGTGCGG	316: GGGGACAAATTATACTTTATTCAGG	
		EST	GDB:4566057	0.123	317: CCATCATCATATTGGTGTGACC	318: TGGCTGCCCAAGAAGAAG	
		EST	GDB:4579290	0.15	319: TTAAGATGCCATTAAACTCATGAC	320: CCAAGGAGATGACCAAGTGG	(DRES9
		Gene	GDB:4576114	0.126	321: CCATCTCTTTTATCAGGGTTGG	322: CTCTGTGCAAGTAAGCATCTTACA	Human VEGF related factor isoform VRF186 precursor (VRF)
	D11S4057	EST	GDB:596509	0.118	323: CGACTGTGTATTTTCCACAG	324: AGAAGCCCATATCAATGCAC	
		EST	GDB:4567386	0.15	325: AGCTTAAAGTAGGACAACCATGG	326: GGATGCTTCACTCCAGAAAG	
		EST	GDB:4578600	0.127	327: TGTTGTTTATTTCCACCTGCC	328: AGAGTGGCTGCAGGCCAG	
		EST	GDB:1222208	0.15	329: TTTTTTTTTTTACACGAATTTGAGG	330: TGAGGAAGTAAAAACAGGTCATC	
		EST	GDB:4574892	0.15	331: ATGAAATCTTAAGCAGAATCCCA	332: CACAGAGTCCCAGGGTCTGT	
		EST	GDB:4584373	0.15	333: AAAGGCCTTTATTTATCTCTCTCTG	334: GCCTCAGAGCTGGTGGGT	
		EST	GDB:4578525	0.125	335: GCTTCTAAGTCTTAGAGTCAGCTGG	336: AGCCCACAGTCAGCCTACC	
		EST	GDB:4578523	0.127	337: TTGGTTAAATGATGCCCAGA	338: TGGTCCCACTCACATCCC	
		EST	GDB:4564415	0.215	339: ACACAGCATGCAGGGAGAG	340: ATCCCTGGTGCTTAGGTGG	
		EST	GDB:4564568	0.137	341: GATGGAAGTAGCTCCTCTCGG	342: GGAAGGCCAGCAAGTACTACC	
		EST	GDB:4565137	0.141	343: CCGGTGCTTGGAAAGATG	344: GAAGTGTCTCTGTTGGGGGA	
		EST	GDB:4559690	0.17	345: TTACAGGCATGAGTCACTACGC	346: ACCACTCTCACAGCCCTTACA	
		EST	GDB:4573113	0.141	347: CCCTCCCTCCACACACAC	348: GCTCACTGAACTTTCAGGGC	
		EST	GDB:4569051	0.171	349: AGATACGGGCAAAACACTGG	350: GTTGAATATAGAGCAGGGCCC	
		EST	GDB:4569063	0.166	351: TTCTGAGGTCAGGGCTGTCT	352: AGCTTGGAAAATCTCGTGTCA	
		EST	GDB:4573137	0.17	353: ACTCAGTCCCCTCCCACCC	354: TCCTCTCACTCCCTTCCCAGA	
		EST	GDB:4569999	0.19	355: GTGATCACGGCTCAACCTG	356: TGGAGGACTGCTTGAGCC	
	D11S4583	Gene	GDB:740246	0.277	357: CTGCAGCTGCCTCAGTTTC	358: TCAAAAGTGCTGGTGACAGC	Human protein kinase (MLK-3)
		Gene	GDB:1231223	0.21	359: ATTTCCAGAGCCAGCTCAAA	360: CTTTAATGTTGTGATGACACAAAGC	FGF3
		EST	GDB:4567878	0.125	361: GATCATGCACTGTTGACCAC	362: TACATTTGAAACATTTAAAACCTGA	
		Gene		0.064	363: AACTGAGCTGTAACCAGACTGGGA	364: TGGAACAGTCTGGTCCTGATGG	FK506-Binding Protein Precursor (FKBP-13)
		EST	GDB:4585099	0.202	365: TTATCCCTTTATTGTTTCTCCTTTG	366: TGGTCACCTGTATTTATTGCTAGG	
		Gene	GDB:4590064	0.859	367: TCTTCAAAGCCTCTGCAGTACC	368: CTCATCTCCAACCTGTCTAACC	4F2 CellL-Surface Antigen Heavy Chain (4F2HC)
	D11S579	STS	GDB:196276	0.108	369: GTGGCTGCAGCTAATGTAAGACAC	370: CAGCAGAGACAATGGCGTAAGTCC	
	D11S3866	STS	GDB:547681	0.135	371: CTGATTGAGAACCAGAACAG	372: TAAAGCCCTATAACCTCTCC	
	D11S3830	STS	GTC:547609	0.118	373: TAGTAAGGGACCTTCACCAG	374: AGATGTTTGGTATGACTTGG	
	D11S2439	STS	GDB:459728	0.123	375: GATGATTAAACTCTCCTGGC	376: GAGACAGCTAAGCACTCATG	
	D11S1137	STS	GDB:197824	0.196	377: GAGGTGGTGGGCACCTGTA	378: AGAGGGGAGGAACACACCTT	Folate receptor2 (FBP2)
	D11S4351	Gene	GDB:676135	0.141	379: GACCAGAGTCTGCCCAGAAG	380: TCCCCAGCTCTATCCCAAC	Collagen binding protein 2, colligin-2 gene (CBP2)
		Gene	GDB:678179	0.1	381: GGAGGGATGGACAAGTCTGA	382: GTCCAGCTCGCTGACTATCC	Retinal outer segment membrane protein 1, ROM1
		Gene		0.172	383: TCAAAACACAGTCATCTCCA	384: GCAAAGGCTTTACCATATTG	ZNF126
	D11S4087	MSAT	GDB:603797	0.158	385: GCTCAGCACCCCCATT	386: TCCCTGCTCGGGAAAC	
	D11S4162	MSAT	GDB:611241	0.263	387: GTTCTCCAGAGAGACAGCAC	388: GAGAGCAACACTATTGCCC	
	D11S4139	MSAT	GDB:609621	0.151	389: TATAGACTTCAGCCCTGCTGC	390: CCTCTGTAGGATGCAGTTGG	
	D11S1314	MSAT	GDB:199292	0.209	391: TTGCTACGCACTCCTCTACT	392: GTGAAGGCAGGAAATGTGAC	
		EST		0.13	393: ATCCTAGACCAGAGGAGCCC	394: CTCCCCCTGGTCCAGTTATT	Serine/threonine kinase
		EST		0.252	395: AACTTTCATTTGCCAAGGGA	396: AGCAGATCTGCTCTTGCGAT	
		EST		0.25	397: bACAGTTGTCATCGGTAGGCA	398: AAAAGTATGAATGGGATGGAGC	
		EST	GDB:4583084	0.142	399: GTGCAGGTGGCGTTTATTTT	400: CCCTATATCTCCGTGTGCTCC	DRES9
		EST	GDB:1223732	0.274	401: GCTCTAGTGGGAAACCTCAGG	402: GAATTCCAGGCTCTTGCTTG	Ultra high-sulphur keratin protein (fKRN1)
		EST		0.273	403: GGTTTGGTCTCAAAGGCAAA	404: CCAGTACATGGTGGTCACCA	
		EST		0.293	405: GCTGCCTTGGAATTTCTGTT	406: GTGCTGTGTGGTGGGGAAAG	
		EST		0.25	407: ATTCAAGCTCATCCAGACCC	408: GGACTGGCCCCTTTGAAACTC	
	D11S4567	STS	GDB:740192	0.225	409: ATATTGACCGTGCACAAATACG	410: AGACCTGGGAAAAAGTGGAGAA	
		STS		0.125	411: ATTGGCAGTGGAAAATGCTT	412: TTAATCTTTTTGTCAACTTCCTGATT	
		Gene		0.242	413: tctgtcctcattcacggaagc	414: ggataaagaaactccgctctgctggtaga	
		Gene	GDB:6262613		415: TCAGGGCCTGTGTTGCCGCACTCTG	416: AGCGATGTAAAGGGTACCAGTGCCGAGG	
		STS			417: AGGCATGCAAGCTTCTTA	418: CCGGGAGGAGACATCTAT	
		STS			419: TGGTAAGCACAGAAAATGC	420: AATGGATGGGGGATTTATT	

TABLE 3: HBM STS Table

STS Name	Locus Name	Type	GDB Access.	Size (kb)	SEQ ID NO: Forward Primer	SEQ ID NO: Reverse Primer	Gene Name
B235G10-HR		STS			421: CTGGACGTTATGTCTGCC	422: AGAGGCCCAGTCACAGAT	
B247F23-HR		STS			423: ATCACTCTGAACTGCCACT	424: CCCTTCTGTTTTTCTGTTTT	
B337H24-HL		STS			425: CAAGCTTTGAAGGAAGAG	426: TAGGACGTTAAGTGAGGAC	
B337L5-HL		STS			427: GCTCTGCAGTGGGTAAAA	428: ACTCTCCAAGACTGTGCG	
B382N10-HR		STS			429: CCCTTTCTGAGGCAAGAT	430: GACCACCTGGGAGAGAAC	
B12I1-HR		STS			431: CGCTATGAGTCCCATCTG	432: GATCAGCTGCAATGAAGG	
B180D17-HR		STS			433: TTGAGTACACGGGGTGAC	434: CGCAGGACTGAAAGATGA	
B236E6-HR		STS			435: ACCTGTCTCCTCTCCTGG	436: TGCTTTTCTTCTGTGGGA	
B278E22-HR		STS			437: ATGACCAGCAAGCATTGT	438: GTACTGGGATTACAGGCG	
B312F21-HR		STS			439: GCAGAAGGTCCTTTGGAT	440: TTTGCAGGATTCATGCTT	
B337H24-HR		STS			441: CGACATTCTTTTCTGGAGG	442: ACCTTTGCATGTTGGTTTT	
B358H9-HR		STS			443: GCACTTTTCCTTCCTTCC	444: TGCTTTGCTTTCTTCTGG	
B148N18-HL		STS			445: ACAGCTCCAGAGAGAAGGA	446: GCAGTCACTTGAAACCAGA	
B172N12-HL		STS			447: AGGCATCAAGCTTTCCTT	448: GGTTTAGAGAACCGAGCC	
B172N12-HR		STS			449: GTGGTGCTGCAAGTTACC	450: GGAATCCCTTTCTTTCCA	
B215J11-HR		STS			451: GACCATTTGTTACGCAGC	452: GATGGGTGTGAATGAACAA	
B223E5-HR		STS			453: CTCAAGCTTCTGTTCATGC	454: GCTGTGAGTGTCTTGGCT	
B312B3-HR		STS			455: TACAGAAAACCGCAGCTC	456: GCCACCAAAGGAAAGATT	
B328G19-HL		STS			457: AAAGGAGGGAATCATGG	458: TCACTTAGCAGGAGGGCAG	
B328G19-HR		STS			459: CTGAGCATCCGATGAGAC	460: GTGCAAAATGAGCAGCTT	
B329I10-HL		STS			461: TCTAACCCCTTACTGGGC	462: TCCTCAAACTGGGAATGA	
B329I10-HR		STS			463: TTTACACAGGACCAGGGA	464: ATCTCCCCCACTCAGAAG	
B368G19-HL		STS			465: GTCCACGGGCTTTATTCT	466: TGAGCATAAATTTCATTAGCTG	
B368G19-HR		STS			467: GGAAGAGCAAAATAAATCCA	468: GGTGCACAGAATTGTTCAT	
B36F16-HL		STS			469: AGCACGCTTATTTCATGG	470: GTAACACCAGCAGGGACA	
B250K11-HR		STS			471: TCCTGCTGCATTATGGAT	472: GGGGGTGAGAAGTAGGAA	
B338D17-HR		STS			473: ATGGGGATTAAATACGGG	474: AGCTAGCATTGGGCTCTT	
B268I23-HL		STS			475: CTGAGGAGAAGAGGCTGG	476: CGCCTTACAAGGCAAGTA	
B268I23-HR		STS			477: AGGATGCTTGCTAGGGTT	478: CACAAGTGTCTGGAAGGC	
B371E15-HR		STS			479: GGTCTCAGGAGCCCTTTA	480: ACATGCCACTCTTCTCACTAA	
B312F21-HL		STS			481: ACTTAACCAAGGATGGGG	482: CAACCCACGAGCATAAGA	
B338D17-HL		STS			483: TAGGCTCTGCACTCTTGG	484: ACCCACGGAGTCTCTCTC	
B369H19-HL		STS			485: TAAAGGCGGTGAAGTGAG	486: CTACCGCTCTCCTAGGCT	
B369H19-HR		STS			487: TGGGCCAGATAATTCTT	488: CTGGTGTTTGGTGGTGTT	
B444M11-HR		STS			489: AAGGAAGAGGTCACCAGG	490: CACAAATTCCATTTCCCA	
B269L23-HL		STS			491: TCAATAGGTGATCCAACATTT	492: AAAGTCCCACAAAGGGTC	
B250K11-HL		STS			493: GGGTAGGGGGATCTTTTT	494: TGTGGAACATTCATTGGC	
B269L23-HR		STS			495: GTCCTGGGAAAGATGGAA	496: TCAAAGCGTCTCCCATAA	
B364H4-HL		STS			497: TCTTTCGCTGTACTTGGC	498: TGGGAGGTCAGAGTGATG	
B364H4-HR		STS			499: GGACAGTGTATGTGTTGGG	500: AGGCAGCTGTTTTTTGTGA	
B473O3-HR		STS			501: CTTCTTGAGTCCCGTGTG	502: CAACCGAGAATCCTCTAGC	
B180D17-HL		STS			503: GCTGGGAGAGAATCACAA	504: GCTTTGCAGAAGAGACCA	
B200E21-HL		STS			505: ACGCTGTCAGGTCACACT	506: GGAGGATGCTCAGGTGAT	
B200E21-HR		STS			507: TAGGGGGATCTTTTTCCA	508: GAGCAATTTGAAAAGCCA	
B14L15-HR		STS			509: ATGGTCCAGCTCCTCTGT	510: ATAGAGCACCCCATCTCC	
B442P6-HR		STS			511: AACATTGCTGTTAGCCCA	512: GCAATCGAAACAGCATTC	
B188N21-HR		STS			513: ATGAGTTGGCAGCTGAAG	514: AATGAAGGTCTTGCCTCC	
GTC-ARRB1		Gene		0.067	515: GAGGAGAAGATCCACAAGCG	516: TCTCTGGGGCATACTGAACC	Beta-arrestin-1
B508A5-HL		STS			517: CTGAGCTTTTGGCACTGT	518: CTGCTAGGTGACAGCAGG	
B36F16-HR		STS			519: TGTATGAGTCTGGAGGGTGT	520: ACACCTGGCTGAGGAAAT	
B117N18-HL		STS			521: GCAGGGGACGTGATAATA	522: TTTTGCTTCCTACCATGC	
B14L15-HL		STS			523: AAAATTGTGAGCACCTCC	524: TTTATATTTAAAGTGGCTTTGTT	
B21K22-HL		STS			525: GTGCAAAGCCCACAGTAT	526: AGGAAAATGCAAGAGCAG	
B21K22-HR		STS			527: CCACTGAATTGCATACTTTTG	528: TCTGGGTCCAGTCTGCTA	
B223E5-HL		STS			529: AGATTTTGGGGAGTCAGG	530: GCGCTCAAGCAATTCTC	
B278E22-HL		STS			531: CAAGCCCCAAAGTAGTCA	532: GAATCATCCAATCCACGA	
B444M11-HL		STS			533: AGCCTCCAGGTGACTACC	534: GAAGGACATGGTCAGCAG	
B543O19-HR		STS			535: ATGCTTTCAGCATTTTCG	536: TGATCCGTGGTAGGGTTA	
B117N18-HR		STS			537: GTCGGATTGGTTTCACAA	538: TTTTATGGGAATTTCAGCC	

TABLE 3: HBM STS Table

STS Name	Locus Nam	Type	GDB Access.	Size (kb	SEQ ID NO: Forward Primer	SEQ ID NO: Reverse Primer	Gene Name
B543O19-HL		STS			539: TTTGGAAAAGAACAGAAATGT	540: GGCTAGTCTTTCCTGAACC	
B442P6-HL		STS			541: CCTTAATGCCCCTGATTC	542: GCGTTTACAAGCTGAGGA	
B367H4-HR		STS			543: TCAAGCTTGCTTTCTCAA	544: GTAGCCCAGCAAGTGTCT	
B250E21-HR		STS			545: CCTGGCTGGAGATAGGAT	546: CTTCCCCTCTGCCTATGT	
B250E21-HL		STS			547: GGCACGTACTTCCTACCA	548: GGTGCTTCTTACAGGCAA	
B248C16-HR		STS			549: ACCCAGGCTGGTGTGT	550: ACTGAGTTAATTATCACTCCCCT	
B248C16-HL		STS			551: GATGCATTTTGCTTCACC	552: TCTGCTTTTAGAGCTGTTAGC	
B160D8-HR		STS			553: TCAAGCTTCAAAGAGCAGA	554: GGAGTACATCCCAGGACC	
B539L7-HR		STS			555: TGGTGCTTTTAAATCCAGA	556: CTCCCTTACTTACTTGCATTG	
B473O3-HL		STS			557: TCTTCTCCCAGGGAATCT	558: TTTATGTCCCCTGAGCAC	
AFMa190xd9	D11S4095	STS	GDB:606064	0.193	559: TCCCTGGCTATCTTGAATC	560: CTTGACTGGGTCCACG	
ARRB1(2)		STS			561: CGAGACGCCAGTAGATACCA	562: CATCCTCCATGCCTTTCAGT	
ARRB1(1)		STS			563: AGTTCCAGAGAACGAGACGC	564: CTTGTCATCCTCCATGCCTT	
P102F3S		STS	GDB:6054145		565: GAGCGTGAGAGGTTGAGGAG	566: AAACAAACTCCAGACGCACC	
N172A		STS	GDB:6054146	0.208	567: CTGAACCACTACCTGTATGACCTG	568: CTAACTACTTACTCCTACAGGGCCC	
N60A		STS	GDB:6054147	0.23	569: GAAGCATTTCAATACTTTAACTG	570: CCACTCCAGTGCACCCAATC	
cCl11-44A		STS	GDB:6054148	0.239	571: CTTCTCCTGGCCACTCTGAC	572: GGTTTACCTTTGAATCCCAGC	
CN1677-2A		STS	GDB:6054149	0.271	573: TGAGGATGAATGAGCACATAGG	574: TTTGTGGTCCATTGAGTAGGC	
cCl11-524B		STS	GDB:6054150	0.221	575: AGGGGAAGGAATGTGCTTGG	576: TTCGGCTGAGCGGGCAGTGT	
P117F3T		STS	GDB:6054151	0.168	577: ATTGAAGGTCCTCCAAAAGAATGCTGCAGC	578: AGAACGTCAACATATCTTTTTTGGGGACAC	
ARRB1(3)		Gene			579: TTGTATTTGAGGACTTTGCTCG	580: CGGTACCATCCTCCTCTTCC	
B215J11-HL		STS		0.122	581: TTTTTGCCTCATCTATGCCC	582: GGGTGACAGAGCAAGACTCC	
B317G1-HR		STS			583: TTGCTCAAGTTCTCCTGG	584: ACCTTGTTTTTGAGGGGAG	
B317G1-HL		STS			585: CTTGGCTATTTGGACAGC	586: GGGCATTTACTCACTTGC	
B292J18-HR		STS			587: CTTGTGTCAGTTGTCAGGG	588: TGGAATTGTTGTGTCTTGG	
B10A18-HL		STS			589: CCAGTTCCACTGGATGTT	590: ATGGGCTGTGTTTCTCAA	
B10A18-HR		STS			591: CTGCCTATCCCCTGGACTT	592: AGTTTGTCCCTAGTGCCC	
B527D12-HL		STS			593: CAACACGTCTGACATCCAT	594: GGATAGTGCACACCCA	
B372J11-HR		STS			595: TGGGTGGTACTATTGTTCCCAT	596: AGTTCCAGCCCCCTTACCAG	
B372J11-HL		STS			597: GGCCACTATCATCCCTGTGT	598: TTTCACATGGGAAGAACACG	
B37E17-HR(GS)		STS			599: ACAGTGACACTAGGGACGGG	600: TGCCAGGATGGAGATAACAA	
B37E17-HL(GS)		STS			601: CCTGTGGCACACATATCACC	602: ACAACCAAGAATGGAGCCAC	
B34F22-HR(GS)		STS			603: TGCTGTGTAACAAGTCCCCA	604: TGAACGGAGGACCTACCAAG	
B34F22-HL(GS)		STS			605: GCAGGGTCCGACTCACTAAG	606: GCTGTGAGTTCCCTTTACGC	
B648P22-HR1		STS			607: ACAGTGGGGACAAAGACAGG	608: TACAGGGCACCTCCCAGTAG	
B82A4-HR2		STS			609: TCTTCTGTTAAGGTTTCCCCCC	610: TGTCTCAAACCTCCCTCTGC	
B648P22-HL		STS			611: AACATATTTCCTCCCCAGCC	612: CAGTCCCAGCCAATGAGAAC	
B82L11-HL (GS)		STS			613: CTCCTCTGCATGGGAGAATC	614: AGACCTGGGACCAGTCTGTG	
B86J13-HL (GS)		STS			615: GGGAGACGACGTCACAAGAT	616: TGATGTTGGGAAGATGGTGA	
144A24-HL		STS			617: CAGGCATCTTCTATGTGCCA	618: GGGAGGCACAAGTTCTTTCA	
B82L11-HR (GS)		STS			619: ACTTCGTGGCACTGAGTGTG	620: CCTTTCTTACGGATGAGGCA	
B86J13-HR (GS)		STS			621: GGCTGCTGAGCTCTTCTGAT	622: TGGGTCTCTCTGCCTGACTT	
B82L11-HL2(GS)		STS			623: TCACCTACTTCCAGCTTCCG	624: AGACCTGGGACCAGTCTGTG	
B82L11-HL3(GS)		STS			625: CTCCTCTGCATGGGAGAATC	626: AATTCAGGAGACCTGGGACC	

Novel STSs were developed either from publicly available genomic sequence or from sequence-derived BAC insert ends. Primers were chosen using a script which automatically performs vector and repetitive sequence masking using Cross_match (P. Green, U. of Washington) and subsequent primer picking using Primer3 (Rozen, Skaletsky (1996, 1997). Primer3 is available at www.genome.wi.mit.edu/genome_software/other/primer3.html.

Polymerase chain reaction (PCR) conditions for each primer pair were initially optimized with respect to $MgCl_2$ concentration. The standard buffer was 10 mM Tris-HCl (pH 8.3), 50 mM KCl, $MgCl_2$, 0.2 mM each dNTP, 0.2 μM each primer, 2.7 ng/μl human DNA, 0.25 units of AmpliTaq (Perkin Elmer) and $MgCl_2$ concentrations of 1.0 mM, 1.5 mM, 2.0 mM or 2.4 mM. Cycling conditions included an initial denaturation at 94°C for 2 minutes followed by 40 cycles at 94°C for 15 seconds, 55°C for 25 seconds, and 72°C for 25 seconds followed by a final extension at 72°C for 3 minutes. Depending on the results from the initial round of optimization the conditions were further optimized if necessary. Variables included increasing the annealing temperature to 58°C or 60°C, increasing the cycle number to 42 and the annealing and extension times to 30 seconds, and using AmpliTaqGold (Perkin Elmer).

BAC clones (Kim et al., *Genomics*, 32:213-218 (1996), Shizuya et al., *Proc. Natl. Acad. Sci. USA*, 89:8794-8797 (1992)) containing STS markers of interest were obtained by PCR-based screening of DNA pools from a total human BAC library purchased from Research Genetics. DNA pools derived from library plates 1-596 were used corresponding to nine genomic equivalents of human DNA. The initial screening process involved PCR reactions of individual markers against superpools, i.e., a mixture of DNA derived from all BAC clones from eight 384-well library plates. For each positive superpool, plate (8), row (16) and column (24) pools were screened to identify a unique library address. PCR products were electrophoresed in 2% agarose gels (Sigma) containing 0.5 μg/ml ethidium bromide in 1X TBE at 150 volts for 45 min. The electrophoresis units used were the Model

A3-1 systems from Owl Scientific Products. Typically, gels contained 10 tiers of lanes with 50 wells/tier. Molecular weight markers (100 bp ladder, Life Technologies, Bethesda, MD) were loaded at both ends of the gel. Images of the gels were captured with a Kodak DC40 CCD camera and processed with Kodak 1D software. The gel data were exported as tab delimited text files; names of the files included information about the library screened, the gel image files and the marker screened. These data were automatically imported using a customized Perl script into Filemaker™ PRO (Claris Corp.) databases for data storage and analysis. In cases where incomplete or ambiguous clone address information was obtained, additional experiments were performed to recover a unique, complete library address.

Recovery of clonal BAC cultures from the library involved streaking out a sample from the library well onto LB agar (Maniatis et al., *Molecular Cloning: A Laboratory Manual.*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982)) containing 12.5 μg/ml chloramphenicol (Sigma). Two individual colonies and a portion of the initial streak quadrant were tested with appropriate STS markers by colony PCR for verification. Positive clones were stored in LB broth containing 12.5 μg/ml chloramphenicol and 15% glycerol at -70°C.

Several different types of DNA preparation methods were used for isolation of BAC DNA. The manual alkaline lysis miniprep protocol listed below (Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982)) was successfully used for most applications, i.e., restriction mapping, CHEF gel analysis, FISH mapping, but was not successfully reproducible in endsequencing. The Autogen and Qiagen protocols were used specifically for BAC DNA preparation for endsequencing purposes.

Bacteria were grown in 15 ml Terrific Broth containing 12.5 μg/ml chloramphenicol in a 50 ml conical tube at 37°C for 20 hrs with shaking at 300 rpm. The cultures were centrifuged in a Sorvall RT 6000 D at 3000 rpm (~1800 g) at 4°C for 15 min. The supernatant was then aspirated as completely as

possible. In some cases cell pellets were frozen at -20°C at this step for up to 2 weeks. The pellet was then vortexed to homogenize the cells and minimize clumping. 250 μl of P1 solution (50 mM glucose, 15 mM Tris-HCl, pH 8, 10 mM EDTA, and 100 μg/ml RNase A) was added and the mixture pipetted up and down to mix. The mixture was then transferred to a 2 ml Eppendorf tube. 350 μl of P2 solution (0.2 N NaOH, 1% SDS) was then added, the mixture mixed gently and incubated for 5 min. at room temperature. 350 μl of P3 solution (3 M KOAc, pH 5.5) was added and the mixture mixed gently until a white precipitate formed. The solution was incubated on ice for 5 min. and then centrifuged at 4°C in a microfuge for 10 min. The supernatant was transferred carefully (avoiding the white precipitate) to a fresh 2 ml Eppendorf tube, and 0.9 ml of isopropanol was added, the solution mixed and left on ice for 5 min. The samples were centrifuged for 10 min., and the supernatant removed carefully. Pellets were washed in 70% ethanol and air dried for 5 min. Pellets were resuspended in 200 μl of TE8 (10 mM Tris-HCl, pH 8.0, 1.0 mM EDTA), and RNase A (Boehringer Mannheim) added to 100 μg/ml. Samples were incubated at 37°C for 30 min., then precipitated by addition of $C_2H_3O_2Na \cdot 3H_2O$ to 0.5 M and 2 volumes of ethanol. Samples were centrifuged for 10 min., and the pellets washed with 70% ethanol followed by air drying and dissolving in 50 μl TE8. Typical yields for this DNA prep were 3-5 μg/15 ml bacterial culture. Ten to 15 μl were used for HindIII restriction analysis; 5 μl was used for NotI digestion and clone insert sizing by CHEF gel electrophoresis.

BACs were inoculated into 15 ml of 2X LB Broth containing 12.5 μg/ml chloramphenicol in a 50 ml conical tube. 4 tubes were inoculated for each clone. Cultures were grown overnight (~16 hr) at 37°C with vigorous shaking (>300 rpm). Standard conditions for BAC DNA isolation were followed as recommended by the Autogen 740 manufacturer. 3 ml samples of culture were placed into Autogen tubes for a total of 60 ml or 20 tubes per clone. Samples were dissolved finally in 100 μl TE8 with 15 seconds of shaking as part of the Autogen protocol. After the Autogen protocol was finished DNA solutions were transferred from each

individual tube and pooled into a 2 ml Eppendorf tube. Tubes with large amounts of debris (carry over from the pelleting debris step) were avoided. The tubes were then rinsed with 0.5 ml of TE8 successively and this solution added to the pooled material. DNA solutions were stored at 4°C; clumping tended to occur upon freezing at -20°C. This DNA was either used directly for restriction mapping, CHEF gel analysis or FISH mapping or was further purified as described below for use in endsequencing reactions.

The volume of DNA solutions was adjusted to 2 ml with TE8, samples were then mixed gently and heated at 65°C for 10 min. The DNA solutions were then centrifuged at 4°C for 5 min. and the supernatants transferred to a 15 ml conical tube. The NaCl concentration was then adjusted to 0.75 M (~0.3 ml of 5 M NaCl to the 2 ml sample). The total volume was then adjusted to 6 ml with Qiagen column equilibration buffer (Buffer QBT). The supernatant containing the DNA was then applied to the column and allowed to enter by gravity flow. Columns were washed twice with 10 ml of Qiagen Buffer QC. Bound DNA was then eluted with four separate 1 ml aliquots of Buffer QF kept at 65°C. DNA was precipitated with 0.7 volumes of isopropanol (~2.8 ml). Each sample was then transferred to 4 individual 2.2 ml Eppendorf tubes and incubated at room temperature for 2 hr or overnight. Samples were centrifuged in a microfuge for 10 min. at 4°C. The supernatant was removed carefully and 1 ml of 70% ethanol was added. Samples were centrifuged again and because the DNA pellets were often loose at this stage, the supernatant removed carefully. Samples were centrifuged again to concentrate remaining liquid which was removed with a micropipet tip. DNA pellets were then dried in a desiccator for 10 min. 20 μl of sterile distilled and deionized H_2O was added to each tube which was then placed at 4°C overnight. The four 20 μl samples for each clone were pooled and the tubes rinsed with another 20 μl of sterile distilled and deionized H_2O for a final volume of 100 μl. Samples were then heated at 65°C for 5 min. and then mixed gently. Typical yields were 2-5 μg/60 ml culture as assessed by NotI digestion and comparison with uncut lambda DNA.

3 ml of LB Broth containing 12.5 μg/ml of chloramphenicol was dispensed into autoclaved Autogen tubes. A single tube was used for each clone. For inoculation, glycerol stocks were removed from -70°C storage and placed on dry ice. A small portion of the glycerol stock was removed from the original tube with a sterile toothpick and transferred into the Autogen tube; the toothpick was left in the Autogen tube for at least two minutes before discarding. After inoculation the tubes were covered with tape making sure the seal was tight. When all samples were inoculated, the tube units were transferred into an Autogen rack holder and placed into a rotary shaker at 37°C for 16-17 hours at 250 rpm. Following growth, standard conditions for BAC DNA preparation, as defined by the manufacturer, were used to program the Autogen. Samples were not dissolved in TE8 as part of the program and DNA pellets were left dry. When the program was complete, the tubes were removed from the output tray and 30 μl of sterile distilled and deionized H_2O was added directly to the bottom of the tube. The tubes were then gently shaken for 2-5 seconds and then covered with parafilm and incubated at room temperature for 1-3 hours. DNA samples were then transferred to an Eppendorf tube and used either directly for sequencing or stored at 4°C for later use.

VI. BAC Clone Characterization for Physical Mapping

DNA samples prepared either by manual alkaline lysis or the Autogen protocol were digested with HindIII for analysis of restriction fragment sizes. This data were used to compare the extent of overlap among clones. Typically 1-2 μg were used for each reaction. Reaction mixtures included: 1X Buffer 2 (New England Biolabs), 0.1 mg/ml bovine serum albumin (New England Biolabs), 50 μg/ml RNase A (Boehringer Mannheim), and 20 units of HindIII (New England Biolabs) in a final volume of 25 μl. Digestions were incubated at 37°C for 4-6 hours. BAC DNA was also digested with NotI for estimation of insert size by CHEF gel analysis (see below). Reaction conditions were identical to those for

HindIII except that 20 units of NotI were used. Six μl of 6X Ficoll loading buffer containing bromphenol blue and xylene cyanol was added prior to electrophoresis.

HindIII digests were analyzed on 0.6% agarose (Seakem, FMC Bioproducts) in 1X TBE containing 0.5 μg/ml ethidium bromide. Gels (20 cm X 25 cm) were electrophoresed in a Model A4 electrophoresis unit (Owl Scientific) at 50 volts for 20-24 hrs. Molecular weight size markers included undigested lambda DNA, HindIII digested lambda DNA, and HaeIII digested _X174 DNA. Molecular weight markers were heated at 65°C for 2 min. prior to loading the gel. Images were captured with a Kodak DC40 CCD camera and analyzed with Kodak 1D software.

NotI digests were analyzed on a CHEF DRII (BioRad) electrophoresis unit according to the manufacturer's recommendations. Briefly, 1% agarose gels (BioRad pulsed field grade) were prepared in 0.5X TBE, equilibrated for 30 minutes in the electrophoresis unit at 14°C, and electrophoresed at 6 volts/cm for 14 hrs with circulation. Switching times were ramped from 10 sec to 20 sec. Gels were stained after electrophoresis in 0.5 μg/ml ethidium bromide. Molecular weight markers included undigested lambda DNA, HindIII digested lambda DNA, lambda ladder PFG ladder, and low range PFG marker (all from New England Biolabs).

BAC DNA prepared either by the manual alkaline lysis or Autogen protocols were labeled for FISH analysis using a Bioprime labeling kit (BioRad) according to the manufacturer's recommendation with minor modifications. Approximately 200 ng of DNA was used for each 50 μl reaction. 3 μl were analyzed on a 2% agarose gel to determine the extent of labeling. Reactions were purified using a Sephadex G50 spin column prior to *in situ* hybridization.

Metaphase FISH was performed as described (Ma et al., *Cytogenet. Cell Genet.*, 74:266-271 (1996)).

VII. BAC Endsequencing

The sequencing of BAC insert ends utilized DNA prepared by either of the two methods described above. The DYEnamic energy transfer primers and Dynamic Direct cycle sequencing kits from Amersham were used for sequencing reactions. Ready made sequencing mix including the M13 -40 forward sequencing primer was used (Catalog # US79730) for the T7 BAC vector terminus; ready made sequencing mix (Catalog # US79530) was mixed with the M13 -28 reverse sequencing primer (Catalog # US79339) for the SP6 BAC vector terminus. The sequencing reaction mixes included one of the four fluorescently labeled dye-primers, one of the four dideoxy termination mixes, dNTPs, reaction buffer, and Thermosequenase. For each BAC DNA sample, 3 μl of the BAC DNA sample was aliquoted to 4 PCR strip tubes. 2 μl of one of the four dye primer/termination mix combinations was then added to each of the four tubes. The tubes were then sealed and centrifuged briefly prior to PCR. Thermocycling conditions involved a 1 minute denaturation at 95°C, 15 second annealing at 45°C, and extension for 1 minute at 70°C for 35 total cycles. After cycling the plates were centrifuged briefly to collect all the liquid to the bottom of the tubes. 5 μl of sterile distilled and deionized H_2O was then added into each tube, the plates sealed and centrifuged briefly again. The four samples for each BAC were then pooled together. DNA was then precipitated by adding 1.5 μl of 7.5 M NH_4OAc and 100 μl of -20°C 100% ethanol to each tube. Samples were mixed by pipetting up and down once. The plates were then sealed and incubated on ice for 10 minutes. Plates were centrifuged in a table top Haraeus centrifuge at 4000 rpm (3,290 g) for 30 minutes at 4°C to recover the DNA. The supernatant was removed and excess liquid blotted onto paper towels. Pellets were washed by adding 100 μl of -20°C 70% ethanol into each tube and re-centrifuging at 4000 rpm (3,290 g) for 10 minutes at 4°C. The supernatant was removed and excess liquid again removed by blotting on a paper towel. Remaining traces of liquid were removed by placing the plates upside down over a paper towel and centrifuging only until the centrifuge reached

800 rpm. Samples were then air dried at room temperature for 30 min. Tubes were capped and stored dry at -20°C until electrophoresis. Immediately prior to electrophoresis the DNA was dissolved in 1.5 μl of Amersham loading dye. Plates were then sealed and centrifuged at 2000 rpm (825 g). The plates were then vortexed on a plate shaker for 1-2 minutes. Samples were then recentrifuged at 2000 rpm (825 g) briefly. Samples were then heated at 65°C for 2 min. and immediately placed on ice. Standard gel electrophoresis was performed on ABI 377 fluorescent sequencers according to the manufacturer's recommendation.

VIII. Sub-cloning and Sequencing of HBM BAC DNA

The physical map of the *LRP5* gene region provides a set of BAC clones that contain within them the *LRP5* gene and the *HBM* gene. DNA sequencing of several of the BACs from the region has been completed. The DNA sequence data is a unique reagent that includes data that one skilled in the art can use to identify the *LRP5* gene and the *HBM* gene, or to prepare probes to identify the gene(s), or to identify DNA sequence polymorphisms that identify the gene(s).

BAC DNA was isolated according to one of two protocols, either a Qiagen purification of BAC DNA (Qiagen, Inc. as described in the product literature) or a manual purification which is a modification of the standard alkaline lysis/Cesium Chloride preparation of plasmid DNA (see e.g., Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons (1997)). Briefly for the manual protocol, cells were pelleted, resuspended in GTE (50 mM glucose, 25 mM Tris-Cl (pH 8), 10 mM EDTA) and lysozyme (50 mg/ml solution), followed by NaOH/SDS (1% SDS/0.2 N NaOH) and then an ice-cold solution of 3 M KOAc (pH 4.5-4.8). RnaseA was added to the filtered supernatant, followed by Proteinase K and 20% SDS. The DNA was then precipitated with isopropanol, dried and resuspended in TE (10 mM Tris, 1 mM EDTA (pH 8.0)). The BAC DNA was further purified by Cesium Chloride density gradient centrifugation (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons (1997)).

Following isolation, the BAC DNA was sheared hydrodynamically using an HPLC (Hengen, *Trends in Biochem. Sci.*, 22:273-274 (1997)) to an insert size of 2000-3000 bp. After shearing, the DNA was concentrated and separated on a standard 1% agarose gel. A single fraction, corresponding to the approximate size, was excised from the gel and purified by electroelution (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring, NY (1989)).

The purified DNA fragments were then blunt-ended using T4 DNA polymerase. The blunt-ended DNA was then ligated to unique BstXI-linker adapters (SEQ ID NOS: 627-628) (5' GTCTTCACCACGGGG and 5' GTGGTGAAGAC in 100-1000 fold molar excess). These linkers were complimentary to the BstXI-cut pMPX vectors (constructed by the inventors), while the overhang was not self-complimentary. Therefore, the linkers would not concatemerize nor would the cut-vector religate itself easily. The linker-adapted inserts were separated from the unincorporated linkers on a 1% agarose gel and purified using GeneClean (BIO 101, Inc.). The linker-adapted insert was then ligated to a modified pBlueScript vector to construct a "shotgun" subclone library. The vector contained an out-of-frame lacZ gene at the cloning site which became in-frame in the event that an adapter-dimer is cloned, allowing these to be avoided by their blue-color.

All subsequent steps were based on sequencing by ABI377 automated DNA sequencing methods. Only major modifications to the protocols are highlighted. Briefly, the library was then transformed into DH5α competent cells (Life Technologies, Bethesda, MD, DH5α transformation protocol). It was assessed by plating onto antibiotic plates containing ampicillin and IPTG/Xgal. The plates were incubated overnight at 37°C. Successful transformants were then used for plating of clones and picking for sequencing. The cultures were grown overnight at 37°. DNA was purified using a silica bead DNA preparation (Ng et al., *Nucl. Acids*

Res., 24:5045-5047 (1996)) method. In this manner, 25 μg of DNA was obtained per clone.

These purified DNA samples were then sequenced using ABI dye-terminator chemistry. The ABI dye terminator sequence reads were run on ABI377 machines and the data was directly transferred to UNIX machines following lane tracking of the gels. All reads were assembled using PHRAP (P. Green, Abstracts of DOE Human Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157) with default parameters and quality scores. The initial assembly was done at 6-fold coverage and yielded an average of 8-15 contigs. Following the initial assembly, missing mates (sequences from clones that only gave one strand reads) were identified and sequenced with ABI technology to allow the identification of additional overlapping contigs. Primers for walking were selected using a Genome Therapeutics program Pick_primer near the ends of the clones to facilitate gap closure. These walks were sequenced using the selected clones and primers. Data were reassembled with PHRAP into sequence contigs.

IX. Gene Identification by Computational Methods

Following assembly of the BAC sequences into contigs, the contigs were subjected to computational analyses to identify coding regions and regions bearing DNA sequence similarity to known genes. This protocol included the following steps.

1. Degap the contigs: the sequence contigs often contain symbols (denoted by a period symbol) that represent locations where the individual ABI sequence reads have insertions or deletions. Prior to automated computational analysis of the contigs, the periods were removed. The original data was maintained for future reference.

2. BAC vector sequences were "masked" within the sequence by using the program cross match (Phil Green, http:\\chimera.biotech.washington.edu\UWGC). Since the shotgun libraries

construction detailed above leaves some BAC vector in the shotgun libraries, this program was used to compare the sequence of the BAC contigs to the BAC vector and to mask any vector sequence prior to subsequent steps. Masked sequences were marked by an "X" in the sequence files, and remained inert during subsequent analyses.

3. *E. coli* sequences contaminating the BAC sequences were masked by comparing the BAC contigs to the entire *E. coli* DNA sequence.

4. Repetitive elements known to be common in the human genome were masked using cross match. In this implementation of crossmatch, the BAC sequence was compared to a database of human repetitive elements (Jerzy Jerka, Genetic Information Research Institute, Palo Alto, CA). The masked repeats were marked by X and remained inert during subsequent analyses.

5. The location of exons within the sequence was predicted using the MZEF computer program (Zhang, *Proc. Natl. Acad. Sci.*, 94:565-568 (1997)).

6. The sequence was compared to the publicly available unigene database (National Center for Biotechnology Information, National Library of Medicine, 38A, 8N905, 8600 Rockville Pike, Bethesda, MD 20894; www.ncbi.nlm.nih.gov) using the blastn2 algorithm (Altschul et al., *Nucl. Acids Res.*, 25:3389-3402 (1997)). The parameters for this search were: E=0.05, v=50, B=50 (where E is the expected probability score cutoff, V is the number of database entries returned in the reporting of the results, and B is the number of sequence alignments returned in the reporting of the results (Altschul et al., *J. Mol. Biol.*, 215:403-410 (1990)).

7. The sequence was translated into protein for all six reading frames, and the protein sequences were compared to a non-redundant protein database compiled from Genpept Swissprot PIR (National Center for Biotechnology Information, National Library of Medicine, 38A, 8N905, 8600 Rockville Pike, Bethesda, MD 20894; www.ncbi.nlm.nih.gov). The parameters for this search were E=0.05, V=50, B= 50, where E, V, and B are defined as above.

8. The BAC DNA sequence was compared to the database of the cDNA clones derived from direct selection experiments (described below) using blastn2 (Altschul et al., *Nucl. Acids. Res.*, 25:3389-3402 (1997)). The parameters for this search were E=0.05, V=250, B=250, where E, V, and B are defined as above.

9. The BAC sequence was compared to the sequences of all other BACs from the HBM region on chromosome 11q12-13 using blastn2 (Altschul et al., *Nucl. Acids. Res.*, 25:3389-3402 (1997)). The parameters for this search were E=0.05, V=50, B=50, where E, V, and B are defined as above.

10. The BAC sequence was compared to the sequences derived from the ends of BACs from the HBM region on chromosome 11q12-13 using blastn2 (Altschul et al., *Nucl. Acids. Res.*, 25:3389-3402 (1997)). The parameters for this search were E=0.05, V=50, B=50, where E, V, and B are defined as above.

11. The BAC sequence was compared to the Genbank database (National Center for Biotechnology Information, National Library of Medicine, 38A, 8N905, 8600 Rockville Pike, Bethesda, MD 20894; www.ncbi.nlm.nih.gov) using blastn2 (Altschul et al., *Nucl. Acids. Res.*, 25:3389-3402 (1997)). The parameters for this search were E=0.05, V=50, B=50, where E, V, and B are defined as above.

12. The BAC sequence was compared to the STS division of Genbank database (National Center for Biotechnology Information, National Library of Medicine, 38A, 8N905, 8600 Rockville Pike, Bethesda, MD 20894; www.ncbi.nlm.nih.gov) using blastn2 (Altschul et al., 1997). The parameters for this search were E=0.05, V=50, B= 50, where E, V, and B are defined as above.

13. The BAC sequence was compared to the Expressed Sequence (EST) Tag Genbank database (National Center for Biotechnology Information, National Library of Medicine, 38A, 8N905, 8600 Rockville Pike, Bethesda, MD 20894; www.ncbi.nlm.nih.gov) using blastn2 (Altschul et al., *Nucl. Acids. Res.*, 25:3389-3402 (1997)). The parameters for this search were E=0.05, V=250, B=250, where E, V, and B are defined as above.

X. Gene Identification by Direct cDNA Selection

Primary linkered cDNA pools were prepared from bone marrow, calvarial bone, femoral bone, kidney, skeletal muscle, testis and total brain. Poly (A) + RNA was prepared from calvarial and femoral bone tissue (Chomczynski et al., *Anal. Biochem.*, 162:156-159 (1987); D'Alessio et al., *Focus*, 9:1-4 (1987)) and the remainder of the mRNA was purchased from Clontech (Palo Alto, California). In order to generate oligo(dT) and random primed cDNA pools from the same tissue, 2.5 μg mRNA was mixed with oligo(dT) primer in one reaction and 2.5 μg mRNA was mixed with random hexamers in another reaction, and both were converted to first and second strand cDNA according to manufacturers recommendations (Life Technologies, Bethesda, MD). Paired phosphorylated cDNA linkers (see sequence below) were annealed together by mixing in a 1:1 ratio (10 μg each) incubated at 65°C for five minutes and allowed to cool to room temperature.

Paired linkers oligo 1/2

OLIGO 1: 5'CTG AGC GGA ATT CGT GAG ACC3' (SEQ ID NO:12)

OLIGO 2: 5'TTG GTC TCA CGT ATT CCG CTC GA3' (SEQ ID NO:13)

Paired linkers oligo3/4

OLIGO 3: 5'CTC GAG AAT TCT GGA TCC TC3' (SEQ ID NO:14)

OLIGO 4: 5'TTG AGG ATC CAG AAT TCT CGA G3' (SEQ ID NO:15)

Paired linkers oligo5/6

OLIGO 5: 5'TGT ATG CGA ATT CGC TGC GCG3' (SEQ ID NO:16)

OLIGO 6: 5'TTC GCG CAG CGA ATT CGC ATA CA3' (SEQ ID NO:17)

Paired linkers oligo7/8

OLIGO 7: 5'GTC CAC TGA ATT CTC AGT GAG3' (SEQ ID NO:18)

OLIGO 8: 5'TTG TCA CTG AGA ATT CAG TGG AC3' (SEQ ID NO:19)

Paired linkers oligo11/12

OLIGO 11:5'GAA TCC GAA TTC CTG GTC AGC3' (SEQ ID NO:20)

OLIGO 12:5'TTG CTG ACC AGG AAT TCG GAT TC3' (SEQ ID NO:21)

Linkers were ligated to all oligo(dT) and random primed cDNA pools (see below) according to manufacturers instructions (Life Technologies, Bethesda, MD).

Oligo 1/2 was ligated to oligo(dT) and random primed cDNA pools prepared from bone marrow. Oligo 3/4 was ligated to oligo(dT) and random primed cDNA pools prepared from calvarial bone. Oligo 5/6 was ligated to oligo(dT) and random primed cDNA pools prepared from brain and skeletal muscle. Oligo 7/8 was ligated to oligo(dT) and random primed cDNA pools prepared from kidney. Oligo 11/12 was ligated to oligo(dT) and random primed cDNA pools prepared from femoral bone.

The cDNA pools were evaluated for length distribution by PCR amplification using 1 μl of a 1:1, 1:10, and 1:100 dilution of the ligation reaction, respectively. PCR reactions were performed in a Perkin Elmer 9600, each 25 μl volume reaction contained 1 μl of DNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 200 mM each dNTPs, 10 μM primer and 1 unit Taq DNA polymerase (Perkin Elmer) and was amplified under the following conditions: 30 seconds at 94°C, 30 seconds at 60°C and 2 minutes at 72°C for 30 cycles. The length distribution of the amplified cDNA pools were evaluated by electrophoresis on a 1% agarose gel. The PCR reaction that gave the best representation of the random primed and oligo(dT) primed cDNA pools was scaled up so that ~2-3 μg of each cDNA pool was produced. The starting cDNA for the

direct selection reaction comprised of 0.5 μg of random primed cDNAs mixed with 0.5 μg of oligo(dT) primed cDNAs.

The DNA from the 54 BACs that were used in the direct cDNA selection procedure was isolated using Nucleobond AX columns as described by the manufacturer (The Nest Group, Inc.).

The BACs were pooled in equimolar amounts and 1 μg of the isolated genomic DNA was labeled with biotin 16-UTP by nick translation in accordance with the manufacturers instructions (Boehringer Mannheim). The incorporation of the biotin was monitored by methods that could be practiced by one skilled in the art (Del Mastro and Lovett, *Methods in Molecular Biology*, Humana Press Inc., NJ (1996)).

Direct cDNA selection was performed using methods that could be practiced by one skilled in the art (Del Mastro and Lovett, *Methods in Molecular Biology*, Humana Press Inc., NJ (1996)). Briefly, the cDNA pools were multiplexed in two separate reactions: In one reaction cDNA pools from bone marrow, calvarial bone, brain and testis were mixed, and in the other cDNA pools from skeletal muscle, kidney and femoral bone were mixed. Suppression of the repeats, yeast sequences and plasmid in the cDNA pools was performed to a Cot of 20. 100 ng of biotinylated BAC DNA was mixed with the suppressed cDNAs and hybridized in solution to a Cot of 200. The biotinylated DNA and the cognate cDNAs was captured on streptavidin-coated paramagnetic beads. The beads were washed and the primary selected cDNAs were eluted. These cDNAs were PCR amplified and a second round of direct selection was performed. The product of the second round of direct selection is referred to as the secondary selected material. A Galanin cDNA clone, previously shown to map to 11q12-13 (Evans, *Genomics*, 18:473-477 (1993)), was used to monitor enrichment during the two rounds of selection.

The secondary selected material from bone marrow, calvarial bone, femoral bone, kidney, skeletal muscle, testis and total brain was PCR amplified using modified primers of oligos 1, 3, 5, 7 and 11, shown below, and cloned into the

UDG vector pAMP10 (Life Technologies, Bethesda, MD), in accordance with the manufacturer's recommendations.

Modified primer sequences:

Oligo1-CUA: 5'CUA CUA CUA CUA CTG AGC GGA ATT CGT GAG ACC3' (SEQ ID NO:22)

Oligo3-CUA: 5'CUA CUA CUA CUA CTC GAG AAT TCT GGA TCC TC3' (SEQ ID NO:23)

Oligo5-CUA: 5'CUA CUA CUA CUA TGT ATG CGA ATT CGC TGC GCG3' (SEQ ID NO:24)

Oligo7-CUA: 5'CUA CUA CUA CUA GTC CAC TGA ATT CTC AGT GAG3' (SEQ ID NO:25)

Oligo11-CUA: 5'CUA CUA CUA CUA GAA TCC GAA TTC CTG GTC AGC3' (SEQ ID NO:26)

The cloned secondary selected material, from each tissue source, was transformed into MAX Efficiency DH5a Competent Cells (Life Technologies, Bethesda, MD) as recommended by the manufacturer. 384 colonies were picked from each transformed source and arrayed into four 96 well microtiter plates.

All secondarily selected cDNA clones were sequenced using M13 dye primer terminator cycle sequencing kit (Applied Biosystems), and the data collected by the ABI 377 automated fluorescence sequencer (Applied Biosystems).

All sequences were analyzed using the BLASTN, BLASTX and FASTA programs (Altschul et al., *J. Mol. Biol.*, 215:403-410 (1990), Altschul et al., *Nucl. Acids. Res.*, 25:3389-3402 (1997)). The cDNA sequences were compared to a

database containing sequences derived from human repeats, mitochondrial DNA, ribosomal RNA, *E. coli* DNA to remove background clones from the dataset using the program cross_match. A further round of comparison was also performed using the program BLASTN2 against known genes (Genbank) and the BAC sequences from the HBM region. Those cDNAs that were >90% homologous to these sequences were filed according to the result and the data stored in a database for further analysis. cDNA sequences that were identified but did not have significant similarity to the BAC sequences from the HBM region or were eliminated by cross_match were hybridized to nylon membranes which contained the BACs from the HBM region, to ascertain whether they hybridized to the target.

Hybridization analysis was used to map the cDNA clones to the BAC target that selected them. The BACs that were identified from the HBM region were arrayed and grown into a 96 well microtiter plate. LB agar containing 25 μg/ml kanamycin was poured into 96 well microtiter plate lids. Once the agar had solidified, pre-cut Hybond N+ nylon membranes (Amersham) were laid on top of the agar and the BACs were stamped onto the membranes in duplicate using a hand held 96 well replica plater (V&P Scientific, Inc.). The plates were incubated overnight at 37°C. The membranes were processed according to the manufacturers recommendations.

The cDNAs that needed to be mapped by hybridization were PCR amplified using the relevant primer (oligos 1, 3, 5, 7 and 11) that would amplify that clone. For this PCR amplification, the primers were modified to contain a linkered digoxigenin molecule at the 5' of the oligonucleotide. The PCR amplification was performed under the same conditions as described in Preparation of cDNA Pools (above). The PCR products were evaluated for quality and quantity by electrophoresis on a 1% agarose gel by loading 5 μl of the PCR reaction. The nylon membranes containing the stamped BACs were individually pre-hybridized in 50 ml conical tubes containing 10 ml of hybridization solution (5x SSPE, 0.5x Blotto, 2.5% SDS and 1 mM EDTA (pH 8.0)). The 50 ml conical tubes were placed in a

rotisserie oven (Robbins Scientific) for 2 hours at 65°C. Twenty-five ng of each cDNA probe was denatured and added into individual 50 ml conical tubes containing the nylon membrane and hybridization solution. The hybridization was performed overnight at 65°C. The filters were washed for 20 minutes at 65°C in each of the following solutions: 3x SSPE, 0.1% SDS; 1x SSPE, 0.1% SDS and 0.1x SSPE, 0.1% SDS.

The membranes were removed from the 50 ml conical tubes and placed in a dish. Acetate sheets were placed between each membrane to prevent them from sticking to each other. The incubation of the membranes with the Anti-DIG-AP and CDP-Star was performed according to manufacturers recommendations (Boehringer Mannheim). The membranes were wrapped in Saran wrap and exposed to Kodak Bio-Max X-ray film for 1 hour.

XI. cDNA Cloning and Expression Analysis

To characterize the expression of the genes identified by direct cDNA selection and genomic DNA sequencing in comparison to the publicly available databases, a series of experiments were performed to further characterize the genes in the HBM region. First, oligonucleotide primers were designed for use in the polymerase chain reaction (PCR) so that portions of a cDNA, EST, or genomic DNA could be amplified from a pool of DNA molecules (a cDNA library) or RNA population (RT-PCR and RACE). The PCR primers were used in a reaction containing genomic DNA to verify that they generated a product of the size predicted based on the genomic (BAC) sequence. A number of cDNA libraries were then examined for the presence of the specific cDNA or EST. The presence of a fragment of a transcription unit in a particular cDNA library indicates a high probability that additional portions of the same transcription unit will be present as well.

A critical piece of data that is required when characterizing novel genes is the length, in nucleotides, of the processed transcript or messenger RNA (mRNA).

One skilled in the art primarily determines the length of an mRNA by Northern blot hybridization (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor NY (1989)). Groups of ESTs and direct-selected cDNA clones that displayed significant sequence similarity to sequenced BACs in the critical region were grouped for convenience into approximately 30 kilobase units. Within each 30 kilobase unit there were from one up to fifty ESTs and direct-selected cDNA clones which comprised one or more independent transcription units. One or more ESTs or direct-selected cDNAs were used as hybridization probes to determine the length of the mRNA in a variety of tissues, using commercially available reagents (Multiple Tissue Northern blot; Clontech, Palo Alto, California) under conditions recommended by the manufacturer.

Directionally cloned cDNA libraries from femoral bone, and calvarial bone tissue were constructed by methods familiar to one skilled in the art (for example, Soares in *Automated DNA Sequencing and Analysis*, Adams, Fields and Venter, Eds., Academic Press, NY, pages 110-114 (1994)). Bones were initially broken into fragments with a hammer, and the small pieces were frozen in liquid nitrogen and reduced to a powder in a tissue pulverizer (Spectrum Laboratory Products). RNA was extracted from the powdered bone by homogenizing the powdered bone with a standard Acid Guanidinium Thiocyanate-Phenol-Chloroform extraction buffer (e.g., Chomczynski and Sacchi, *Anal. Biochem.*, 162:156-159 (1987)) using a polytron homogenizer (Brinkman Instruments). Additionally, human brain and lung total RNA was purchased from Clontech. PolyA RNA was isolated from total RNA using dynabeads-dT according to the manufacturer's recommendations (Dynal, Inc.). First strand cDNA synthesis was initiated using an oligonucleotide primer with the sequence:

5'-AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT-3'

(SEQ ID NO:27). This primer introduces a *Not*I restriction site (underlined) at the 3' end of the cDNA. First and second strand synthesis were performed using the

"one-tube" cDNA synthesis kit as described by the manufacturer (Life Technologies, Bethesda, MD). Double stranded cDNAs were treated with T4 polynucleotide kinase to ensure that the ends of the molecules were blunt (Soares in *Automated DNA Sequencing and Analysis*, Adams, Fields and Venter, Eds., Academic Press, NY, pages 110-114 (1994)), and the blunt ended cDNAs were then size selected by a Biogel column (Huynh et al in *DNA Cloning*, Vol. 1, Glover, Ed., IRL Press, Oxford, pages 49-78 (1985)) or with a size-sep 400 sepharose column (Pharmacia, catalog # 27-5105-01). Only cDNAs of 400 base pairs or longer were used in subsequent steps. EcoRI adapters (sequence: 5' OH-AATTCGGCACGAG-OH 3' (SEQ ID NO:28), and 5' p-CTCGTGCCG-OH 3' (SEQ ID NO:29)) were then ligated to the double stranded cDNAs by methods familiar to one skilled in the art (Soares, 1994). The EcoRI adapters were then removed from the 3' end of the cDNA by digestion with *Not*I (Soares, 1994). The cDNA was then ligated into the plasmid vector pBluescript® II KS+ (Stratagene, La Jolla, California), and the ligated material was transformed into *E. coli* host DH10B or DH12S by electroporation methods familiar to one skilled in the art (Soares, 1994). After growth overnight at 37°C, DNA was recovered from the *E. coli* colonies after scraping the plates by processing as directed for the Mega-prep kit (Qiagen, Chatsworth, California). The quality of the cDNA libraries was estimated by counting a portion of the total numbers of primary transformants and determining the average insert size and the percentage of plasmids with no cDNA insert. Additional cDNA libraries (human total brain, heart, kidney, leukocyte, and fetal brain) were purchased from Life Technologies, Bethesda, MD.

cDNA libraries, both oligo (dT) and random hexamer (N_6) primed, were used for isolating cDNA clones transcribed within the HBM region: human bone, human brain, human kidney and human skeletal muscle (all cDNA libraries were made by the inventors, except for skeletal muscle (dT) and kidney (dT) cDNA libraries). Four 10 x 10 arrays of each of the cDNA libraries were prepared as follows: the cDNA libraries were titered to 2.5 x 10^6 using primary transformants.

The appropriate volume of frozen stock was used to inoculate 2 L of LB/ampicillin (100 mg/ml). This inoculated liquid culture was aliquotted into 400 tubes of 4 ml each. Each tube contained approximately 5000 cfu. The tubes were incubated at 30°C overnight with gentle agitation. The cultures were grown to an OD of 0.7-0.9. Frozen stocks were prepared for each of the cultures by aliquotting 100 μl of culture and 300 μl of 80% glycerol. Stocks were frozen in a dry ice/ethanol bath and stored at -70°C. The remaining culture was DNA prepared using the Qiagen (Chatsworth, CA) spin miniprep kit according to the manufacturer's instructions. The DNAs from the 400 cultures were pooled to make 80 column and row pools. The cDNA libraries were determined to contain HBM cDNA clones of interest by PCR. Markers were designed to amplify putative exons. Once a standard PCR optimization was performed and specific cDNA libraries were determined to contain cDNA clones of interest, the markers were used to screen the arrayed library. Positive addresses indicating the presence of cDNA clones were confirmed by a second PCR using the same markers.

Once a cDNA library was identified as likely to contain cDNA clones corresponding to a specific transcript of interest from the HBM region, it was manipulated to isolate the clone or clones containing cDNA inserts identical to the EST or direct-selected cDNA of interest. This was accomplished by a modification of the standard "colony screening" method (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor NY (1989)). Specifically, twenty 150 mm LB+ampicillin agar plates were spread with 20,000 colony forming units (cfu) of cDNA library and the colonies allowed to grow overnight at 37°C. Colonies were transferred to nylon filters (Hybond from Amersham, or equivalent) and duplicates prepared by pressing two filters together essentially as described (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor NY (1989)). The "master" plate was then incubated an additional 6-8 hours to allow the colonies to grow back. The DNA from the bacterial colonies was then affixed to the nylon

filters by treating the filters sequentially with denaturing solution (0.5 N NaOH, 1.5 M NaCl) for two minutes, neutralization solution (0.5 M Tris-Cl pH 8.0, 1.5 M NaCl) for two minutes (twice). The bacterial colonies were removed from the filters by washing in a solution of 2X SSC/0.1% SDS for one minute while rubbing with tissue paper. The filters were air dried and baked under vacuum at 80°C for 1-2 hours.

A cDNA hybridization probe was prepared by random hexamer labeling (Fineberg and Vogelstein, *Anal. Biochem.*, 132:6-13 (1983)) or by including gene-specific primers and no random hexamers in the reaction (for small fragments). Specific activity was calculated and was $>5X10^8$ cpm/10^8 μg of cDNA. The colony membranes were then prewashed in 10 mM Tris-Cl pH 8.0, 1 M NaCl, 1 mM EDTA, 0.1% SDS for 30 minutes at 55°C. Following the prewash, the filters were prehybridized in > 2 ml/filter of 6X SSC, 50 % deionized formamide, 2% SDS, 5X Denhardt's solution, and 100 mg/ml denatured salmon sperm DNA, at 42°C for 30 minutes. The filters were then transferred to hybridization solution (6X SSC, 2% SDS, 5X Denhardt's, 100 mg/ml denatured salmon sperm DNA) containing denatured α^{32}P-dCTP-labeled cDNA probe and incubated at 42°C for 16-18 hours.

After the 16-18 hour incubation, the filters were washed under constant agitation in 2X SSC, 2% SDS at room temperature for 20 minutes, followed by two washes at 65°C for 15 minutes each. A second wash was performed in 0.5 X SSC, 0.5% SDS for 15 minutes at 65°C. Filters were then wrapped in plastic wrap and exposed to radiographic film for several hours to overnight. After film development, individual colonies on plates were aligned with the autoradiograph so that they could be picked into a 1 ml solution of LB Broth containing ampicillin. After shaking at 37°C for 1-2 hours, aliquots of the solution were plated on 150 mm plates for secondary screening. Secondary screening was identical to primary screening (above) except that it was performed on plates containing ~250 colonies so that individual colonies could be clearly identified for picking.

After colony screening with radiolabeled probes yielded cDNA clones, the clones were characterized by restriction endonuclease cleavage, PCR, and direct sequencing to confirm the sequence identity between the original probe and the isolated clone. To obtain the full-length cDNA, the novel sequence from the end of the clone identified was used to probe the library again. This process was repeated until the length of the cDNA cloned matches that estimated to be full-length by the northern blot analysis.

RT-PCR was used as another method to isolate full length clones. The cDNA was synthesized and amplified using a "Superscript One Step RT-PCR" kit (Life Technologies, Gaithersburg, MD). The procedure involved adding 1.5 μg of RNA to the following: 25 μl of reaction mix provided which is a proprietary buffer mix with $MgSO_4$ and dNTP's, 1 μl sense primer (10 μM) and 1 μl anti-sense primer (10 μM), 1 μl reverse transcriptase and Taq DNA polymerase mix provided and autoclaved water to a total reaction mix of 50 μl. The reaction was then placed in a thermocycler for 1 cycle at 50°C for 15 to 30 minutes, then 94°C for 15 seconds, 55-60°C for 30 seconds and 68-72°C for 1 minute per kilobase of anticipated product and finally 1 cycle of 72°C for 5-10 minutes. The sample was analyzed on an agarose gel. The product was excised from the gel and purified from the gel (GeneClean, Bio 101). The purified product was cloned in pCTNR (General Contractor DNA Cloning System, 5 Prime - 3 Prime, Inc.) and sequenced to verify that the clone was specific to the gene of interest.

Rapid Amplification of cDNA ends (RACE) was performed following the manufacturer's instructions using a Marathon cDNA Amplification Kit (Clontech, Palo Alto, CA) as a method for cloning the 5' and 3' ends of candidate genes. cDNA pools were prepared from total RNA by performing first strand synthesis, where a sample of total RNA sample was mixed with a modified oligo (dT) primer, heated to 70°C, cooled on ice and followed by the addition of: 5x first strand buffer, 10 mM dNTP mix, and AMV Reverse Transcriptase (20 U/μl). The tube was incubated at 42°C for one hour and then the reaction tube was placed on ice.

For second strand synthesis, the following components were added directly to the reaction tube: 5x second strand buffer, 10 mM dNTP mix, sterile water, 20x second strand enzyme cocktail and the reaction tube was incubated at 16°C for 1.5 hours. T4 DNA Polymerase was added to the reaction tube and incubated at 16°C for 45 minutes. The second-strand synthesis was terminated with the addition of an EDTA/Glycogen mix. The sample was subjected to a phenol/chloroform extraction and an ammonium acetate precipitation. The cDNA pools were checked for quality by analyzing on an agarose gel for size distribution. Marathon cDNA adapters (Clontech) were then ligated onto the cDNA ends. The specific adapters contained priming sites that allowed for amplification of either 5' or 3' ends, depending on the orientation of the gene specific primer (GSP) that was chosen. An aliquot of the double stranded cDNA was added to the following reagents: 10 μM Marathon cDNA adapter, 5x DNA ligation buffer, T4 DNA ligase. The reaction was incubated at 16°C overnight. The reaction was heat inactivated to terminate the reaction. PCR was performed by the addition of the following to the diluted double stranded cDNA pool: 10x cDNA PCR reaction buffer, 10 μM dNTP mix, 10 μM GSP, 10 μM AP1 primer (kit), 50x Advantage cDNA Polymerase Mix. Thermal Cycling conditions were 94°C for 30 seconds, 5 cycles of 94°C for 5 seconds, 72°C for 4 minutes, 5 cycles of 94°C for 5 seconds, 70°C for 4 minutes, 23 cycles of 94°C for 5 seconds, 68°C for 4 minutes. After the first round of PCR was performed using the GSP to extend to the end of the adapter to create the adapter primer binding site, exponential amplification of the specific cDNA of interest was observed. Usually a second nested PCR is performed to confirm the specific cDNA. The RACE product was analyzed on an agarose gel and then excised and purified from the gel (GeneClean, BIO 101). The RACE product was then cloned into pCTNR (General Contractor DNA Cloning System, 5' - 3', Inc.) and the DNA sequence determined to verify that the clone is specific to the gene of interest.

XII. Mutation Analysis

Comparative genes were identified using the above procedures and the exons from each gene were subjected to mutation detection analysis. Comparative DNA sequencing was used to identify polymorphisms in HBM candidate genes from chromosome 11q12-13. DNA sequences for candidate genes were amplified from patient lymphoblastoid cell lines.

The inventors developed a method based on analysis of direct DNA sequencing of PCR products amplified from candidate regions to search for the causative polymorphism. The procedure consisted of three stages that used different subsets of HBM family to find segregating polymorphisms and a population panel to assess the frequency of the polymorphisms. The family resources result from a single founder leading to the assumption that all affected individuals will share the same causative polymorphism.

Candidate regions were first screened in a subset of the HBM family consisting of the proband, daughter, and her mother, father and brother. Monochromosomal reference sequences were produced concurrently and used for comparison. The mother and daughter carried the HBM polymorphism in this nuclear family, providing the ability to monitor polymorphism transmission. The net result is that two HBM chromosomes and six non-HBM chromosomes were screened. This allowed exclusion of numerous frequent alleles. Only alleles exclusively present in the affected individuals passed to the next level of analysis.

Polymorphisms that segregated exclusively with the HBM phenotype in this original family were then re-examined in an extended portion of the HBM pedigree consisting of two additional nuclear families. These families consisted of five HBM and three unaffected individuals. The HBM individuals in this group included the two critical crossover individuals, providing the centromeric and telomeric boundaries of the critical region. Tracking the heredity of polymorphisms between these individuals and their affected parents allowed for further refining of the

critical region. This group brought the total of HBM chromosomes screened to seven and the total of non-HBM chromosomes to seventeen.

When a given polymorphism continued to segregate exclusively with the HBM phenotype in the extended group, a population panel was then examined. This panel of 84 persons consisted of 42 individuals known to have normal bone mineral density and 42 individuals known to be unrelated but with untyped bone mineral density. For this purpose, normal bone mineral density is within two standard deviations of a BMD Z score of 0. The second group was from the widely used CEPH panel of individuals. Any segregating polymorphisms found to be rare in this population were subsequently examined on the entire HBM pedigree and a larger population.

Polymerase chain reaction (PCR) was used to generate sequencing templates from the HBM family's DNA and monochromosomal controls. Enzymatic amplification of genes within the HBM region on 11q12-13 was accomplished using the PCR with oligonucleotides flanking each exon as well as the putative 5' regulatory elements of each gene. The primers were chosen to amplify each exon as well as 15 or more base pairs within each intron on either side of the splice. All PCR primers were made as chimeras to facilitate dye primer sequencing. The M13-21F (5'- GTA A CGA CGG CCA GT -3') (SEQ ID NO:30) and -28REV (5'- AAC AGC TAT GAC CAT G -3') (SEQ ID NO:31) primer binding sites were built on to the 5' end of each forward and reverse PCR primer, respectively, during synthesis. 150 ng of genomic DNA was used in a 50 μl PCR with 2 U AmpliTaq, 500 nM primer and 125 μM dNTP. Buffer and cycling conditions were specific to each primer set. TaqStart antibody (Clontech) was used for hot start PCR to minimize primer dimer formation. 10% of the product was examined on an agarose gel. The appropriate samples were diluted 1:25 with deionized water before sequencing.

Each PCR product was sequenced according to the standard Energy Transfer primer (Amersham) protocol. All reactions took place in 96 well trays. 4 separate reactions, one each for A, C, G and T were performed for each template. Each

reaction included 2 μl of the sequencing reaction mix and 3 μl of diluted template. The plates were then heat sealed with foil tape and placed in a thermal cycler and cycled according to the manufacturer's recommendation. After cycling, the 4 reactions were pooled. 3 μl of the pooled product was transferred to a new 96 well plate and 1 μl of the manufacturer's loading dye was added to each well. All 96 well pipetting procedures occurred on a Hydra 96 pipetting station (Robbins Scientific, USA). 1 μl of pooled material was directly loaded onto a 48 lane gel running on an ABI 377 DNA sequencer for a 10 hour, 2.4 kV run.

Polyphred (University of Washington) was used to assemble sequence sets for viewing with *Consed* (University of Washington). Sequences were assembled in groups representing all relevant family members and controls for a specified target region. This was done separately for each of the three stages. Forward and reverse reads were included for each individual along with reads from the monochromosomal templates and a color annotated reference sequence. *Polyphred* indicated potential polymorphic sites with a purple flag. Two readers independently viewed each assembly and assessed the validity of the purple-flagged sites.

A total of 23 exons present in the mature mRNA and several other portions of the primary transcript were evaluated for heterozygosity in the nuclear family of two HBM-affected and two unaffected individuals. 25 SNPs were identified, as shown in the table below.

TABLE 4

Single Nucleotide Polymorphisms in the *LRP5* gene and Environs

Exon Name	Location	Base Change
b200e21-h_Contig1_1.nt	69169 (309G)	C/A
b200e21-h_Contig4_12.nt	27402 (309G)	A/G
b200e21-h_Contig4_13.nt	27841 (309G)	T/C
b200e21-h_Contig4_16.nt	35600 (309G)	A/G
b200e21-h_Contig4_21.nt	45619 (309G)	G/A
b200e21-h_Contig4_22.nt-a	46018 (309G)	T/G

Exon Name	Location	Base Change
b200e21-h_Contig4_22.nt-b	46093 (309G)	T/G
b200e21-h_Contig4_22.nt-c	46190 (309G)	A/G
b200e21-h_Contig4_24.nt-a	50993 (309G)	T/C
b200e21-h_Contig4_24.nt-b	51124 (309G)	C/T
b200e21-h_Contig4_25.nt	55461 (309G)	C/T
b200e21-h_Contig4_33.nt-a	63645 (309G)	C/A
b200e21-h_Contig4_33.nt-b	63646 (309G)	A/C
b200e21-h_Contig4_61.nt	24809 (309G)	T/G
b200e21-h_Contig4_62.nt	27837 (309G)	T/C
b200e21-h_Contig4_63.nt-a	31485 (309G)	C/T
b200e21-h_Contig4_63.nt-b	31683 (309G)	A/G
b200e21-h_Contig4_9.nt	24808 (309G)	T/G
b527d12-h_Contig030g_1.nt-a	31340 (308G)	T/C
b527d12-h_Contig030g_1.nt-b	32538 (308G)	A/G
b527d12-h_Contig080C_2.nt	13224 (308G)	A/G
b527d12-h_Contig087C_1.nt	21119 (308G)	C/A
b527d12-h_Contig087C_4.nt	30497 (308G)	G/A
b527d12-h_Contig088C_4.nt	24811 (309G)	A/C
b527d12-h_Contig089_1HP.nt	68280 (309G)	G/A

In addition to the polymorphisms presented in Table 4, two additional polymorphisms can also be present in SEQ ID NO:2. These is a change at position 2002 of SEQ ID NO:2. Either a guanine or an adenine can appear at this position. This polymorphism is silent and is not associated with any change in the amino acid sequence. The second change is at position 4059 of SEQ ID NO:2 corresponding in a cytosine (C) to thymine (T) change. This polymorphism results in a corresponding amino acid change from a valine (V) to an alanine (A). Other polymorphisms were found in the candidate gene exons and adjacent intron

sequences. Any one or combination of the polymorphisms listed in Table 4 or the two discussed above could also have a minor effect on bone mass when present in SEQ ID NO:2.

The present invention encompasses the nucleic acid sequences having the nucleic acid sequence of SEQ ID NO: 1 with the above-identified point mutations.

Preferably, the present invention encompasses the nucleic acid of SEQ ID NO: 2. Specifically, a base-pair substitution changing G to T at position 582 in the coding sequence of LRP5 (the *HBM* gene) was identified as heterozygous in all HBM individuals, and not found in the unaffected individuals (i.e., b527d12-h_Contig087C_1.nt). Fig. 5 shows the order of the contigs in B527D12. The direction of transcription for the *HBM* gene is from left to right. The sequence of contig308G of B527D12 is the reverse complement of the coding region to the *HBM* gene. Therefore, the relative polymorphism in contig 308G shown in Table 4 as a base change substitution of C to A is the complement to the G to T substitution in the *HBM* gene. This mutation causes a substitution of glycine 171 with valine (G171V).

The HBM polymorphism was confirmed by examining the DNA sequence of different groups of individuals. In all members of the HBM pedigree (38 individuals), the HBM polymorphism was observed in the heterozygous form in affected (i.e., elevated bone mass) individuals only (N=18). In unaffected relatives (N=20) (BMDZ<2.0) the HBM polymorphism was never observed. To determine whether this polymorphism was ever observed in individuals outside of the HBM pedigree, 297 phenotyped individuals were characterized at the site of the *HBM* gene. None were heterozygous at the site of the HBM polymorphism. In an unphenotyped control group, none of 64 individuals were observed to be heterozygous at position 582. Taken together, these data prove that the polymorphism observed in the kindred displaying the high bone mass phenotype is strongly correlated with the G→T polymorphism at position 582 of *LRP5*. Furthermore, these results coupled with the ASO results described below, establish

that the HBM polymorphism genetically segregates with the HBM phenotype, and that both the HBM polymorphism and phenotype are rare in the general population.

XIII. Allele Specific Oligonucleotide (ASO) Analysis

The amplicon containing the HBM polymorphism was PCR amplified using primers specific for the exon of interest. The appropriate population of individuals was PCR amplified in 96 well microtiter plates as follows. PCR reactions (20 μl) containing 1X Promega PCR buffer (Cat. # M1883 containing 1.5 mM $MgCl_2$), 100mM dNTP, 200 nM PCR primers (SEQ ID NOS: 629-630) (1863F: CCAAGTTCTGAGAAGTCC and 1864R: AATACCTGAAACCATACCTG), 1 U Amplitaq, and 20 ng of genomic DNA were prepared and amplified under the following PCR conditions: 94°C, 1 minute, (94°C, 30 sec.; 58°C, 30 sec.; 72°C, 1 min. X35 cycles), 72°C, 5' min, 4°C, hold. Loading dye was then added and 10 μl of the products was electrophoresed on 1.5% agarose gels containing 1 μg/ml ethidium bromide at 100-150 V for 5-10 minutes. Gels were treated 20 minutes in denaturing solution (1.5 M NaCl, 0.5 N NaOH), and rinsed briefly with water. Gels were then neutralized in 1 M Tris-HCl, pH 7.5, 1.5 M NaCl, for 20 minutes and rinsed with water. Gels were soaked in 10 X SSC for 20 minutes and blotted onto nylon transfer membrane (Hybond N+- Amersham) in 10X SSC overnight. Filters were the rinsed in 6X SSC for 10 minutes and UV crosslinked.

The allele specific oligonucleotides (ASO) were designed with the polymorphism approximately in the middle. Oligonucleotides were phosphate free at the 5'end and were purchased from Gibco BRL. Sequences of the oligonucleotides are (SEQ ID NOS: 631-632) :

2326 ZMAX.ASO.g: AGACTGGGGTGAGACGC

2327 ZMAX.ASO.t: CAGACTGGGTTGAGACGCC

The polymorphic nucleotides are underlined. To label the oligos, 1.5 μl of 1 μg/μl ASO oligo (2326.ZMAX.ASO.g or 2327.ZMAX.ASO.t), 11 μl ddH_2O, 2 μl 10X kinase forward buffer, 5 μl γ^{32}P-ATP (6000 Ci/mMole), and 1 μl T4

polynucleotide kinase (10 U/μl) were mixed, and the reaction incubated at 37°C for 30-60 minutes. Reactions were then placed at 95°C for 2 minutes and 30 ml H_2O was added. The probes were purified using a G25 microspin column (Pharmacia).

Blots were prehybridized in 10 ml 5X SSPE, 5X Denhardt's, 2% SDS, and 100 μg/ml, denatured, sonicated salmon sperm DNA at 40°C for 2 hr. The entire reaction mix of kinased oligo was then added to 10 ml fresh hybridization buffer (5X SSPE, 5X Denhardt's, 2% SDS) and hybridized at 40°C for at least 4 hours to overnight.

All washes done in 5X SSPE, 0.1 % SDS. The first wash was at 45°C for 15 minutes; the solution was then changed and the filters washed 50°C for 15 minutes. Filters were then exposed to Kodak biomax film with 2 intensifying screens at –70°C for 15 minutes to 1 hr. If necessary the filters were washed at 55°C for 15 minutes and exposed to film again. Filters were stripped by washing in boiling 0.1X SSC, 0.1% SDS for 10 minutes at least 3 times.

The two films that best captured the allele specific assay with the 2 ASOs were converted into digital images by scanning them into Adobe PhotoShop. These images were overlaid against each other in Graphic Converter and then scored and stored in FileMaker Pro 4.0 (see Fig. 9).

In order to determine the HBM allele frequency in ethnically diverse populations, 672 random individuals from various ethnic groups were typed by the allele specific oligonucleotide (ASO) method. This population included 96 CEPH grandparents (primarily Caucasian), 192 Caucasian, 192 African-American, 96 Hispanic, and 96 Asian individuals. No evidence was obtained for the presence of the HBM polymorphism in any of these individuals. Overall, a total of 911 individuals were typed either by direct sequencing or ASO hybridization; all were homozygous GG at the site of the HBM polymorphism (Fig. 14). This information illustrates that the HBM allele is rare in various ethnic populations.

Thus this invention provides a rapid method of identifying individuals with the HBM allele. This method could be used in the area of diagnostics and screening

of an individual for susceptibility to osteoporosis or other bone disorder. The assay could also be used to identify additional individuals with the HBM allele or the additional polymorphisms described herein.

XIV. Cellular Localization of LRP5

Gene Expression in Rat tibia by non isotopic In Situ Hybridization

In situ hybridization was conducted by Pathology Associates International (PAI), Frederick, MD. This study was undertaken to determine the specific cell types that express the *LRP5* gene in rat bone with particular emphasis on areas of bone growth and remodeling. LRP5 probes used in this study were generated from both human (HuZmax1) and mouse (MsZmax1) cDNAs, which share an 87% sequence identity. The homology of human and mouse LRP5 with rat LRP5 is unknown.

For example, gene expression by non-isotopic *in situ* hybridization was performed as follows, but other methods would be known to the skilled artisan. Tibias were collected from two 6 to 8 week old female Sprague Dawley rats euthanized by carbon dioxide asphyxiation. Distal ends were removed and proximal tibias were snap frozen in OCT embedding medium with liquid nitrogen immediately following death. Tissues were stored in a –80°C freezer.

Probes for amplifying PCR products from cDNA were prepared as follows. The primers to amplify PCR products from a cDNA clone were chosen using published sequences of both human LRP5 (Genbank Accession No. ABO17498) and mouse LRP5 (Genbank Accession No. AF064984). In order to minimize cross reactivity with other genes in the LDL receptor family, the PCR products were derived from an intracellular portion of the protein coding region. PCR was

performed in a 50 μl reaction volume using cDNA clone as template. PCR reactions contained 1.5 mM $MgCl_2$, 1 unit Amplitaq, 200 μM dNTPs and 2 μM each primer. PCR cycling conditions were 94°C for 1 min., followed by 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds; followed by a 5 minute extension at 72°C. The reactions were then run on a 1.5% agarose Tris-Acetate gel. DNA was eluted from the agarose, ethanol precipitated and resuspended in 10 mM Tris, pH 8.0. Gel purified PCR products were prepared for both mouse and human cDNAs and supplied to Pathology Associates International for *in situ* hybridizations.

The sequence of the human and mouse PCR primers and products were as follows:

Human LRP5 sense primer (SEQ ID NO:633) (HBM253)

CCCGTGTGCTCCGCCGCCCAGTTC

Human LRP5 antisense primer (SEQ ID NO:634) (HBM465)

GGCTCACGGAGCTCATCATGGACTT

Human LRP5 PCR product (SEQ ID NO:635)

CCCGTGTGCTCCGCCGCCCAGTTCCCCTGCGCGCGGGGTCAGTGTGTGGA
CCTGCGCCTGCGCTGCGACGGCGAGGCAGACTGTCAGGACCGCTCAGAC
GAGGTGGACTGTGACGCCATCTGCCTGCCCAACCAGTTCCGGTGTGCGAG
CGGCCAGTGTGTCCTCATCAAACAGCAGTGCGACTCCTTCCCCGACTGTA
TCGACGGCTCCGACGAGCTCATGTGTGAAATCACCAAGCCGCCCTCAGAC
GACAGCCCGGCCCACAGCAGTGCCATCGGGCCCGTCATTGGCATCATCCT
CTCTCTCTTCGTCATGGGTGGTGTCTATTTTGTGTGCCAGCGCGTGGTGTG
CCAGCGCTATGCGGGGGCCAACGGGCCCTTCCCGCACGAGTATGTCAGC
GGGACCCCGCACGTGCCCCTCAATTTCATAGCCCCGGGCGGTTCCCAGCA

TGGCCCCTTCACAGGCATCGCATGCGGAAAGTCCATGATGAGCTCCGTGA

GCC

Mouse LRP5 Sense primer (SEQ ID NO:636) (HBM655)

AGCGAGGCCACCATCCACAGG

Mouse LRP5 antisense primer (SEQ ID NO:637) (HBM656)

TCGCTGGTCGGCATAATCAAT

Mouse LRP5 PCR product (SEQ ID NO:638)

AGCAGAGCCACCATCCACAGGATCTCCCTGGAGACTAACAACAACGATG

TGGCTATCCCACTCACGGGTGTCAAAGAGGCCTCTGCACTGGACTTTGAT

GTGTCCAACAATCACATCTACTGGACTGATGTTAGCCTCAAGACGATCAG

CCGAGCCTTCATGAATGGGAGCTCAGTGGAGCACGTGATTGAGTTTGGCC

TCGACTACCCTGAAGGAATGGCTGTGGACTGGATGGGCAAGAACCTCTAT

TGGGCGGACACAGGGACCAACAGGATTGAGGTGGCCCGGCTGGATGGGC

AGTTCCGGCAGGTGCTTGTGTGGAGAGACCTTGACAACCCCAGGTCTCTG

GCTCTGGATCCTACTAAAGGCTACATCTACTGGACTGAGTGGGGTGGCAA

GCCAAGGATTGTGCGGGCCTTCATGGATGGGACCAATTGTATGACACTGG

TAGACAAGGTGGGCCGGGCCAACGACCTCACCATTGATTATGCCGACCA

GCGA

Riboprobes were synthesized as follows. The PCR products were reamplified with chimeric primers designed to incorporate either a T3 promoter upstream, or a T7 promoter downstream of the reamplification products. The resulting PCR products were used as template to synthesize digoxigenin-labeled riboprobes by *in vitro* transcription (IVT). Antisense and sense riboprobes were synthesized using T7 and T3 RNA polymerases, respectively, in the presence of digoxigenin-11-UTP (Boehringer-Mannheim) using a MAXIscript IVT kit (Ambion)

according to the manufacturer. The DNA was then degraded with Dnase-1, and unincorporated digoxigenin was removed by ultrafiltration. Riboprobe integrity was assessed by electrophoresis through a denaturing polyacrylamide gel. Molecular size was compared with the electrophoretic mobility of a 100–1000 base pair (bp) RNA ladder (Ambion). Probe yield and labeling was evaluated by blot immunochemistry. Riboprobes were stored in 5 μl aliquots at –80°C.

The *in situ* hybridization was performed as follows. Frozen rat bone was cut into 5 μM sections on a Jung CM3000 cryostat (Leica) and mounted on adhesive slides (Instrumedics). Sections were kept in the cryostat at –20°C until all the slides were prepared in order to prevent mRNA degradation prior to post-fixation for 15 minutes in 4% paraformaldehyde. Following post-fixation, sections were incubated with 1 ng/μl of either antisense or sense riboprobe in Pathology Associates International (PAI) customized hybridization buffer for approximately 40 hours at 58°C. Following hybridization, slides were subjected to a series of post-hybridization stringency washes to reduce nonspecific probe binding. Hybridization was visualized by immunohistochemistry with an anti-digoxigenin antibody (FAB fragment) conjugated to alkaline phosphatase. Nitroblue tetrazolium chloride/bromochloroindolyl phosphate (Boehringer-Mannheim), a precipitating alkaline phosphatase substrate, was used as the chromogen to stain hybridizing cells purple to nearly black, depending on the degree of staining. Tissue sections were counter-stained with nuclear fast red. Assay controls included omission of the probe, omission of probe and anti-digoxigenin antibody.

Specific cell types were assessed for demonstration of hybridization with antisense probes by visualizing a purple to black cytoplasmic and/or peri-nuclear staining indicating a positive hybridization signal for mRNA. Each cell type was

compared to the replicate sections, which were hybridized with the respective sense probe. Results were considered positive if staining was observed with the antisense probe and no staining or weak background with the sense probe.

The cellular localization of the hybridization signal for each of the study probes is summarized in Table 5. Hybridization for LRP5 was primarily detected in areas of bone involved in remodeling, including the endosteum and trabecular bone within the metaphysis. Hybridization in selected bone lining cells of the periosteum and epiphysis were also observed. Positive signal was also noted in chondrocytes within the growth plate, particularly in the proliferating chondrocytes. See Figs. 10, 11 and 12 for representative photomicrographs of *in situ* hybridization results.

TABLE 5

Summary of LRP5 *in situ* hybridization in rat tibia

PROBE	SITE	ISH SIGNAL
HuZmax1	Epiphysis	
	Osteoblasts	+
	Osteoclasts	-
	Growth Plate	
	resting chondrocytes	-
	proliferating chondrocytes	+
	hypertrophic chondrocytes	-
	Metaphysis	
	osteoblasts	+
	osteoclasts	+
	Diaphysis	-
	Endosteum	
	osteoblasts	+
	osteoclasts	+
	Periosteum	-
MsZmax1	Epiphysis	
	Osteoblasts	+
	Osteoclasts	-
	Growth Plate	
	resting chondrocytes	-
	proliferating chondrocytes	+
	hypertrophic chondrocytes	+
	Metaphysis	
	osteoblasts	+
	osteoclasts	+
	Diaphysis	-
	Endosteum	
	osteoblasts	+
	osteoclasts	+
	Periosteum	+

Legend: "+" = hybridization signal detected "-" = no hybridization signal detected

"ISH" – *In situ* hybridization

These studies confirm the positional expression of LRP5 in cells involved in bone remodeling and bone formation. LRP5 expression in the zone of proliferation and in the osteoblasts and osteoclasts of the proximal metaphysis, suggests that the *LRP5* gene is involved in the process of bone growth and mineralization. The activity and differentiation of osteoblasts and osteoclasts are closely coordinated during development as bone is formed and during growth as well as in adult life as bone undergoes continuous remodeling. The formation of internal bone structures and bone remodeling result from the coupling of bone resorption by activated osteoclasts with subsequent deposition of new material by osteoblasts. LRP5 is related to the LDL receptor gene, and thus may be a receptor involved in mechanosensation and subsequent signaling in the process of bone remodeling. Therefore, changes in the level of expression of this gene could impact on the rate of remodeling and degree of mineralization of bone.

XV. Antisense

Antisense oligonucleotides are short synthetic nucleic acids that contain complementary base sequences to a targeted RNA. Hybridization of the RNA in living cells with the antisense oligonucleotide interferes with RNA function and ultimately blocks protein expression. Therefore, any gene for which the partial sequence is known can be targeted by an antisense oligonucleotide.

Antisense technology is becoming a widely used research tool and will play an increasingly important role in the validation and elucidation of therapeutic targets identified by genomic sequencing efforts.

Antisense technology was developed to inhibit gene expression by utilizing an oligonucleotide complementary to the mRNA that encodes the target gene. There are several possible mechanisms for the inhibitory effects of antisense oligonucleotides. Among them, degradation of mRNA by RNase H is considered to be the major mechanism of inhibition of protein function. This technique was

originally used to elucidate the function of a target gene, but may also have therapeutic applications, provided it is designed carefully and properly.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethlaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, I-methylguanine, I-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), t-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

In addition, the use of morpholino oligonucleotides could be employed. Morpholinos are oligomers with modification of the ribose moiety to a morpholino group. This technology is covered by U.S. Patent No. 5,185,444 and is described in Summerton and Weller *Antisense Nucleic Acid Drug Dev*. 1997 Jun; 7(3):

187-95. The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an HBM or LRP5 protein or a protein which interacts with LRP5 and/or HBM to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of an antisense nucleic acid molecule of the invention includes direct injection at a tissue site. Alternatively, an antisense nucleic acid molecule can be modified to target selected cells and then administered systemically. For example, for systemic administration, an antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An μ-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual γ-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330). In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature*

334:585-591)) can be used to catalytically cleave LRP5 or HBM mRNA transcripts to thereby inhibit translation of LRP5 or HBM mRNA. A ribozyme having specificity for a LRP5- or HBM-encoding nucleic acid can be designed based upon the nucleotide sequence of a LRP5 or HBM cDNA disclosed herein (i.e., SEQ ID NO:1 or 3). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an HBM or LRP5-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071 and Cech et al. U.S. Patent No. 5,116,742 both incorporated herein by reference. Alternatively, LRP5 or HBM mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) Science 261:1411-1418. Alternatively *LRP5* or *HBM* gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the *LRP5* or *HBM* gene (e.g., the *LRP5* or *HBM* gene promoter and/or enhancers) to form triple helical structures that prevent transcription of the *LRP5* *HBM* gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15. LRP5 or HBM gene expression can also be inhibited using RNA interference (RNAi). This is a technique for post-transcriptional gene silencing (PTGS), in which target gene activity is specifically abolished with cognate double-stranded RNA (dsRNA). RNAi resembles in many aspects PTGS in plants and has been detected in many invertebrates including trypanosome, hydra, planaria, nematode and fruit fly (*Drosophila melanogaster*). It may be involved in the modulation of transposable element mobilizaiton and antiviral state formation. RNAi in mammalian systems is disclosed in PCT application WO 00/63364 which is incorporated by reference herein in its entirety. Basically, dsRNA, homologous to the target (LRP5 or HBM) is introduced into the cell and a sequence specific reduction in gene activity is observed. Both small and interfering RNAs (siRNAs) and short hairpin RNAs (shRNAs) are contemplated.

See for example Yu et al., (2002) *PNAS*, 99, 6047-6052; Paddison et al., (2002) *Genes and Development*, 16, 948-58; Brummelkamp et al., (2002) *Science* 296, 550-553 ; Tuschl, (2002) *Nature Biotechnology* 20, 446-448; and, references therein.

As an example, preparing antisense oligonucleotides can be performed as follows. Studies have been undertaken using antisense technology in the osteoblast-like murine cell line, MC3T3. These cells can be triggered to develop along the bone differentiation sequence. An initial proliferation period is characterized by minimal expression of differentiation markers and initial synthesis of collagenous extracellular matrix. Collagen matrix synthesis is required for subsequent induction of differentiation markers. Once the matrix synthesis begins, osteoblast marker genes are activated in a clear temporal sequence: alkaline phosphatase is induced at early times while bone sialoprotein and osteocalcin appear later in the differentiation process. This temporal sequence of gene expression is useful in monitoring the maturation and mineralization process. Matrix mineralization, which does not begin until several days after maturation has started, involves deposition of mineral on and within collagen fibrils deep within the matrix near the cell layer-culture plate interface. The collagen fibril–associated mineral formed by cultured osteoblasts resembles that found in woven bone in vivo and therefore is used frequently as a study reagent.

MC3T3 cells were transfected with antisense oligonucleotides for the first week of the differentiation, according to the manufacturer's specifications (U.S. Patent No. 5,849,902).

The oligonucleotides designed for LRP5 (Zmax1) are given below (SEQ ID NOS:639-641):

10875: AGUACAGCUUCUUGCCAACCCAGUC

10876: UCCUCCAGGUCGAUGGUCAGCCCAU

10877: GUCUGAGUCCGAGUUCAAAUCCAGG

Fig. 13 shows the results of antisense inhibition of LRP5 in MC3T3 cells. The three oligonucleotides shown above were transfected into MC3T3 and RNA was isolated according to standard procedures. Northern analysis clearly shows markedly lower steady state levels of the LRP5 transcript while the control gene GAPDH remained unchanged. Thus, antisense technology using the primers described above allows for the study of the role of LRP5 expression on bone biology.

XVI. Yeast Two Hybrid

In order to identify the signaling pathway that LRP5 participates in to modulate bone density, the yeast two hybrid protein interaction technology was utilized. This technique facilitates the identification of proteins that interact with one another by coupling tester proteins to components of a yeast transcription system (Fields and Song, 1989, *Nature* 340: 245-246; U.S. Pat. No. 5,283,173 by Fields and Song; Johnston, 1987, *Microbiol. Rev.* 51: 458-476; Keegan et al., 1986, *Science* 231: 699-704; Durfee et al., 1993, *Genes Dev.* 7: 555-569; Chien et al., 1991, *Proc. Natl. Acad. Sci USA* 88: 9578-9582; Fields et al., 1994, *Trends in Genetics* 10: 286-292; and Gyuris et al., 1993, *Cell* 75: 791-803). First a "bait" protein, the protein for which one seeks interacting proteins, is fused to the DNA binding domain of a yeast transcription factor. Second, a cDNA library is constructed that contains cDNAs fused to the transcriptional activation domain of the same yeast transcription factor; this is termed the prey library. The bait construct and prey library are transformed into yeast cells and then mated to produce diploid cells. If the bait interacts with a specific prey from the cDNA library, the activation domain is brought into the vicinity of the promoter via this interaction. Transcription is then driven through selectable marker genes and growth on selective media indicates the presence of interacting proteins.

The amino acid sequence used in the yeast two hybrid experiments discussed herein consisted of the entire cytoplasmic domain and a portion of the

transmembrane domain and is shown below (amino to carboxy orientation) (SEQ ID NO:765):

RVVCQRYAGA NGPFPHEYVS GTPHVPLNFI APGGSQHGPF TGIACGKSMM SSVSLMGGRG GVPLYDRNHV TGASSSSSS TKATLYPPIL NPPPSPATDP SLYNMDMFYS SNIPATVRPY RPYIIRGMAP PTTPCSTDVC DSDYSASRWK ASKYYLDLNS DSDPYPPPPT PHSQYLSAED SCPPSPATER SYFHLFPPPP SPCTDSS

The last 6 amino acids of the putative transmembrane domain are indicated in bold. Putative SH3 domains are underlined. Additional amino acid sequences of 50 amino acids or greater in either the proteins encoded by the LRP5 or HBM alleles can also be used as bait. The upper size of the polypeptide used as bait is limited only by the presence of a complete transmembrane domain (see Fig. 4), which will render the bait to be nonfunctional in a yeast two hybrid system. These additional bait proteins can be used to identify additional proteins which interact with the proteins encoded by HBM or LRP5 in the focal adhesion signaling pathway or in other pathways in which these HBM or LRP5 proteins may act. Once identified, methods of identifying agents which regulate the proteins in the focal adhesion signaling pathway or other pathways in which HBM acts can be performed as described herein for the HBM and LRP5 proteins.

In order to identify cytoplasmic LRP5 signaling pathways, the cytoplasmic domain of LRP5 was subcloned into two bait vectors. The first vector was pDBleu, which was used to screen a brain, and Hela prey cDNA library cloned into the vector pPC86 (Clontech). The second bait vector used was pDBtrp, which was used to screen a cDNA prey library derived from the TE85 osteosarcoma cell line in vector pOP46. Another suitable vector which is widely available, is p86 (Gibco ,iest™ System). Standard techniques known to those skilled in the art were used as described in Fields and Song, 1989, *Nature* 340: 245-246; U.S. Pat. No. 5,283,173 by Fields and Song; Johnston, 1987, *Microbiol. Rev.* 51: 458-476; Keegan et al., 1986, *Science* 231: 699-704; Durfee et al., 1993, *Genes Dev.* 7: 555-569; Chien et al., 1991, *Proc. Natl. Acad. Sci USA* 88: 9578-9582; Fields et al., 1994, *Trends in*

Genetics 10: 286-292; and Gyuris et al., 1993, *Cell* 75: 791-803. The bait construct and prey cDNA libraries were transformed into yeast using standard procedures.

To perform the protein interaction screen, an overnight culture of the bait yeast strain was grown in 20 ml SD selective medium with 2% glucose (pDBLeu, SD -Leu medium, pDBtrp, SD –trp medium). The cultures were shaken vigorously at 30°C overnight. The cultures were diluted 1 : 10 with complete medium (YEPD with 2% glucose) and the cultures then incubated with shaking for 2 hrs at 30°C.

The frozen prey library was thawed, and the yeast cells reactivated by growing them in 150 ml YEPD medium with 2% glucose for 2 hrs at 30°C. A filter unit was sterilized with 70% ethanol and washed with sterile water to remove the ethanol. The cell densities of both bait and prey cultures were measured by determining the OD at 600 nm. An appropriate volume of yeast cells that corresponded to a cell number of 1 ml of OD 600 = 4 of each yeast strain, bait and prey (library) was placed in a 50 ml Falcon tube. The mixture was then filtered through the sterilized filter unit. The filter was then transferred onto a prewarmed YEPD agar plate with the cell side up, removing all air bubbles underneath the filter. Plates were then incubated at 30°C for 6 hrs. One filter was transferred into a 50 ml Falcon tube, and 10 ml of SD with 2% Glucose was added; cells were resuspended by vortexing for 10 sec.

The number of primary diploid cells (growth on SD -Leu, -Trp plates) versus the numbers of colony forming units growing on SD -Trp and SD -Leu plates only was then titered. Different dilutions were plated and incubated at 30°C for two days. The number of colony forming units was then counted. The number of diploid colonies (colonies on SD -Leu -Trp plates) permits the calculation of whether or not the whole library of prey constructs was mated to the yeast expressing the bait. This information is important to judge the quality of the screen.

A. Indirect selection

Resuspended cells from 5 filtermatings were then pooled and the cells sedimented by centrifugation in a 50 ml Falcon tube. Cells were then resuspended in 16 ml SD medium with 2% Glc. Two ml of this cell suspension was plated onto 8 square plates each (SD -Leu, -Trp) with sterile glass beads and selected for diploid cells by incubating at 30°C for 18 - 20 hrs.

Cells were then scraped off the square plates, the cells centrifuged and combined into one 50 ml Falcon tube. The cell pellet was then resuspended in 25 ml of SD medium with 2% glucose. The cell number was then determined by counting of an appropriate dilution (usually 1:100 to 1:1000) with a Neugebauer chamber. Approximately 5 x 10^7 diploid cells were plated onto the selective medium. The observations about the growth of the bait strain together with irrelevant prey vectors helps to determine which selective plates will have to be chosen for the library screen. Generally, all screens were plated on one square plate each with SD -Leu, -Trp, -His; SD -Leu, -Trp, His, 5 mM 3AT, and SD -Leu, -Trp, -His, -Ade is recommended.

The yeast cells were then spread homogeneously with sterile glass beads and incubated at 30°C for 4 days. The number of colony forming yeast cells was titered by plating different dilutions of the scraped cell suspension onto SD -Leu, -Trp plates. Usually, plating of 100 μl of a 10^{-3} and 10^{-4} dilution gave 100 - 1000 colonies per plate.

B. Direct selection

Five filters with the mated yeast cells were each transferred into separate 50 ml Falcon tubes and the cells resuspended with 10 ml SD medium with 2% Glc, each, followed by vortexing for 10 sec. The resuspended cells were combined and centrifuged in a Beckman centrifuge at 3000 rpm. The supernatant was discarded and the cells resuspended in 6 ml of SD medium with 2% Glc. Two ml of the suspension was spread onto each selective square plate and incubated at 30°C for 4 - 5 days.

C. Isolation of Single Colonies

Yeast cells from an isolated colony were picked with a sterile tooth pick and transferred into individual wells of a 96 well plate. The cells were resuspended in 50 μl of SD -Leu, -Trp, -His medium and incubated at 30°C for one day. The yeast cells were then stamped onto a SD -Leu, -Trp, -His plate in 96 well format and incubated at 30°C for 2 days. Yeast cells were also stamped onto a Nylon filter covering a YEPD plate and incubated at 30°C for one day. The cells on the Nylon filter were used for the analysis of the ß - Gal reporter activity.

Yeast colonies were scraped from the SD -Leu, -Trp, -His plate with a sterile tooth pick, and reconfigured, if necessary, according to the ß - Gal activity and then resuspended in 20 % glycerol. This served as a master plate for storage at -80°C.

For DNA preparation, yeast cells from the glycerol stock were stamped onto a SD -Trp plate and incubated at 30°C for 2 days. After two days of incubation, the yeast colonies were ready for colony PCR and sequencing. Standard colony PCR conditions were used to amplify inserts from preys recovered from the interaction screen. Sequencing was done using standard sequencing reactions and ABI377 (Perkin Elmer) fluorescent sequencing machines.

D. Verification of bait/prey interaction

Glycerol stocks of the prey of interest were thawed and inoculated in a 10 ml overnight culture of SD with glucose -Trp. After overnight growth, plasmid DNA preparation was performed using the BIO 101 RPM Yeast Plasmid Isolation Kit with 10 ml of culture. The culture was centrifuged and transfered to a 1.5 ml microcentrifuge tube. Yeast Lysis Matrix was then added to the pellet followed by 250 μl of Alkaline Lysis Solution. Samples were then vortexed for 5 minutes. 250 μl Neutralizing Solution was added and the sample mixed briefly. Samples were centrifuged for 2 minutes at room temperature in a microcentrifuge. The

supernatant was transferred to a Spin Filter avoiding debris and Lysis Matrix. 250 μl of Glassmilk Spin Buffer was added, and the tubes inverted to mix. Samples were centrifuged for 1 min and the liquid in the Catch Tube was discarded. 500 μl of Wash Solution was added, the samples were centrifuged for 1 min, and the wash solution was discarded. The wash step was repeated once followed by a 1 min dry centrifugation to drive the remaining liquid out of the Spin Filter. The filter was transferred to a new Catch Tube and 100 μl of sterile H_2O was added; samples were then vortexed briefly to resuspend and centrifuged for 30 seconds to collect the DNA in the bottom of the Catch Tube.

Five μl of DNA was then transformed into DH10B Electromax cells using standard procedures and glycerol stocks prepared. Miniprep DNA was prepared using the Qiagen QIAprep Spin Miniprep Kit. DNA was finally eluted with 30 μl of Qiagen EB buffer. One μl of the plasmid DNA samples was then used to transform yeast cells using standard procedures. After 2 days of growth on SD –trp media, colonies were picked and patched onto fresh media. Similarly, bait colonies were patched onto SD –Leu media. Both were grown overnight at 30°C.

For mating, cells from bait and prey patches were spread together on YAPD media and incubated at 30°C for 12 hr. This plate was then replicaplated onto an SD Agar-Leu-Trp plate and grown for 2 days at 30°C. To test the strength of interaction these plates were replicaplated onto SD Agar-Leu-Trp-His, SD Agar-Leu-Trp-His with 5 mM 3AT and 10 mM 3AT, SD Agar-Leu-Trp-His-Ade, and SD Agar-Leu-Trp-Ura media and grown for 2 days at 30°C.

E. Galacton Star β-Galactosidase Activity Assay

After streaking and replica plating positive interactors on selection plates, colonies were placed in a 96 well dish with 200 μl of SD-medium, leaving wells 1 and 96 blank. Ten microliters from the first 96 well dish was plated into another flat bottom 96 well dish containing 100 μl of SD-medium. Controls consisted of a negative control and a very weak positive control. The cell density was measured

at OD_{600} (a value of 1 corresponds to $1x10^7$ cells utilizing a 96 well spectrophotometer). The OD was usually between 0.03 and 0.10. Using microplates specifically for the luminometer, 50 μl of reaction mixture were pipetted into each well. Fifty microliters of culture were then added and mixed by pipetting up and down twice. The reaction was incubated for 30 minutes at room temperature followed by measurement of Relative Light Units using a luminometer.

Table 6 lists the genes identified in the yeast two hybrid screens from the 3 prey libraries tested. Two genes, zyxin and axin, were found to interact with the cytoplasmic domain of LRP5 in all three screens. Three genes, alpha-actinin, TCB and S1-5 interacted in two of the three screens.

A variety of proteins found at sites of cell-cell and cell-matrix contact (focal contacts/adesion plaques) were shown to interact with the cytoplasmic domain of LRP5. These include alpha-actinin, Trio, Pinch-like protein, and Zyxin. PINCH is a LIM domain-containing protein that is known to interact with integrin-linked kinase, an early signaler in integrin and growth factor signaling pathways. The finding of a closely related gene in the yeast two hybrid screen raises the possibility of a novel pathway linked to integrin signaling from extracellular matrix signals. Trio, also known to localize to focal adhesions, is thought to play a key role in coordinating cell-matrix interactions and cytoskeletal rearrangements involved in cell movement. Zyxin, another LIM domain-containing protein, is also localized to adhesion plaques and is thought to be involved in reorganization of the cytoskeleton when triggers are transmitted via integrin signaling pathways. Zyxin also interacts with alpha actinin, which we identified as interacting with LRP5. Other LIM domain containing proteins identified include the human homologue of mouse ajuba, LIMD1, and a novel LIMD1-like protein.

Axin was also identified from the two hybrid experiments. This protein is involved in inhibition of the Wnt signaling pathway and interacts with the tumor suppressor APC. There is a link here with the focal adhesion signaling described above: one common step in the two pathways involves inhibition of glycogen

synthase kinase 3, which in turn results in the activation of ß-catenin/Lef-1 and AP-1 transcription factors. Axin/APC are involved in this as well as integrin linked kinase. The Wnt pathway has a role in determining cell fates during embryogenesis. If inappropriately activated, the Wnt pathway may also lead to cancer. The Wnt pathway also seems to have a role in cytoskeletal rearrangements. In a Xenopus embryo assay, the combination of HBM and Wnt5a preoteins stimulated the Wnt pathway to a much greater extent than the combination of LRP5 and Wnt5a, which was modestly above the control and Wnt5a alone scores. The HBM and LRP5 extracellular domains (ECD) caused a modest stimulation of Wnt signaling in the absence of Wnt5a which was slightly increased by the presence of Wnt5a in the presence of HBM ECD. A model depicting LRP5 involvement in focal adhesion signaling is depicted in Fig. 15.

This data coupled with other studies suggest that integrin signaling pathways have a role in cellular responses to mechanical stress and adhesion. This provides an attractive model for the mechanism of action of LRP5 in bone biology. It is possible that LRP5 is involved in sensing either mechanical stress directly or binding a molecule in the extracellular matrix that is related to mechanical sensation. Signaling through subsequent pathways may be involved in bone remodeling due to effects on cell morphology, cell adhesion, migration, proliferation, differentiation, and apoptosis in bone cells.

Table 6: Yeast Two Hybrid Results

Gene Symbol	Gene	Genbank Accession #	NT SEQ ID NO:	AA SEQ ID NO:
ACTN1	alpha-actinin	NM_001102	642	
AES	amino-terminal enhancer of	NM_001130.	643	
AIP4	atrophin-1 interacting protein	AF038564.1	644	
Novel	Ajuba		645	
AXIN	Wnt signaling	AF009674.1	646	

Gene Symbol	Gene	Genbank Accession #	NT SEQ ID NO:	AA SEQ ID NO:
CDC23	cell division cycle 23, yeast,	NM_004661.	647	
HSM800944	Similar to TRIO	AL117435.1	648	
HSM800936		AL117427.1	649	
Novel	Similar to LIM domains		650	
DEEPEST	mitotic spindle coiled-coil	NM_006461.	651	
ECM1	extracellular matrix protein 1	U65932.1	652	
EF1A	elongation factor 1-alpha	X16869.1	653	
FN	fibronectin	X02761.1	654	
HOXB13	homeodomain protein	U81599.1	655	
Novel	Glu-Lys Rich protein		656	
LIMD1	LIM domains containing 1	NM_014240.	567	
Novel	PINCH-like		568	
RANBPM	centrosomal protein	NM_005493.	659	
S1-5	extracellular protein	U03877.1	660	
TCB	gene encoding cytosolic	M26252.1	661	
TID	tumorous imaginal discs	NM_005147.	662	
ZYX	Zyxin	NM_003461.	663	
TRIO	GTPase	U42390.1	664	
HUMPITPB	phosphatidylinositol transfer	D30037.1	665	
ACTN1	alpha-actinin	NP_001093.1		666
AES	amino-terminal enhancer of	NP_001121.2		667
AIP4	atrophin-1 interacting protein	AAC04845.1		668
Novel	Ajuba			669
AXIN	Wnt signalling	AAC51624.1		670
CDC23	cell division cycle 23, yeast	NP_004652.1		671
Novel	Similar to TRIO CAB55923.1			672
Novel	Similar to LIM domains			673
DEEPEST	mitotic spindle coiled-coil	NP_006452.1		674

Gene Symbol	Gene	Genbank Accession #	NT SEQ ID NO:	AA SEQ ID NO:
ECM1	extracellular matrix protein 1	AAB05933.1		675
EF1A	elongation factor 1-alpha	CAA34756.1		676
FN	fibronectin	CAA26536.1		677
Novel	Glu-Lys rich protein			678
HOXB13	homeodomain protein B13	AAB39863.1		679
LIMD1	LIM domains containing 1	NP_055055.1		680
Novel	PINCH-like			681
RANBPM	centrosomal protein	NP_005484.1		682
S1-5	extracellular protein	AAA65590.1		683
TCB	cytosolic thyroid hormone-	AAA36672.1		684
TID	tumorous imaginal discs	NP_005138.1		685
ZYX	Zyxin	NP_003452.1		686
TRIO	GTPase	AAC34245.1		687
PTDINSTP	phosphatidylinositol transfer	P48739		688

In light of the model depicted in Fig. 15 and the results shown in Table 6, another aspect contemplated by the invention would be to regulate bone density and bone mass disorders by the regulating focal adhesion signaling. The regulation can occur by regulating the DNA, mRNA transcript or protein encoded by any of the members involved in the focal adhesion signaling pathway as identified by the yeast two hybrid system.

Also contemplated are the novel nucleic acids and proteins identified by the HBM yeast two hybrid system. These include but are not limited to SEQ ID NO: 645 (Ajuba), SEQ ID NO: 651 (a gene similar to a gene encoding LIM domains containing protein 1), SEQ ID NO: 656 (Glu-Lys Rich protein), SEQ ID NO: 658 (PINCH-like gene), SEQ ID NO: 669 (Ajuba protein), SEQ ID NO: 672 (protein similar to TRIO), SEQ ID NO: 673, SEQ ID NO: 678 (Glu-Lys rich protein) and SEQ ID NO: 681 (PINCH-like protein).

XVII. LRP5 / LRP6 and HBM Function

Recent studies have indicated that LRP5 participates in the Wnt signal transduction pathway. Gong et al. have also recently published results which further support the role of LRP5 in bone development (Gong et al., *Cell*, 107:513-23, 2001). The study by Gong and co-workers describes mutations of LRP5 which cause the autosomal recessive disorder osteoporosis-pseudoglioma syndrome (OPPG). They conclude that OPPG is caused by loss of LRP5 function and implicate LRP6 as a redundant receptor in the Wnt pathway. Loss of LRP5 function has recently been shown to result in a low bone mass phenotype. (Kato *et al., J. Cell. Biol.*, 157:303-14 (2002)).

The Wnt pathway is critical in limb early embryological development. Nusse, *Nature* 411:255-6 (2001); and Mao *et al.*, *Nature* 411:321-5 (2001)). Wnt proteins are secreted proteins which interact with the transmembrane protein Frizzled (Fz). LRP proteins, such as LRP5 and LRP6, are believed to modulate the Wnt signal in a complex with Fz (Tamai *et al.*, *Nature* 407:530-5 (2000)). The Wnt pathway acts intracellularly through the Disheveled protein (Dsh) which in turn inhibits glycogen synthetase kinase-3 (GSK3) from phosphorylating β-catenin. Phosphorylated β-catenin is rapidly degraded following ubiquitination. However, the stabilized β-catenin accumulates and translocates to the nucleus where it acts as a cofactor of the T-cell factor (TCF) transcription activator complex.

The protein dickkopf-1 (Dkk-1) is an antagonist of the Wnt pathway required for head formation in early development. (Glinka *et al.*, *Nature*, 391:357-62 (1998)) Dkk-1 and its function in the Wnt pathway are described in *e.g.*, Krupnik, *et al.*, *Gene* 238:301-13 (1999); Fedi *et al.*, *J. Biol. Chem.* 274:19465-72 (1999); *see also* for Dkk-1 and the Wnt pathway, Wu *et al.*, *Curr. Biol.* 10:1611-4 (2000), Shinya *et al.*, *Mech. Dev.* 98:3-17 (2000), Mukhopadhyay *et al.*, *Dev Cell* 1:423-434 (2001) and in PCT Patent Application No. WO 00/52047, and in references cited in each. It has been known that Dkk-1 acts upstream of Dsh, however the nature of the mechanism of inhibition by Dkk-1 is just beginning to be

elucidated. Dkk-1 is expressed in the mouse embryonic limb bud and its disruption results in abnormal limb morphogensis, among other developmental defects (Gotewold *et al., Mech. Dev.* 89:151-3 (1999); and, Mukhopadhyay *et al., Dev Cell* 1:423-434 (2001)).

The interaction between Dkk-1 and LRP5 was discovered by a yeast two hybrid (Y2H) screen for proteins which interact with the ligand binding domain of LRP5 in experiments disclosed by the present inventors in U.S. Applications 60/291,311 filed May 17, 2001; 60/353,058 filed February 1, 2002, and 60/361,293 filed March 4, 2002. The two-hybrid screen is a common procedure in the art, which is described, for example, by Gietz *et al., Mol. Cell. Biochem.* 172:67-79 (1997); Young, *Biol. Reprod.* 58:302-11 (1998); Brent and Finley, *Ann. Rev. Genet.* 31:663-704 (1997); and Lu and Hannon, eds., Yeast Hybrid Technologies, Eaton Publishing, Natick MA, (2000). More recently, other studies confirm that Dkk-1 is a binding partner for LRP and modulates the Wnt pathway via direct binding with LRP (R. Nusse, *Nature* 411:255-256 (2001); A. Bafico *et al., Nat. Cell Biol.* 3:683-686 (2001); M. Semënov, *Curr. Biol.* 11:951-961 (2001); B. Mao, *Nature* 411:321-325 (2001), Zorn, *Curr. Biol.* 11:R592-5 (2001)); and, L. Li *et al., J. Biol Chem.* 277:5977-81 (2002)).

Mao and colleagues (2001) identified Dkk-1 as a ligand for LRP6 and suggest that Dkk-1 and LRP6 interact antagonistically in that Dkk proteins inhibit the Wnt coreceptor functions of LRP6. Using co-immunoprecipitation, the group verified that the Dkk-1/LRP6 interaction was direct. Dkk-2 was also found to directly bind LRP6. However, Mao *et al.* report that no interaction was detected between any Dkk protein and LRP5 nor do they find an interaction with LDLR, VLDLR, ApoER, or LRP. Additionally, Mao *et al.* demonstrated that LRP6 can titrate Dkk-1's effects of inhibiting Wnt signaling using the commercial TCF-luciferase reporter gene assay (TOPFLASH). A similar conclusion was drawn from analogous studies in *Xenopus* embryos. Deletion analysis of LRP6 functional domains revealed that EGF repeats (beta-propellers) 3 and 4 were necessary for

Dkk-1 binding and that the ligand binding domains of LRP6 had no effect on Dkk-1 binding. The findings of Mao *et al.* contrast with data obtained by the present inventors indicating that the ligand binding domains of LRP5 were necessary and sufficient for Dkk-1 binding in yeast. Using classical biochemical ligand-receptor studies, Mao *et al.* determined a Kd=0.34 nM for Dkk-1/LRP6 and a Kd=0.73 nM for Dkk-2/LRP6.

Semenov *et al.* (2001) verified the Mao group's results and confirmed by coimmunoprecipitation that Dkk-1 does not directly bind to Wnt or Frizzled but rather interacts with LRP6. Their Scatchard analysis found Kd=0.5 nM for Dkk-1/LRP6. Semenov *et al.* also demonstrated that Dkk-1 could abolish an LRP5/Frizzled8 complex implying that Dkk-1 can also repress Wnt signaling via interactions with LRP5. A Dkk-1 mutant having cysteine 220 changed to alanine abolished LRP6 binding and was unable to repress Wnt signaling. Studies in *Xenopus* embryos confirmed the results and revealed a functional consequence of Dkk-1/LRP6: repression of Wnt signaling. Their *Xenopus* work also suggested that LRP6/Dkk-1 may be specific for the canonical, β-catenin-mediated, Wnt pathways as opposed to the Wnt Planar Cell Polarity pathway.

Bafico *et al.* (2001) employed a ^{125}I-labeled Dkk-1 molecule to identify LRP6 as its sole membrane receptor with a Kd=0.39 nM. Again, the functional consequences of the Dkk-1/LRP6 interaction was a repression of the canonical Wnt signaling even when Dkk-1 was added at extremely low concentrations (30 pM).

Dkk-1 is able to repress LRP5-mediated Wnt signaling but is less effective in repressing HBM-mediated Wnt signaling as first disclosed by the present inventors in U.S. Applications 60/291,311 filed May 17, 2001; 60/353,058 filed February 1, 2002, and 60/361,293 filed March 4, 2002. This observation is of particular interest because the HBM mutation in LRP5 is a gain of function or activation mutation. That is, Wnt signaling, via the canonical pathway, is enhanced with HBM versus LRP5. Not wishing to be bound by theory, it is believed that this

interaction provides an explanation of the developemental signaling differences between HBM and the more common LRP5 / LRP6.

Further investigations of additional Wnt or Dkk family members show subtle differences in activities through the Wnt pathway and demonstrate the complexity and variability in Wnt signaling that can be achieved as a function of the LRP/Dkk/Wnt/Frizzled repertoire that is expressed in a particular cell or tissue. This may attest to the apparent bone specificity of the HBM phenotype in humans and in the HBM transgenic animals. These subtle variations should be considered in the development of potential therapies and or drug interventions. It is not desireable to simply on or off LRP5 signaling.

It may be that the reduced effectiveness of Dkk inhibition of LRP5 which is observed for HBM is not necessarily mediated by enhancing or preventing the binding of Dkk to LRP5/LRP6/HBM. More than one mechanism may be involved. Indeed, the inventors have observed that Dkk-1 binds LRP5, LRP6, and HBM. Further, has been observed that different members of the Dkk family differentially affect LRP5/LRP6/HBM activity. Rather, the more preferred approach is to develop a drug or therapy which results in the desired protective benefits by reproducing as nearly as possible the subtle effect of HBM. The ability to refine and test potential drugs and therapies by comparing their effects to an animal model of HBM is among the major features of the present invention.

The transgenic animals and methods of the invention may be utilized to screen potential therapeutic compounds and methods in the context of an animal model of the HBM phenotype. As such the animals and methods of the invention represent an invaluable tool in the development of future drugs and therapies. The present invention provides important research tools to develop an effective model of osteoporosis, to increase understanding of bone mass and lipid modulation, and to modulate bone mass an lipid metabolism.

The protein encoded by LRP5 is structurally related to the Low Density Lipoprotein receptor (LDL receptor). *See*, Goldstein et al., *Ann. Rev. Cell*

Biology, 1:1-39 (1985); Brown et al., *Science*, 232:34-47 (1986). The LDL receptor is responsible for uptake of low density lipoprotein, a lipid-protein aggregate that includes cholesterol. Individuals with a defect in the LDL receptor are deficient in cholesterol removal and tend to develop artherosclerosis. In addition, cells with a defective LDL receptor show increased production of cholesterol, in part because of altered feedback regulation of cholesterol synthetic enzymes and in part because of increased transcription of the genes for these enzymes. In some cell types, cholesterol is a precursor for the formation of steroid hormones. Thus, the LDL receptor may also function as an indirect signal transduction protein and may regulate gene expression.

The glycine 171 amino acid is likely to be important for the function of LRP5 because this amino acid is also found in the mouse homolog of LRP5. The closely related LRP6 (Genbank Accession No. JE0272) protein also contains glycine at the corresponding position (Brown et al., *Biochemical and Biophysical Research Comm.*, 248:879-888 (1988)). Amino acids that are important in a protein's structure or function tend to be conserved between species, because natural selection prevents mutations with altered amino acids at important positions from arising.

In addition, the extracellular domain of LRP5 contains four repeats consisting of five YWTD motifs followed by an EFG motif. This 5YWTD+EGF repeat is likely to form a distinct folded protein domain, as this repeat is also found in the LDL receptor and other LDL receptor-related proteins. The first three 5YWTD+EGF repeats are very similar in their structure, while the fourth is highly divergent. Glycine 171 occurs in the central YWTD motif of the first 5YWTD+EGF repeat in LRP5. The other two similar 5YWTD+EGF repeats of LRP5 also contain glycine at the corresponding position, as does the 5YWTD+EGF repeat in the LDL receptor protein. However, only 17.6% of the amino acids are identical among the first three 5YWTD+EGF repeats in LRP5 and the single repeat in the LDL receptor. These observations indicate that glycine 171

is essential to the function of this repeat, and mutation of glycine 171 causes a functional alteration of LRP5. The cDNA and peptide sequences are shown in Figs. 6A-6E. The critical base at nucleotide position 582 is indicated in bold and is underlined.

Northern blot analysis (Figs. 7A-B) reveals that LRP5 is expressed in human bone tissue as well as numerous other tissues. A multiple-tissue Northern blot (Clontech, Palo Alto, CA) was probed with exons from LRP5. As shown in Fig. 7A, the 5.5 kb LRP5 transcript was highly expressed in heart, kidney, lung, liver and pancreas and is expressed at lower levels in skeletal muscle and brain. A second northern blot, shown in Fig. 7B, confirmed the transcript size at 5.5 kb, and indicated that LRP5 is expressed in bone, bone marrow, calvaria and human osteoblastic cell lines.

Taken together, these results coupled with the yeast two hybrid results indicate that the HBM polymorphism in the *LRP5* gene is responsible for the HBM phenotype, and that the *LRP5* gene is important in bone development. In addition, because mutation of LRP5 can alter bone mineralization and development, it is likely that molecules that bind to LRP5 may usefully alter bone development. Such molecules may include, for example, small molecules, proteins, RNA aptamers, peptide aptamers, and the like.

XVIII. Preparation of Nucleic Acids, Vectors, Transformations and Host Cells

Large amounts of the nucleic acids of the present invention may be produced by replication in a suitable host cell. Natural or synthetic nucleic acid fragments coding for a desired fragment will be incorporated into recombinant nucleic acid constructs, usually DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the nucleic acid constructs will be suitable for replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to (with and without integration within the genome) cultured mammalian or plant or other eukaryotic cell lines. The purification of

nucleic acids produced by the methods of the present invention is described, for example, in Sambrook et al., *Molecular Cloning. A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) or Ausubel et al., *Current Protocols in Molecular Biology*, J. Wiley and Sons, NY (1992).

The nucleic acids of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage et al., *Tetra. Letts.*, 22:1859-1862 (1981) or the triester method according to Matteucci,et al., J. *Am. Chem. Soc.*, 103:3185 (1981), and may be performed on commercial, automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Nucleic acid constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended nucleic acid fragment encoding the desired protein, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the protein encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate, whether from a native HBM or LRP5 protein or from other receptors or from secreted proteins of the same or related species, which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook et al., *Molecular Cloning. A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold

Spring Harbor, NY (1989) or Ausubel et al., *Current Protocols in Molecular Biology*, J. Wiley and Sons, NY (1992).

An appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may include, when appropriate, those naturally associated with LRP5 or *HBM* genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook et al., *Molecular Cloning. A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) or Ausubel et al., *Current Protocols in Molecular Biology*, J. Wiley and Sons, NY (1992). Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England BioLabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Vectors and promoters suitable for use in yeast expression are further described in EP 73,675A. Appropriate non-native mammalian promoters might include the early and late promoters from SV40 (Fiers et al., *Nature*, 273:113 (1978)) or promoters derived from murine Moloney leukemia virus, mouse tumor virus, avian sarcoma viruses, adenovirus II, bovine papilloma virus or polyoma. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences, see also *Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1983).

While such expression vectors may replicate autonomously, they may also replicate by being inserted into the genome of the host cell, by methods well known in the art.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for survival or growth of a host cell transformed with the vector. The presence of this gene ensures growth of only those host cells which express the inserts. Typical selection genes encode proteins that a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; b) complement auxotrophic deficiencies, or c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for Bacilli. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed *in vitro*, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection (see, Kubo et al., *FEBS Letts*. 241:119 (1988)), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook et al., 1989 and Ausubel et al., 1992. The introduction of the nucleic acids into the host cell by any method known in the art, including those described above, will be referred to herein as "transformation." The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and proteins of the present invention may be prepared by expressing the LRP5 or HBM nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* or *Pseudomonas* may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, or amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is per se well known. See, Jakoby and Pastan (eds.), *Cell Culture. Methods in Enzymology*, volume 58, Academic Press, Inc., Harcourt Brace Jovanovich, NY, (1979)). Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other cell lines may be appropriate, e.g., to provide higher expression desirable glycosylation patterns, or other features.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. In prokaryotic hosts, the transformant may be selected, e.g., by resistance to ampicillin, tetracycline or other antibiotics. Production of a particular product based on temperature sensitivity may also serve as an appropriate marker.

Prokaryotic or eukaryotic cells transformed with the nucleic acids of the present invention will be useful not only for the production of the nucleic acids and proteins of the present invention, but also, for example, in studying the characteristics of LRP5 or HBM proteins.

Antisense nucleic acid sequences are useful in preventing or diminishing the expression of LRP5 or HBM, as will be appreciated by one skilled in the art. For example, nucleic acid vectors containing all or a portion of the LRP5 or *HBM* gene or other sequences from the LRP5 or HBM region may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with LRP5 or HBM transcription and/or translation and/or replication. Also contemplated are RNA interference methodologies including siRNAs or shRNAs.

The probes and primers based on the LRP5 and *HBM* gene sequences disclosed herein are used to identify homologous LRP5 and *HBM* gene sequences and proteins in other species. These LRP5 and *HBM* gene sequences and proteins are used in the diagnostic/prognostic, therapeutic and drug screening methods described herein for the species from which they have been isolated.

XIX. Protein Expression and Purification

Expression and purification of the HBM protein of the invention can be performed essentially as outlined below. To facilitate the cloning, expression and purification of membrane and secreted protein from the *HBM* gene, a gene expression system, such as the pET System (Novagen), for cloning and expression of recombinant proteins in *E. coli* was selected. Also, a DNA sequence encoding a peptide tag, the His-Tap, was fused to the 3' end of DNA sequences of interest to facilitate purification of the recombinant protein products. The 3' end was selected for fusion to avoid alteration of any 5' terminal signal sequence.

Nucleic acids chosen, for example, from the nucleic acids set forth in SEQ ID NOS: 1, 3 and 5-12 for cloning HBM were prepared by polymerase chain reaction (PCR). Synthetic oligonucleotide primers specific for the 5' and 3' ends of the HBM nucleotide sequence were designed and purchased from Life Technologies (Gaithersburg, MD). All forward primers (specific for the 5' end of the sequence) were designed to include an *Nco*I cloning site at the 5' terminus. These primers were designed to permit initiation of protein translation at the methionine residue encoded within the *Nco*I site followed by a valine residue and the protein encoded by the HBM DNA sequence. All reverse primers (specific for the 3' end of the sequence) included an *Eco*RI site at the 5' terminus to permit cloning of the HBM sequence into the reading frame of the pET-28b. The pET-28b vector provided a sequence encoding an additional 20 carboxyl-terminal amino acids including six histidine residues (at the C-terminus), which comprised the histidine affinity tag.

Genomic DNA prepared from the *HBM* gene was used as the source of template DNA for PCR amplification (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons (1994)). To amplify a DNA sequence containing the HBM nucleotide sequence, genomic DNA (50 ng) was introduced into a reaction vial containing 2 mM $MgCl_2$, 1 μM synthetic oligonucleotide primers (forward and reverse primers) complementary to and flanking a defined HBM, 0.2 mM of each of deoxynucleotide triphosphate, dATP, dGTP, dCTP, dTTP and 2.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ) in a final volume of 100 microliters.

Upon completion of thermal cycling reactions, each sample of amplified DNA was purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD). All amplified DNA samples were subjected to digestion with the restriction endonucleases, e.g., NcoI and EcoRI (New England BioLabs, Beverly, MA) (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994)). DNA samples were then subjected to electrophoresis on 1.0% NuSeive (FMC BioProducts, Rockland, ME) agarose gels. DNA was visualized by exposure to ethidium bromide and long wave UV irradiation. DNA contained in slices isolated from the agarose gel was purified using the Bio 101 GeneClean Kit protocol (Bio 101, Vista, CA).

The pET-28b vector was prepared for cloning by digestion with restriction endonucleases, e.g., NcoI and EcoRI (New England BioLabs, Beverly, MA) (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994)). The pET-28a vector, which encodes the histidine affinity tag that can be fused to the 5' end of an inserted gene, was prepared by digestion with appropriate restriction endonucleases.

Following digestion, DNA inserts were cloned (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994)) into the previously digested pET-28b expression vector. Products of the ligation reaction were then used to transform the BL21 strain of *E. coli* (Ausubel et al., *Current*

Protocols in Molecular Biology, John Wiley & Sons, Inc. (1994)) as described below.

Competent bacteria, *E. coli* strain BL21 or *E. coli* strain BL21 (DE3), were transformed with recombinant pET expression plasmids carrying the cloned HBM sequence according to standard methods (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994)). Briefly, 1 μl of ligation reaction was mixed with 50 μl of electrocompetent cells and subjected to a high voltage pulse, after which samples were incubated in 0.45 ml SOC medium (0.5% yeast extract, 2.0% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) at 37°C with shaking for 1 hour. Samples were then spread on LB agar plates containing 25 μg/ml kanamycin sulfate for growth overnight. Transformed colonies of BL21 were then picked and analyzed to evaluate cloned inserts, as described below.

Individual BL21 clones transformed with recombinant pET-28b HBM nucleotide sequences were analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers specific for the HBM sequences that were used in the original PCR amplification cloning reactions. Successful amplification verifies the integration of the HBM sequence in the expression vector (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994)).

Individual clones of recombinant pET-28b vectors carrying properly cloned HBM nucleotide sequences were picked and incubated in 5 ml of LB broth plus 25 μg/ml kanamycin sulfate overnight. The following day plasmid DNA was isolated and purified using the Qiagen plasmid purification protocol (Qiagen Inc., Chatsworth, CA).

The pET vector can be propagated in any *E. coli* K-12 strain, e.g., HMS174, HB101, JM109, DH5 and the like, for purposes of cloning or plasmid preparation. Hosts for expression include *E. coli* strains containing a chromosomal copy of the gene for T7 RNA polymerase. These hosts were lysogens of

bacteriophage DE3, a lambda derivative that carries the lacI gene, the lacUV5 promoter and the gene for T7 RNA polymerase. T7 RNA polymerase was induced by addition of isopropyl-β-D-thiogalactoside (IPTG), and the T7 RNA polymerase transcribes any target plasmid containing a functional T7 promoter, such as pET-28b, carrying its gene of interest. Strains include, for example, BL21(DE3) (Studier et al., *Meth. Enzymol.*, 185:60-89 (1990)).

To express the recombinant HBM sequence, 50 ng of plasmid DNA are isolated as described above to transform competent BL21(DE3) bacteria as described above (provided by Novagen as part of the pET expression kit). The lacZ gene (β-galactosidase) is expressed in the pET-System as described for the HBM recombinant constructions. Transformed cells were cultured in SOC medium for 1 hour, and the culture was then plated on LB plates containing 25 μg/ml kanamycin sulfate. The following day, the bacterial colonies were pooled and grown in LB medium containing kanamycin sulfate (25 μg/ml) to an optical density at 600 nM of 0.5 to 1.0 O.D. units, at which point 1 mM IPTG was added to the culture for 3 hours to induce gene expression of the HBM recombinant DNA constructions.

After induction of gene expression with IPTG, bacteria were collected by centrifugation in a Sorvall RC-3B centrifuge at 3500 x g for 15 minutes at 4°C. Pellets were resuspended in 50 ml of cold mM Tris-HCl, pH 8.0, 0.1 M NaCl and 0.1 mM EDTA (STE buffer). Cells were then centrifuged at 2000 x g for 20 minutes at 4°C. Wet pellets were weighed and frozen at -80°C until ready for protein purification.

Chinese Hamster Ovary (CHO) Expression System

Alternatively, HBM and LRP5 may be expressed in eukaryotic cells. Eukaryotic cells, such as mammalian derived cell lines, are more capable of expressing properly folded proteins containing cystine rich domains such as the EGF and LDLR modules.

Development of Constructs

HBM and LRP5 extracellular domain fusions (ECD) to IgG-Fc were prepared. These ECD fusions to the IgG-Fc domain remove the endogenous transmembrane and cytoplasmic portion of the LRP5/HBM receptor and should produce a secreted fusion protein. The Fc region is separated from the LRP5/HBM ECD by an enterokinase recognition site so that purified LRP5 or HBM ECD protein can be obtained without the Fc domain present. The vector used for this construct was pHTop, a derivative of pED (Kaufman et al., 1991 *Nuc. Acids Res.* 19: 4485-4490) in which the majority of the adenomajor late promoter was replaced by six repeats of the tet operator (Gossen et al., 1992, *Proc. Natl. Acad. Sci. USA* 89:5547-5551). This vector contains the dihydrofolate reductase (*dhfr*) gene, and when introduced in the cell line CHO/A2 (see description below), functions very efficiently. Clones with high expression can be selected by isolating cells which survive in high methotrexate (MTX) concentrations.

The CHO expression vector pHTOP-Fc was digested with *Sal*I and *Not*I. The intervening sequence was purified away from the rest of the vector by electroelution from an acrylamide gel slice. *Sal*I cuts 5' to the intrinsic honey bee mellitin signal sequence in pHTOP-Fc, and *Not*I cuts just 5' to the coding sequence IgG1-Fc. The resulting SalI-NotI pHTOP-Fc vector has the signal sequence removed and the NotI cloning site is amenable to creating a 5' fusion to IgG-Fc. Full-length LRP5 in pCMVSPORT6 and full-length HBM in pCMVSPORT6 were digested individually with *Xma*1 which cuts within the region of the ORF that encodes the signal sequence) and *Bam*HI (that cuts internally in the ORF) to generate a 2286 bp 5' fragment of LRP5 and HBM. The mutation which distinguishes LRP5 from HBM lies on this fragment. Separately, the LRP5 DNA was digested with BamHI and *Sac*I to isolate an 1800 bp 3' fragment which is common to both the LRP5 and the HBM genes. Together, these two fragments constitute the coding sequence for the HBM and LRP5 extracellular domains, less the coding sequence for the first 6 amino acids of the signal sequence and ending 18

amino acids prior to the end of the extracellular domain, which we estimated from Kyte-Doolittle plots to end at the sequence (SEQ ID NO:698) "SPAHSS."

A synthetic duplex was designed to recreate the coding sequence of the LRP5/HBM signal sequence 5' of the native Xmal site, which included the initiator methionine and Kozak sequence. This duplex was designed to contain *Sal*I (5') and *Xma*l (3') cohesive ends to adapt ends to adapt the gene fragments described above to the pHTOP-Fc vector. This synthetic duplex was constructed from two partially complementary oligonucleotides as given below (SEQ ID NO:699-700):

5'-TCGACCACCATGGAGGCAGCGCCGC-3'

3'-GGTGGTACCTCCGTCGCGGCGGGCC-5'

A second synthetic duplex was designed to recreate the 3' coding sequence from a native *Sac*I site to the estimated end of the extracellular domain following the serine in the sequence ". . .SPAHSS", and to also encode a cloning site to allow in-frame fusion to the downstream IgG-Fc. This duplex was designed to contain *Sac*I (5') and *Not*I (3') cohesive ends to adapt the gene fragments described above to the pHTOP-Fc vector. This synthetic duplex was constructed from two partially complementary oligonucleotides whose sequences are given below (SEQ ID NO:701-702):

5'-CATGTGTGAAATCACCAAGCCGCCCTCAGACGACAGCCC GGCCCACAGCAGTGGC-3'

3'-TCGAGTACACACTTTAGTGGTTCGGCGGGAGTCTGCTGTC GGGCCGGGTGTCGTCACCGCCGG-5'

The fragments, synthetic duplexes, and vector were ligated together in a single reaction. Separate reactions were performed for LRP5 and HBM. The ligation mixtures were used to transform electrocompetent *E. coli* DH10B cells, and the resulting colonies were screened by radioactive colony hybridization using the common *Sac*I-*Bam*HI 3' fragment as a probe. Colonies containing plasmids with the LRP5 or HBM fragment inserted were identified, and plasmids were isolated

from multiple candidates and their sequences were verified by DNA sequencing. Verified constructs were then used for transfection into CHO cells.

Establishment of CHO stable cell lines

The CHO/A2 cell line is derived from CHO DUKX B11 (Urlaub and Chasin, 1980, *Proc. Natl. Acad. Sci. USA* 77: 4216-4220) by stably integrating a transcriptional activator (tTA), a fusion protein, between the Tet repressor and the herpes virus VP16 transcriptional domain (Gossen *et al.*) CHO cell lines expressing extracellular HBM-1.Fc and LRP5.Fc were established by transfecting (using lipofection) pHTopHBM-1.Fc into CHO/A2 cells and pHTopZmax1.Fc into CHO/A2 cells. Clones were selected using by culturing the cells in 0.02 μM methotrexate. Clones were later amplified step-wise to a final concentration of 0.5 μM methotrexate.

Screening of CHO stable cell lines

Multiple clones were screened by a variety of techniques. Clones were screened by Western blot assay using a (mouse) anti-human IgG.Fc horseradish peroxidase (HRP) antibody. The same clones were also metabolically labeled with ^{35}S-Met/Cys) for a 6 hour pulse, or a 15 minute pulse, followed by a 1 hour, 4 hour, or 24 hour chase in media without radiolabeled Met/Cys. Immunoprecipitations were performed on proteins obtained from conditioned media, as well as from cell extracts. Purification is then performed followed by sequencing of the proteins using N-terminal sequencing as known in the art.

Fusion Protein Purification

LRP5-IgG or HBM-IgG fusion protein can be purified from conditioned media or cell extracts of CHO stable cells. The fusion protein is isolated by affinity binding to protein A (for example using protein A coated beads or columns). The IgG-FC domain can then subsequently be cleaved from the Zmax/HBM1 ECD protein by enterokinase digestion.

Potential uses for cell lines and protein

Stable cell lines may be used for generation of purified protein for use in ligand hunting, antibody generation, determination of crystal structure, and competitive binding assays.

A variety of methodologies known in the art can be used to purify the isolated proteins (Coligan et al., *Current Protocols in Protein Science,* John Wiley & Sons (1995)). For example, the frozen cells can be thawed, resuspended in buffer and ruptured by several passages through a small volume microfluidizer (Model M-110S, Microfluidics International Corp., Newton, MA). The resultant homogenate is centrifuged to yield a clear supernatant (crude extract) and, following filtration, the crude extract is fractioned over columns. Fractions are monitored by absorbance at OD_{280} nm and peak fractions may be analyzed by SDS-PAGE.

The concentrations of purified protein preparations are quantified spectrophotometrically using absorbance coefficients calculated from amino acid content (Perkins, *Eur. J. Biochem.*, 157:169-180 (1986)). Protein concentrations are also measured by the method of Bradford, *Anal. Biochem.*, 72:248-254 (1976) and Lowry et al., *J. Biol. Chem.*, 193:265-275 (1951) using bovine serum albumin as a standard.

SDS-polyacrylamide gels of various concentrations were purchased from BioRad (Hercules, CA), and stained with Coomassie blue. Molecular weight markers may include rabbit skeletal muscle myosin (200 kDa), *E. coli* β-galactosidase (116 kDa), rabbit muscle phosphorylase B (97.4 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), bovine carbonic anyhdrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), egg white lysozyme (14.4 kDa) and bovine aprotinin (6.5 kDa).

Once a sufficient quantity of the desired protein has been obtained, it may be used for various purposes. A typical use is the production of antibodies specific for binding. These antibodies may be either polyclonal or monoclonal, and may be produced by *in vitro* or *in vivo* techniques well known in the art. Monoclonal antibodies to epitopes of any of the peptides identified and isolated as described can

be prepared from murine hybridomas (Kohler, *Nature*, 256:495 (1975)). In summary, a mouse is inoculated with a few micrograms of HBM protein over a period of two weeks. The mouse is then sacrificed. The cells that produce antibodies are then removed from the mouse's spleen. The spleen cells are then fused with polyethylene glycol with mouse myeloma cells. The successfully fused cells are diluted in a microtiter plate and growth of the culture is continued. The amount of antibody per well is measured by immunoassay methods such as ELISA (Engvall, *Meth. Enzymol.*, 70:419 (1980)). Clones producing antibody can be expanded and further propagated to produce HBM antibodies. Other suitable techniques involve *in vitro* exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse et al., *Science,* 246:1275-1281 (1989). For additional information on antibody production see Davis et al., *Basic Methods in Molecular Biology*, Elsevier, NY, Section 21-2 (1989).

LRP5 and LRP6 Polyclonal Antibodies

Polyclonal Antibodies were developed to both human LRP5 (SEQ ID NO:3) and LRP6 (GenBank Accession No. AF074264). Peptides from the LRP5 amino acid sequence were selected as immunogens based on five goals. 1) Maximize differences between LRP5 and LRP6 amino acid sequences (71% amino acid identity). See Figure 27. For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s), relative to the reference sequence, based on the designated program parameters. 2) Minimize potential cross reactivity with other known genes by performing sequence alignment and similarity searches. Optimal alignment of sequences for comparison can be

conducted, e.g., by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2, 482 (1981), by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.*, 48, 443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85, 2444 (1988), by computerized implementations of these algorithms and others in programs contained in the Wisconsin genetics software package, Genetics Computer Group, 585 Science Dr., Madison, WI, or by visual inspection (see generally Ausubel et al., *Current Protocols in Molecular Biology,* John Wiley & Sons (1997). Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215, 403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. 3) Obtain peptides with the highest antigenicity index as possible as determined by PeptideStructure protein analysis using software programs contained in the Wisconsin Genetics software package, Genetics Computer Group, 575 Science Dr., Madison, WI. 4) Locating peptides relative to highly homologous domains (*e.g.*, EGF-like domains and LDL receptor repeats) within the gene family and the location relative to the extracellular and cytoplasmic regions of the gene. 5) And, for human LRP5 specific antibodies, the human amino acid sequence (SEQ ID NO:3) was compared to the mouse LRP5 sequence (GenBank Accession No. AF064984) and peptides were selected based on the above criteria in addition to minimizing the sequence similarity between the two species (See Figure 26).

Using the same criteria above, LRP6 specific peptides were selected for polyclonal antibody production. The table lists the amino acid sequences that were chosen, the amino acid differences within the peptide between the human and mouse sequences. All the peptide sequences (ranging from 12-18 amino acids) were provided to Sigma/Genosys (St. Louis, MO) for peptide synthesis and subsequent polyclonal antibody production in New Zealand White Rabbits. The IgG fraction from the serum of each immunized rabbit was isolated using Protein G Sepharose

(Amersham). Polyclonal antibody generation using these peptides may be done in other species as well, for example, chickens. This is often advantageous when there is a high degree of similarity between the human (reference) and murine/rodent sequence.

Amino Acids	Amino Acids	H/M* Differences	Comments	SEQ ID NO:
171-187	VETPRIERAGMDGSTRK	5	Contains HBM polymorphism	703
264-278	NKRTGGKRKEILSAL	3	Extracellular	704
290-301	ERQPFFHTRCEE	2	Adjacent to EGF-I, extracellular	705
532-546	VDGTKRRTLLEDKLP	5	Extracellular	706
901-915	DGLNDCMHNNGQCGQ	2	In EGF-III, extracellular	707
1010-1021	PFVLTSLSQGQN	6	Extracellular, human specific	708
1415-1429	YAGANGPFPHEYVSG	3	Cytoplasmic	709
1452-1464	ACGKSMMSSVSLM	5	Cytoplasmic, human specific	710
1556-1573	RWKASKYYLDLNSDSDPY	1	Cytoplasmic	711
888-902	SGWNECASSNGHCSH		LRP6 specific	712
1308-1321	NGDANCQDKSDEKN		LRP6 specific	713

* H/M-differences between human and mouse sequences

Single chain Fv molecules developed by phage display

Peptides were chosen from the LRP5 sequence (SEQ ID NO:3) to screen for single chain Fv molecules by phage display. A total of 17 peptides from the LRP5 sequence were selected for synthesis and subsequent phage display screen for scFv molecules. All peptide synthesis and phage display work was performed at Cambridge Antibody Technology (CaT) in Cambridge, UK. Peptides were selected based on criteria as described above.

Protein Domain	LRP5 Residues	LRP6 Residues	% Identity
Spacer 1 (+G171V)	161-181	148-168	76%
Spacer 1 (-G171V)	161-181	148-168	76%
EGF 1	301-321	288-308	76%
Spacer 2	401-421	388-408	52%
EGF 2	611-631	598-618	62%
Spacer 3	781-801	768-788	62%
EGF 3	921-941	908-928	10%
Spacer 4	1000-1021	988-1008	26%
EGF 4	1229-1249	1219-1239	76%
LDLR 1	1261-1282	1252-1272	81%
LDLR 2	1300-1320	1290-1310	57%
LDLR 3	1338-1358	1328-1348	48%
Cytosolic 1	1418-1438	1405-1425	14%
Cytosolic 1	1516-1536	1503-1525	52%
Cytosolic 1	1535-1555	1524-1544	81%
Cytosolic 1	1595-1615	1592-1613	82%
Spacer 2 (cross reactive)	421-441	408-428	100%

Note that a number of these regions (e.g. 401-421, 421-441, 781-801, and 1229-1249) share 100% identity with mouse LRP5 (see Figure 26). Therefore, these may be used against both mouse and human forms of the protein. The peptide

421-441 was included to facilitate the generation of an antibody that would recognize both LRP5 and LRP6 (see Figure 27). Two peptides were synthesized spanning the HBM mutation site (LRP5 residues 161-181), one with the LRP5 sequence and the other containing the HBM sequence.

Once scFv molecules were isolated, they were used as reagents in immunochemistry to detect LRP5 protein expression in a variety of human normal and diseased tissues. The details of the scFV antibody immunohistochemical analysis of three phage clones against peptide 1000-1021 (IKRAKDDGTQPFVLTSLSQGQN) (SEQ ID NO:714) of the extracellular domain of LRP5 showed positive staining with cardiac muscle, kidney, lung and liver. Expression was also detected in prostate carcinoma. These results are consistent with mRNA tissue distribution profiles as well as with the published reports of LRP5 mRNA localization (Kim *et al.*, *J. Biochem.* 124: 1072-6, 1998). The resulting phage clones arise from pools and will be sequenced to identify potential variants in the Fv region of the molecules. Once identified, the suitable scFVs can then be subcloned into variable heavy chain and variable light chain DNA constructs for cotransfection into COS cells for final assembly of an intact and functional immunoglobulin gamma (IgG) molecule. The IgG that is expressed by the cells can then be further characterized for specificity and reactivity as would be known in the art.

Monoclonal antibody development

Monoclonal antibodies can be developed to LRP5 and HBM, as well as variants thereof, by complete cell and adenovirus immunization of, for example, Balb/c mice. Dendritic cells can be isolated from spleens of Balb/c mice, for example, and the cells expanded *in vitro* in the presence of growth factors IL-4 and GM-CSF. The dendritic cells can then be infected with HBM or LRP5 adenovirus particles. The cells are then cultured for 24 hours prior to intravenous injection into Balb/c mice. Dentritic cells ($1x10^6$ cells/mouse) are injected 2-3 times every 3-4 weeks over a three month period.

Alternatively, purified HBM and LRP5 DNAs in, for example, the pcDNA3.1 expression vector, can be coated on colloidal gold particles. These particles can then be injected subcutaneously into the desired mouse using gene gun technology. Approximately, 5 μg cDNA/mouse can be injected. Injections are performed 4-6 times every 2 weeks over approximately a 3 month period.

Another option is that cells (any species of animal, but preferably Balb/c mouse strain or the same species as the mouse strain being used which is related to limit antigen response to non-specific protein) overexpressing HBM and LRP5 and their respective adenovirus will be injected into the mice every 2-3 weeks for a period of about 1.5 to 3 months, as necessary. The bleeds from the mice can be tested for reactivity with the native and denatured protein by ELISA (using purified protein or protein-fusions), cell based ELISA, immunohistochemical staining and Western blotting. Serum samples from the animals can be screened by FACS (fluorescent activated cell sorting) using cells infected with LRP5 or HBM adenovirus. The spleen cells (antibody producing cells) from the mice with the strongest reactivity can then be fused with a myeloma to generate the hybridoma cells. The conditioned media from the hybridoma is then screened for the positive cell colonies for subsequent cloning. These cloned cells can then be injected into the intraperitoneal space in mice for ascites production.

Polyclonal Antibody Applications

Polyclonal antibodies directed against LRP5 and LRP6 were developed to determine the function of these proteins, analyze the expressed pattern and levels in various tissues, cells or any biological sample. Uses for polyclonal antibodies against LRP5, HBM, LRP6 and related variants include, but are not limited to: analysis of bone cross-sectional mounts, tissue distribution, evaluation of expression of the protein from bone biopsy samples of affected/non-affected family members (e.g., bone cell digests, explants of bone marrow stromal cell cultures), evaluation of protein expression levels in transiently or stably transfected cells, evaluating protein concentration in tissues, serum or body fluid, purification of full length or

fragments of these proteins for ligand hunting and functional assay development, identification of ligands or proteins which interact with these proteins, and elucidations of the signaling pathways of LRP6, LRP5 and HBM, and related variants.

For example, LRP5 cloned in pcDNA3.1 (Invitrogen, Carlsbad, CA). This was used to generate ^{35}S-labeled *in vitro* translated (Promega, Madison, WI) LRP5. Antibody (10 μg/ml) 3109 and 3110, which are directed against peptide immunogen RWKASKYYLDLNSDSDPY (SEQ ID NO:711) , was combined with 20 μl of the *in vitro* translated product in the presence of either 10 μg/ml specific peptide (*i.e.*, RWKASKYYLDLNSDSDPY)(SEQ ID NO:711) or non-specific peptide (*i.e.*, SGWNECASSNGHCSH)(SEQ ID NO:712) or no peptide and incubated for 1.5 hr at 4°C. Protein A Sepharose was then added to the samples (previously blocked for about 1.5 hr with reticulocyte lysate), and the samples were shaken for 1 hour at 4°C. The protein A Sepharose was washed 3 times with 0.5 ml of PBS. The bound protein was subsequently separated on a 4-12% gradient NuPAGE gel (Invitrogen) according to manufacturer's instructions. The gel was dried at 80°C for 30 min and then exposed to Kodak X-OMAT-AR film for 24 to 48 hr. The specific peptide was observed to significantly compete for the ^{35}S-labeled LRP5 immunoprecipitated protein with either antibody. The competition was not observed with a non-specific peptide.

These antibodies can also be used for immunohistochemistry. For example, HBM transgenic and wild-type mice were sacrificed using CO_2 narcosis. Mouse calvariae were removed intact, and the soft tissues gently dissected. The bones were fixed in 10% phosphate buffered formalin for 24 hours for further processing and analysis. After fixation, calvariae were decalcified in TBD-2 decalcifying agent (Shandon, Pittsburgh, PA) for about 7-8 hours and then dehydrated in graded alcohol. Calvariae were then bisected perpendicular to the sagittal suture through the central portion of the parietal bones parallel to the lambdoidal and coronal

sutures and embedded in paraffin. Four to six 5 μm thick representative sections were cut.

For example, the rabbit polyclonal antibody, LRP5/HBM (i.e., antibody 3109 and 3110) recognize LRP5 in both HBM transgenic and wild-type mouse calvariae. An anti-LRP5 or anti-HBM antibody can be used to detect LRP5 or HBM protein in order to evaluate its abundance and pattern of protein expression. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, P-galactosidase, or acetylcholinesterase; example of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; and example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin and acquorien; and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S, and ^{3}H. Alternatively, a secondary antibody can be employed that detects the presence of the primary LRP5 polyclonal antibody. An example would be an antibody that recognized all rabbit immunoglobulins. This secondary antibody could be coupled in an identical manner as described above to facilitate detection. Controls comprised samples with the avidin peroxidase, but without antibody. Intensive positive staining of stroma cells and mesenchymal cells was observed in the suture area. Pre-osteoblasts and osteoblasts were observed to stain within the periosteum and some osteocytes with antibody 3109 and 3110 in the calvariae of the HBM compared to wild-type mice. High magnification of tissue calvaria sections of the HBM transgenic mice showed a pronounced cell membrane staining of the osteocytes and the cells within the suture area.

XX. Methods of Use: Gene Therapy

In recent years, significant technological advances have been made in the area of gene therapy for both genetic and acquired diseases. (Kay et al., *Proc. Natl. Acad. Sci. USA*, 94:12744-12746 (1997)) Gene therapy can be characterized as the deliberate transfer of DNA for therapeutic purposes. Improvement in gene transfer methods has allowed for development of gene therapy protocols for the treatment of diverse types of diseases. Gene therapy has also taken advantage of recent advances in the identification of new therapeutic genes, improvement in both viral and nonviral gene delivery systems, better understanding of gene regulation, and improvement in cell isolation and transplantation.

The preceding experiments identify the *HBM* gene as a dominant mutation conferring elevated bone mass. The fact that this mutation is dominant indicates that expression of the HBM protein causes elevated bone mass. Older individuals carrying the *HBM* gene, and, therefore expressing the HBM protein, do not suffer from osteoporosis. These individuals are equivalent to individuals being treated with the HBM protein. These observations are a strong experimental indication that therapeutic treatment with the HBM protein prevents osteoporosis. The bone mass elevating activity of the *HBM* gene is termed "HBM function."

Therefore, according to the present invention, a method is also provided of supplying HBM function to mesenchymal stem cells (Onyia et al., *J. Bone Miner. Res.*, 13:20-30 (1998); Ko et al., *Cancer Res.*, 56:4614-4619 (1996)). Supplying such a function provides protection against osteoporosis. The *HBM* gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location.

Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation, and viral transduction are known in the art, and the

choice of method is within the competence of one skilled in the art (Robbins, Ed., *Gene Therapy Protocols*, Human Press, NJ (1997)). Cells transformed with the *HBM* gene can be used as model systems to study osteoporosis and drug treatments that promote bone growth.

As generally discussed above, the *HBM* gene or fragment, where applicable, may be used in gene therapy methods in order to increase the amount of the expression products of such genes in mesenchymal stem cells. It may be useful also to increase the level of expression of a given HBM protein, or a fragment thereof, even in those cells in which the wild type gene is expressed normally. Gene therapy would be carried out according to generally accepted methods as described by, for example, Friedman, *Therapy for Genetic Diseases*, Friedman, Ed., Oxford University Press, pages 105-121 (1991).

A virus or plasmid vector containing a copy of the *HBM* gene linked to expression control elements and capable of replicating inside mesenchymal stem cells, is prepared. Suitable vectors are known and described, for example, in U.S. Patent No. 5,252,479 and WO 93/07282, the disclosures of which are incorporated by reference herein in their entirety. The vector is then injected into the patient, either locally into the bone marrow or systemically (in order to reach any mesenchymal stem cells located at other sites, i.e., in the blood). If the transfected gene is not permanently incorporated into the genome of each of the targeted cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and non-viral transfer methods. A number of viruses have been used as gene transfer vectors, including polyoma, i.e., SV40 (Madzak et al., *J. Gen. Virol.*, 73:1533-1536 (1992)), adenovirus (Berkner, *Curr. Top. Microbiol. Immunol.*, 158:39-61 (1992); Berkner et al., *Bio Techniques*, 6:616-629 (1988); Gorziglia et al., *J. Virol.*, 66:4407-4412 (1992); Quantin et al., *Proc. Natl. Acad. Sci. USA*, 89:2581-2584 (1992); Rosenfeld et al., *Cell*, 68:143-155 (1992); Wilkinson et al., *Nucl. Acids*

Res., 20:2233-2239 (1992); Stratford-Perricaudet et al., *Hum. Gene Ther.*, 1:241-256 (1990)), vaccinia virus (Mackett et al., *Biotechnology*, 24:495-499 (1992)), adeno-associated virus (Muzyczka, *Curr. Top. Microbiol. Immunol.*, 158:91-123 (1992); Ohi et al., *Gene*, 89:279-282 (1990)), herpes viruses including HSV and EBV (Margolskee, *Curr. Top. Microbiol. Immunol.*, 158:67-90 (1992); Johnson et al., *J. Virol.*, 66:2952-2965 (1992); Fink et al., *Hum. Gene Ther.*, 3:11-19 (1992); Breakfield et al., *Mol. Neurobiol.*, 1:337-371 (1987;) Fresse et al., *Biochem. Pharmacol.*, 40:2189-2199 (1990)), and retroviruses of avian (Brandyopadhyay et al., *Mol. Cell Biol.*, 4:749-754 (1984); Petropouplos et al., *J. Virol.*, 66:3391-3397 (1992)), murine (Miller, *Curr. Top. Microbiol. Immunol.*, 158:1-24 (1992); Miller et al., *Mol. Cell Biol.*, 5:431-437 (1985); Sorge et al., *Mol. Cell Biol.*, 4:1730-1737 (1984); Mann et al., *J. Virol.*, 54:401-407 (1985)), and human origin (Page et al., *J. Virol.*, 64:5370-5276 (1990); Buchschalcher et al., *J. Virol.*, 66:2731-2739 (1992)). Most human gene therapy protocols have been based on disabled murine retroviruses.

Non-viral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham et al., *Virology*, 52:456-467 (1973); Pellicer et al., *Science*, 209:1414-1422 (1980)), mechanical techniques, for example microinjection (Anderson et al., *Proc. Natl. Acad. Sci. USA*, 77:5399-5403 (1980); Gordon et al., *Proc. Natl. Acad. Sci. USA*, 77:7380-7384 (1980); Brinster et al., *Cell*, 27:223-231 (1981); Constantini et al., *Nature*, 294:92-94 (1981)), membrane fusion-mediated transfer via liposomes (Felgner et al., *Proc. Natl. Acad. Sci. USA*, 84:7413-7417 (1987); Wang et al., *Biochemistry*, 28:9508-9514 (1989); Kaneda et al., *J. Biol. Chem.*, 264:12126-12129 (1989); Stewart et al., *Hum. Gene Ther.*, 3:267-275 (1992); Nabel et al., *Science*, 249:1285-1288 (1990); Lim et al., *Circulation*, 83:2007-2011 (1992)), and direct DNA uptake and receptor-mediated DNA transfer (Wolff et al., *Science*, 247:1465-1468 (1990); Wu et al., *BioTechniques*, 11:474-485 (1991); Zenke et al., *Proc. Natl. Acad. Sci. USA*, 87:3655-3659 (1990); Wu et al., *J. Biol. Chem.*, 264:16985-

16987 (1989); Wolff et al., *BioTechniques*, 11:474-485 (1991); Wagner et al., 1990; Wagner et al., *Proc. Natl. Acad. Sci. USA*, 88:4255-4259 (1991); Cotten et al., *Proc. Natl. Acad. Sci. USA*, 87:4033-4037 (1990); Curiel et al., *Proc. Natl. Acad. Sci. USA*, 88:8850-8854 (1991); Curiel et al., *Hum. Gene Ther.*, 3:147-154 (1991)). Viral-mediated gene transfer can be combined with direct *in vivo* vectors to the mesenchymal stem cells and not into the surrounding cells (Romano et al., *In Vivo*, 12(1):59-67 (1998); Gonez et al., *Hum. Mol. Genetics*, 7(12):1913-9 (1998)). Alternatively, the retroviral vector producer cell line can be injected into the bone marrow (Culver et al., *Science*, 256:1550-1552 (1992)). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumors.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct *in vivo* gene transfer. While in standard liposome preparations the gene transfer process is non-specific, localized *in vivo* uptake and expression have been reported in tumor deposits, for example, following direct *in situ* administration (Nabel, *Hum. Gene Ther.*, 3:399-410 (1992)).

XXI. Methods of Use: Transformed Hosts and Transgenic Animals as Research Tools and for the Development of Pharmaceuticals

Cells and animals that carry the *HBM, LRP5*, or *LRP6* gene, used as model systems, are valuable research tools to study and test for substances that have potential as therapeutic agents (Onyia et al., *J. Bone Miner. Res.*, 13:20-30 (1998); Broder et al., *Bone*, 21:225-235 (1997)). Cells for this purpose are typically

cultured mesenchymal stem cells. These may be isolated from individuals with somatic or germline *HBM* genes. Alternatively, the cell line can be engineered to carry the *HBM* gene, as described above. After a test substance is applied to the cells, the transformed phenotype of the cell is determined. Any trait of transformed cells can be assessed, including, for example, formation of bone matrix in culture (Broder et al., *Bone*, 21:225-235 (1997)), mechanical properties (Kizer et al., *Proc. Natl. Acad. Sci. USA*, 94:1013-1018 (1997)), expression of marker genes and response to application of putative therapeutic agents.

Transgenic modified animals and cell lines may be used to test therapeutic agents. Transgenic modifications include, for example, insertion of the *LRP5* gene, as well as insertion of the *HBM* gene and disrupted homologous genes. Alternatively, the inserted *LRP5* gene(s) and/or *HBM* gene(s) of the animals may be disrupted by insertion or deletion mutation of other genetic alterations using conventional techniques, such as those described by, for example, Capecchi, *Science*, 244:1288 (1989); Valancuis et al., *Mol. Cell Biol.*, 11:1402 (1991); Hasty et al., *Nature*, 350:243 (1991); Shinkai et al., *Cell*, 68:855 (1992); Mombaerts et al., *Cell*, 68:869 (1992); Philpott et al., *Science*, 256:1448 (1992); Snouwaert et al., *Science*, 257:1083 (1992); Donehower et al., *Nature*, 356:215 (1992). After test substances have been administered to the animals, the growth of bone must be assessed. If the test substance modulates (e.g., enhances) the growth of bone, then the test substance is a candidate therapeutic agent. These animal models provide an extremely important vehicle for potential therapeutic products.

The present invention also provides animals and cell lines wherein the expression of endogenous genes are activated, and may be further amplified, which does not require *in vitro* manipulation and transfection of exogenous DNA encoding LRP5 or HBM proteins. These methods are described for example in PCT Application WO 94/12650 and U.S. Patent No. 5,968,502, both of which are herein incorporated in their entirety by reference. In addition, the present invention includes methods wherein endogenous activation or over-expression is achieved by

in situ homologous recombination, non-homologous recombination, or illegitimate recombination methods. These methods are described for example in PCT Applications WO 99/15659 and WO 00/49162, both of which are incorporated herein in their entirety.

Creating transgenic and gene-targeted animals.

The present invention provides genetically modified animals that recapitulate the human HBM phenotype. The approaches taken involve the creation of both transgenic and gene-targeted animals that have the human G to T nucleotide substitution incorporated into the genome, animals which express human LRP5 (Zmax1) or express a variant which produces a bone mass altering phenotype. Subsequent to the making of the present invention, Kato et al., (*Journal of Cell Biology*, 157:303-14, 2002) have recently described the creation of LRP5 knock-out mice which demonstrated a low bone mass phenotype. The results described by Kato and coworkers is consistent with the hypothesis that LRP5 is a determinant of peak bone mass and demonstrates aspects of the utility of the present invention.

Transgenic mice over-expressing the HBM polymorphism.

Plasmid constructs were prepared that utilized either the CMVbActin or type I collagen promoters to drive expression of the human HBM cDNA. The most commonly-used promoters for mammalian expression are from cytomegalovirus (CMV), Rous sarcoma virus (RSV), Simian virus 40 (SV40), and EF-1a (human elongation factor 1a-subunit). CMV is derived from the human cytomegalovirus immediate-early viral promoter. CMV is a stronger promoter in most cell lines than either RSV or SV40. The RSV promoter is derived from an avian virus and tends to be a strong promoter in avian cell lines. The SV40 promoter expresses well in cell lines that carry the large T antigen, such as COS-1. In these cell lines, the plasmid is replicated to higher copy numbers. EF-1a is beginning to be more widely used for recombinant protein expression. EF-1a is the promoter from the

human elongation factor 1a-subunit, a gene that is highly expressed and conserved in mammalian cell lines.

The chimeric CMVbActin promoter is a strong promoter that has been shown to produce ubiquitous gene expression in transgenic mice including bone. This promoter was chosen to drive expression of HBM in a manner consistent with the reported widespread expression of the endogenous mouse *LRP5* gene. Although the HBM phenotype is observed in bone, the *HBM* gene may have direct or indirect effects in other tissues. Therefore, other strong ubiquitous promoters may be utilized as would be known to those skilled in the art.

Type I collagen promoters provide tissue-restricted gene expression wherein expression is primarily limited to bone. Other bone-specific promoters are available that could result in expression of HBM in bone. For example, promoters associated with proliferation of osteoblasts include histone, type I collagen, TGFβ1, MSX2, cfos/cJun and Cbfa1 may be used. Promoters associated with bone matrix maturation including alkaline phosphatase, MGP, Cbfa1, Fra/Jun and Dlx5 also can be used. Promoters associated with bone mineralization such as osteocalcin, osteopontin, bone sialoprotein and collagenase also can be used. The promoter chosen would be determined by, for example, the tissue expression, the degree of regulatable control and the like as would be known to the skilled artisan. For example, the type I collagen promoters were chosen to insure that HBM would be expressed in bone in a temporal, spatial and bone cell-specific pattern resembling the endogenous pattern of LRP5 expression in bone.

Transgenic mice over-expressing the wild-type LRP5 gene.

Plasmid constructs were prepared using the CMVbActin and type I collagen promoters driving expression of LRP5. These animals can serve as a control animal model for the HBM transgenic mice. Additional controls include non-transgenic littermates and wild-type animals of an identical genetic background.

Methods for preparing these animals would be similar to what is discussed for mice which over express the HBM polymorphism.

Gene-targeted mice expressing the HBM polymorphism.

A gene-targeting construct was prepared that could be used to create animals containing a HBM knock-in (KI) allele and a LRP5 knock-out (KO) allele. The gene-targeting construct contained the HBM polymorphism in exon 3 and included a neomycin selection cassette that was linked to a transcriptional stop sequence and was flanked with lox P sites. The HBM polymorphism in mouse LRP5 results in a glycine to valine change in the amino acid sequence at position 170 of the mouse LRP5 homolog (Genbank Accession No. AF064984). Homologous recombination is used to stably introduce the construct into the mouse genome. If the transcriptional stop sequence functioned to completely block transcription of the modified *LRP5* allele, then a functional *LRP5* knock-out allele would be generated. This would facilitate production of a homozygous knock-out animal for the *LRP5* gene.

To create the knock-in allele, Cre recombinase could be used to excise the neomycin selection cassette leaving behind the modified exon 3 and one copy of the loxP site. Cre could be introduced into single-cell fertilized embryos to facilitate ubiquitous expression of HBM or by crossing animals with transgenic mice to obtain bone-specific HBM expression. Homozygous animals could be made for the HBM knock-in allele. Alternatively, animals could be created by nuclear transfer techniques, wherein nuclei from homozygous animals is transferred into a prepared oocyte (e.g., enucleated) as is known in the art. See, e.g., Campbell et al., *Nature* 380: 64-68 (1996). Additional methods of creating knock out mice include engineering a homologous recombination vector wherein the ATG start codon is deleted or mutated, engineering a frame-shift mutation into the vector, engineering deletions of critical portions of the promoter region, and/ or engineering a vector to delete critical regions of the gene.

Materials and Methods

Construction of the LRP5 plasmid ZMAXGI_3AS

The full-length *LRP5* cDNA construct has been engineered into the *Xba*I-*Not*I sites of the pCMVSPORT6.0 vector from Life Technologies (part of the Gateway cloning system) to create ZMAXGI_3AS. The insert (5,278 bp) can be released from the vector by digestion with either *Hind*III or *Xba*I on the 5' side together with either *Eco*RV or *Eco*RI on the 3' end.

The LRP5 construct was generated from four independent partial clones. These clones were isolated from a LRP5 specific primed cDNA library. A partial LRP5 cDNA clone existed in the internal survey sequencing clone set as L236B_P0049E08 isolated from an oligo-dT primed HeLa cell cDNA library. This clone was truncated at the 5-primed end. In order to isolate more 5-prime containing fragments necessary to generate a full length cDNA, a *LRP5* gene-specific cDNA library was generated from Clontech human liver poly-A mRNA (catalog #6510-1, lot# 9060032) and Life Technologies SuperScript® Plasmid System for cDNA Synthesis and Plasmid Cloning kit (catalog no. 18248-013). This library was designated as L401. The manufacturer's protocol was carried out with the following modifications. 1)In both first and second strand synthesis reactions, DEPC-treated water was substituted for α-^{32}P-dCTP. 2)Reverse transcription was primed using oligonucleotides that were selected to be specific for the *LRP5* gene at approximately 1 kb intervals. These sequences were checked using the program BLAST against the public databases to ensure *LRP5* specificity. 3)Two separate reverse transcription reactions were performed. The first reaction, (A), was primed with oligonucleotides which annealed to the more 3' regions of *LRP5* as follows (SEQ ID NO:715-718):

47114: 5'-CGTACGTAAAGCGGCCGCTTGGCAATACAGATGTGGGA-3',

47116:5'-CGTACGTAAAGCGGCCGCAGTAGCTCCTCTCGGTGGC-3',

47118:5'-CGTACGTAAAGCGGCCGCGCTCATCATGGACTTTCCG-3' and

47120:5'-CGTACGTAAAGCGGCCGCGCACTGCTGTTTGATGAGG-3'.. The

second reaction, (B), used the previously mentioned four oligonucleotides, as well as (SEQ ID NO:719-721)
47108:5'-CGTACGTAAAGCGGCCGCGAGTGTGGAAGAAAGGCTGC-3',
47110:5'-CGTACGTAAAGCGGCCGCAGTAGAGCTTCCCCTCCTGC-3' and
47112:5'-CGTACGTAAAGCGGCCGCGTCCATCACGAAGTCCAGGT-3'. All oligonucleotides contained a *Not*I linker sequence and were used at a concentration of 0.02 ug/ul. 4)The *Sal*I-adapted cDNA from both reverse transcription reactions was size-fractionated by electrophoresis on 1% agarose, 0.1 ug/ml ethidium bromide, 1x TAE gels. The ethidium bromide-stained cDNA between 0.6 and 8.0 kb was excised from the gel. The cDNA was purified from the agarose gel by electroelution (ISCO Little Blue Tank Electroelutor®) using the manufacturer's protocol. The purified cDNA from reactions A and B were then pooled together. 5) The size-fractionated *Sal*I-adapted cDNA was ligated to *Not*I-*Sal*I digested pBluescript® (Stratagene, La Jolla, CA).

Ligated library cDNA (3 µl) was used to transform electrocompetent *E.coli* cells (ElectroMAX® DH10B cells and protocol, Life Technologies catalog no. 18290-015, BioRad *E. coli* pulser, voltage 1.8KV, 3-5 msec pulse). The transformed cells were plated on LB-ampicillin (100 ug/ml) agar plates and incubated overnight at 37°C. Approximately 10^6 colony forming units (cfu) were plated at a density of 50,000 cfu/150 mm plate. Cells were washed off the plates with LB media, and collected by centrifugation. Plasmid DNA was purified from the cells using the QIAGEN Plasmid Giga Kit and protocol (catalog no. 12191) at a final concentration of 2.05 ug/ul.

Two probes for use in library screening were generated by the polymerase chain reaction (PCR) using 100 ng of library L401 as template. Standard PCR techniques were used. A reaction mixture contained 10 pmol of each oligo primer; 0.2 mM each dATP, dTTP, dCTP and dGTP (PE Applied Biosystems catalog no. N808-0260); 1.5 units Expand™ High Fidelity Taq DNA polymerase and 1x PCR reaction buffer (Roche Molecular Biochemicals, catalog no. 732-641; 10 mM Tric-

HCl, 1.5 mM $MgCl^2$, 50 mM KCl, pH 8.3). The mixture was incubated at 99 °C for one minute, followed by 30 cycles of 96° C for 15 seconds, 50° C for 30 seconds, 72° C for 1 minute with a final incubation at 72° C for 7 minutes (MJResearch DNA Engine® Tetrat PTC-225). The first was generated using oligos (SEQ ID NOS:722-723) 107335:5'-CAGCGGCCTGGAGGATGC-3' and 107338:5'-CGGTCCAGTAGAGGTTTCG-3', which amplify a *Not*I-*Sal*I fragment of the *LRP5* gene. The second was generated using oligos (SEQ ID NOS:724-725) 107341:5'-CATCAGCCGCGCCTTCATG-3' and 107342:5'-CCTGCATGTTGGTGAAGTAC-3', which amplify a *Sac*I-*Kpn*I fragment of the *LRP5* gene. Both PCR products were purified using the Qiaquick® kit and the manufacturer's protocol (Qiagen catalog 28106) then subcloned into the vector pCRII-TOPO® (Invitrogen catalog no. K4600) following the manufacturer's protocol. Positive subclones were identified by restriction digestion of purified plasmid DNA (using standard molecular biology techniques) and subsequent DNA sequence analysis (ABI Prism BigDye Terminator Cycle® sequencing, catalog no. 4303154, ABI 377 instruments). Probe DNA was isolated by *Eco*RI restriction digestion (New England Biolabs, catalog no. R0101L) of the respective sequence-verified pCRII-TOPO® clone. Restriction fragments were size-fractionated by gel electrophoresis on 1% agarose, 0.1 ug/ml ethidium bromide, 1x TAE gels. Insert DNA was excised from the gel and purified using the Clontech NucleoSpin® Nucleic Acid Purification Kit (catalog no. K3051-2) following the manufacturer's protocol. The purified DNA fragment (25-50 ng) was labeled with Redivue ®$(a^{32}P)$-dCTP (Amersham Pharmacia, catalog no. AA0005) using the Prime-It II® Random Primer labeling Kit and protocol (Stratagene, catalog no. 300385). Unincorporated dCTP was removed with Amersham's NICK® column and protocol (catalog no. 17-0855-02).

Two rounds of screening library L401 were initiated to isolate fragments of the *LRP5* gene. In the first, forty-three 150 cm LB -100 ug/ml ampicillin agar plates were plated with primary transformants from L401 at a density of about

3,000-4,000 colonies per plate. This library was screened using the ^{32}P-labeled probe (*Not*I-*Sal*I fragment) as described above at 500,000-1,000,000 cpm/ml hybridization buffer, using standard molecular biology protocols. From this primary screen, 13 single colonies were identified based on positive hybridization to the LRP5 probe. Plasmid DNA, prepared using the QIAprep Spin Miniprep Kit and protocol (Qiagen Inc., catalog no. 27106), was analyzed by restriction digestion and sequence analysis as described above. cDNA clone # 44 was isolated from this screen and sequence verified to contain a partial LRP5 clone.

In the second library screen, one hundred-and-four 150 cm LB ampicillin 100 ug/ml agar plates were plated with primary transformants from L401 at a density of 3000-4000 colonies per plate. This library was screened using the ^{32}P-labeled probe (*Sac*I-*Kpn*I) exactly as previously described. From this primary screen, 48 colonies were identified based on positive hybridization to the *LRP5* probe. Since these colonies were not single colony isolates, a secondary screen was initiated where each of the 48 primary isolates was plated at a density of approximately 500 colonies per plate. These colonies were then screened exactly as the primaries using the labeled *Sac*I-*Kpn*I fragment as probe. Thirty-four of the 48 primary clones resulted in positive hybridization to the *LRP5* probe and were isolated as single colonies. Plasmid DNA was prepared and analyzed as described above. cDNA isolate #71_2 was isolated from this screen and sequence verified to contain a partial *LRP5* clone.

In all cases, the sequence of any *LRP5* isolate was compared to a reference sequence (i.e., the sequence of the wildtype *LRP5* allele from an affected member of the HBM kindred). This analysis was important since DNA polymorphisms had been reported for this gene in the literature. This reference sequence is predicted to encode a polypeptide of Genbank Accession No. AF077820.

The four independent partial clones used to prepare ZmaxIGI_3AS are as follows:

1) Bases 1-1366: A *Xba*I-*Sal*I fragment was obtained from a LRP5 cDNA construct, GTC.Zmax1_13. GTC.Zmax1_13 contains a 5075 BP insert containing the entire ORF of LRP5. The clone was blunt end cloned in the *Eco*RV site of pSTBlue-1. This clone was generated by fusing a 5' clone derived from screening a bone random primed cDNA library in a pBluescript™ II derivative with a 3' clone derived from a PCR product from a bone dT primed cDNA library in pBluescript™ II. PCR was performed using LRP5 specific forward primer (SEQ ID NO:726) 5'-GCCCGAAACCTCTACTGGACCGAC-3' and reverse primer (SEQ ID NO:727) 5'-GCCCACCCCATCACAGTTCACATT-3' using DNAzyme polymerase. The resultant 3.7 kb PCR product was cloned into PCR-XL-TOPO. To generate the full length clone the 5' and 3' plasmids were transformed into DM1 (dam-) from Gibco/BRL. The 5' plasmid was digested with *Xba*I and the 3' plasmid was digested with *Hind*III. The digested plasmids were filled in with T4 polymerase to generate blunt ends and cut with *Bcl*I. the 1.7 kb 5' fragment and 3.5 kb 3' fragments were gel purified, ligated together, and cloned in the *Eco*RV site of pSTBlue-1. It provides a short 5′ UTR, with coding sequence beginning at base 100. Furthermore, it carries some additional restriction sites at the 5′ multiple cloning site. This fragment also contains a DNA polymorphism relative to the Genbank Accession No. AF077820 sequence at position 558 resulting in an A (AF077820) to a G change; this mutation does not result in an amino acid difference (Pro).

2) Bases 1367-2403: This clone was obtained from a *LRP5*-gene primed cDNA library made from commercial human liver RNA described above. This fragment is a *Sal*I-*BamH*I piece of DNA obtained from isolate #44. The sequence is identical to Genbank Accession No. AF077820.

3) Bases 2404-4013: This *BamH*I-*Bss*HII fragment was obtained from isolate #71-2 from the same library as described above. At position 3456, there is a DNA polymorphism resulting in a G (AF077820) to an A. This nucleotide difference does not change the encoded amino acid (Val).

4) Bases 4014-5278: This *Bss*HII-*Not*I fragment came from an internal clone, L236B_P0049E08. It was obtained from an oligo-dT primed HeLa cell cDNA library. The stop codon occurs at base 4947. The clone contains 331 bp of 3′ UTR sequence, including a 120 bp poly-A tail followed by the *Not*I site. The 3′ *Not*I site used in this subcloning step is a result of an added linker that was introduced at the end of the poly-A tail during library construction. A DNA polymorphism is present at base 4515 resulting in a G (AF077820) to C change that is silent at the amino acid level (Leu).

To generate the 5' section of the LRP5 gene, the *Xba*I-*Sal*I fragment and *Sal*I-*Bam*HI fragment were ligated into the *Xba*I-*Bam*HI sites of pBluescript® (Stratagene). The 3' section of the gene was obtained by ligating the 1.61 kb *Bam*HI-*Bss*HII fragment from LRP5 isolate #71_2 to the 1.26 kb *Bss*HII-*Not*I fragment from L236B_P0049E08. These two fragments were ligated into the *Bam*HI-*Not*I sites of pBluescript®. The full length *LRP5* cDNA was engineered into the *Xba*I-*Not*I sites of the vector pCMVSPORT6.0 (Life Technologies) by ligation of the *Xba*I-*Bam*HI 5' section and the *Bam*HI-*Not*I 3' section. The resulting plasmid, ZmaxGI_3AS, contains an insert of 5278 bp from the *Xba*I site to the *Not*I site of the vector's multiple cloning site. This clone is in the antisense orientation with respect to the CMV promoter present in the vector. *LRP5* coding sequence begins at base 100 and ends at base 4947, followed by 331 bp of 3-primed UTR sequence including a 120 bp poly-A tail. This full length cDNA contains three DNA polymorphisms from the reference sequence (GenBank Accession No. AF077820) that do not alter the predicted amino acid sequence. These polymorphisms are at position 558 resulting in an A to G that maintains the proline residue; at position 3456 resulting in a G to A that maintains the valine residue; and, at position 4515 resulting in a G to C that maintains the leucine residue.

The sequence of ZMAXGI_3AS (Figure 25) also contains a DNA polymorphism relative to SEQ ID NO. 1 at base 4088 resulting in a C (SEQ ID NO: 1 and Genbank Accession No: AB017498) to T change that results in an amino

acid change at position 1330 of alanine to valine. This is consistent with the sequence determined in the wild-type allele from an affected member of the HBM kindred as well as with the published sequence of Genbank Accession No. AF077820. ZMAXGI_3AS also has 29 additional bases at the 5' end relative to SEQ ID NO: 1, as well as 129 bases at the 3' end consisting of an extra G, 120 bases of poly-A tract, and the *Not*I site.

Creation of the HBM mutation G171V

The HBM mutation that results in a predicted amino acid change from glycine to valine at amino acid 171 was introduced into the full length human LRP5 cDNA (plasmid ZMAXGI_3AS) using PCR to change the G at position 611 to a T. Introduction of the HBM mutation was done using oligos (SEQ ID NO:728) 107335: (5'-CAGCGGCCTGGAGGATGC-3') and (SEQ ID NO:729) 49513: (5'-CGGGTACATGTACTGGACAGCTGATTAGC-3'), which flank the endogenous *Not*I site of the *LRP5* gene. This method creates a new *PvuII* site at the 3' end of the PCR product. A second PCR reaction was completed using oligos which introduce a *Sca*I site at the 5′ end of the product and contains the endogenous *Sal*I site of *LRP5* in the 3-primed end. PCR products were purified using the QiaQuick® procedure (Qiagen Inc.); subcloned into the vector pCRII-TOPO (Invitrogen) as described above. Plasmid DNA was purified from single bacterial colonies and analyzed by restriction digest and subsequent sequence analysis, all as described above. The sequence-verified pCRII-TOPO clones were restriction digested with *Not*I-*Pvu*II and *Sca*I-*Sal*I, respectively. The resulting DNA fragments were size fractionated and purified as described above. These two fragments were then subcloned into the vector, pBluescript® that had been prepared by *Not*I-*Sal*I digestion. Both *Pvu*II and *Sca*I produce blunt ends when used to digest double stranded nucleic acids. Thus, the resulting ligated fragment fails to recreate either the *Pvu*II or *Sca*I site and contains only the consensus LRP5 sequence, with the exception of the newly introduced HBM mutation. To introduce the mutation into

the full length *LRP5* gene, this resulting plasmid was digested with *Msc*I and *Sal*I, while the 5' region of *LRP5* was obtained by *Xba*I-*Msc*I digestion of LRP5 plasmid GTC.Zmax1_13. These two fragments were ligated together into *Xba*I-*Sal*I digested pBluescript®, in effect creating a similar 1.366 kb *Xba*I-*Sal*I fragment. The only difference being that this construct contains the HBM mutation described above. The full length HBM cDNA then was assembled into pCMVSPORT6.0 exactly as described above for the *LRP5* gene, with the substitution of this newly created *Xba*I-*Sal*I fragment containing the HBM mutation. The entire cDNA insert was verified by DNA sequence analysis and the introduction of the HBM mutation was confirmed.

The resulting plasmid, HBMGI_2AS (Figure 24), contains an insert of 5,278 bp from the *Xba*I site to the *Not*I site of the vector's multiple cloning site. This clone is in the antisense orientation with respect to the CMV promoter in the vector. HBM coding sequence begins at base 100 and ends at base 4947, followed by 331 bp of 3-primed UTR sequence which includes a 120 bp poly-A tail. This full length cDNA contains three DNA polymorphisms from the reference sequence, which do not alter the amino acid sequence. These polymorphisms occur at position 558, resulting in an A to G change that maintains the proline residue; at position 3456 resulting in a G to A change that maintains the valine residue; and, at position 4515 resulting in a G to C change that maintains the leucine residue. Additionally, the HBM mutation is present at position 611 (G in LRP5 to T in HBM) which results in a predicted amino acid change of glycine to valine at amino acid position 171, as found in affected members of the HBM kindred. This insert sequence was used to generate the construct used for HBM over-expressing transgenic mice.

Transgene preparation

The examples provided herein are illustrations of how transgenic animals can be prepared. Additional transgenic animals can be prepared as would be known

in the art. See, for example, Glenn Monastersley et al., ed, Strategies in Transgenic Animal Studies (Amer. Soc. Microbiology 1995) and the references cited therein.

CMVβActin promoter-HBM cDNA (HBMMCBA)

To prepare the CMV*β*actin–HBM construct, pCX-EGFP, a plasmid containing the chimeric CMV*β*actin promoter, was purified as a 4778 bp *Eco*RI fragment. Subsequently, the HBM cDNA was excised from HBMGI_2AS as a 4994 bp *Xba*I/*Dra*I fragment, treated with Klenow fragment of DNA polymerase, ligated to *Eco*RI linkers, and digested with *Eco*RI. This fragment was then inserted into the *Eco*RI site of pCX-EGFP. A *Spe*I/*Hind*III 7265 bp CMV*β*actin–HBM fragment was purified for microinjection into mouse embryos.

Type I collagen promoter–HBM cDNA (HBMMTIC)

The rat type I collagen promoter–HBM construct was created by first replacing the pBS(SK-) (Stratagene) polylinker with another polylinker (i.e., comprising *Kpn*1 – *Spe*I – *Hind*III - *Bg*lII – *Nde*I – *Sal*I – *Sm*aI – *Eco*RI – *Pst*I – *Bam*HI – *Xba*I – *Sca*I – *Nco*I – *Cla*I – *Not*I – *Sac*II - *Sac*I), that is referred to as BS(SK-)A/D. The SV40 splice and poly $(A)_n$ *Xba*I-*Nco*I region (750 bp) from pcDNA I (Invitrogen, Inc.) was directionally cloned into BS(SK-)A/D. Next, a 4994 bp *Eco*RI HBM cDNA fragment (above) was cloned into the EcoRI site. A 3640 bp *Xba*I, type I collagen promoter fragment was subcloned into the *Xba*I site of BS(SK-) (Stratagene). The promoter fragment was then excised from BS(SK-) with *Sac*II, blunt-ended with T4 DNA polymerase, digested with *Spe*I, and ligated into the HBM BS(SK-) A/D construct, which was digested with *Nde* I, blunted with T4 DNA polymerase, and digested with *Spe*I. A *Spe*I/*Cla*I 9435 bp type I collagen–HBM fragment was purified for microinjection into mouse embryos.

CMVβActin promoter–LRP5 cDNA (Zmax1WTCBA)

The CMV*β*Actin promoter-HBM cDNA construct from above was used to generate the final plasmid. The following three fragments were ligated together: 1) a 6.34 kb *Xba*I-*Kpn*I backbone fragment from HBMMCBA; 2) a 0.64 kb *Xba*I-*Sap*I fragment from HBMMCBA containing the 3' end of the bActin promoter and the 5' end of the HBM cDNA; and 3) a 2.8 kb *Sap*I-*Kpn*I fragment derived from the LRP5 cDNA that contains the wild-type sequence. A 7.2 kb *Spe*I – *Hind*III CMV*β*Actin – LRP5 fragment was purified for micro-injection into mouse embryos.

Type I collagen promoter–LRP5 cDNA (Zmax1WTTIC)

The type I collagen – HBM cDNA construct from above was used to generate the final plasmid. The HBMMTIC plasmid was linearized with *Hind*III and cut with either *Sal*I to yield a 8.52 kb *Hind*III – *Sal*I fragment or *Sap*I to yield a 2.98 kb *Sap*I – *Hind*III fragment. A 2.8kb *Sap*I – *Sal*I fragment from the LRP5 cDNA containing the wild-type sequence was then ligated to the above two fragments to yield the final plasmid. A 9.4 kb *Spe*I – *Cla*I type I collagen – LRP5 fragment was purified for micro-injection into mouse embryos.

<u>Confirmation of transgene expression *in vitro*</u>

Plasmid constructs for HBMMCBA, HBMMTIC, Zmax1WTCBA and Zmax1WTTIC were transiently transfected into human osteoblast (HOB) cells to measure mRNA expression as a test for functionality.

Transient Transfections

HOB-02-02 cells are a clonal, post-senescent, cell line derived from the HOB-02-C1 cells (Bodine et. al, 1996, *J. Bone Miner. Res.* 11: 806-819). Like the parental cell line, the HOB-02-02 cells express the temperature-sensitive SV40 large T-antigen mutant, *tsA*209. Consequently, these cells proliferate at the permissive

temperature of 34°C, but stop dividing at non-permissive temperatures of 37°C or above. Also like the parental cell line, the HOB-02-02 cells are cultured with Growth Medium (D-MEM/F-12 containing 10% heat inactivated fetal bovine serum, 1% penicillin-streptomycin and 2 mM GlutaMAX-1) at 34°C in a 5% $C0_2$/95% humidified air incubator (Forma Scientific, Marietta, OH).

For the transient transfections, the HOB-02-02 cells were seeded with Growth Medium at 400,000 cells/well into 6-well plates and incubated overnight at 34°C. The cells were transfected with 0.3 mg/well of either the CMVβActin-HBM expression plasmid, the Type I Collagen-HBM expression plasmid or the corresponding empty vectors using LipofectAMINE 2000 transfection reagent according to the manufacturer's instructions (Life Technologies, Rockville, Md). After a 24 hr incubation at 34°C, the medium was changed, and the cells were incubated for an additional 24 hr at 39°C. At the end of this last incubation, the cells were rinsed with Hank's buffered salt solution. Total cellular RNA was then isolated using TRIzol® according to the manufacturer's instructions (GibcoBRL, Grand Island, NY). The RNA was treated with RNase-free DNase in order to remove contaminating DNA as previously described (Bodine et al., 1997, *J. Cell. Biochem.* 65: 368-387).

TaqMan® assay for mRNA expression

TaqMan® primers and probes were chosen based on human and mouse LRP5 cDNA sequences. The selected sequences were designed to be gene-specific by analysis of an alignment of human and mouse LRP5 (Zmax1) sequences as illustrated in Fig. 26.

TaqMan® quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) analysis of RNA isolated from human cells was performed as described by the manufacturer (PE Applied Biosystems, Foster City, CA) using the following primers and probe set:

Human Zmax1-1/HBM: (SEQ ID NOS:730-732)

Forward Primer: 5'-GTCAGCCTGGAGGAGTTCTCA-3'

Reverse Primer: 5'-TCACCCTTGGCAATACAGATGT-3'

Probe: 6-FAM-5'-CCCACCCATGTGCCCGTGACA-3'

Results from the experimental primers/probe set were normalized to human GAPDH levels using the multiplex protocol with the human GAPDH control kit from PE Applied Biosystems. Species-specific TaqMan® quantitative RT-PCR analysis of RNA isolated from murine cells and tissues was performed as described by the manufacturer (PE Applied Biosystems, Foster City, CA) using the following primers and probes sets:

Human Zmax-1/HBM-1: (SEQ ID NOS:733-735)

Forward Primer: 5'-CGTGATTGCCGACGATCTC-3'

Reverse Primer: 5'-TTCCGGCCGCTAGTCTTGT-3'

Probe: 6-FAM-5'-CGCACCCGTTCGGTCTGACGCAGTAC-3'

Mouse Zmax-1/HBM-1:(SEQ ID NOS:736-738)

Forward Primer: 5'-CTTTCCCCACGAGTATGTTGGT-3'

Reverse Primer: 5'-AAGGGACCGTGCTGTGAGC-3'

Probe: 6-FAM-5'-AGCCCCTCATGTGCCTCTCAACTTCATAG-3'

Results from the experimental primers/probe sets were normalized to 18S ribosomal RNA levels using the multiplex protocol with the 18S ribosomal RNA control kit from PE Applied Biosystems. A summary of these results is presented in Figures 17-20.

Production of transgenic mice

DNA microinjection

Transgene fragments for micro-injection were first purified on 1% agarose gels according to the GELase protocol from Epicentre Technologies. Fragments were then further purified on cesium chloride density gradients and extensively dialyzed against 5 mM Tris (pH 7.4), and 0.1mM EDTA.

Linearized DNA was microinjected into mouse embryos according to standard procedures. DNA was injected into primarily the male pronucleus of fertilized C57BL/6T mouse embryos. Injected embryos (n=20-35) were transferred to the oviduct (unilaterally) of day 0.5 post coitum pseudopregnant Swiss Webster embryo recipients. Offspring were tail-biopsied and genotyped at age 10-14 days.

Production of gene-targeted transgenic mice

Gene targeting vectors and probes

Two gene-targeting vectors were constructed for modification of the *LRP5* gene in embryonic stem (ES) cells. The two constructs, illustrated in Figure 16, designated as Zmax1-KI/KO A&B were designed to generate two types of mutations, a knock-out (KO) of the *LRP5* gene and a Cre recombinase dependent knock-in (KI) of a nucleotide substitution in order to create a mouse model (i.e., glycine 170 to valine amino acid substitution in mouse LRP5, of the HBM kindred.

Both gene-targeting vectors were constructed using genomic DNA of the mouse genomic DNA BAC clone 473P5 (Genbank Accession No. AZ095413) containing the first five exons of the mouse *LRP5* gene. This clone was isolated by Research Genetics (Huntsville, AL) from their mouse 129SvJ genomic BAC library using a polymerase chain reaction (PCR) screen for exon 3. A forward primer of the sequence (5'-GAGCGGGCAGGGATGGATGG-3')(SEQ ID NO:739) and a reverse primer of the sequence(5'-AGGTTGGCACGGTGGATGAAGC-3')(SEQ ID NO:740) were used to amplify exon 3 by PCR; the following thermal cycling

conditions were employed: for thirty cycles, 95°C for 0.5 minute, 55°C for 1 minute and 72°C for 1 minute. Identity of this clone with mouse *LRP5* was confirmed by sequencing exon 3 using the BAC clone DNA as template. PCR products were cloned using the pGEM-T-easy T/A cloning kit.

LRP5 knock-in/knock-out vector

The organization of the genomic BAC clone 473P5 was characterized by Southern blot analysis using subcloned exon 1, exon 2 and exon 3 as probes and by sequencing the region spanning exon 1 through exon 5. Two different constructs were prepared for the *LRP5 (Zmax1)* KI/KO targeting. These constructs (A and B) differ only in flanking arms of homology. Construct (A) contains a 6.5 kb *Bst*EII-*Xba*I 5' arm of homology and a 1 kb *Xba*I-*Eco*RI 3' arm of homology; whereas, construct (B) contains a 1 kb 5' arm of homology and a 6.0 kb 3' arm of homology. The constructs were prepared by ligating short and long arms of homology to a LoxP flanked cassette containing the neomycin resistance gene (MC1-Neo, Stratagene) and a synthetic transcriptional pause sequence (Promega).

Both *Zmax1*-KI/KO- targeting vectors (A and B) were modified to a G-to-T nucleotide substitution, encoding the G170V amino acid substitution, in exon 3. These modifications were introduced into by overlapping PCR mutagenesis using the wild type sequence of the short arm of homology as template. In addition, the 1 kb short arm of the *Zmax1*-KI/KO (B) targeting vector was modified to include a 5' terminal *Pme*I restriction recognition site. The 5' overlapping fragment was made using the forward primer of the sequence (5'-AAGCTTGTTTAAACTGGGCATGGTGGCACATGGTTGTAAT-3') (SEQ ID NO:741) and a reverse (mutagenic) primer of the sequence (5'-GGGCTTCCACCCAGTCAGTCCAGTACATGTACCT -3') (SEQ ID NO:742). The thermal cycling conditions utilized for thirty cycles are 95°C for 0.5 minute, 55°C for 1 minute and 72°C for 1 minute. The 3' fragment was made using the forward primer of the sequence

(5'-CTGACTGGGTGGAAGCACCCCGGATCGAGC-3') (SEQ ID NO:743) and a reverse (mutagenic) primer of the sequence (5'-GAATTCATCGGTACCTGTGCGGCCGCTTCATTG -3') (SEQ ID NO:744). The thermal cycling conditions utilized for thirty cycles are 95°C for 0.5 minute, 55°C for 1 minute and 72°C for 1 minute. The final overlapping PCR used 1 ml each of the 5' fragment and 3' fragment PCR reactions as template and amplification was performed using the forward and reverse primers of the 5' and 3' fragments respectively and the same thermal cycling parameters. The final PCR product was cloned using the pGEM-T-easy T/A cloning kit. The mutagenized exon 3 was excised from *Zmax1*-KI/KO (B) and transferred to *Zmax1*-KI/KO (A) as a 600 bp *Bsm*BI-*Xba*I fragment.

Probes for screening for and characterization of *Zmax1*-KI/KO (A) gene targeted ES cell clones are prepared by subcloning restriction fragments of BAC clone 473P5. The 5' outside probe is a 400 bp *Nde*-*Bst*EII fragment, and the 3' outside probe is a 500bp *Eco*RI-*Bst*XI fragment.

The outside probes for Zmax1-KI/KO (B) are prepared by PCR cloning genomic fragments flanking and immediately adjacent to the targeting vector region of homology. The 5' outside probe used for Zmax1-KI/KO (B) is a 498 bp fragment generated using the forward primer of the sequence (5'-TGAGATGTCCTGTCTGTGGC-3') (SEQ ID NO:745) and a reverse primer of the sequence (5'-TCCTTCCTTCCCTACAGTTG-3')(SEQ ID NO:746). The thermal cycling conditions utilized with these probes for thirty cycles are: 95°C for 0.5 minute, 55°C for 1 minute and 72°C for 1 minute. The 3' outside probe is a 600 bp fragment generated using the forward primer of the sequence (5'-CCTAAGGATCTCCTTGTGTCTGTGG-3')(SEQ ID NO:747) and a reverse primer of the sequence (5'-CTGCAGCAGGTCAGTAGCCTGC-3')(SEQ ID NO:748). The thermal cycling conditions utilized with these probes for thirty cycles were: 95°C for 0.5 minute, 55°C for 1 minute and 72°C for 1 minute.

Both probes are specific for the *LRP5* gene in genomic southern analysis. PCR products are cloned using the pGEM-T-easy T/A cloning kit.

A probe for ribonuclease protection analysis of *LRP5* mRNA structure and transcription levels was prepared by PCR cloning a cDNA fragment containing exon 3 through exon 4. The PCR reaction used a complete cDNA as template, a forward primer of the sequence (5'-TGAGATGTCCTGTCTGTGGC-3')(SEQ ID NO:749), a reverse primer of the sequence (5'-TCCTTCCTTCCCTACAGTTG-3') (SEQ ID NO:750) and the following thermal cycling conditions for thirty cycles; 95°C for 0.5 minute, 55°C for 1 minute and 72°C for 1 minute. The PCR product is cloned using the pGEM-T-easy T/A cloning kit.

Gene targeting in ES cells

For gene targeting, embryonic stem (ES) cells are electroporated with 50 mg of linearized targeting vector and selected in 200 mg/ml G418 for 7-10 days beginning the day after electroporation. G418 resistant clones are picked, expanded and cryopreserved. Resistant clones were screened for homologous recombination by an *Eco*RI genomic Southern restriction fragment length analysis using the 5' outside probe, which detects the wild type and targeted alleles of LRP5 as 4 kb and 5 kb fragments, respectively. Gene targeted ES cell clones are thawed, expanded, and characterized by *Sca*I genomic restriction fragment length analysis using the 3' outside probe, which detects the wild type and targeted alleles of LRP5 as 9 kb and 8 kb fragments, respectively. Gene targeted clones are also characterized by sequence analysis of LRP5 exon 3 to ensure that the G to T substitution was included in homologous recombination.

Production of gene targeted mice by blastocyst injection

To generate chimeric mice, gene targeted ES cell clones are thawed, expanded and 9-14 ES cells injected into the blastocoel of 3.5 post coitum (p.c.) host C57BL/6 blastocysts. Injected blastocysts (12-17) are then transferred

unilaterally into the uterus of 2.5 p.c. pseudopregnant Swiss Webster embryo recipients and allowed to develop to term. Chimeric males are back-crossed to 129SvEv females and tested for transmission of the targeted allele by PCR geneotyping with primers specific to the neomycin resistance gene.

In vitro deletion of the Neomycin resistance cassette via Cre recombinase

To generate LRP5 KI mice from the LRP5 KI/KO mice the Neomycin resistance (KO) cassette was deleted by micro injection of a Cre expressing plasmid (2 mg/ml) into the male pronucleus of LRP5 KI/KO pre-fusion zygotes. Deletion of the KO cassette was confirmed by PCR analysis of the cassette insertion site.

Genotyping transgenic mice

Genomic DNA was isolated from mouse tail snips by digestion in 500 ul buffer containing 50 mM Tris-HCl, (pH 7.2), 50 mM EDTA, (pH 8.0), 0.5% SDS and 0.8 mg/ml proteinase K. Samples are incubated at 55^0 C with shaking overnight. A 10 μl aliquot was heat-inactivated at 99° C for 5 minutes and diluted 1:20 in water. For PCR, 1 μl of the diluted DNA was amplified under the following conditions: Denature: 96^0 C for 4.5 min; 45 cycles: 96 ^{0} C for 30 sec; 63^0 C for 1 min; 72^0 C for 1 min; Extension: 72^0 C for 5 min; 4^0 C hold.

The following primer sets are used for genotyping:

HBMMCBA:

<u>5' primers: 296 bp fragment</u>

forward: 5′-GCT TCT GGC GTG TGA CCG GCG-3′ (SEQ ID NO:751)

reverse: 5′-GCC GCACAG CGC CAG CAG CAG C-3′ (SEQ ID NO:752)

<u>3' primers: 400 bp fragment</u>

forward: 5′-CAC CCA CGC CCC ACA GCC AGT A-3′ (SEQ ID NO:753)

reverse: 5′-ATT TGC CCT CCC ATA TGT CCT TCC-3′ (SEQ ID NO:754)

HBMMTIC:

5' primers: 382 bp fragment

forward: 5′-TTC CTC CCA GCC CTC CTC CAT CAG-3′ (SEQ ID NO:755)

reverse: 5′-GCC GCA CAG CGC CAG CAG CAG C-3′ (SEQ ID NO:756)

3' primers: 524 bp fragment

forward: 5′-GAA TGG CGC CCC CGA CGA C-3′ (SEQ ID NO:757)

reverse: 5′-GCT CCC ATT CAT CAG TTC CAT AGG-3′ (SEQ ID NO:758)

Confirmation of genotype by Southern analysis

Mouse genomic DNA was digested with *Eco*RI and probed with a 1.0 kb *Sal*I-*Bam*HI restriction fragment from the *LRP5* cDNA. The probe hybridizes to a 5 kb fragment in transgene positive animals.

Phenotyping

Both *in vivo* and *ex vivo* assays are used to evaluate the phenotype in transgenic mice. Two strains of wild-type mice, namely C57BL/6 and 129 SvEv, are studied to provide control data for phenotypic evaluation in transgenic and gene-targeted mice. In addition, non-transgenic littermate animals are used as controls.

In vivo analysis

pDXA: Wild-type and transgenic mice are anesthetized, weighed and whole-body X-ray scans of the skeleton generated using the LUNAR small animal PIXImus device. Scans are begun when the mice are weaned (i.e., at 3 weeks of age) and repeated at 2 week intervals. Wild-type animals are scanned at 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 27, and 29 weeks. Scanning of transgenic animals would be performed for periods up to 17 weeks. Scans are analysed for BMD (bone mineral density), BMC (bone mineral content), TTM (total tissue mass), and % fat for various body regions.

Faxitron radiographs: Following pDXA scanning of anesthetized animals, an additional X-ray was taken using a Faxitron device allowing measurement of bone size.

Calcein labeling: Animals are dosed with 15 mg/kg calcein intraperitoneally on two consecutive occasions. The first dose was given 9 days before euthanasia and the second given 2 days before euthanasia allowing measurement of bone formation rate.

Ex vivo analysis

RNA isolation: Total RNA was isolated from tibia and other tissues using TRIzol® to determine mRNA expression.

pQCT: The right femur was cleaned of soft tissue and stored in 70% ethanol for determination of total and trabecular density of the distal metaphysis and cortical density of the mid-shaft.

MicroCT: The right femur was used to determine trabecular indices of the distal metaphysis.

Histology: The right femur was used to determine bone area and static and dynamic parameters of the distal metaphysis.

Bending strength: The left femur was cleaned of soft tissue and stored at -20^0 C prior to analysis of 3-point bending strength of the mid-shaft.

Compressive strength of vertebra: The entire spine was removed from T10-L6-7. Soft tissue was left on and the spine frozen at -20°C until analysis. Compressive strength was measured at the L5 vertebra.

Serum: Animals are euthanized and serum prepared from blood to measure total cholesterol, triglycerides, osteocalcin, and other biochemical surrogate markers.

Histological analysis: Examples include immunocytochemistry such as *in situ* hybridization of osteogenic markers and TUNEL staining of cells undergoing apoptosis.

Results

Confirmation of expression from transgenic plasmid constructs

The HBM (HBMMCBA and HBMMTIC) and wild-type (Zmax1WTCBA and Zmax1WTTIC) plasmid constructs were transiently transfected into HOB-02-02 cells, which have a very low endogenous level of LRP5 expression. Two days after transfection, RNA was isolated and TaqMan® quantitative RT-PCR was performed to determine the mRNA levels of LRP5/HBM in the cells. To control for contaminating plasmid DNA, PCR was performed with or without the prior RT step, in the absence of the RT step, only very low levels of LRP5/HBM mRNA were detected. However, with the RT step, a 1000-fold increase in HBM and LRP5 mRNA was observed in cells transfected with CMV-bActin-promoter constructs as compared to those transfected with the CMV-b-Galactosidase control. The type I collagen promoter constructs showed approximately 10-fold increases in HBM and LRP5 mRNAs, which is consistent with the weaker nature of this promoter compared to the CMVβActin promoter. See Figure 17.

Species specific Taqman® reagents for HBM/LRP5 expression

Species specific TaqMan® primer and probe sets for LRP5/HBM were developed. In a series of experiments using HOB cells and mouse MC-3T3-E1 osteoblastic cells, LRP5/HBM mRNA was measured in a mouse background, and vise versa. These reagents useful for the detection and quantization of species-specific expression. As demonstrated in Fig. 18, the primer sets are species specific in the mouse and human cell lines. Further, Fig. 19 demonstrates the quantitative measurement of human LRP5 RNA in a background of mouse RNA. These TaqMan® sets can be used to determine the levels of human or mouse HBM or LRP5 (or other HBM-like variant) message that are being expressed in the mouse transgenic lines.

The species-specific TaqMan® reagents are novel tools for the characterization of both endogenous LRP5 mRNA levels and human LRP5/HBM mRNA levels in the transgenic mouse tissues. These tools have several advantages over other conventional methods, such as Northern hybridization and standard RT-PCR. Some of these advantages are as follows: (1) specificity, since only a small region (<100 bps) is amplified primers and probes are chosen to sequence regions predicted to have no or minimal cross-reactivity; (2) speed, since the procedure is less labor intensive; (3) accuracy, since it is truly quantitative; and (4) sensitivity, since it requires only small amounts of starting material (i.e., RNA) and the signal-to-noise ratio is high. These advantages are especially important for analyzing mRNA levels in bone, because it is difficult to obtain large amounts of RNA from bone. Thus, the primer sets developed for TaqMan® analysis of a HBM and LRP5 expression are important embodiments of the present invention. One skilled in the art will recognize that the primers described here are preferred embodiments; modifications such as extension or truncation of a primer or base substitution are encompassed by the present invention so long as the resultant nucleic acid continues to perform substantially the same function.

HBM expression in transgenic mice

Eight transgenic founder animals were produced for the CMVbActin (HBMMCBA) construct and a breeding program initiated to establish lines. Expression of mRNA determined by Taqman® analysis, shown in Fig. 20, showed variable levels of bone expression in 4 lines. In tibia, expression levels (relative to endogenous LRP5 in HOB-03-C5 cells) showed the following range: line 18 (x10-11 fold); line 2 (x7-10 fold); line 13 (x1-2 fold) and line 28 (x1 fold). Expression was also detected in other tissues as expected based on the known activity of the promoter. For lines 2 and 13, the highest levels of HBM expression were found in the heart. A Taqman® genotyping assay will screen for potential homozygous animals.

Six transgenic founder animals were produced for the type I collagen (HBMMTIC) construct, and a breeding program initiated to establish lines. Expression of mRNA was found in two lines initially tested. In line 19, expression was 7-8 fold and 19-20 fold greater than LRP5 in HOB-03-C5 cells in tibia and femur, respectively. In line 35, a low level of expression was detected in tibia and femur.

In vivo pDXA measurements of HBM transgenic mice

HBMMCBA construct

Analysis of transgenic mice, illustrated in Fig. 21 (A-C), at the 5 week and 9 week time-points showed that one line tested to date had greater BMD values compared to control. At 5 weeks, HBMMCBA line 2 (n=11) BMD in femur, spine and total body was 21%, 24% and 10% greater respectively, than wild-type control. At 9 weeks (n=3), these increases in BMD amounted to 19%, 32% and 12%, respectively. Over 17 weeks, the percent increase in line 2 relative to wild-type controls was 10%, 11%, and 8%, respectively.

HBMMTIC construct

Analysis of transgenic mice, illustrated in Fig. 21 (D-F), at the 5 week and 9 week timepoints showed that two lines tested had significantly greater BMD values as compared to control animals. At 5 weeks, HBMMTIC line 19 (n=5) BMD in femur, spine and total body was 63%, 70% and 41% greater respectively than the wild-type control. At 9 weeks (n=2), these increases in BMD amounted to 52%, 64% and 37% respectively. Over 17 weeks, 35%, 40%, and 28%, respectively. At 5 weeks, HBMMTIC line 35 (n=1) BMD in femur, spine and total body was 4%, 47% and 6% greater, respectively than the wild-type control. At 9 weeks (n=3), these increases in BMD amounted to 32%, 43% and 18% respectively. At 17 weeks, the percent increases in BMD were 20%, 33% and 19%, respectively. Two additional HBM transgenic lines 188 and 189 where HBM expression is driven by

the type I collegen promoter have been shown to have a HBM phenotype. Over 17 weeks, line 188 demonstrates a 23%, 35%, and 22% increase in BMD for femur, spine and total bone.

Overall, the BMD results from the transgenic mice show similarities in magnitude to the phenotype observed in the HBM affected kindred (Johnson et al., 1997, *Am. J. Hum. Genetics*, 60:1326-1332). For example, spinal BMD measured in affected individuals is approximately 34-63% greater than non-affected family members. The data for spinal BMD from the transgenic animals ranges from ~30-70% greater than normal at 9 weeks of age.

Ex vivo analysis of transgenic mice

In order to further examine increases in bone density that were detected in select transgenic lines through monitoring of the animals by non-invasive bone imaging, necropsies were performed on animals of these lines at 5 and 9 weeks of age for direct bone densitometric and histologic analysis. The left femur was isolated, cleaned and positioned in an XCT Research peripheral Quantitative Computed Tomograph (pQCT; Stradtec Medizintechnik, Pforzheim, Germany). The distal end of the femur was located and pQCT scanning was initiated 2.5 mm proximal from this point for total and trabecular measurements. The pQCT scan for cortical measurements was initiated 3.5 mm proximal from the first scan (i.e., 6 mm proximal from the distal end). The pQCT scans were 0.5 mm thick, had a voxel (i.e., three dimensional pixel) size of 0.07 mm, and consisted of 360 projections through the slice. After the scans were completed, the images were displayed on the monitor and a region of interest, including the entire femur for each scan, was outlined. The soft tissue was automatically removed using an iterative algorithm, and the density of the remaining bone (total density) in the first slice was determined. The outer 55% of the bone was then peeled away in a concentric spiral and the density of the remaining bone (trabecular density) of the first slice was reported in mg/cm^3. In the second slice, the boundary between

cortical and trabecular bone was determined using an iterative algorithm, and the density of the cortical bone was determined.

Analysis of Line 2 F1 CMVβActin-HBM 5 week old transgenic animals revealed that total density, trabecular density and cortical density, were 20%, 37% and 4% higher, respectively, in the transgenic male mice versus the non-transgenic males. At later timepoints, the difference between the transgenic and non-transgenic animals in this line is diminished. However, in type I collagen-HBM transgenic males at 5 weeks old from Line 19, an even more dramatic increase in bone density over their non-transgenic littermates was evident. Total density, trabecular density and cortical density were 53%, 104 % and 5% higher, respectively. In the Line 19 animals, the phenotype was found to be maintained beyond 9 weeks of age with elevated total and trabecular bone density as seen in Table 7. At 17 weeks, total and trabecular density were increased 46%, 202%, respectively. The effects on the total trabecular parameters in line 19 at all three time points were statistically significantly higher ($p<0.001$). A somewhat different pattern of bone phenotypic expression was evident from males of type I collagen-HBM transgenic Line 35. At 5 weeks of age total, trabecular and cortical density were only marginally higher (7%, 4% and 4%, respectively). However, at 9 weeks of age a clear and statistically significant increase in these parameters became evident as seen in Table 7. Total and trabecular bone densities remain elevated in Line 35 through 17 weeks of age.

Two additional HBM transgenic lines with the type I collagen promoter have been studied that show dramatic high bone density phenotypes similar to Line 19. In males from both line 188 and 189, total bone density was increased by 40% relative to non-transgenic animals. Trabecular bone density was increased 75% at 9 weeks of age. At 17 weeks, total and trabecular density of Line 188 males was 42% and 161% above control animals such as non-transgenic littermates. These values are consistent with the effects seen in Line 19 at this age. Females in Line 188 show 36% and 144% increases at 9 weeks and 26% and 148% increases in

total and trabecular bone density respectively. Females from Line 189 had total and trabecular densities that were increased by 27% and 64% at 9 weeks, and 15% and 84% at 26 weeks of age.

The occurrence of different patterns of age-related expression of the phenotype is not unexpected, particularly with the "bone specific" type I collagen transgene, which is influenced by stage of bone cell differentiation. Both Line 2 and Line 19 animals at 5 weeks of age express comparable levels of HBM mRNA in tibia samples, and these levels are significantly greater (>7-8-fold) than other lines that show no apparent bone phenotype at this age. Line 19, which is driven by the type I collagen promoter, unlike line 2, shows very low expression of the transgene in tissues other than bone. At 5 weeks of age, Line 35 animals show low level expression in bone and none in other tissues. Immunohistochemistry of calvarial bone sections using an HBM/LRP5 specific antibody reveals much more intense staining in bone cells of transgenic animals from Lines 2 and 19 at 5 weeks of age and from Line 35 at 9 weeks of age versus their non-transgenic littermates.

The findings revealed by pQCT analysis were further examined under greater resolution using μCT instrumentation (Scanco). The femur was positioned such that the region being imaged includes the distal end of the femur extending approximately 4 mm proximally with the view being perpendicular to the axis of the articulating cartilage. The reference line for beginning the μCT measurement was placed to minimally overlap the growth plate and extends proximally for 200 scan slices (9 mm thickness). After completing the μCT measurement, the first slice in which the condyles have fully merged was identified. A region of interest was outlined to include a maximum amount of the trabecular space, while excluding the cortex. For the first thirty slices, regions of interest were drawn every five slices and merged. For the remaining 105 slices, regions of interest were drawn every 10-20 slices. The more regular the trabecular space, the less frequently a region of interest needed to be drawn. Each region of interest was merged with its predecessor after it was drawn. After regions of interest had been established for

all 135 slices, three dimensional evaluation was performed using a threshold setting of 350.

The increased bone densities identified by pQCT were confirmed and extended by μCT to include elements of bone architecture. In the Line 2 transgenic animals, μCT bone volume/total volume, connectivity density and trabecular thickness were 50%, 83% and 12% higher, respectively. Both the connectivity density and trabecular thickness indices suggest that the increased density is also associated with increased structural strength. Bone surface/ bone volume was lower by 17% in the transgenic males, which may suggest that there may be fewer resorptive surfaces and pits. The trabecular bone response was further confirmed by histological evaluation of non-decalcified, Goldner's stained sections, which revealed 36% greater bone mineral area in the distal femoral metaphysis of the transgenic males. Dynamic histomorphometric analysis revealed that a substantial increase in bone mineral apposition rate (+100%), as determined by calcein double labeling, may be partially responsible for the increased bone in the transgenics.

The dramatic effects evident by pQCT on trabecular bone in Line 19 were supported by μCT evaluation where bone volume/total volume, trabecular number, trabecular thickness and connectivity density were found to be 130%, 45%, 30% and 121% higher, respectively, in the transgenic males. All of these effects were statistically significant with $p < 0.01$. The bone phenotype seen at 5 weeks of age in Line 19 was maintained in 9 week-old animals where bone volume/total volume, trabecular number and connectivity density were significantly higher than in the non-transgenic littermates as seen in Table 7.

μCT analysis of the Line 35 transgenics revealed a somewhat different pattern than the other two lines. In contrast to only modestly increased density indicated by pQCT in 5 week-old females from Line 35, a statistically significant effect ($p < 0.01$) was seen with μCT, which has greater image resolution and encompasses a larger volumetric sample. Bone volume/total volume, trabecular thickness and connectivity density were 35%, 9% and 27% higher. A similar result

was seen in 5 week-old males from Line 35 where bone volume/total volume and connectivity density increases of 37% and 45%, respectively, were evident by μCT analysis, where only slight increases were revealed by pQCT. The differences between the Line 35 transgenic males and their non-transgenic littermates appeared to increase with age such that statistically significant increases in total density (28%) and trabecular density (52%) were evident by pQCT at 9 weeks of age. The μCT results support an age-related divergence in bone phenotype in this line and show that differences between transgenic and non-transgenic animals. In terms of bone volume/total volume and connectivity density, these parameters more than doubled those seen at 5 weeks to 97% and 188%, respectively. The bone volume increases seen in the transgenic animals is in agreement with a significant increase in this parameter that was detected in a bone biopsy sample from an adult male affected member of the HBM kindred. The other parameters that were found to be affected in the transgenic lines may reflect changes that lead to an increased peak/adult bone mass, which in this strain of mice occurs between the ages of 17-20 weeks.

Immunohistochemistry of the calvaria from Line 19 has revealed strong expression of the transgene in pre-osteoblasts and osteoblastic cells lining the periosteum, as well as in osteocytes present in mineralized bone. Periosteal osteoblasts in the transgenics appeared plump and cuboidal, indicative of cells actively secreting extracellular matrix. In contrast, periosteal cells of the nontransgenic littermates appeared as flat, lining cells. Staining for alkaline phosphatase, an osteoblast differentiation and functional marker, was elevated confirming the active secretory status of the cells in the transgenics compared to the controls. Further analysis has revealed a reduced number of TUNEL-positive osteocytes, osteoblasts and stromal cells in transgenic mouse calvaria suggesting a reduction in apoptosis. In calvariae from 9 week-old male non-transgenic mice there were 30.9 $\pm$ 1.8 ((n=9) apoptotic osteoblasts/stromal cells per mm^2 were whereas in calvariae from HBM transgenics there were 11.6 $\pm$ 2.8 (n=9) apoptotic

osteoblast/stromal cells per mm^2. Taken together these results indicate that the increased BMD in the transgenics is due to increased osteoblast number and activity, which could in part be due to their increased functional lifespan.

The bone density and bone architectural changes seen in the over-expressing transgenic lines would suggest potentially greater bio-mechanical strength. This was tested directly by evaluating 3-point bending strength of femurs from 5 week old Line 19 males. The femora were cleaned of soft tissue and the femoral length measured using a digital caliper. Periosteal and endosteal circumferences, as well as cortical thickness, were measured 6 mm from the distal end of the bone using pQCT. The femur was then placed on a fixture so that the center of mid-shaft was at an equal distance from fixed supports located 5 mm apart. The cross bar of an Instron 5543 load device was placed over the mid-shaft and a force applied at a speed of 1 mm/minute until fracture occurred. A force vs. displacement curve was generated and peak load determined using Instron Merlin software. There was a 75% increase ($p<0.01$) in strength that appears to be due to an increase in periosteal circumference leading to an increase in cortical thickness. Thus, it appears that the changes in bone density and bone geometry, as seen in the HBM transgenic animals, do translate into increases in biomechanical strength.

In view of the association of HBM/LRP5 within the class of LDL related receptor proteins, it was of interest to determine whether the mutation might affect lipid profiles. Indeed, lipid studies in the HBM kindred (i.e., 8 affected and 7 unaffected members) have revealed that triglyceride and VLDL levels are statistically lower in the affected members. Serum samples from the transgenic lines were analyzed on a Hitachi 911 instrument using Boehringer Mannheim (for Cholesterol) and Roche (for triglycerides) reagents. The cholesterol was measured via *o*-quinone imine dye (which is formed following enzymatic reactions with cholesterol) photometrically at 505 nm at 37°C. Enzymatic methods for triglyceride measurements are based on determination of the glycerol part of triglyceride after hydrolysis of triglycerides and fatty acids. The end dye product of

enzymatic reaction was measured at 505 nm. In 5 week old male Line 2 transgenics, although serum cholesterol was only slightly reduced, serum triglyceride levels were reduced by 26% in the transgenics versus their non-transgenic littermates. In a limited sample of Line 2 animals at 9 weeks of age, triglyceride levels remained 20% lower. Similarly, at 5 weeks of age triglyceride levels in male transgenics from Line 19 were 32% lower. In contrast, at 5 weeks of age both male and female transgenics of Line 35 did not have lower triglyceride levels. The fact that the 5 week old Line 35 animals did have statistically greater bone volume/total volume suggests that the lipid change may not be directly related to the skeletal phenotype. This would appear to be supported by the fact that the Line 35 animals at 9 weeks of age had only slightly reduced triglyceride levels (11%) but exhibited substantially higher bone density than at 5 weeks of age. Due to the different levels and sites of expression of the transgene in these lines we can not rule out the possibility that serum lipid levels could serve as a surrogate marker for agents favorably affecting a bone phenotype through HBM/LRP5.

These and other transgenic lines based on HBM or HBM-like genes will serve as valuable models for exploring the nature of bone homeostasis. Bone density in all species accommodates to its customary loading conditions. In the HBM kindred and in the transgenic lines, the sensor/effector systems of the skeleton appear to perceive greater load signals resulting in greater bone density. Experimental models have been established showing that increased bone loading can lead to increased bone density and that unloading or disuse leads to a loss of bone density. Evaluating the histological, biochemical and genetic responses of the skeleton of the transgenic animals in these experimental paradigms will yield much information regarding the sensor/effector system responsible for bone homeostasis. The application of the transgenic animals in other established models of altered bone turnover, including but not limited to steroid deficiency-induced osteopenia and aging-related osteopenia will provide further insight into the role of LRP5 in

bone homeostasis and the nature of the favorable changes induced by the HBM mutation.

LRP5 over-expression in transgenic mice

In order to evaluate the role of overexpression of wild-type LRP5 and for contrast with the effects of HBM, transgenic mice have been created that express LRP5 driven by the type I collagen promoter. Statistically increased total and trabecular femoral bone density is observed at 9 weeks of age in one of these lines (LRPWWTTIC-19). Although not as great as seen in the HBM transgenic lines, the observations are supported by μCT measurements that show significant increases in bone volume and connectivity density. A comparison of percentage changes in skeletal parameters for HBM and LRP5 transgenic mice relative to non-transgenic mice at 9 weeks is shown in Table 7 below:

Table 7

Line (Gender)	Total Density	Trabecular Density	BV/ TV	Connectivity Density	Trabecular #	Trabecular Thickness
HBM-19 (M)	60	146	252	348	55	47
HBM-19 (F)	59	222	206	193	56	44
HBM-35 (M)	28	52	97	188	31	11
LRP5-19 (F)	10	41	35	47	15	4.3

Further evaluation of another LRP5 transgenic line did not show significant pQCT values as a group has revealed on individual analysis that the level of expression of the transgene is associated with the skeletal phenotype parameters. While overexpression of the wild type receptor produces an anabolic bone phenotype, the phenotypic change is greatly magnified by the HBM mutation.

HBM gene-targeting

The LRP5 KI/KO gene-targeting vector is electroporated into 129 SvEv, C57BL/6 ES and 129 ES cells. Restriction fragment length analysis of genomic DNA and sequencing of PCR amplified fragments can be used to identify gene targeted clones. The knock-in version of the gene-targeting vector allows for the introduction of the HBM mutation into the endogenous LRP5 genomic locus with minimal impact on the mouse genome. It permits the production of the HBM protein in a more natural environment, i.e. not in an over-expression model such as the transgenic mice or transfected cell lines. The knock-out version of the gene-targeting vector was engineered to contain a transcriptional stop sequence that has the potential to result in loss of one functional LRP5 allele. Breeding heterozygous animals with this mutation leads to the production of embryos homozygous for the null allele. In a different design of the gene-targeting vector, lox P sites can be positioned to facilitate production of a conditional knock-out of the endogenous LRP5 gene. In the presence of Cre recombinase, a critical region of the LRP5 gene would be deleted in between the lox P sites, thus resulting in the potential loss of one functional allele. Animal breeding would then be used to create homozygotes with a null allele. Other recombinase enzyme systems, such as flp recombinase in combination with cognate frt sites, could be used to create the deletion. The recombinase could be administered in a number of ways as described earlier, including plasmid injection into embryos and use of transgenic animals expressing Cre. The promoter used to drive expression of Cre could be chosen in a manner that would result in ubiquitous or tissue-specific deletion of the LRP5 gene thus resulting in a conditional knockout. In a further embodiment expression of the Cre enzyme itself could be made conditional using inducible systems such as GeneSwitch and Tetracycline paradigms.

LRP6 gene targeted knock out mice

LRP6 knock-out mice were generated using Omnibank ® embryonic stem (ES) cells carrying a gene trap vector which inserted into the first intron of the LRP6 gene. The insert location was determined to be the LRP6 gene by an Omnibank Sequence Tag (OST) generated by reverse transcription PCR (RT-PCR) of a fusion transcript comprised of 5' gene trap vector sequence spliced to the host gene transcript 3' of the insertion site. The gene trap vector functionally knocks out the mouse LRP6 gene by forced spicing of LRP6 exon 1 to the IRES-LacZ-PolyA element of the gene trap, preventing transcrition of LRP.

Chimeric mice were generated with ES cells, identified as OST38808, by injection into C57BL/6 albino host blastocyts which were then transferred to pseudopregnant females and allowed to develop through birth. Germline chimeras were backcrossed to 129SvEVBrd strain mice to maintain the knockout allele of LRP6 on an inbred 1298SvEvBrd genetic background. Germline transmission of the LRP6-KO allele was identified by PCR amplification of a gene trap specific sequence. Heterozygous LRP6-KO mating pairs were used for continued breeding. The genotype of wt and LRP6-KO progeny is determined by tail DNA PCR.

Measurements of bone density at 9 weeks of age in female heterozygous knock-out mice has shown significant ($p < 0.05$) decreases in bone volume, trabecular number, and trabecular thickness as measured by uCT. These results are consistent with the hypothesis that LRP6 is also involved in modulating bone density and is a target for development of therapies and drugs. Accordingly, LRP6 transgenic animals and transgenic animals expressing bone modulating variants of LRP6 are contemplated within the scope of the invention.

Uses of transgenic animals and cells

The transgenic animals and cells of the present invention are useful tools in methods for identifying surrogate markers for the HBM phenotype. The surrogate markers provided by the present invention are also useful tools for the assessment and screening of prospective treatments. Individuals carrying the *HBM* gene have elevated bone mass. The *HBM* gene causes this phenotype by altering the activities, levels, expression patterns, and modification states of other molecules involved in bone development. Using a variety of established techniques, it is possible to identify molecules, preferably proteins or mRNAs, whose activities, levels, expression patterns, and modification states are different between systems containing the *LRP5* gene and systems containing the *HBM* gene. Such systems can be, for example, cell-free extracts, cells, tissues or living organisms, such as mice or humans. For a mutant form of LRP5, a complete deletion of LRP5, mutations lacking the extracellular or intracellular portion of the protein, or any other mutation in the *LRP5* gene may be used. It is also possible to use expression of antisense LRP5 RNA or oligonucleotides to inhibit production of the LRP5 protein. For a mutant form of HBM, a complete deletion of HBM, mutations lacking the extracellular or intracellular portion of the HBM protein, or any other mutation in the *HBM* gene may be used. It is also possible to use expression of antisense HBM RNA or oligonucleotides or RNA interference methodologies to inhibit production of the HBM protein.

Molecules identified by comparison of LRP5 systems and HBM systems can be used as surrogate markers in pharmaceutical development or in diagnosis of human or animal bone disease. Alternatively, such molecules may be used in treatment of bone disease. *See*, Schena et al., *Science*, 270:467-470 (1995).

For example, a transgenic mouse carrying the *HBM* gene in the mouse homologue locus is constructed. A mouse of the genotype HBM/+ is viable, healthy and has elevated bone mass. To identify surrogate markers for elevated bone mass, HBM/+ (i.e., heterozygous) and isogenic +/+ (i.e., wild-type) mice

are sacrificed. Bone tissue mRNA is extracted from each animal, and a "gene chip" corresponding to mRNAs expressed in the +/+ individual is constructed. mRNA from different tissues is isolated from animals of each genotype, reverse-transcribed, fluorescently labeled, and then hybridized to gene fragments affixed to a solid support. The ratio of fluorescent intensity between the two populations is indicative of the relative abundance of the specific mRNAs in the +/+ and HBM/+ animals. Alternatively, mRNA may be isolated from wild-type and transgenic animals. cDNA prepared from these samples is transcribed *in vitro* to obtain labeled mRNA for use on custom made or commercially available gene array chips such as are manufactured by Affymetrix. Sets of genes with altered expression as a function of phenotype may be identified be a variety of routine computational analyses. Genes encoding mRNA over- and under-expressed relative to the wild-type control are candidates for genes coordinately regulated by the *HBM* gene.

One standard procedure for identification of new proteins that are part of the same signaling cascade as an already-discovered protein is as follows. Cells are treated with radioactive phosphorous, and the already-discovered protein is manipulated to be more or less active. The phosphorylation state of other proteins in the cell is then monitored by polyacrylamide gel electrophoresis and autoradiography, or similar techniques. Levels of activity of the known protein may be manipulated by many methods, including, for example, comparing wild-type mutant proteins using specific inhibitors such as drugs or antibodies, simply adding or not adding a known extracellular protein, or using antisense inhibition of the expression of the known protein (Tamura et al., *Science*, 280(5369):1614-7 (1998); Meng, *EMBO J.*, 17(15):4391-403 (1998); Cooper et al., *Cell*, 1:263-73 (1982)).

In another example, proteins with different levels of phosphorylation are identified in TE85 osteosarcoma cells expressing either a sense or antisense cDNA for LRP5. TE85 cells normally express high levels of LRP5 (Dong et al.,

Biochem. & Biophys. Res. Comm., 251:784-790 (1998)). Cells containing the sense construct express even higher levels of LRP5, while cells expressing the antisense construct express lower levels. Cells are grown in the presence of ^{32}P, harvested, lysed, and the lysates run on SDS polyacrylamide gels to separate proteins, and the gels subjected to autoradiography (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons (1997)). Bands that differ in intensity between the sense and antisense cell lines represent phosphoproteins whose phosphorylation state or absolute level varies in response to levels of LRP5. As an alternative to the ^{32}P-labeling, unlabeled proteins may be separated by SDS-PAGE and subjected to immunoblotting, using the commercially available anti-phosphotyrosine antibody as a probe (Thomas et al., *Nature*, 376(6537):267-71 (1995)). As an alternative to the expression of antisense RNA, transfection with chemically modified antisense oligonucleotides can be used (Woolf et al., *Nucleic Acids Res.*, 18(7):1763-9 (1990)).

Many bone disorders, such as osteoporosis, have a slow onset and a slow response to treatment. It is therefore useful to develop surrogate markers for bone development and mineralization. Such markers can be useful in developing treatments for bone disorders, and for diagnosing patients who may be at risk for later development of bone disorders. Examples of preferred markers are N- and C-terminal telopeptide markers described, for example, in U.S. Patent Nos. 5,455,179, 5,641,837 and 5,652,112, the disclosures of which are incorporated by reference herein in their entirety. In the area of HIV disease, CD4 counts and viral load are useful surrogate markers for disease progression (Vlahov et al., *JAMA*, 279(1):35-40 (1998)). There is a need for analogous surrogate markers in the area of bone disease.

A surrogate marker can be any characteristic that is easily tested and relatively insensitive to non-specific influences. For example, a surrogate marker can be a molecule such as a protein or mRNA in a tissue or in blood serum.

Alternatively, a surrogate marker may be a diagnostic sign such as sensitivity to pain, a reflex response or the like.

In yet another example, surrogate markers for elevated bone mass are identified using a pedigree of humans carrying the *HBM* gene. Blood samples are withdrawn from three individuals that carry the *HBM* gene, and from three closely related individuals that do not. Proteins in the serum from these individuals are electrophoresed on a two dimensional gel system, in which one dimension separates proteins by size, and another dimension separates proteins by isoelectric point (Epstein et al., *Electrophoresis*, 17(11):1655-70 (1996)). Spots corresponding to proteins are identified. A few spots are expected to be present in different amounts or in slightly different positions for the HBM individuals compared to their normal relatives. These spots correspond to proteins that are candidate surrogate markers. The identities of the proteins are determined by microsequencing, and antibodies to the proteins can be produced by standard methods for use in diagnostic testing procedures. Diagnostic assays for HBM proteins or other candidate surrogate markers include using antibodies described in this invention and a reporter molecule to detect HBM in human body fluids, membranes, bones, cells, tissues or extracts thereof. The antibodies can be labeled by joining them covalently or noncovalently with a substance that provides a detectable signal. In many scientific and patent literature, a variety of reporter molecules or labels are described including radionuclides, enzymes, fluorescent, chemi luminescent or chromogenic agents (U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241). The transgenic or genetically modified animals can also serve in a method for surrogate marker identification.

Using these antibodies, the levels of candidate surrogate markers are measured in normal individuals and in patients suffering from a bone disorder, such as osteoporosis, osteoporosis pseudoglioma, Engelmann's disease, Ribbing's disease, hyperphosphatasemia, Van Buchem's disease, melorheostosis, osteopetrosis, pychodysostosis, sclerosteosis, osteopoikilosis, acromegaly, Paget's

disease, fibrous dysplasia, tubular stenosis, osteogenesis imperfecta, hypoparathyroidism, pseudohypoparathyroidism, pseudopseudohypoparathyroidism, primary and secondary hyperparathyroidism and associated syndromes, hypercalciuria, medullary carcinoma of the thyroid gland, osteomalacia and other diseases. Techniques for measuring levels of protein in serum in a clinical setting using antibodies are well established. A protein that is consistently present in higher or lower levels in individuals carrying a particular disease or type of disease is a useful surrogate marker.

A surrogate marker can be used in diagnosis of a bone disorder. For example, consider a child that presents to a physician with a high frequency of bone fracture. The underlying cause may be child abuse, inappropriate behavior by the child, or a bone disorder. To rapidly test for a bone disorder, the levels of the surrogate marker protein are measured using the antibody described above.

Levels of modification states of surrogate markers can be measured as indicators of the likely effectiveness of a drug that is being developed. It is especially convenient to use surrogate markers in creating treatments for bone disorders, because alterations in bone development or mineralization may require a long time to be observed. For example, a set of bone mRNAs, termed the "HBM-inducible mRNA set" is found to be overexpressed in HBM/+ mice as compared to +/+ mice, as described above. Expression of this set can be used as a surrogate marker. Specifically, if treatment of +/+ mice with a compound results in overexpression of the HBM-inducible mRNA set, then that compound is considered a promising candidate for further development.

This invention is particularly useful for screening compounds by using the LRP5 or HBM protein or binding fragment thereof in any of a variety of drug screening techniques.

The LRP5 or HBM protein or fragment employed in such a test may either be free in solution, affixed to a solid support, or borne on a cell surface. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are

stably transformed with recombinant nucleic acids expressing the protein or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, for the formation of complexes between a LRP5 or HBM protein or fragment and the agent being tested, or examine the degree to which the formation of a complex between a LRP5 or HBM protein or fragment and a known ligand is interfered with by the agent being tested.

Thus, the present invention provides methods of screening for drugs comprising contacting such an agent with a LRP5 or HBM protein or fragment thereof and assaying (i) for the presence of a complex between the agent and the LRP5 or HBM protein or fragment, or (ii) for the presence of a complex between the LRP5 or HBM protein or fragment and a ligand, by methods well known in the art. In such competitive binding assays the LRP5 or HBM protein or fragment is typically labeled. Free LRP5 or HBM protein or fragment is separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label is a measure of the binding of the agent being tested to LRP5 or HBM or its interference with LRP5 or HBM: ligand binding, respectively.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the LRP5 or HBM proteins and is described in detail in WO 84/03564. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with LRP5 or HBM proteins and washed. Bound LRP5 or HBM protein is then detected by methods well known in the art. Purified LRP5 or HBM can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the protein can be used to capture antibodies to immobilize the LRP5 or HBM protein on the solid phase.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the LRP5 or

HBM protein compete with a test compound for binding to the LRP5 or HBM protein or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide that shares one or more antigenic determinants of the LRP5 or HBM protein.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) that have a nonfunctional LRP5 or *HBM* gene. These host cell lines or cells are defective at the LRP5 or HBM protein level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of LRP5 or HBM defective cells.

The goal of rational drug design is to produce structural analogs of biologically active proteins of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the protein, or which, e.g., enhance or interfere with the function of a protein *in vivo*. See, e.g., Hodgson, *Bio/Technology*, 9:19-21 (1991). In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., LRP5 or HBM protein) or, for example, of the LRP5- or HBM-receptor or ligand complex, by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a protein may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., *Science,* 249:527-533 (1990)). In addition, peptides (e.g., LRP5 or HBM protein) are analyzed by an alanine scan (Wells, *Methods in Enzymol.*, 202: 390-411 (1991)). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, for example, desired LRP5 or HBM protein activity or stability, or which act as inhibitors, agonists, antagonists, etc. of LRP5 or HBM protein activity. By virtue of the availability of cloned LRP5 or HBM sequences, sufficient amounts of the LRP5 or HBM protein may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the LRP5 or HBM protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

Identified drug candidates (known as "leads") may be further studied by use of transgenic animals. The transgenic animals of the present invention are useful for creating an animal model of bone density modulation which may be used to test and refine drug leads. The transgenic animals described above represent a single example of the LRP5 and HBM or HBM-like transgenic animals contemplated herein. One skilled in the art is aware of variations and the considerations which will be routinely applied in modifying the present invention to a specific purpose. Examples of the development of transgenic animal models are given for example in Strategies in Transgenic Animal Science (1995, Monastersky and Robl Eds., Washington, DC: American Society for Microbiology) and references therein which are all incorporated herein by reference in their entirety.

As an example, at least two groups of transgenic animals can be created as described, so that one group expresses HBM and another group expresses LRP5. These animals can be treated with the candidate drug for some time spanning from a few days to the remainder of the animal's life-span. The animals are monitored for changes in bone mass and/or surrogate markers for the HBM phenotype. The transgenic animals used in such a study may express human HBM protein and LRP5 protein or the homologous HBM and LRP5 proteins defined for each species or variants thereof. Expression may be driven by a ubiquitous promoter or a bone specific promoter as would be known. It will be informative to compare groups of animals utilizing different promoters.

The transgenic animals of the method according to the present inventions may also comprise knock-in (KI) and/or knock-out (KO) animals, such as mice, which express HBM, LRP5, or neither under the control of the animal's native promoter. Such animals may be created by homologous recombination in ES cells as described above (and elsewhere in the literature of the art such as, for example, U.S. Patent Nos. 6,187,991 and 6,187,992 and references cited therein which are incorporated herein in their entirety). The experimental groups of transgenic animals treated with candidate drugs may be monitored by non-invasive means, by the monitoring of surrogate markers as described above, and/or by *ex vivo* analysis of bones from sacrificed animals at given time-points.

Likewise the effect of such treatments as dietary control (e.g. varying intake of vitamins, minerals, proteins, lipids, etc.), ovariectomy, direct administration of all or part of purified HBM or LRP5 proteins, administration of antisense nucleotides, antibodies against LRP5 or gene therapy in adults may be investigated by systematic administration of the treatment to transgenic animals according to the invention. Such treatments may include, administration of estrogens, tamoxifen, raloxifene, (or other selective estrogen modulators, SERMs), vitamin D analogs, calcitonin, cathepsin K inhibitors, statins (e.g. simvastatin, pravastatin, and lovastatin), bis-phosphonates, parathyroid hormone (PTH), bone morphogenetic

proteins (BMP) as described in U.S. Patent Nos. 6,190,880 and 5,866,364, and combinations of the above compounds.

In view of the homology between LRP5 and LRP6 and to the LDL receptor and the further observations of markers for cardiac health being modulated in HBM subjects, it is an aspect of the present invention to use the novel research methods disclosed herein to screen known cardio-protective treatments for bone modulating effects. Thereby, the present invention provides therapeutic methods which are both cardio-protective and which improve bone quality. The models are useful for testing drugs and researching lipid modulation effects related to LRP5 and HBM.

The effect of various mutations of *LRP5* and *HBM* genes may be investigated by creation of additional lines of transgenic animals according to the invention, wherein these animals comprise such mutations. By comparison of direct measures of bone development or surrogate markers, an embodiment of the invention provides a useful research tool for screening gene therapy reagents, candidate drug therapies, and elucidating molecular mechanisms of bone development modulation. One skilled in the art knows how to use the methods of the present invention to achieve these goals.

The present invention provides a method and useful research tools for testing prospective gene therapies. Transgenic knock-out mice are useful for testing prospective gene therapies. As an example, a transgenic knock-out animal such as a mouse is created as described above which does not express endogenous LRP5 or HBM. A prospective gene therapy, such as intravenous injection of a recombinant replication-defective adenovirus encoding the human HBM protein driven by the CMVβActin promoter, is administered. Parameters of bone density and/or surrogate markers are monitored over time following therapy (Ishibashi et al., 1993 *J. Clin. Invest.* 92:883-93) A TaqMan® primer set such as that described above may be used to measure expression of transgenic HBM. One skilled in the art knows alternative methods such as the Northern blot method. Comparison of treated and untreated animals both within and between groups of germ-line

transgenic animals, knock-out (null allele) background, and wild-type endogenous LRP5 background animals provides complementary controls for assessing the relative effectiveness of various modalities of gene therapy.

Uses for the transgenic animals models contemplated herein also include, but are not limited to: (1) sources for generating bone cell cultures from the calvaria of the transgenic animals to study bone cell (*e.g.*, osteoblast and osteoclast) function and number; (2) models for studying the effects of estrogen loss by ovariectomizing (ovx) the transgenic animals; (3) models for testing mechanical loading on the bones and other stress/strength tests; (4) breeding models with which to breed to other genetically modified or naturally occurring mutant animals that display bone abnormalities; (Chipman et al., *PNAS*, 90:1701-05 (1993); Phillips et al., *Bone*, 27:219-226 (2000); Kajkenova et al., *J. Bone Min. Res.*, 12:1772-79 (1997); Jilka et al., *J. Clin. Invest.* 97:1732-40 (1996); Takahashi et al., *Bone and Mineral*, 24:245-255 (1994); (5) bone mis/disuse models to test the effects of weight bearing or gravity; (6) models for identifying and screening reagents which may or are known to modulate bone metabolism (e.g., PTH, estrogen, vitamin D analogs, bisphosphonates, statins, leptin, BMP, apoE, SERMS); (7) models for investigating prospective treatments to improve fracture repair. Transgenic animals may be cross bread with other genetic (or genetically modified) mouse models of bone disease, lipid disease, Wnt signaling, and the like. Examples of these: osteogenesis imperfecta (oi) mice, spontaneous fracture (sfx) mice, animals with abnormal ApoE, transgenic animals that monitor Wnt signaling with TCF-LacZ or some other reporter gene (GFP, luciferase, CAT), and the like.

The transgenic animal models can be analyzed using, but not limited to, such methods as bone densitometry by pDEXA, pQCT and microCT; histology, molecular marker analysis, apoptosis, cell proliferation, cell cycle, mineralization, serum biochemistry, transcriptional profiling, and the like.

XXII. Methods of Use: Avian and Mammalian Animal Husbandry

The LRP5 DNA and LRP5 protein and/or the HBM DNA and HBM protein can be used for vertebrate and preferably human therapeutic agents and for avian and mammalian veterinary agents, including for livestock breeding. Birds, including, for example, chickens, roosters, hens, turkeys, ostriches, ducks, pheasants and quails, can benefit from the identification of the gene and pathway for high bone mass. In many examples cited in literature (for example, McCoy et al., *Res. Vet. Sci.*, 60(2): 185-186 (1996)), weakened bones due to husbandry conditions cause cage layer fatigue, osteoporosis and high mortality rates. Additional therapeutic agents to treat osteoporosis or other bone disorders in birds can have considerable beneficial effects on avian welfare and the economic conditions of the livestock industry, including, for example, meat and egg production.

XXIII. Methods of use: Diagnostic assays using LRP5-specific oligonucleotides for detection of genetic alterations affecting bone development.

In cases where an alteration or disease of bone development is suspected to involve an alteration of the *LRP5* gene or the *HBM* gene, specific oligonucleotides may be constructed and used to assess the level of LRP5 mRNA or HBM mRNA, respectively, in bone tissue or in another tissue that affects bone development.

For example, to test whether a person has the *HBM* gene, which affects bone density, polymerase chain reaction can be used. Two oligonucleotides are synthesized by standard methods or are obtained from a commercial supplier of custom-made oligonucleotides. The length and base composition are determined by standard criteria using the Oligo 4.0 primer Picking program (Wojchich Rychlik, 1992) or any suitable alternative. One of the oligonucleotides is designed so that it will hybridize only to HBM DNA under the PCR conditions used. The other oligonucleotide is designed to hybridize a segment of LRP5 genomic DNA such that amplification of DNA using these oligonucleotide primers produces a conveniently identified DNA fragment. For example, the pair of primers

CCAAGTTCTGAGAAGTCC (SEQ ID NO:32) and AATACCTGAAACCATACCTG (SEQ ID NO:33) will amplify a 530 base pair DNA fragment from a DNA sample when the following conditions are used: step 1 at 95°C for 120 seconds; step 2 at 95°C for 30 seconds; step 3 at 58°C for 30 seconds; step 4 at 72°C for 120 seconds; where steps 2-4 are repeated 35 times. Tissue samples may be obtained from hair follicles, whole blood, or the buccal cavity.

The fragment generated by the above procedure is sequenced by standard techniques. Individuals heterozygous for the *HBM* gene will show an equal amount of G and T at the second position in the codon for glycine 171. Normal or homozygous wild-type individuals will show only G at this position. Similar routine procedures may be used to develop assays for other polymorphisms and variants according to the invention.

Other amplification techniques besides PCR may be used as alternatives, such as ligation-mediated PCR or techniques involving Q-beta replicase (Cahill et al., *Clin. Chem.*, 37(9):1482-5 (1991)). For example, the oligonucleotides AGCTGCTCGTAGCTG TCTCTCCCTGGATCACGGGTACATGTACTGGACAGACTGGGT (SEQ ID NO:34) and TGAGACGCCCCGGATTGAGCGGGCAGGGATAGCTTA TTCCCTGTGCCGCATTACGGC (SEQ ID NO:35) can be hybridized to a denatured human DNA sample, treated with a DNA ligase, and then subjected to PCR amplification using the primer oligonucleotides AGCTGCTCGTAGCTGTCT CTCCCTGGA (SEQ ID NO:36) and GCCGTAATGCGGCACAGGGAATAAGCT (SEQ ID NO:37). In the first two oligonucleotides, the outer 27 bases are random sequence corresponding to primer binding sites, and the inner 30 bases correspond to sequences in the *LRP5* gene. The T at the end of the first oligonucleotide corresponds to the *HBM* gene. The first two oligonucleotides are ligated only when hybridized to human DNA carrying the *HBM* gene, which results in the formation of an amplifiable 114 bp DNA fragment.

Products of amplification can be detected by agarose gel electrophoresis, quantitative hybridization, or equivalent techniques for nucleic acid detection known to one skilled in the art of molecular biology (Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring, NY (1989)).

Other alterations in the *LRP5* gene or the *HBM* gene may be diagnosed by the same type of amplification-detection procedures, by using oligonucleotides designed to identify those alterations. These procedures can be used in animals as well as humans to identify alterations in LRP5 or HBM that affect bone development.

Expression of LRP5 or HBM in bone tissue may be accomplished by fusing the cDNA of LRP5 or HBM, respectively, to a bone-specific promoter in the context of a vector for genetically engineering vertebrate cells. DNA constructs are introduced into cells by packaging the DNA into virus capsids, by the use of cationic liposomes, electroporation, or by calcium phosphate transfection. Transfected cells, preferably osteoblasts, may be studied in culture or may be introduced into bone tissue in animals by direct injection into bone or by intravenous injection of osteoblasts, followed by incorporation into bone tissue (Ko et al., *Cancer Research*, 56(20):4614-9 (1996)). For example, the osteocalcin promoter, which is specifically active in osteoblasts, may be used to direct transcription of the *LRP5* gene or the *HBM* gene. Any of several vectors and transfection methods may be used, such as retroviral vectors, adenovirus vectors, or vectors that are maintained after transfection using cationic liposomes, or other methods and vectors described herein.

Alteration of the level of functional LRP5 protein or HBM protein affects the level of bone mineralization. By manipulating levels of functional LRP5 protein or HBM protein, it is possible to affect bone development and to increase or decrease levels of bone mineralization. For example, it may be useful to increase bone mineralization in patients with osteoporosis. Alternatively, it may be useful to decrease bone mineralization in patients with osteopetrosis or Paget's disease.

Alteration of LRP5 levels or HBM levels can also be used as a research tool. Specifically, it is possible to identify proteins, mRNA and other molecules whose level or modification status is altered in response to changes in functional levels of LRP5 or HBM. The pathology and pathogenesis of bone disorders is known and described, for example, in Rubin and Farber (Eds.), *Pathology*, 2nd Ed., S.B. Lippincott Co., Philadelphia, PA (1994).

A variety of techniques can be used to alter the levels of functional LRP5 or HBM. For example, intravenous or intraosseous injection of the extracellular portion of LRP5 or mutations thereof, or HBM or mutations thereof, will alter the level of LRP5 activity or HBM activity, respectively, in the body of the treated human, animal or bird. Truncated versions of the LRP5 protein or HBM protein can also be injected to alter the levels of functional LRP5 protein or HBM protein, respectively. Certain forms of LRP5 or HBM enhance the activity of endogenous protein, while other forms are inhibitory.

In a preferred embodiment, the HBM protein is used to treat osteoporosis, fracture, or other bone disorder. In a further preferred embodiment, the extracellular portion of the HBM protein is used. This HBM protein may be optionally modified by the addition of a moiety that causes the protein to adhere to the surface of cells. The protein is prepared in a pharmaceutically acceptable solution and is administered by injection or another method that achieves acceptable pharmacokinetics and distribution.

In a second embodiment of this method, LRP5 or HBM levels are increased or decreased by gene therapy techniques. To increase LRP5 or HBM levels, osteoblasts or another useful cell type are genetically engineered to express high levels of LRP5 or HBM as described above. Alternatively, to decrease LRP5 or HBM levels, antisense constructs that specifically reduce the level of translatable LRP5 or HBM mRNA can be used. In general, a tissue-nonspecific promoter may be used, such as the CMV promoter or another commercially available promoter found in expression vectors (Wu et al., *Toxicol. Appl. Pharmacol.*, 141(1):330-9

(1996)). In a preferred embodiment, a LRP5 cDNA or its antisense is transcribed by a bone-specific promoter, such as the osteocalcin or another promoter, to achieve specific expression in bone tissue. In this way, if a LRP5-expressing DNA construct or HBM-expressing construct is introduced into non-bone tissue, it will not be expressed.

In a third embodiment of this method, antibodies against LRP5 or HBM are used to inhibit its function. Such antibodies are identified herein.

In a fourth embodiment of this method, drugs that are agonists or antagonists of LRP5 function or HBM function are used. Such drugs are described herein and optimized according to techniques of medicinal chemistry well known to one skilled in the art of pharmaceutical development.

LRP5 and HBM interact with several proteins, such as ApoE. Molecules that inhibit the interaction between LRP5 or HBM and ApoE or another binding partner are expected to alter bone development and mineralization. Such inhibitors may be useful as drugs in the treatment of osteoporosis, osteopetrosis, or other diseases of bone mineralization. Such inhibitors may be low molecular weight compounds, proteins or other types of molecules. *See*, Kim et al., *J. Biochem.* (Tokyo), 124(6):1072-1076 (1998).

Inhibitors of the interaction between LRP5 or HBM and interacting proteins may be isolated by standard drug-screening techniques. For example, LRP5 protein, (or a fragment thereof) or HBM protein (or a fragment thereof) can be immobilized on a solid support such as the base of microtiter well. A second protein or protein fragment, such as ApoE is derivatized to aid in detection, for example with fluorescein. Iodine, or biotin, then added to the LRP5 or HBM in the presence of candidate compounds that may specifically inhibit this protein-protein domain of LRP5 or HBM, respectively, and thus avoid problems associated with its transmembrane segment. Drug screens of this type are well known to one skilled in the art of pharmaceutical development.

Because LRP5 and HBM are involved in bone development, proteins that bind to LRP5 and HBM are also expected to be involved in bone development. Such binding proteins can be identified by standard methods, such as co-immunoprecipitation, co-fractionation, or the two-hybrid screen (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons (1997)). For example, to identify LRP5-interacting proteins or HBM-interacting proteins using the two-hybrid system, the extracellular domain of LRP5 or HBM is fused to LexA and expressed for the yeast vector pEG202 (the "bait") and expressed in the yeast strain EGY48. The yeast strain is transformed with a "prey" library in the appropriate vector, which encodes a galactose-inducible transcription-activation sequence fused to candidate interacting proteins. The techniques for initially selecting and subsequently verifying interacting proteins by this method are well known to one skilled in the art of molecular biology (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons (1997)).

In a preferred embodiment, proteins that interact with HBM, but not LRP5, are identified using a variation of the above procedure (Xu *et al.*, *Proc. Natl. Acad. Sci. USA,* 94(23):12473-8 (Nov. 1997)). This variation of the two-hybrid system uses two baits, and LRP5 and HBM are each fused to LexA and TetR, respectively. Alternatively, proteins that interact with the HBM but not LRP5 may be isolated. These procedures are well known to one skilled in the art of molecular biology, and are a simple variation of standard two-hybrid procedures.

As an alternative method of isolating substances interacting with LRP5 or HBM, a biochemical approach is used. The LRP5 protein or a fragment thereof, such as the extracellular domain, or the HBM protein or a fragment thereof, such as the extracellular domain, is chemically coupled to Sepharose beads. The LRP5- or HBM-coupled beads are poured into a column. A biological extract, such as a lipid fraction, serum proteins, proteins in the supernatant of a bone biopsy, or cellular contents from gently lysed osteoblast cells, is added to the column. Non-specifically bound compounds are eluted, the column is washed several times with a

low-salt buffer, and then tightly binding compounds may be eluted with a high-salt buffer. These are candidate compounds that bind to LRP5 or HBM, and can be tested for specific binding by standard tests and control experiments. Sepharose beads used for coupling proteins and the methods for performing the coupling are commercially available (Sigma), and the procedures described here are well known to one skilled in the art of protein biochemistry.

As a variation of the above procedure, proteins that are eluted by high salt from the LRP5- or HBM-sepharose column are then added to an HBM-LRP5-sepharose column. Proteins that flow through without sticking are proteins that bind to LRP5 but not to HBM. Alternatively, proteins that bind to the HBM protein and not to the LRP5 protein can be isolated by reversing the order in which the columns are used.

Isolated compounds may be identified by standard methods such as 2D gel electrophoresis, chromatography, and mass spectroscopy.

XXIV. Method of Use: Transformation-Associated Recombination (TAR) Cloning

Essential for the identification of novel allelic variants of LRP5 is the ability to examine the sequence of both copies of the gene in an individual. To accomplish this, two "hooks," or regions of significant similarity, are identified within the genomic sequence such that they flank the portion of DNA that is to be cloned. Most preferably, the first of these hooks is derived from sequences 5' to the first exon of interest and the second is derived from sequences 3' to the last exon of interest. These two "hooks" are cloned into a bacterial/yeast shuttle vector such as that described by Larionov et al., *Proc. Natl. Acad. Sci. USA*, 94:7384-7387 (1997). Other similar vector systems may also be used. To recover the entire genomic copy of the *LRP5* gene, the plasmid containing the two "hooks" is linearized with a restriction endonuclease or is produced by another method such as PCR. This linear DNA fragment is introduced into yeast cells along with human

genomic DNA. Typically, the yeast *Saccharomyces cerevisiae* is used as a host cell, although Kouprina et al. (*Genome Res.*, 8:666-72, 1998) have reported using chicken host cells as well. During and after the process of transformation, the endogenous host cell converts the linear plasmid to a circle by a recombination event whereby the region of the human genomic DNA homologous to the "hooks" is inserted into the plasmid. This plasmid can be recovered and analyzed by methods well known to one skilled in the art. Obviously, the specificity for this reaction requires the host cell machinery to recognize sequences similar to the "hooks" present in the linear fragment. However, 100% sequence identity is not required, as shown by Kouprina et al., *Genomics*, 53(1):21-28 (October 1998), where the author describes using degenerate repeated sequences common in the human genome to recover fragments of human DNA from a rodent/human hybrid cell line.

In another example, only one "hook" is required, as described by Larionov et al., *Proc. Natl. Acad. Sci. USA*, 95(8):4469-74 (April 1998). For this type of experiment, termed "radial TAR cloning," the other region of sequence similarity to drive the recombination is derived from a repeated sequence from the genome. In this way, regions of DNA adjacent to the *LRP5* gene coding region can be recovered and examined for alterations that may affect function.

XXV. Methods of Use: Genomic Screening

The use of polymorphic genetic markers linked to the *HBM* gene or to LRP5 is very useful in predicting susceptibility to osteoporosis or other bone diseases. Koller et al., *Amer. J. Bone Min. Res.*, 13:1903-1908 (1998) have demonstrated that the use of polymorphic genetic markers is useful for linkage analysis. Similarly, the identification of polymorphic genetic markers within the high bone mass gene will allow the identification of specific allelic variants that are in linkage disequilibrium with other genetic lesions that affect bone development. Using the DNA sequence from the BACs, a dinucleotide CAn repeat was identified and two

unique PCR primers that will amplify the genomic DNA containing this repeat were designed, as shown below:

B200E21C16_L: GAGAGGCTATATCCCTGGGC (SEQ ID NO:38)

B200E21C16_R: ACAGCACGTGTTTAAAGGGG (SEQ ID NO:39)

and used in the genetic mapping study.

This method has been used successfully by others skilled in the art (e.g., Sheffield et al., *Genet.*, 4:1837-1844 (1995); LeBlanc-Straceski et al., *Genomics*, 19:341-9 (1994); Chen et al., *Genomics*, 25:1-8 (1995)). Use of these reagents with populations or individuals will predict their risk for osteoporosis. Similarly, single nucleotide polymorphisms (SNPs), such as those shown in Table 4 above, can be used as well to predict risk for developing bone diseases or resistance to osteoporosis in the case of the *HBM* gene.

XXVI. Methods of Use: Modulators of Tissue Calcification

The calcification of tissues in the human body is well documented. Towler et al., *J. Biol. Chem.*, 273:30427-34 (1998) demonstrated that several proteins known to regulate calcification of the developing skull in a model system are expressed in calcified aorta. The expression of Msx2, a gene transcribed in osteoprogenitor cells, in calcified vascular tissue indicates that genes which are important in bone development are involved in calcification of other tissues. Treatment with HBM protein, agonists or antagonists is likely to ameliorate calcification (such as the vasculature, dentin and bone of the skull visera) due to its demonstrated effect on bone mineral density. In experimental systems where tissue calcification is demonstrated, the over-expression or repression of LRP5 activity permits the identification of molecules that are directly regulated by the *LRP5* gene. These genes are potential targets for therapeutics aimed at modulating tissue calcification. For example, an animal, such as the LDLR -/-, mouse is fed a high fat diet and is observed to demonstrate expression of markers of tissue calcification, including LRP5. These animals are then treated with antibodies to LRP5 or HBM

protein, antisense oligonucleotides directed against LRP5 or HBM cDNA, or with compounds known to bind the LRP5 or HBM protein or its binding partner or ligand. RNA or proteins are extracted from the vascular tissue and the relative expression levels of the genes expressed in the tissue are determined by methods well known in the art. Genes that are regulated in the tissue are potential therapeutic targets for pharmaceutical development as modulators of tissue calcification.

The nucleic acids, proteins, peptides, amino acids, small molecules or other pharmaceutically useful compounds of the present invention that are to be given to an individual may be administered in the form of a composition with a pharmaceutically acceptable carrier, excipient or diluent, which are well known in the art. The individual may be a mammal or a bird, preferably a human, a rat, a mouse or bird. Such compositions may be administered to an individual in a pharmaceutically effective amount. The amount administered will vary depending on the condition being treated and the patient being treated. The compositions may be administered alone or in combination with other treatments.

EXAMPLES

The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

Example 1

The propositus was referred by her physicians to the Creighton Osteoporosis Center for evaluation of what appeared to be unusually dense bones. She was 18 years old and came to medical attention two years previous because of back pain, which was precipitated by an auto accident in which the car in which she was riding as a passenger was struck from behind. Her only injury was soft tissue injury to

her lower back that was manifested by pain and muscle tenderness. There was no evidence of fracture or subluxation on radiographs. The pain lasted for two years, although she was able to attend school full time. By the time she was seen in the Center, the pain was nearly resolved and she was back to her usual activities as a high school student. Physical exam revealed a normal healthy young woman standing 66 inches and weighing 128 pounds. Radiographs of the entire skeleton revealed dense looking bones with thick cortices. All bones of the skeleton were involved. Most importantly, the shapes of all the bones were entirely normal. The spinal BMC was 94.48 grams in L1-4, and the spinal BMD was 1.667 gm/cm^2 in L1-4. BMD was 5.62 standard deviations (SD) above peak skeletal mass for women. These were measured by DXA using a Hologic 2000~. Her mother was then scanned and a lumbar spinal BMC of 58.05 grams and BMD of 1.500 gm/cm^2 were found. Her mother's values place her 4.12 SD above peak mass and 4.98 SD above her peers. Her mother was 51 years old, stood 65 inches and weighed 140 pounds. Her mother was in excellent health with no history of musculoskeletal or other symptoms. Her father's lumbar BMC was 75.33 grams and his BMD was 1.118 gm/cm^2. These values place him 0.25 SD above peak bone mass for males. He was in good health, stood 72 inches tall, and weighed 187 pounds.

These clinical data suggested that the propositus inherited a trait from her mother, which resulted in very high bone mass, but an otherwise normal skeleton, and attention was focused on the maternal kindred. In U.S. Patent No. 5,691,153, twenty- two of these members had measurement of bone mass by DXA. In one case, the maternal grandfather of the propositus, was deceased, however, medical records, antemortem skeletal radiographs and a gall bladder specimen embedded in paraffin for DNA genotyping were obtained. His radiographs showed obvious extreme density of all of the bones available for examination including the femur and the spine, and he was included among the affected members. In this invention, the pedigree has been expanded to include 37 informative individuals. These additions are a significant improvement over the original kinship (Johnson et al.,

Am. J. Hum. Genet., 60:1326-1332 (1997)) because, among the fourteen individuals added since the original study, two individuals harbor key crossovers. X-linkage is ruled out by the presence of male-to-male transmission from individual 12 to 14 and 15.

Example 2

The present invention describes DNA sequences derived from two BAC clones from the *HBM* gene region, as evident in Table 8 below, which is an assembly of these clones. Clone b200e21-h (ATCC No. 980812; SEQ ID NOS: 10-11) was deposited at the American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110-2209 U.S.A., on December 30, 1997. Clone b527d12-h (ATCC No. 980720; SEQ ID NOS: 5-9) was deposited at the American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110-2209 U.S.A., on October 2, 1998. These sequences are unique reagents that can be used by one skilled in the art to identify DNA probes for the *LRP5* gene, PCR primers to amplify the gene, nucleotide polymorphisms in the *LRP5* gene, or regulatory elements of the *LRP5* gene.

TABLE 8

Contig	ATCC No.	SEQ ID NO.	Length (base pairs)
b527d12-h_contig302G	980720	5	3096
b527d12-h_contig306G	980720	6	26928
b527d12-h_contig307G	980720	7	29430
b527d12-h_contig308G	980720	8	33769
b527d12-h_contig309G	980720	9	72049
b200e21-h_contig1	980812	10	8705
b200e21-h_contig4	980812	11	66933

Example 3

Transcriptional profiling of calvaria and tibia explant cultures from HBM overexpressing transgenic and non-transgenic mice.

The use of transgenic animals of the invention for the identification of surrogate markers for the HBM phenotype and putative targets for bone mass modulation therapies and drugs by the methods of the invention and the identification and characterization of genes related to HBM through transcriptional profiling is demonstrated.

Calvaria and tibia were obtained from neonatal (12-day-old) mice, including transgenic mice expressing HBM under the bone specific type I collagen promoter (Line 19). Calvaria were pooled from 4 transgenic and 4 non-transgenic mice and digested with collegenase. The digests were plated in culture. Calvaria cultures were maintained with or without ascorbic acid and beta glycerol phosphate for 19 days. RNA was isolated at day 19.

Bone marrow stromal cells were flushed out of tibia and the tibias from individual animals were then subjected to two consecutive collagenase digests. Following collagenase digestion the bone chips were plated in culture and three consecutive seedings were obtained. Cells from seeding 3 were much slower growing than those from seedings 1 and 2. RNA was isolated from confluent cells of seedings 1 through 3 (passage 1). RNA from both calvaria and tibia cultures were analyzed on U74Av2 transcriptional profiling arrays.

Treatment with ascorbic acid and beta glycerol phosphate resulted in the set of genes differentially expressed being quite different. Alkaline phosphatase (AKP) gene expression increased following treatment indicating differentiation of the treated cells. The data obtained from treated cells was of different quality than that obtained from untreated cells, and there was also variation in gene expression between the culture replicates.

The transcriptional profile from non-transgenic and transgenic mice showed differences in the expression of several relevant genes, for example, S100A1, MMP9, and MT1.

There is mouse-to-mouse variability in the transcriptional profile of tibia explant cultures from the 4 mice in each group (i.e., non-transgenic and HBM transgenic). This variation can still be seen following normalization. However, the variability does not affect the interpretation of the data to any significant extent. Variability is also seen in measurements of alkaline phosphatase activity in these cells.

The results may be summarized as follows: Seeding 1 and seeding 2 were similar (and different from seeding 3) in their transcriptional profiles. This is likely be due to differences in growth characteristics of these cells. Seeding 1 also shows greater differences between transgenic and non-transgenic profiles, probably because it is a relatively more mixed population of cells than either seeding 2 or 3.

Several of the differences in gene expression between the non-transgenic and HBM transgenic mice are consistent with differences seen between the affected and unaffected individuals from the human HBM1 kindred. As one example, S100A1 (GenBank # AF087687) is upregulated in transgenic osteoblast cultures. The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins are localized in the cytoplasm and/or nucleus of a wide range of cells, and involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. Matrix metalloproteinase 9 (MMP9) (GenBank # X72794) is also upregulated in transgenic osteoblast cultures. Proteins of the matrix metalloproteinase (MMP) family are involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis. Most MMP's are secreted as inactive proproteins which are activated when cleaved by extracellular proteinases. The enzyme encoded by this gene degrades type IV and V collagens.

Among genes downregulated in HBM transgenic osteoblast cultures is metallothionein 1 (MT1) (GenBank # S62785) a cysteine-rich, metal-binding protein that has been shown to play an important role as antioxidant. Its activity mediates cytotoxicity from inflammatory processes. It is expressed in both bone and cartilage.

Additional genes which are differentially expressed in transgenic bone tissue may be determined by one of skill in the art using the methods herein described. From the transcriptional profiling data, it can be seen that the profile of transgenic mouse tibia resembles that of the affected members of the human HBM kindred in several ways.

Example 4

Loading Studies using Transgenic Mice

Effects of transgenic modifications, such as LRP5, LRP6, and HBM expression, over-expression or knock-out, on bone development and can be assessed using a loading or unloading protocol. Bone growth rates subject to loading or unloading, gene expression response profiling, and biomechanical parameters of the HBM phenotype can be further characterized by these methods. These methods of using transgenic animals of the invention are valuable tools in the development of treatments and drugs which recapitulate desired characteristics of the HBM phenotype.

Mechanical loads are delivered to the tibiae of transgenic or non-transgenic mice with the four-point bending device. The device is calibrated for accurate, *in vivo*, external force application. The device applies force through four rounded pads composed of balsa wood and covered by 1 mm thick surgical tubing. The upper pads are 4.5 mm apart and centered between the lower pads that are 12 mm apart. With four-point bending, a constant bending moment is delivered throughout the bone tissue between the two inner pads with the lateral side of the tibia in compression and the medial in tension. An illustration of the device is seen in

Figure 28. The upper distal pad contacts the leg 1 mm proximal to the tibia-fibular junction (TFJ) and the lower distal pad contacts the medial surface at 2.5 mm distal to the TFJ. The region of maximal bending is from 1-6 mm proximal to the TFJ or 8.5 to 13.5 mm from distal end. This region has been radiographically defined.

The loading device is calibrated before each experiment, and loads are recorded for each animal daily. The machine is zeroed and adjusted if there is any drift in the load. Leg positioning and applied loads are consistent between animals and days with less than 10% variation in strain due to leg positioning and less than 1.6% variation in loads (Hagino et al., J. Bone Miner. Res. 8, 347-57 , 1993).

Mechanical loads are applied to the right lower leg while the mouse is under light isoflurane anesthesia (2%). Reliable leg positioning will be attained by standard positioning of the mouse on a platform, placing the right foot in a stirrup, and aligning the knee with the loading device. The isoflurane is short acting so it prevents movement during loading, but normal weight bearing activity returns within seconds after loading. Activity is monitored for proper recovery from the isoflurane and that no injury to the leg has occurred. De-loading may be accomplished by unilateral neurectomy (Kodama et al., Bone 25, 183-90, 1999). In addition, a strain gauge may be applied to the tibia in vivo during the application of four point load.

Fluorochrome labels: All mice to be studied for histomorphometry receive a double calcein label administered 10 and 3 days before tissue collection. Mice on the longterm loading study receive a baseline injection of tetracycline before loading in addition to the final double calcein injection. These injections are prepared in a dilution suitable for injection at 1 ml/kg. The volume of injection for a 25 g mouse would be .025 ml. All injections are subcutaneous and given under mild sedation (isoflurane 2%).

Calcein labels- (Sigma, St. Louis, MO) is injected at 6.2 mg/kg. Two calcein labels are administered on two different days, i.e. 3 and 10 days, before autopsy. This fluorochrome label is used to identify mineralizing

surfaces in undecalcified tissue and quantify the rate of bone formation during the final week of the treatment. The label is not used for BrdU or *in situ* hybridization studies that examine decalcified bone.

Tetracycline- (Pfizer, CT) is injected at 25mg/kg. A single tetracycline label is administered on Day 0 to all animals in long term studies (greater than 5 weeks) This fluorochrome label marks the mineralizing surface at the start of the study and allows quantification of total bone formation during the experiment.

BrDU- (bromodeoxyuridine, Boehringer Mannheim): is injected at 40mg/kg and the vehicle is bacteriostat. Mice are given 5 injections at 6 hr intervals to label DNA synthesis over a 24 hr period. The last injection is one hour before tissue collection.

Death is induced by CO inhalation, except when animals are perfused with fixative. The right and left tibia are excised for all loading and disuse studies. The right leg is the loaded or treated leg and the left the treated control. Tissue is collected from the loaded region of the right tibia and from a similar region on the left tibia. For mice, we have determined the average loaded region to be from 1 to 6 mm proximal to the tibial fibula junction (TFJ).

Undecalcified Cortical Bone Samples: Tibia, femur, and vertebra are collected for standard histomorphometry of undecalcified bone sections. The majority of the muscle is removed and the bone placed in 70% EtOH for 48 hours. The bones are cut with a saw to create the following samples for analysis a) tibial shaft including the TFJ, b) the distal femur, and c) vertebral body free of the disks. The tibial diaphysis is placed in Villanueva stain for 72 hrs and then returned to 80% ethanol. All other bones move directly into dehydration. During the next 14 days, the specimens are dehydrated in graded ethanols and acetone, then embedded individually in modified methyl methacrylate. The embedded tibial cortical samples are cross-sectioned at 70μm on a saw-microtome (Model 1600, Leica, Germany) with sections collected from the region. Sections are taken from the loaded region

to produce a section 5-7 mm proximal to the TFJ in rats and 9-13 mm from the distal end with a 0.8 mm inter-section distance. Two sections from each tibia will be mounted, given a random number, and analyzed.

Decalcified Cortical Bone Samples: The animal tissue is perfused with 4% paraformaldehyde until the soft tissue in the leg is rigid. The tibiae is excised and muscle trimmed with scissors while the bone is submerged in cold 4% paraformaldehyde. The periosteum and a small muscle layer are left intact. For *in situ* hybridization studies all work is done with RNAse free materials. A 4-5 mm section from the loaded region of the tibia is excised from the intact tibia with a saw and fixed in 4% paraformaldehyde at 4°C for 24 hours. After fixation the bones are decalcified in 7%EDTA (Sigma) at a pH of 6.5 for 2-3 weeks. The bones are then placed in 1% MgCl for 6 hours to restore alkaline phosphatase activity. The diaphyseal segment is embedded in JB-4 plus (Polyscience) or paraffin. Cross sections are cut on a microtome (Reichert Jung) at 5μm thickness using a tungsten carbide knife. The sections are mounted on poly l-lysine (Sigma) or coated slides.

As seen in Figure 29, calcein staining of mice subjected to loading shows greater growth in heterozygous HBM transgenic mice (Line 19) than in non-transgenic control mice. A single strain (5N or 7N, 36 cycles at 2 Hz) was administered for 5 days. Calcein labeling occurred on days 5 and 12 with tissues harvested on day 15.

Although the invention has been set forth in detail, one skilled in the art will recognize that numerous changes and modifications can be made, and that such changes and modifications may be made without departing from the spirit and scope of the invention.

The patents, patent applications and publications cited in the specification are hereby incorporated by reference herein in their entirety for all purposes. Further, U.S. Application Nos. 09/543,771 and 09/544,398 filed on April 5, 2000, Application No. 09/229,319, filed January 13, 1999, U.S. Provisional Application

No. 60/071,449, filed January 13, 1998, and U.S. Provisional Application No. 60/105,511, filed October 23, 1998, are herein incorporated by reference in their entirety for all purposes.

Additionally, this application claims priority of Application Nos. 60/290,071 filed May 11, 2001; 60/291,311 filed May 17, 2001; 60/353,058 filed February 1, 2002, and 60/361,293 filed March 4, 2002; the disclosures of each are herein incorporated by reference in their entirety for all purposes.

SEQUENCE LISTING

<110> Babij, Philip
Yaworsky, Paul
Bex, Frederick J. III
Bodine, Peter Van Nest

<120> Transgenic Animal Model of Bone Mass Modulation

<130> 032796-133

<150> US 60/290,071
<151> 2001-05-11

<150> US 60/291,311
<151> 2001-05-17

<150> US 60/353,058
<151> 2002-02-01

<150> US 60/361,293
<151> 2002-03-04

<160> 765

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 5120
<212> DNA
<213> Homo sapiens

<400> 1

actaaagcgc cgccgccgcg ccatggagcc cgagtgagcg cggcgcgggc ccgtccggcc 60
gccggacaac atg gag gca gcg ccg ccc ggg ccg ccg tgg ccg ctg ctg 109
Met Glu Ala Ala Pro Pro Gly Pro Pro Trp Pro Leu Leu
1 5 10
ctg ctg ctg ctg ctg ctg ctg gcg ctg tgc ggc tgc ccg gcc ccc gcc 157
Leu Leu Leu Leu Leu Leu Leu Ala Leu Cys Gly Cys Pro Ala Pro Ala
15 20 25
gcg gcc tcg ccg ctc ctg cta ttt gcc aac cgc cgg gac gta cgg ctg 205
Ala Ala Ser Pro Leu Leu Leu Phe Ala Asn Arg Arg Asp Val Arg Leu
30 35 40 45
gtg gac gcc ggc gga gtc aag ctg gag tcc acc atc gtg gtc agc ggc 253
Val Asp Ala Gly Gly Val Lys Leu Glu Ser Thr Ile Val Val Ser Gly
50 55 60
ctg gag gat gcg gcc gca gtg gac ttc cag ttt tcc aag gga gcc gtg 301
Leu Glu Asp Ala Ala Ala Val Asp Phe Gln Phe Ser Lys Gly Ala Val
65 70 75

```
tac tgg aca gac gtg agc gag gag gcc atc aag cag acc tac ctg aac        349
Tyr Trp Thr Asp Val Ser Glu Glu Ala Ile Lys Gln Thr Tyr Leu Asn
        80                  85                  90
cag acg ggg gcc gcc gtg cag aac gtg gtc atc tcc ggc ctg gtc tct        397
Gln Thr Gly Ala Ala Val Gln Asn Val Val Ile Ser Gly Leu Val Ser
    95                  100                 105
ccc gac ggc ctc gcc tgc gac tgg gtg ggc aag aag ctg tac tgg acg        445
Pro Asp Gly Leu Ala Cys Asp Trp Val Gly Lys Lys Leu Tyr Trp Thr
110                 115                 120                 125
gac tca gag acc aac cgc atc gag gtg gcc aac ctc aat ggc aca tcc        493
Asp Ser Glu Thr Asn Arg Ile Glu Val Ala Asn Leu Asn Gly Thr Ser
                130                 135                 140
cgg aag gtg ctc ttc tgg cag gac ctt gac cag ccg agg gcc atc gcc        541
Arg Lys Val Leu Phe Trp Gln Asp Leu Asp Gln Pro Arg Ala Ile Ala
            145                 150                 155
ttg gac ccc gct cac ggg tac atg tac tgg aca gac tgg ggt gag acg        589
Leu Asp Pro Ala His Gly Tyr Met Tyr Trp Thr Asp Trp Gly Glu Thr
        160                 165                 170
ccc cgg att gag cgg gca ggg atg gat ggc agc acc cgg aag atc att        637
Pro Arg Ile Glu Arg Ala Gly Met Asp Gly Ser Thr Arg Lys Ile Ile
    175                 180                 185
gtg gac tcg gac att tac tgg ccc aat gga ctg acc atc gac ctg gag        685
Val Asp Ser Asp Ile Tyr Trp Pro Asn Gly Leu Thr Ile Asp Leu Glu
190                 195                 200                 205
gag cag aag ctc tac tgg gct gac gcc aag ctc agc ttc atc cac cgt        733
Glu Gln Lys Leu Tyr Trp Ala Asp Ala Lys Leu Ser Phe Ile His Arg
                210                 215                 220
gcc aac ctg gac ggc tcg ttc cgg cag aag gtg gtg gag ggc agc ctg        781
Ala Asn Leu Asp Gly Ser Phe Arg Gln Lys Val Val Glu Gly Ser Leu
            225                 230                 235
acg cac ccc ttc gcc ctg acg ctc tcc ggg gac act ctg tac tgg aca        829
Thr His Pro Phe Ala Leu Thr Leu Ser Gly Asp Thr Leu Tyr Trp Thr
        240                 245                 250
gac tgg cag acc cgc tcc atc cat gcc tgc aac aag cgc act ggg ggg        877
Asp Trp Gln Thr Arg Ser Ile His Ala Cys Asn Lys Arg Thr Gly Gly
    255                 260                 265
aag agg aag gag atc ctg agt gcc ctc tac tca ccc atg gac atc cag        925
Lys Arg Lys Glu Ile Leu Ser Ala Leu Tyr Ser Pro Met Asp Ile Gln
270                 275                 280                 285
gtg ctg agc cag gag cgg cag cct ttc ttc cac act cgc tgt gag gag        973
Val Leu Ser Gln Glu Arg Gln Pro Phe Phe His Thr Arg Cys Glu Glu
                290                 295                 300
gac aat ggc ggc tgc tcc cac ctg tgc ctg ctg tcc cca agc gag cct       1021
Asp Asn Gly Gly Cys Ser His Leu Cys Leu Leu Ser Pro Ser Glu Pro
            305                 310                 315
ttc tac aca tgc gcc tgc ccc acg ggt gtg cag ctg cag gac aac ggc       1069
Phe Tyr Thr Cys Ala Cys Pro Thr Gly Val Gln Leu Gln Asp Asn Gly
        320                 325                 330
agg acg tgt aag gca gga gcc gag gag gtg ctg ctg ctg gcc cgg cgg       1117
Arg Thr Cys Lys Ala Gly Ala Glu Glu Val Leu Leu Leu Ala Arg Arg
```

```
    335                 340                 345
acg gac cta cgg agg atc tcg ctg gac acg ccg gac ttc acc gac atc        1165
Thr Asp Leu Arg Arg Ile Ser Leu Asp Thr Pro Asp Phe Thr Asp Ile
350                 355                 360                 365
gtg ctg cag gtg gac gac atc cgg cac gcc att gcc atc gac tac gac        1213
Val Leu Gln Val Asp Asp Ile Arg His Ala Ile Ala Ile Asp Tyr Asp
                370                 375                 380
ccg cta gag ggc tat gtc tac tgg aca gat gac gag gtg cgg gcc atc        1261
Pro Leu Glu Gly Tyr Val Tyr Trp Thr Asp Asp Glu Val Arg Ala Ile
            385                 390                 395
cgc agg gcg tac ctg gac ggg tct ggg gcg cag acg ctg gtc aac acc        1309
Arg Arg Ala Tyr Leu Asp Gly Ser Gly Ala Gln Thr Leu Val Asn Thr
        400                 405                 410
gag atc aac gac ccc gat ggc atc gcg gtc gac tgg gtg gcc cga aac        1357
Glu Ile Asn Asp Pro Asp Gly Ile Ala Val Asp Trp Val Ala Arg Asn
    415                 420                 425
ctc tac tgg acc gac acg ggc acg gac cgc atc gag gtg acg cgc ctc        1405
Leu Tyr Trp Thr Asp Thr Gly Thr Asp Arg Ile Glu Val Thr Arg Leu
430                 435                 440                 445
aac ggc acc tcc cgc aag atc ctg gtg tcg gag gac ctg gac gag ccc        1453
Asn Gly Thr Ser Arg Lys Ile Leu Val Ser Glu Asp Leu Asp Glu Pro
                450                 455                 460
cga gcc atc gca ctg cac ccc gtg atg ggc ctc atg tac tgg aca gac        1501
Arg Ala Ile Ala Leu His Pro Val Met Gly Leu Met Tyr Trp Thr Asp
            465                 470                 475
tgg gga gag aac cct aaa atc gag tgt gcc aac ttg gat ggg cag gag        1549
Trp Gly Glu Asn Pro Lys Ile Glu Cys Ala Asn Leu Asp Gly Gln Glu
        480                 485                 490
cgg cgt gtg ctg gtc aat gcc tcc ctc ggg tgg ccc aac ggc ctg gcc        1597
Arg Arg Val Leu Val Asn Ala Ser Leu Gly Trp Pro Asn Gly Leu Ala
    495                 500                 505
ctg gac ctg cag gag ggg aag ctc tac tgg gga gac gcc aag aca gac        1645
Leu Asp Leu Gln Glu Gly Lys Leu Tyr Trp Gly Asp Ala Lys Thr Asp
510                 515                 520                 525
aag atc gag gtg atc aat gtt gat ggg acg aag agg cgg acc ctc ctg        1693
Lys Ile Glu Val Ile Asn Val Asp Gly Thr Lys Arg Arg Thr Leu Leu
                530                 535                 540
gag gac aag ctc ccg cac att ttc ggg ttc acg ctg ctg ggg gac ttc        1741
Glu Asp Lys Leu Pro His Ile Phe Gly Phe Thr Leu Leu Gly Asp Phe
            545                 550                 555
atc tac tgg act gac tgg cag cgc cgc agc atc gag cgg gtg cac aag        1789
Ile Tyr Trp Thr Asp Trp Gln Arg Arg Ser Ile Glu Arg Val His Lys
        560                 565                 570
gtc aag gcc agc cgg gac gtc atc att gac cag ctg ccc gac ctg atg        1837
Val Lys Ala Ser Arg Asp Val Ile Ile Asp Gln Leu Pro Asp Leu Met
    575                 580                 585
ggg ctc aaa gct gtg aat gtg gcc aag gtc gtc gga acc aac ccg tgt        1885
Gly Leu Lys Ala Val Asn Val Ala Lys Val Val Gly Thr Asn Pro Cys
590                 595                 600                 605
gcg gac agg aac ggg ggg tgc agc cac ctg tgc ttc ttc aca ccc cac        1933
```

```
Ala Asp Arg Asn Gly Gly Cys Ser His Leu Cys Phe Phe Thr Pro His
                610                 615                 620
gca acc cgg tgt ggc tgc ccc atc ggc ctg gag ctg ctg agt gac atg        1981
Ala Thr Arg Cys Gly Cys Pro Ile Gly Leu Glu Leu Leu Ser Asp Met
            625                 630                 635
aag acc tgc atc gtg cct gag gcc ttc ttg gtc ttc acc agc aga gcc        2029
Lys Thr Cys Ile Val Pro Glu Ala Phe Leu Val Phe Thr Ser Arg Ala
        640                 645                 650
gcc atc cac agg atc tcc ctc gag acc aat aac aac gac gtg gcc atc        2077
Ala Ile His Arg Ile Ser Leu Glu Thr Asn Asn Asn Asp Val Ala Ile
    655                 660                 665
ccg ctc acg ggc gtc aag gag gcc tca gcc ctg gac ttt gat gtg tcc        2125
Pro Leu Thr Gly Val Lys Glu Ala Ser Ala Leu Asp Phe Asp Val Ser
670                 675                 680                 685
aac aac cac atc tac tgg aca gac gtc agc ctg aag acc atc agc cgc        2173
Asn Asn His Ile Tyr Trp Thr Asp Val Ser Leu Lys Thr Ile Ser Arg
                690                 695                 700
gcc ttc atg aac ggg agc tcg gtg gag cac gtg gtg gag ttt ggc ctt        2221
Ala Phe Met Asn Gly Ser Ser Val Glu His Val Val Glu Phe Gly Leu
            705                 710                 715
gac tac ccc gag ggc atg gcc gtt gac tgg atg ggc aag aac ctc tac        2269
Asp Tyr Pro Glu Gly Met Ala Val Asp Trp Met Gly Lys Asn Leu Tyr
        720                 725                 730
tgg gcc gac act ggg acc aac aga atc gaa gtg gcg cgg ctg gac ggg        2317
Trp Ala Asp Thr Gly Thr Asn Arg Ile Glu Val Ala Arg Leu Asp Gly
    735                 740                 745
cag ttc cgg caa gtc ctc gtg tgg agg gac ttg gac aac ccg agg tcg        2365
Gln Phe Arg Gln Val Leu Val Trp Arg Asp Leu Asp Asn Pro Arg Ser
750                 755                 760                 765
ctg gcc ctg gat ccc acc aag ggc tac atc tac tgg acc gag tgg ggc        2413
Leu Ala Leu Asp Pro Thr Lys Gly Tyr Ile Tyr Trp Thr Glu Trp Gly
                770                 775                 780
ggc aag ccg agg atc gtg cgg gcc ttc atg gac ggg acc aac tgc atg        2461
Gly Lys Pro Arg Ile Val Arg Ala Phe Met Asp Gly Thr Asn Cys Met
            785                 790                 795
acg ctg gtg gac aag gtg ggc cgg gcc aac gac ctc acc att gac tac        2509
Thr Leu Val Asp Lys Val Gly Arg Ala Asn Asp Leu Thr Ile Asp Tyr
        800                 805                 810
gct gac cag cgc ctc tac tgg acc gac ctg gac acc aac atg atc gag        2557
Ala Asp Gln Arg Leu Tyr Trp Thr Asp Leu Asp Thr Asn Met Ile Glu
    815                 820                 825
tcg tcc aac atg ctg ggt cag gag cgg gtc gtg att gcc gac gat ctc        2605
Ser Ser Asn Met Leu Gly Gln Glu Arg Val Val Ile Ala Asp Asp Leu
830                 835                 840                 845
ccg cac ccg ttc ggt ctg acg cag tac agc gat tat atc tac tgg aca        2653
Pro His Pro Phe Gly Leu Thr Gln Tyr Ser Asp Tyr Ile Tyr Trp Thr
                850                 855                 860
gac tgg aat ctg cac agc att gag cgg gcc gac aag act agc ggc cgg        2701
Asp Trp Asn Leu His Ser Ile Glu Arg Ala Asp Lys Thr Ser Gly Arg
            865                 870                 875
```

```
aac cgc acc ctc atc cag ggc cac ctg gac ttc gtg atg gac atc ctg        2749
Asn Arg Thr Leu Ile Gln Gly His Leu Asp Phe Val Met Asp Ile Leu
        880                 885                 890
gtg ttc cac tcc tcc cgc cag gat ggc ctc aat gac tgt atg cac aac        2797
Val Phe His Ser Ser Arg Gln Asp Gly Leu Asn Asp Cys Met His Asn
    895                 900                 905
aac ggg cag tgt ggg cag ctg tgc ctt gcc atc ccc ggc ggc cac cgc        2845
Asn Gly Gln Cys Gly Gln Leu Cys Leu Ala Ile Pro Gly Gly His Arg
910                 915                 920                 925
tgc ggc tgc gcc tca cac tac acc ctg gac ccc agc agc cgc aac tgc        2893
Cys Gly Cys Ala Ser His Tyr Thr Leu Asp Pro Ser Ser Arg Asn Cys
                930                 935                 940
agc ccg ccc acc acc ttc ttg ctg ttc agc cag aaa tct gcc atc agt        2941
Ser Pro Pro Thr Thr Phe Leu Leu Phe Ser Gln Lys Ser Ala Ile Ser
            945                 950                 955
cgg atg atc ccg gac gac cag cac agc ccg gat ctc atc ctg ccc ctg        2989
Arg Met Ile Pro Asp Asp Gln His Ser Pro Asp Leu Ile Leu Pro Leu
        960                 965                 970
cat gga ctg agg aac gtc aaa gcc atc gac tat gac cca ctg gac aag        3037
His Gly Leu Arg Asn Val Lys Ala Ile Asp Tyr Asp Pro Leu Asp Lys
    975                 980                 985
ttc atc tac tgg gtg gat ggg cgc cag aac atc aag cga gcc aag gac        3085
Phe Ile Tyr Trp Val Asp Gly Arg Gln Asn Ile Lys Arg Ala Lys Asp
990                 995                 1000                1005
gac ggg acc cag ccc ttt gtt ttg acc tct ctg agc caa ggc caa aac        3133
Asp Gly Thr Gln Pro Phe Val Leu Thr Ser Leu Ser Gln Gly Gln Asn
                1010                1015                1020
cca gac agg cag ccc cac gac ctc agc atc gac atc tac agc cgg aca        3181
Pro Asp Arg Gln Pro His Asp Leu Ser Ile Asp Ile Tyr Ser Arg Thr
            1025                1030                1035
ctg ttc tgg acg tgc gag gcc acc aat acc atc aac gtc cac agg ctg        3229
Leu Phe Trp Thr Cys Glu Ala Thr Asn Thr Ile Asn Val His Arg Leu
        1040                1045                1050
agc ggg gaa gcc atg ggg gtg gtg ctg cgt ggg gac cgc gac aag ccc        3277
Ser Gly Glu Ala Met Gly Val Val Leu Arg Gly Asp Arg Asp Lys Pro
    1055                1060                1065
agg gcc atc gtc gtc aac gcg gag cga ggg tac ctg tac ttc acc aac        3325
Arg Ala Ile Val Val Asn Ala Glu Arg Gly Tyr Leu Tyr Phe Thr Asn
1070                1075                1080                1085
atg cag gac cgg gca gcc aag atc gaa cgc gca gcc ctg gac ggc acc        3373
Met Gln Asp Arg Ala Ala Lys Ile Glu Arg Ala Ala Leu Asp Gly Thr
                1090                1095                1100
gag cgc gag gtc ctc ttc acc acc ggc ctc atc cgc cct gtg gcc ctg        3421
Glu Arg Glu Val Leu Phe Thr Thr Gly Leu Ile Arg Pro Val Ala Leu
            1105                1110                1115
gtg gtg gac aac aca ctg ggc aag ctg ttc tgg gtg gac gcg gac ctg        3469
Val Val Asp Asn Thr Leu Gly Lys Leu Phe Trp Val Asp Ala Asp Leu
        1120                1125                1130
aag cgc att gag agc tgt gac ctg tca ggg gcc aac cgc ctg acc ctg        3517
Lys Arg Ile Glu Ser Cys Asp Leu Ser Gly Ala Asn Arg Leu Thr Leu
```

```
    1135                1140                1145
gag gac gcc aac atc gtg cag cct ctg ggc ctg acc atc ctt ggc aag        3565
Glu Asp Ala Asn Ile Val Gln Pro Leu Gly Leu Thr Ile Leu Gly Lys
1150                1155                1160                1165
cat ctc tac tgg atc gac cgc cag cag cag atg atc gag cgt gtg gag        3613
His Leu Tyr Trp Ile Asp Arg Gln Gln Gln Met Ile Glu Arg Val Glu
                1170                1175                1180
aag acc acc ggg gac aag cgg act cgc atc cag ggc cgt gtc gcc cac        3661
Lys Thr Thr Gly Asp Lys Arg Thr Arg Ile Gln Gly Arg Val Ala His
            1185                1190                1195
ctc act ggc atc cat gca gtg gag gaa gtc agc ctg gag gag ttc tca        3709
Leu Thr Gly Ile His Ala Val Glu Glu Val Ser Leu Glu Glu Phe Ser
        1200                1205                1210
gcc cac cca tgt gcc cgt gac aat ggt ggc tgc tcc cac atc tgt att        3757
Ala His Pro Cys Ala Arg Asp Asn Gly Gly Cys Ser His Ile Cys Ile
    1215                1220                1225
gcc aag ggt gat ggg aca cca cgg tgc tca tgc cca gtc cac ctc gtg        3805
Ala Lys Gly Asp Gly Thr Pro Arg Cys Ser Cys Pro Val His Leu Val
1230                1235                1240                1245
ctc ctg cag aac ctg ctg acc tgt gga gag ccg ccc acc tgc tcc ccg        3853
Leu Leu Gln Asn Leu Leu Thr Cys Gly Glu Pro Pro Thr Cys Ser Pro
                1250                1255                1260
gac cag ttt gca tgt gcc aca ggg gag atc gac tgt atc ccc ggg gcc        3901
Asp Gln Phe Ala Cys Ala Thr Gly Glu Ile Asp Cys Ile Pro Gly Ala
            1265                1270                1275
tgg cgc tgt gac ggc ttt ccc gag tgc gat gac cag agc gac gag gag        3949
Trp Arg Cys Asp Gly Phe Pro Glu Cys Asp Asp Gln Ser Asp Glu Glu
        1280                1285                1290
ggc tgc ccc gtg tgc tcc gcc gcc cag ttc ccc tgc gcg cgg ggt cag        3997
Gly Cys Pro Val Cys Ser Ala Ala Gln Phe Pro Cys Ala Arg Gly Gln
    1295                1300                1305
tgt gtg gac ctg cgc ctg cgc tgc gac ggc gag gca gac tgt cag gac        4045
Cys Val Asp Leu Arg Leu Arg Cys Asp Gly Glu Ala Asp Cys Gln Asp
1310                1315                1320                1325
cgc tca gac gag gtg gac tgt gac gcc atc tgc ctg ccc aac cag ttc        4093
Arg Ser Asp Glu Val Asp Cys Asp Ala Ile Cys Leu Pro Asn Gln Phe
                1330                1335                1340
cgg tgt gcg agc ggc cag tgt gtc ctc atc aaa cag cag tgc gac tcc        4141
Arg Cys Ala Ser Gly Gln Cys Val Leu Ile Lys Gln Gln Cys Asp Ser
            1345                1350                1355
ttc ccc gac tgt atc gac ggc tcc gac gag ctc atg tgt gaa atc acc        4189
Phe Pro Asp Cys Ile Asp Gly Ser Asp Glu Leu Met Cys Glu Ile Thr
        1360                1365                1370
aag ccg ccc tca gac gac agc ccg gcc cac agc agt gcc atc ggg ccc        4237
Lys Pro Pro Ser Asp Asp Ser Pro Ala His Ser Ser Ala Ile Gly Pro
    1375                1380                1385
gtc att ggc atc atc ctc tct ctc ttc gtc atg ggt ggt gtc tat ttt        4285
Val Ile Gly Ile Ile Leu Ser Leu Phe Val Met Gly Gly Val Tyr Phe
1390                1395                1400                1405
gtg tgc cag cgc gtg gtg tgc cag cgc tat gcg ggg gcc aac ggg ccc        4333
```

```
Val Cys Gln Arg Val Val Cys Gln Arg Tyr Ala Gly Ala Asn Gly Pro
                1410                1415                1420
ttc ccg cac gag tat gtc agc ggg acc ccg cac gtg ccc ctc aat ttc      4381
Phe Pro His Glu Tyr Val Ser Gly Thr Pro His Val Pro Leu Asn Phe
            1425                1430                1435
ata gcc ccg ggc ggt tcc cag cat ggc ccc ttc aca ggc atc gca tgc      4429
Ile Ala Pro Gly Gly Ser Gln His Gly Pro Phe Thr Gly Ile Ala Cys
        1440                1445                1450
gga aag tcc atg atg agc tcc gtg agc ctg atg ggg ggc cgg ggc ggg      4477
Gly Lys Ser Met Met Ser Ser Val Ser Leu Met Gly Gly Arg Gly Gly
    1455                1460                1465
gtg ccc ctc tac gac cgg aac cac gtc aca ggg gcc tcg tcc agc agc      4525
Val Pro Leu Tyr Asp Arg Asn His Val Thr Gly Ala Ser Ser Ser Ser
1470                1475                1480                1485
tcg tcc agc acg aag gcc acg ctg tac ccg ccg atc ctg aac ccg ccg      4573
Ser Ser Ser Thr Lys Ala Thr Leu Tyr Pro Pro Ile Leu Asn Pro Pro
                1490                1495                1500
ccc tcc ccg gcc acg gac ccc tcc ctg tac aac atg gac atg ttc tac      4621
Pro Ser Pro Ala Thr Asp Pro Ser Leu Tyr Asn Met Asp Met Phe Tyr
            1505                1510                1515
tct tca aac att ccg gcc act gcg aga ccg tac agg ccc tac atc att      4669
Ser Ser Asn Ile Pro Ala Thr Ala Arg Pro Tyr Arg Pro Tyr Ile Ile
        1520                1525                1530
cga gga atg gcg ccc ccg acg acg ccc tgc agc acc gac gtg tgt gac      4717
Arg Gly Met Ala Pro Pro Thr Thr Pro Cys Ser Thr Asp Val Cys Asp
    1535                1540                1545
agc gac tac agc gcc agc cgc tgg aag gcc agc aag tac tac ctg gat      4765
Ser Asp Tyr Ser Ala Ser Arg Trp Lys Ala Ser Lys Tyr Tyr Leu Asp
1550                1555                1560                1565
ttg aac tcg gac tca gac ccc tat cca ccc cca ccc acg ccc cac agc      4813
Leu Asn Ser Asp Ser Asp Pro Tyr Pro Pro Pro Pro Thr Pro His Ser
                1570                1575                1580
cag tac ctg tcg gcg gag gac agc tgc ccg ccc tcg ccc gcc acc gag      4861
Gln Tyr Leu Ser Ala Glu Asp Ser Cys Pro Pro Ser Pro Ala Thr Glu
            1585                1590                1595
agg agc tac ttc cat ctc ttc ccg ccc cct ccg tcc ccc tgc acg gac      4909
Arg Ser Tyr Phe His Leu Phe Pro Pro Pro Pro Ser Pro Cys Thr Asp
        1600                1605                1610
tca tcc tgacctcggc cgggccactc tggcttctct gtgcccctgt aaatagtttt      4965
Ser Ser
    1615
aaatatgaac aaagaaaaaa atatatttta tgatttaaaa aataaatata attgggattt      5025
taaaaacatg agaaatgtga actgtgatgg ggtgggcagg gctgggagaa ctttgtacag      5085
tggagaaata tttataaact taattttgta aaaca                                 5120
```

<210> 2
<211> 5120
<212> DNA
<213> Homo sapiens

<400> 2

```
actaaagcgc cgccgccgcg ccatggagcc cgagtgagcg cggcgcgggc ccgtccggcc        60
gccggacaac  atg gag gca gcg ccg ccc ggg ccg ccg tgg ccg ctg ctg          109
            Met Glu Ala Ala Pro Pro Gly Pro Pro Trp Pro Leu Leu
            1               5                   10
 ctg ctg ctg ctg ctg ctg ctg gcg ctg tgc ggc tgc ccg gcc ccc gcc         157
 Leu Leu Leu Leu Leu Leu Leu Ala Leu Cys Gly Cys Pro Ala Pro Ala
     15                  20                  25
gcg gcc tcg ccg ctc ctg cta ttt gcc aac cgc cgg gac gta cgg ctg          205
Ala Ala Ser Pro Leu Leu Leu Phe Ala Asn Arg Arg Asp Val Arg Leu
30                  35                  40                  45
gtg gac gcc ggc gga gtc aag ctg gag tcc acc atc gtg gtc agc ggc          253
Val Asp Ala Gly Gly Val Lys Leu Glu Ser Thr Ile Val Val Ser Gly
                    50                  55                  60
ctg gag gat gcg gcc gca gtg gac ttc cag ttt tcc aag gga gcc gtg          301
Leu Glu Asp Ala Ala Ala Val Asp Phe Gln Phe Ser Lys Gly Ala Val
                65                  70                  75
tac tgg aca gac gtg agc gag gag gcc atc aag cag acc tac ctg aac          349
Tyr Trp Thr Asp Val Ser Glu Glu Ala Ile Lys Gln Thr Tyr Leu Asn
            80                  85                  90
cag acg ggg gcc gcc gtg cag aac gtg gtc atc tcc ggc ctg gtc tct          397
Gln Thr Gly Ala Ala Val Gln Asn Val Val Ile Ser Gly Leu Val Ser
    95                  100                 105
ccc gac ggc ctc gcc tgc gac tgg gtg ggc aag aag ctg tac tgg acg          445
Pro Asp Gly Leu Ala Cys Asp Trp Val Gly Lys Lys Leu Tyr Trp Thr
110                 115                 120                 125
gac tca gag acc aac cgc atc gag gtg gcc aac ctc aat ggc aca tcc          493
Asp Ser Glu Thr Asn Arg Ile Glu Val Ala Asn Leu Asn Gly Thr Ser
                130                 135                 140
cgg aag gtg ctc ttc tgg cag gac ctt gac cag ccg agg gcc atc gcc          541
Arg Lys Val Leu Phe Trp Gln Asp Leu Asp Gln Pro Arg Ala Ile Ala
            145                 150                 155
ttg gac ccc gct cac ggg tac atg tac tgg aca gac tgg gtt gag acg          589
Leu Asp Pro Ala His Gly Tyr Met Tyr Trp Thr Asp Trp Val Glu Thr
        160                 165                 170
ccc cgg att gag cgg gca ggg atg gat ggc agc acc cgg aag atc att          637
Pro Arg Ile Glu Arg Ala Gly Met Asp Gly Ser Thr Arg Lys Ile Ile
    175                 180                 185
gtg gac tcg gac att tac tgg ccc aat gga ctg acc atc gac ctg gag          685
Val Asp Ser Asp Ile Tyr Trp Pro Asn Gly Leu Thr Ile Asp Leu Glu
190                 195                 200                 205
gag cag aag ctc tac tgg gct gac gcc aag ctc agc ttc atc cac cgt          733
Glu Gln Lys Leu Tyr Trp Ala Asp Ala Lys Leu Ser Phe Ile His Arg
                210                 215                 220
gcc aac ctg gac ggc tcg ttc cgg cag aag gtg gtg gag ggc agc ctg          781
Ala Asn Leu Asp Gly Ser Phe Arg Gln Lys Val Val Glu Gly Ser Leu
            225                 230                 235
acg cac ccc ttc gcc ctg acg ctc tcc ggg gac act ctg tac tgg aca          829
Thr His Pro Phe Ala Leu Thr Leu Ser Gly Asp Thr Leu Tyr Trp Thr
        240                 245                 250
```

```
gac tgg cag acc cgc tcc atc cat gcc tgc aac aag cgc act ggg ggg        877
Asp Trp Gln Thr Arg Ser Ile His Ala Cys Asn Lys Arg Thr Gly Gly
    255                 260                 265
aag agg aag gag atc ctg agt gcc ctc tac tca ccc atg gac atc cag        925
Lys Arg Lys Glu Ile Leu Ser Ala Leu Tyr Ser Pro Met Asp Ile Gln
270                 275                 280                 285
gtg ctg agc cag gag cgg cag cct ttc ttc cac act cgc tgt gag gag        973
Val Leu Ser Gln Glu Arg Gln Pro Phe Phe His Thr Arg Cys Glu Glu
                290                 295                 300
gac aat ggc ggc tgc tcc cac ctg tgc ctg ctg tcc cca agc gag cct       1021
Asp Asn Gly Gly Cys Ser His Leu Cys Leu Leu Ser Pro Ser Glu Pro
            305                 310                 315
ttc tac aca tgc gcc tgc ccc acg ggt gtg cag ctg cag gac aac ggc       1069
Phe Tyr Thr Cys Ala Cys Pro Thr Gly Val Gln Leu Gln Asp Asn Gly
        320                 325                 330
agg acg tgt aag gca gga gcc gag gag gtg ctg ctg ctg gcc cgg cgg       1117
Arg Thr Cys Lys Ala Gly Ala Glu Glu Val Leu Leu Leu Ala Arg Arg
    335                 340                 345
acg gac cta cgg agg atc tcg ctg gac acg ccg gac ttc acc gac atc       1165
Thr Asp Leu Arg Arg Ile Ser Leu Asp Thr Pro Asp Phe Thr Asp Ile
350                 355                 360                 365
gtg ctg cag gtg gac gac atc cgg cac gcc att gcc atc gac tac gac       1213
Val Leu Gln Val Asp Asp Ile Arg His Ala Ile Ala Ile Asp Tyr Asp
                370                 375                 380
ccg cta gag ggc tat gtc tac tgg aca gat gac gag gtg cgg gcc atc       1261
Pro Leu Glu Gly Tyr Val Tyr Trp Thr Asp Asp Glu Val Arg Ala Ile
            385                 390                 395
cgc agg gcg tac ctg gac ggg tct ggg gcg cag acg ctg gtc aac acc       1309
Arg Arg Ala Tyr Leu Asp Gly Ser Gly Ala Gln Thr Leu Val Asn Thr
        400                 405                 410
gag atc aac gac ccc gat ggc atc gcg gtc gac tgg gtg gcc cga aac       1357
Glu Ile Asn Asp Pro Asp Gly Ile Ala Val Asp Trp Val Ala Arg Asn
    415                 420                 425
ctc tac tgg acc gac acg ggc acg gac cgc atc gag gtg acg cgc ctc       1405
Leu Tyr Trp Thr Asp Thr Gly Thr Asp Arg Ile Glu Val Thr Arg Leu
430                 435                 440                 445
aac ggc acc tcc cgc aag atc ctg gtg tcg gag gac ctg gac gag ccc       1453
Asn Gly Thr Ser Arg Lys Ile Leu Val Ser Glu Asp Leu Asp Glu Pro
                450                 455                 460
cga gcc atc gca ctg cac ccc gtg atg ggc ctc atg tac tgg aca gac       1501
Arg Ala Ile Ala Leu His Pro Val Met Gly Leu Met Tyr Trp Thr Asp
            465                 470                 475
tgg gga gag aac cct aaa atc gag tgt gcc aac ttg gat ggg cag gag       1549
Trp Gly Glu Asn Pro Lys Ile Glu Cys Ala Asn Leu Asp Gly Gln Glu
        480                 485                 490
cgg cgt gtg ctg gtc aat gcc tcc ctc ggg tgg ccc aac ggc ctg gcc       1597
Arg Arg Val Leu Val Asn Ala Ser Leu Gly Trp Pro Asn Gly Leu Ala
    495                 500                 505
ctg gac ctg cag gag ggg aag ctc tac tgg gga gac gcc aag aca gac       1645
Leu Asp Leu Gln Glu Gly Lys Leu Tyr Trp Gly Asp Ala Lys Thr Asp
```

```
510                 515                 520                 525
aag atc gag gtg atc aat gtt gat ggg acg aag agg cgg acc ctc ctg        1693
Lys Ile Glu Val Ile Asn Val Asp Gly Thr Lys Arg Arg Thr Leu Leu
                530                 535                 540
gag gac aag ctc ccg cac att ttc ggg ttc acg ctg ctg ggg gac ttc        1741
Glu Asp Lys Leu Pro His Ile Phe Gly Phe Thr Leu Leu Gly Asp Phe
            545                 550                 555
atc tac tgg act gac tgg cag cgc cgc agc atc gag cgg gtg cac aag        1789
Ile Tyr Trp Thr Asp Trp Gln Arg Arg Ser Ile Glu Arg Val His Lys
        560                 565                 570
gtc aag gcc agc cgg gac gtc atc att gac cag ctg ccc gac ctg atg        1837
Val Lys Ala Ser Arg Asp Val Ile Ile Asp Gln Leu Pro Asp Leu Met
    575                 580                 585
ggg ctc aaa gct gtg aat gtg gcc aag gtc gtc gga acc aac ccg tgt        1885
Gly Leu Lys Ala Val Asn Val Ala Lys Val Val Gly Thr Asn Pro Cys
590                 595                 600                 605
gcg gac agg aac ggg ggg tgc agc cac ctg tgc ttc ttc aca ccc cac        1933
Ala Asp Arg Asn Gly Gly Cys Ser His Leu Cys Phe Phe Thr Pro His
                610                 615                 620
gca acc cgg tgt ggc tgc ccc atc ggc ctg gag ctg ctg agt gac atg        1981
Ala Thr Arg Cys Gly Cys Pro Ile Gly Leu Glu Leu Leu Ser Asp Met
            625                 630                 635
aag acc tgc atc gtg cct gag gcc ttc ttg gtc ttc acc agc aga gcc        2029
Lys Thr Cys Ile Val Pro Glu Ala Phe Leu Val Phe Thr Ser Arg Ala
        640                 645                 650
gcc atc cac agg atc tcc ctc gag acc aat aac aac gac gtg gcc atc        2077
Ala Ile His Arg Ile Ser Leu Glu Thr Asn Asn Asn Asp Val Ala Ile
    655                 660                 665
ccg ctc acg ggc gtc aag gag gcc tca gcc ctg gac ttt gat gtg tcc        2125
Pro Leu Thr Gly Val Lys Glu Ala Ser Ala Leu Asp Phe Asp Val Ser
670                 675                 680                 685
aac aac cac atc tac tgg aca gac gtc agc ctg aag acc atc agc cgc        2173
Asn Asn His Ile Tyr Trp Thr Asp Val Ser Leu Lys Thr Ile Ser Arg
                690                 695                 700
gcc ttc atg aac ggg agc tcg gtg gag cac gtg gtg gag ttt ggc ctt        2221
Ala Phe Met Asn Gly Ser Ser Val Glu His Val Val Glu Phe Gly Leu
            705                 710                 715
gac tac ccc gag ggc atg gcc gtt gac tgg atg ggc aag aac ctc tac        2269
Asp Tyr Pro Glu Gly Met Ala Val Asp Trp Met Gly Lys Asn Leu Tyr
        720                 725                 730
tgg gcc gac act ggg acc aac aga atc gaa gtg gcg cgg ctg gac ggg        2317
Trp Ala Asp Thr Gly Thr Asn Arg Ile Glu Val Ala Arg Leu Asp Gly
    735                 740                 745
cag ttc cgg caa gtc ctc gtg tgg agg gac ttg gac aac ccg agg tcg        2365
Gln Phe Arg Gln Val Leu Val Trp Arg Asp Leu Asp Asn Pro Arg Ser
750                 755                 760                 765
ctg gcc ctg gat ccc acc aag ggc tac atc tac tgg acc gag tgg ggc        2413
Leu Ala Leu Asp Pro Thr Lys Gly Tyr Ile Tyr Trp Thr Glu Trp Gly
                770                 775                 780
ggc aag ccg agg atc gtg cgg gcc ttc atg gac ggg acc aac tgc atg        2461
```

```
Gly Lys Pro Arg Ile Val Arg Ala Phe Met Asp Gly Thr Asn Cys Met
                785                 790                 795
acg ctg gtg gac aag gtg ggc cgg gcc aac gac ctc acc att gac tac        2509
Thr Leu Val Asp Lys Val Gly Arg Ala Asn Asp Leu Thr Ile Asp Tyr
        800                 805                 810
gct gac cag cgc ctc tac tgg acc gac ctg gac acc aac atg atc gag        2557
Ala Asp Gln Arg Leu Tyr Trp Thr Asp Leu Asp Thr Asn Met Ile Glu
    815                 820                 825
tcg tcc aac atg ctg ggt cag gag cgg gtc gtg att gcc gac gat ctc        2605
Ser Ser Asn Met Leu Gly Gln Glu Arg Val Val Ile Ala Asp Asp Leu
830                 835                 840                 845
ccg cac ccg ttc ggt ctg acg cag tac agc gat tat atc tac tgg aca        2653
Pro His Pro Phe Gly Leu Thr Gln Tyr Ser Asp Tyr Ile Tyr Trp Thr
                850                 855                 860
gac tgg aat ctg cac agc att gag cgg gcc gac aag act agc ggc cgg        2701
Asp Trp Asn Leu His Ser Ile Glu Arg Ala Asp Lys Thr Ser Gly Arg
            865                 870                 875
aac cgc acc ctc atc cag ggc cac ctg gac ttc gtg atg gac atc ctg        2749
Asn Arg Thr Leu Ile Gln Gly His Leu Asp Phe Val Met Asp Ile Leu
        880                 885                 890
gtg ttc cac tcc tcc cgc cag gat ggc ctc aat gac tgt atg cac aac        2797
Val Phe His Ser Ser Arg Gln Asp Gly Leu Asn Asp Cys Met His Asn
    895                 900                 905
aac ggg cag tgt ggg cag ctg tgc ctt gcc atc ccc ggc ggc cac cgc        2845
Asn Gly Gln Cys Gly Gln Leu Cys Leu Ala Ile Pro Gly Gly His Arg
910                 915                 920                 925
tgc ggc tgc gcc tca cac tac acc ctg gac ccc agc agc cgc aac tgc        2893
Cys Gly Cys Ala Ser His Tyr Thr Leu Asp Pro Ser Ser Arg Asn Cys
                930                 935                 940
agc ccg ccc acc acc ttc ttg ctg ttc agc cag aaa tct gcc atc agt        2941
Ser Pro Pro Thr Thr Phe Leu Leu Phe Ser Gln Lys Ser Ala Ile Ser
            945                 950                 955
cgg atg atc ccg gac gac cag cac agc ccg gat ctc atc ctg ccc ctg        2989
Arg Met Ile Pro Asp Asp Gln His Ser Pro Asp Leu Ile Leu Pro Leu
        960                 965                 970
cat gga ctg agg aac gtc aaa gcc atc gac tat gac cca ctg gac aag        3037
His Gly Leu Arg Asn Val Lys Ala Ile Asp Tyr Asp Pro Leu Asp Lys
    975                 980                 985
ttc atc tac tgg gtg gat ggg cgc cag aac atc aag cga gcc aag gac        3085
Phe Ile Tyr Trp Val Asp Gly Arg Gln Asn Ile Lys Arg Ala Lys Asp
990                 995                 1000                1005
gac ggg acc cag ccc ttt gtt ttg acc tct ctg agc caa ggc caa aac        3133
Asp Gly Thr Gln Pro Phe Val Leu Thr Ser Leu Ser Gln Gly Gln Asn
                1010                1015                1020
cca gac agg cag ccc cac gac ctc agc atc gac atc tac agc cgg aca        3181
Pro Asp Arg Gln Pro His Asp Leu Ser Ile Asp Ile Tyr Ser Arg Thr
            1025                1030                1035
ctg ttc tgg acg tgc gag gcc acc aat acc atc aac gtc cac agg ctg        3229
Leu Phe Trp Thr Cys Glu Ala Thr Asn Thr Ile Asn Val His Arg Leu
        1040                1045                1050
```

```
agc ggg gaa gcc atg ggg gtg gtg ctg cgt ggg gac cgc gac aag ccc        3277
Ser Gly Glu Ala Met Gly Val Val Leu Arg Gly Asp Arg Asp Lys Pro
    1055                1060                1065
agg gcc atc gtc gtc aac gcg gag cga ggg tac ctg tac ttc acc aac        3325
Arg Ala Ile Val Val Asn Ala Glu Arg Gly Tyr Leu Tyr Phe Thr Asn
1070                1075                1080                1085
atg cag gac cgg gca gcc aag atc gaa cgc gca gcc ctg gac ggc acc        3373
Met Gln Asp Arg Ala Ala Lys Ile Glu Arg Ala Ala Leu Asp Gly Thr
                1090                1095                1100
gag cgc gag gtc ctc ttc acc acc ggc ctc atc cgc cct gtg gcc ctg        3421
Glu Arg Glu Val Leu Phe Thr Thr Gly Leu Ile Arg Pro Val Ala Leu
            1105                1110                1115
gtg gtg gac aac aca ctg ggc aag ctg ttc tgg gtg gac gcg gac ctg        3469
Val Val Asp Asn Thr Leu Gly Lys Leu Phe Trp Val Asp Ala Asp Leu
        1120                1125                1130
aag cgc att gag agc tgt gac ctg tca ggg gcc aac cgc ctg acc ctg        3517
Lys Arg Ile Glu Ser Cys Asp Leu Ser Gly Ala Asn Arg Leu Thr Leu
    1135                1140                1145
gag gac gcc aac atc gtg cag cct ctg ggc ctg acc atc ctt ggc aag        3565
Glu Asp Ala Asn Ile Val Gln Pro Leu Gly Leu Thr Ile Leu Gly Lys
1150                1155                1160                1165
cat ctc tac tgg atc gac cgc cag cag cag atg atc gag cgt gtg gag        3613
His Leu Tyr Trp Ile Asp Arg Gln Gln Gln Met Ile Glu Arg Val Glu
                1170                1175                1180
aag acc acc ggg gac aag cgg act cgc atc cag ggc cgt gtc gcc cac        3661
Lys Thr Thr Gly Asp Lys Arg Thr Arg Ile Gln Gly Arg Val Ala His
            1185                1190                1195
ctc act ggc atc cat gca gtg gag gaa gtc agc ctg gag gag ttc tca        3709
Leu Thr Gly Ile His Ala Val Glu Glu Val Ser Leu Glu Glu Phe Ser
        1200                1205                1210
gcc cac cca tgt gcc cgt gac aat ggt ggc tgc tcc cac atc tgt att        3757
Ala His Pro Cys Ala Arg Asp Asn Gly Gly Cys Ser His Ile Cys Ile
    1215                1220                1225
gcc aag ggt gat ggg aca cca cgg tgc tca tgc cca gtc cac ctc gtg        3805
Ala Lys Gly Asp Gly Thr Pro Arg Cys Ser Cys Pro Val His Leu Val
1230                1235                1240                1245
ctc ctg cag aac ctg ctg acc tgt gga gag ccg ccc acc tgc tcc ccg        3853
Leu Leu Gln Asn Leu Leu Thr Cys Gly Glu Pro Pro Thr Cys Ser Pro
                1250                1255                1260
gac cag ttt gca tgt gcc aca ggg gag atc gac tgt atc ccc ggg gcc        3901
Asp Gln Phe Ala Cys Ala Thr Gly Glu Ile Asp Cys Ile Pro Gly Ala
            1265                1270                1275
tgg cgc tgt gac ggc ttt ccc gag tgc gat gac cag agc gac gag gag        3949
Trp Arg Cys Asp Gly Phe Pro Glu Cys Asp Asp Gln Ser Asp Glu Glu
        1280                1285                1290
ggc tgc ccc gtg tgc tcc gcc gcc cag ttc ccc tgc gcg cgg ggt cag        3997
Gly Cys Pro Val Cys Ser Ala Ala Gln Phe Pro Cys Ala Arg Gly Gln
    1295                1300                1305
tgt gtg gac ctg cgc ctg cgc tgc gac ggc gag gca gac tgt cag gac        4045
Cys Val Asp Leu Arg Leu Arg Cys Asp Gly Glu Ala Asp Cys Gln Asp
```

```
1310                1315                1320                1325
cgc tca gac gag gtg gac tgt gac gcc atc tgc ctg ccc aac cag ttc      4093
Arg Ser Asp Glu Val Asp Cys Asp Ala Ile Cys Leu Pro Asn Gln Phe
                1330                1335                1340
cgg tgt gcg agc ggc cag tgt gtc ctc atc aaa cag cag tgc gac tcc      4141
Arg Cys Ala Ser Gly Gln Cys Val Leu Ile Lys Gln Gln Cys Asp Ser
            1345                1350                1355
ttc ccc gac tgt atc gac ggc tcc gac gag ctc atg tgt gaa atc acc      4189
Phe Pro Asp Cys Ile Asp Gly Ser Asp Glu Leu Met Cys Glu Ile Thr
        1360                1365                1370
aag ccg ccc tca gac gac agc ccg gcc cac agc agt gcc atc ggg ccc      4237
Lys Pro Pro Ser Asp Asp Ser Pro Ala His Ser Ser Ala Ile Gly Pro
    1375                1380                1385
gtc att ggc atc atc ctc tct ctc ttc gtc atg ggt ggt gtc tat ttt      4285
Val Ile Gly Ile Ile Leu Ser Leu Phe Val Met Gly Gly Val Tyr Phe
1390                1395                1400                1405
gtg tgc cag cgc gtg gtg tgc cag cgc tat gcg ggg gcc aac ggg ccc      4333
Val Cys Gln Arg Val Val Cys Gln Arg Tyr Ala Gly Ala Asn Gly Pro
                1410                1415                1420
ttc ccg cac gag tat gtc agc ggg acc ccg cac gtg ccc ctc aat ttc      4381
Phe Pro His Glu Tyr Val Ser Gly Thr Pro His Val Pro Leu Asn Phe
            1425                1430                1435
ata gcc ccg ggc ggt tcc cag cat ggc ccc ttc aca ggc atc gca tgc      4429
Ile Ala Pro Gly Gly Ser Gln His Gly Pro Phe Thr Gly Ile Ala Cys
        1440                1445                1450
gga aag tcc atg atg agc tcc gtg agc ctg atg ggg ggc cgg ggc ggg      4477
Gly Lys Ser Met Met Ser Ser Val Ser Leu Met Gly Gly Arg Gly Gly
    1455                1460                1465
gtg ccc ctc tac gac cgg aac cac gtc aca ggg gcc tcg tcc agc agc      4525
Val Pro Leu Tyr Asp Arg Asn His Val Thr Gly Ala Ser Ser Ser Ser
1470                1475                1480                1485
tcg tcc agc acg aag gcc acg ctg tac ccg ccg atc ctg aac ccg ccg      4573
Ser Ser Ser Thr Lys Ala Thr Leu Tyr Pro Pro Ile Leu Asn Pro Pro
                1490                1495                1500
ccc tcc ccg gcc acg gac ccc tcc ctg tac aac atg gac atg ttc tac      4621
Pro Ser Pro Ala Thr Asp Pro Ser Leu Tyr Asn Met Asp Met Phe Tyr
            1505                1510                1515
tct tca aac att ccg gcc act gcg aga ccg tac agg ccc tac atc att      4669
Ser Ser Asn Ile Pro Ala Thr Ala Arg Pro Tyr Arg Pro Tyr Ile Ile
        1520                1525                1530
cga gga atg gcg ccc ccg acg acg ccc tgc agc acc gac gtg tgt gac      4717
Arg Gly Met Ala Pro Pro Thr Thr Pro Cys Ser Thr Asp Val Cys Asp
    1535                1540                1545
agc gac tac agc gcc agc cgc tgg aag gcc agc aag tac tac ctg gat      4765
Ser Asp Tyr Ser Ala Ser Arg Trp Lys Ala Ser Lys Tyr Tyr Leu Asp
1550                1555                1560                1565
ttg aac tcg gac tca gac ccc tat cca ccc cca ccc acg ccc cac agc      4813
Leu Asn Ser Asp Ser Asp Pro Tyr Pro Pro Pro Pro Thr Pro His Ser
                1570                1575                1580
cag tac ctg tcg gcg gag gac agc tgc ccg ccc tcg ccc gcc acc gag      4861
```

```
Gln Tyr Leu Ser Ala Glu Asp Ser Cys Pro Pro Ser Pro Ala Thr Glu
                1585                1590                1595
agg agc tac ttc cat ctc ttc ccg ccc cct ccg tcc ccc tgc acg gac      4909
Arg Ser Tyr Phe His Leu Phe Pro Pro Pro Pro Ser Pro Cys Thr Asp
        1600                1605                1610
tca tcc tgacctcggc cgggccactc tggcttctct gtgcccctgt aaatagtttt      4965
Ser Ser
    1615
aaatatgaac aaagaaaaaa atatatttta tgatttaaaa ataaatata attgggattt      5025
taaaaacatg agaaatgtga actgtgatgg ggtgggcagg gctgggagaa ctttgtacag      5085
tggagaaata tttataaact taattttgta aaaca                                5120

<210> 3
<211> 1615
<212> PRT
<213> Homo sapiens

<400> 3
Met Glu Ala Ala Pro Pro Gly Pro Pro Trp Pro Leu Leu Leu Leu Leu
 1               5                  10                  15
 Leu Leu Leu Leu Ala Leu Cys Gly Cys Pro Ala Pro Ala Ala Ala Ser
            20                  25                  30
 Pro Leu Leu Leu Phe Ala Asn Arg Arg Asp Val Arg Leu Val Asp Ala
        35                  40                  45
 Gly Gly Val Lys Leu Glu Ser Thr Ile Val Val Ser Gly Leu Glu Asp
    50                  55                  60
Ala Ala Ala Val Asp Phe Gln Phe Ser Lys Gly Ala Val Tyr Trp Thr
65                  70                  75                  80
Asp Val Ser Glu Glu Ala Ile Lys Gln Thr Tyr Leu Asn Gln Thr Gly
                85                  90                  95
Ala Ala Val Gln Asn Val Val Ile Ser Gly Leu Val Ser Pro Asp Gly
            100                 105                 110
Leu Ala Cys Asp Trp Val Gly Lys Lys Leu Tyr Trp Thr Asp Ser Glu
        115                 120                 125
Thr Asn Arg Ile Glu Val Ala Asn Leu Asn Gly Thr Ser Arg Lys Val
    130                 135                 140
Leu Phe Trp Gln Asp Leu Asp Gln Pro Lys Ala Ile Ala Leu Asp Pro
145                 150                 155                 160
Ala His Gly Tyr Met Tyr Trp Thr Asp Trp Gly Glu Thr Pro Arg Ile
                165                 170                 175
Glu Arg Ala Gly Met Asp Gly Ser Thr Arg Lys Ile Ile Val Asp Ser
            180                 185                 190
Asp Ile Tyr Trp Pro Asn Gly Leu Thr Ile Asp Leu Glu Glu Gln Lys
        195                 200                 205
Leu Tyr Trp Ala Asp Ala Lys Leu Ser Phe Ile His Arg Ala Asn Leu
    210                 215                 220
Asp Gly Ser Phe Arg Gln Lys Val Val Glu Gly Ser Leu Thr His Pro
225                 230                 235                 240
Phe Ala Leu Thr Leu Ser Gly Asp Thr Leu Tyr Trp Thr Asp Trp Gln
                245                 250                 255
```

Thr Arg Ser Ile His Ala Cys Asn Lys Arg Thr Gly Gly Lys Arg Lys
260 265 270
Glu Ile Leu Ser Ala Leu Tyr Ser Pro Met Asp Ile Gln Val Leu Ser
275 280 285
Gln Glu Arg Gln Pro Phe Phe His Thr Arg Cys Glu Glu Asp Asn Gly
290 295 300
Gly Trp Ser His Leu Cys Leu Leu Ser Pro Ser Glu Pro Phe Tyr Thr
305 310 315 320
Cys Ala Cys Pro Thr Gly Val Gln Met Gln Asp Asn Gly Arg Thr Cys
325 330 335
Lys Ala Gly Ala Glu Glu Val Leu Leu Leu Ala Arg Arg Thr Asp Leu
340 345 350
Arg Arg Ile Ser Leu Asp Thr Pro Asp Phe Thr Asp Ile Val Leu Gln
355 360 365
Val Asp Asp Ile Arg His Ala Ile Ala Ile Asp Tyr Asp Pro Leu Glu
370 375 380
Gly Tyr Val Tyr Trp Thr Asp Asp Glu Val Arg Ala Ile Arg Arg Ala
385 390 395 400
Tyr Leu Asp Gly Ser Gly Ala Gln Thr Leu Val Asn Thr Glu Ile Asn
405 410 415
Asp Pro Asp Gly Ile Ala Val Asp Trp Val Ala Arg Asn Leu Tyr Trp
420 425 430
Thr Asp Thr Gly Thr Asp Arg Ile Glu Val Thr Arg Leu Asn Gly Thr
435 440 445
Ser Arg Lys Ile Leu Val Ser Glu Asp Leu Asp Glu Pro Arg Ala Ile
450 455 460
Ala Leu His Pro Val Met Gly Leu Met Tyr Trp Thr Asp Trp Gly Glu
465 470 475 480
Asn Pro Lys Ile Glu Cys Ala Asn Leu Asp Gly Gln Glu Arg Arg Val
485 490 495
Leu Val Asn Ala Ser Leu Gly Trp Pro Asn Gly Leu Ala Leu Asp Leu
500 505 510
Gln Glu Gly Lys Leu Tyr Trp Gly Asp Ala Lys Thr Asp Lys Ile Glu
515 520 525
Val Ile Asn Val Asp Gly Thr Lys Arg Arg Thr Leu Leu Glu Asp Lys
530 535 540
Leu Pro His Ile Phe Gly Phe Thr Leu Leu Gly Asp Phe Ile Tyr Trp
545 550 555 560
Thr Asp Trp Gln Arg Arg Ser Ile Glu Arg Val His Lys Val Lys Ala
565 570 575
Ser Arg Asp Val Ile Ile Asp Gln Leu Pro Asp Leu Met Gly Leu Lys
580 585 590
Ala Val Asn Val Ala Lys Val Val Gly Thr Asn Pro Cys Ala Asp Arg
595 600 605
Asn Gly Gly Cys Ser His Leu Cys Phe Phe Thr Pro His Ala Thr Arg
610 615 620
Cys Gly Cys Pro Ile Gly Leu Glu Leu Leu Ser Asp Met Lys Thr Cys
625 630 635 640
Ile Val Pro Glu Ala Phe Leu Val Phe Thr Ser Arg Ala Ala Ile His
645 650 655

Arg Ile Ser Leu Glu Thr Asn Asn Asn Asp Val Ala Ile Pro Leu Thr
660 665 670
Gly Val Lys Glu Ala Ser Ala Leu Asp Phe Asp Val Ser Asn Asn His
675 680 685
Ile Tyr Trp Thr Asp Val Ser Leu Lys Asn Ile Ser Arg Ala Phe Met
690 695 700
Asn Gly Ser Ser Val Glu His Val Val Glu Phe Gly Leu Asp Tyr Pro
705 710 715 720
Glu Gly Met Ala Val Asp Trp Met Gly Lys Asn Leu Tyr Trp Ala Asp
725 730 735
Thr Gly Thr Asn Arg Ile Glu Val Ala Arg Leu Asp Gly Gln Phe Arg
740 745 750
Gln Val Leu Val Trp Arg Asp Leu Asp Asn Pro Arg Ser Leu Ala Leu
755 760 765
Asp Pro Thr Lys Gly Tyr Ile Tyr Trp Thr Glu Trp Gly Gly Lys Pro
770 775 780
Arg Ile Val Arg Ala Phe Met Asp Gly Thr Asn Cys Met Thr Leu Val
785 790 795 800
Asp Lys Val Gly Arg Ala Asn Asp Leu Thr Ile Asp Tyr Ala Asp Gln
805 810 815
Arg Leu Tyr Trp Thr Asp Leu Asp Thr Asn Met Ile Glu Ser Ser Asn
820 825 830
Met Leu Gly Gln Glu Arg Val Val Ile Ala Asp Asp Leu Pro His Pro
835 840 845
Phe Gly Leu Thr Gln Tyr Ser Asp Tyr Ile Tyr Trp Thr Asp Trp Asn
850 855 860
Leu His Ser Ile Glu Arg Ala Asp Lys Thr Ser Gly Arg Asn Arg Thr
865 870 875 880
Leu Ile Gln Gly His Leu Asp Phe Val Met Asp Ile Leu Val Phe His
885 890 895
Ser Ser Arg Gln Asp Gly Leu Asn Asp Cys Met His Asn Asn Gly Gln
900 905 910
Cys Gly Gln Leu Cys Leu Ala Ile Pro Gly Gly His Arg Cys Gly Cys
915 920 925
Ala Ser His Tyr Thr Leu Asp Pro Ser Ser Arg Asn Cys Ser Pro Pro
930 935 940
Thr Thr Phe Leu Leu Phe Ser Gln Lys Ser Ala Ile Ser Arg Met Ile
945 950 955 960
Pro Asp Asp Gln His Ser Pro Asp Leu Ile Leu Pro Leu His Gly Leu
965 970 975
Arg Asn Val Lys Ala Ile Asp Tyr Asp Pro Leu Asp Lys Phe Ile Tyr
980 985 990
Trp Val Asp Gly Arg Gln Asn Ile Lys Arg Ala Lys Asp Asp Gly Thr
995 1000 1005
Gln Pro Phe Val Leu Thr Ser Leu Ser Gln Gly Gln Asn Pro Asp Arg
1010 1015 1020
Gln Pro His Asp Leu Ser Ile Asp Ile Tyr Ser Arg Thr Leu Phe Trp
1025 1030 1035 1040
Thr Cys Glu Ala Thr Asn Thr Ile Asn Val His Arg Leu Ser Gly Glu
1045 1050 1055

Ala Met Gly Val Val Leu Arg Gly Asp Arg Asp Lys Pro Arg Ala Ile
1060 1065 1070
Val Val Asn Ala Glu Arg Gly Tyr Leu Tyr Phe Thr Asn Met Gln Asp
1075 1080 1085
Arg Ala Ala Lys Ile Glu Arg Ala Ala Leu Asp Gly Thr Glu Arg Glu
1090 1095 1100
Val Leu Phe Thr Thr Gly Leu Ile Arg Pro Val Ala Leu Val Val Asp
1105 1110 1115 1120
Asn Thr Leu Gly Lys Leu Phe Trp Val Asp Ala Asp Leu Lys Arg Ile
1125 1130 1135
Glu Ser Cys Asp Leu Ser Gly Ala Asn Arg Leu Thr Leu Glu Asp Ala
1140 1145 1150
Asn Ile Val Gln Pro Leu Gly Leu Thr Ile Leu Gly Lys His Leu Tyr
1155 1160 1165
Trp Ile Asp Arg Gln Gln Gln Met Ile Glu Arg Val Glu Lys Thr Thr
1170 1175 1180
Gly Asp Lys Arg Thr Arg Ile Gln Gly Arg Val Ala His Leu Thr Gly
1185 1190 1195 1200
Ile His Ala Val Glu Glu Val Ser Leu Glu Glu Phe Ser Ala His Pro
1205 1210 1215
Cys Ala Arg Asp Asn Gly Gly Cys Ser His Ile Cys Ile Ala Lys Gly
1220 1225 1230
Asp Gly Thr Pro Arg Cys Ser Cys Pro Val His Leu Val Leu Leu Gln
1235 1240 1245
Asn Leu Leu Thr Cys Gly Glu Pro Pro Thr Cys Ser Pro Asp Gln Phe
1250 1255 1260
Ala Cys Ala Thr Gly Glu Ile Asp Cys Ile Pro Gly Ala Trp Arg Cys
1265 1270 1275 1280
Asp Gly Phe Pro Glu Cys Asp Asp Gln Ser Asp Glu Glu Gly Cys Pro
1285 1290 1295
Val Cys Ser Ala Ala Gln Phe Pro Cys Ala Arg Gly Gln Cys Val Asp
1300 1305 1310
Leu Arg Leu Arg Cys Asp Gly Glu Ala Asp Cys Gln Asp Arg Ser Asp
1315 1320 1325
Glu Val Asp Cys Asp Ala Ile Cys Leu Pro Asn Gln Phe Arg Cys Ala
1330 1335 1340
Ser Gly Gln Cys Val Leu Ile Lys Gln Gln Cys Asp Ser Phe Pro Asp
1345 1350 1355 1360
Cys Ile Asp Gly Ser Asp Glu Leu Met Cys Glu Ile Thr Lys Pro Pro
1365 1370 1375
Ser Asp Asp Ser Pro Ala His Ser Ser Ala Ile Gly Pro Val Ile Gly
1380 1385 1390
Ile Ile Leu Ser Leu Phe Val Met Gly Gly Val Tyr Phe Val Cys Gln
1395 1400 1405
Arg Val Val Cys Gln Arg Tyr Ala Gly Ala Asn Gly Pro Phe Pro His
1410 1415 1420
Glu Tyr Val Ser Gly Thr Pro His Val Pro Leu Asn Phe Ile Ala Pro
1425 1430 1435 1440
Gly Gly Ser Gln His Gly Pro Phe Thr Gly Ile Ala Cys Gly Lys Ser
1445 1450 1455

```
Met Met Ser Ser Val Ser Leu Met Gly Gly Arg Gly Gly Val Pro Leu
            1460                1465                1470
Tyr Asp Arg Asn His Val Thr Gly Ala Ser Ser Ser Ser Ser Ser Ser
        1475                1480                1485
Thr Lys Ala Thr Leu Tyr Pro Pro Ile Leu Asn Pro Pro Pro Ser Pro
    1490                1495                1500
Ala Thr Asp Pro Ser Leu Tyr Asn Met Asp Met Phe Tyr Ser Ser Asn
1505                1510                1515                1520
Ile Pro Ala Thr Ala Arg Pro Tyr Arg Pro Tyr Ile Ile Arg Gly Met
                1525                1530                1535
Ala Pro Pro Thr Thr Pro Cys Ser Thr Asp Val Cys Asp Ser Asp Tyr
            1540                1545                1550
Ser Ala Ser Arg Trp Lys Ala Ser Lys Tyr Tyr Leu Asp Leu Asn Ser
        1555                1560                1565
Asp Ser Asp Pro Tyr Pro Pro Pro Pro Thr Pro His Ser Gln Tyr Leu
    1570                1575                1580
Ser Ala Glu Asp Ser Cys Pro Pro Ser Pro Ala Thr Glu Arg Ser Tyr
1585                1590                1595                1600
Phe His Leu Phe Pro Pro Pro Pro Ser Pro Cys Thr Asp Ser Ser
                1605                1610                1615
```

<210> 4
<211> 1615
<212> PRT
<213> Homo sapiens

<400> 4

```
Met Glu Ala Ala Pro Pro Gly Pro Pro Trp Pro Leu Leu Leu Leu Leu
1               5                  10                  15
Leu Leu Leu Leu Ala Leu Cys Gly Cys Pro Ala Pro Ala Ala Ala Ser
            20                  25                  30
Pro Leu Leu Leu Phe Ala Asn Arg Arg Asp Val Arg Leu Val Asp Ala
        35                  40                  45
Gly Gly Val Lys Leu Glu Ser Thr Ile Val Val Ser Gly Leu Glu Asp
    50                  55                  60
Ala Ala Ala Val Asp Phe Gln Phe Ser Lys Gly Ala Val Tyr Trp Thr
65                  70                  75                  80
Asp Val Ser Glu Glu Ala Ile Lys Gln Thr Tyr Leu Asn Gln Thr Gly
                85                  90                  95
Ala Ala Val Gln Asn Val Val Ile Ser Gly Leu Val Ser Pro Asp Gly
            100                 105                 110
Leu Ala Cys Asp Trp Val Gly Lys Lys Leu Tyr Trp Thr Asp Ser Glu
        115                 120                 125
Thr Asn Arg Ile Glu Val Ala Asn Leu Asn Gly Thr Ser Arg Lys Val
    130                 135                 140
Leu Phe Trp Gln Asp Leu Asp Gln Pro Lys Ala Ile Ala Leu Asp Pro
145                 150                 155                 160
Ala His Gly Tyr Met Tyr Trp Thr Asp Trp Val Glu Thr Pro Arg Ile
                165                 170                 175
```

Glu Arg Ala Gly Met Asp Gly Ser Thr Arg Lys Ile Ile Val Asp Ser
180 185 190
Asp Ile Tyr Trp Pro Asn Gly Leu Thr Ile Asp Leu Glu Glu Gln Lys
195 200 205
Leu Tyr Trp Ala Asp Ala Lys Leu Ser Phe Ile His Arg Ala Asn Leu
210 215 220
Asp Gly Ser Phe Arg Gln Lys Val Val Glu Gly Ser Leu Thr His Pro
225 230 235 240
Phe Ala Leu Thr Leu Ser Gly Asp Thr Leu Tyr Trp Thr Asp Trp Gln
245 250 255
Thr Arg Ser Ile His Ala Cys Asn Lys Arg Thr Gly Gly Lys Arg Lys
260 265 270
Glu Ile Leu Ser Ala Leu Tyr Ser Pro Met Asp Ile Gln Val Leu Ser
275 280 285
Gln Glu Arg Gln Pro Phe Phe His Thr Arg Cys Glu Glu Asp Asn Gly
290 295 300
Gly Trp Ser His Leu Cys Leu Leu Ser Pro Ser Glu Pro Phe Tyr Thr
305 310 315 320
Cys Ala Cys Pro Thr Gly Val Gln Met Gln Asp Asn Gly Arg Thr Cys
325 330 335
Lys Ala Gly Ala Glu Glu Val Leu Leu Leu Ala Arg Arg Thr Asp Leu
340 345 350
Arg Arg Ile Ser Leu Asp Thr Pro Asp Phe Thr Asp Ile Val Leu Gln
355 360 365
Val Asp Asp Ile Arg His Ala Ile Ala Ile Asp Tyr Asp Pro Leu Glu
370 375 380
Gly Tyr Val Tyr Trp Thr Asp Asp Glu Val Arg Ala Ile Arg Arg Ala
385 390 395 400
Tyr Leu Asp Gly Ser Gly Ala Gln Thr Leu Val Asn Thr Glu Ile Asn
405 410 415
Asp Pro Asp Gly Ile Ala Val Asp Trp Val Ala Arg Asn Leu Tyr Trp
420 425 430
Thr Asp Thr Gly Thr Asp Arg Ile Glu Val Thr Arg Leu Asn Gly Thr
435 440 445
Ser Arg Lys Ile Leu Val Ser Glu Asp Leu Asp Glu Pro Arg Ala Ile
450 455 460
Ala Leu His Pro Val Met Gly Leu Met Tyr Trp Thr Asp Trp Gly Glu
465 470 475 480
Asn Pro Lys Ile Glu Cys Ala Asn Leu Asp Gly Gln Glu Arg Arg Val
485 490 495
Leu Val Asn Ala Ser Leu Gly Trp Pro Asn Gly Leu Ala Leu Asp Leu
500 505 510
Gln Glu Gly Lys Leu Tyr Trp Gly Asp Ala Lys Thr Asp Lys Ile Glu
515 520 525
Val Ile Asn Val Asp Gly Thr Lys Arg Arg Thr Leu Leu Glu Asp Lys
530 535 540
Leu Pro His Ile Phe Gly Phe Thr Leu Leu Gly Asp Phe Ile Tyr Trp
545 550 555 560
Thr Asp Trp Gln Arg Arg Ser Ile Glu Arg Val His Lys Val Lys Ala
565 570 575

```
Ser Arg Asp Val Ile Ile Asp Gln Leu Pro Asp Leu Met Gly Leu Lys
            580                 585                 590
Ala Val Asn Val Ala Lys Val Val Gly Thr Asn Pro Cys Ala Asp Arg
        595                 600                 605
Asn Gly Gly Cys Ser His Leu Cys Phe Phe Thr Pro His Ala Thr Arg
    610                 615                 620
Cys Gly Cys Pro Ile Gly Leu Glu Leu Leu Ser Asp Met Lys Thr Cys
625                 630                 635                 640
Ile Val Pro Glu Ala Phe Leu Val Phe Thr Ser Arg Ala Ala Ile His
                645                 650                 655
Arg Ile Ser Leu Glu Thr Asn Asn Asn Asp Val Ala Ile Pro Leu Thr
            660                 665                 670
Gly Val Lys Glu Ala Ser Ala Leu Asp Phe Asp Val Ser Asn Asn His
        675                 680                 685
Ile Tyr Trp Thr Asp Val Ser Leu Lys Asn Ile Ser Arg Ala Phe Met
    690                 695                 700
Asn Gly Ser Ser Val Glu His Val Val Glu Phe Gly Leu Asp Tyr Pro
705                 710                 715                 720
Glu Gly Met Ala Val Asp Trp Met Gly Lys Asn Leu Tyr Trp Ala Asp
                725                 730                 735
Thr Gly Thr Asn Arg Ile Glu Val Ala Arg Leu Asp Gly Gln Phe Arg
            740                 745                 750
Gln Val Leu Val Trp Arg Asp Leu Asp Asn Pro Arg Ser Leu Ala Leu
        755                 760                 765
Asp Pro Thr Lys Gly Tyr Ile Tyr Trp Thr Glu Trp Gly Gly Lys Pro
    770                 775                 780
Arg Ile Val Arg Ala Phe Met Asp Gly Thr Asn Cys Met Thr Leu Val
785                 790                 795                 800
Asp Lys Val Gly Arg Ala Asn Asp Leu Thr Ile Asp Tyr Ala Asp Gln
                805                 810                 815
Arg Leu Tyr Trp Thr Asp Leu Asp Thr Asn Met Ile Glu Ser Ser Asn
            820                 825                 830
Met Leu Gly Gln Glu Arg Val Val Ile Ala Asp Asp Leu Pro His Pro
        835                 840                 845
Phe Gly Leu Thr Gln Tyr Ser Asp Tyr Ile Tyr Trp Thr Asp Trp Asn
    850                 855                 860
Leu His Ser Ile Glu Arg Ala Asp Lys Thr Ser Gly Arg Asn Arg Thr
865                 870                 875                 880
Leu Ile Gln Gly His Leu Asp Phe Val Met Asp Ile Leu Val Phe His
                885                 890                 895
Ser Ser Arg Gln Asp Gly Leu Asn Asp Cys Met His Asn Asn Gly Gln
            900                 905                 910
Cys Gly Gln Leu Cys Leu Ala Ile Pro Gly Gly His Arg Cys Gly Cys
        915                 920                 925
Ala Ser His Tyr Thr Leu Asp Pro Ser Ser Arg Asn Cys Ser Pro Pro
    930                 935                 940
Thr Thr Phe Leu Leu Phe Ser Gln Lys Ser Ala Ile Ser Arg Met Ile
945                 950                 955                 960
Pro Asp Asp Gln His Ser Pro Asp Leu Ile Leu Pro Leu His Gly Leu
                965                 970                 975
```

Arg Asn Val Lys Ala Ile Asp Tyr Asp Pro Leu Asp Lys Phe Ile Tyr
980 985 990
Trp Val Asp Gly Arg Gln Asn Ile Lys Arg Ala Lys Asp Asp Gly Thr
995 1000 1005
Gln Pro Phe Val Leu Thr Ser Leu Ser Gln Gly Gln Asn Pro Asp Arg
1010 1015 1020
Gln Pro His Asp Leu Ser Ile Asp Ile Tyr Ser Arg Thr Leu Phe Trp
1025 1030 1035 1040
Thr Cys Glu Ala Thr Asn Thr Ile Asn Val His Arg Leu Ser Gly Glu
1045 1050 1055
Ala Met Gly Val Val Leu Arg Gly Asp Arg Asp Lys Pro Arg Ala Ile
1060 1065 1070
Val Val Asn Ala Glu Arg Gly Tyr Leu Tyr Phe Thr Asn Met Gln Asp
1075 1080 1085
Arg Ala Ala Lys Ile Glu Arg Ala Ala Leu Asp Gly Thr Glu Arg Glu
1090 1095 1100
Val Leu Phe Thr Thr Gly Leu Ile Arg Pro Val Ala Leu Val Val Asp
1105 1110 1115 1120
Asn Thr Leu Gly Lys Leu Phe Trp Val Asp Ala Asp Leu Lys Arg Ile
1125 1130 1135
Glu Ser Cys Asp Leu Ser Gly Ala Asn Arg Leu Thr Leu Glu Asp Ala
1140 1145 1150
Asn Ile Val Gln Pro Leu Gly Leu Thr Ile Leu Gly Lys His Leu Tyr
1155 1160 1165
Trp Ile Asp Arg Gln Gln Gln Met Ile Glu Arg Val Glu Lys Thr Thr
1170 1175 1180
Gly Asp Lys Arg Thr Arg Ile Gln Gly Arg Val Ala His Leu Thr Gly
1185 1190 1195 1200
Ile His Ala Val Glu Glu Val Ser Leu Glu Glu Phe Ser Ala His Pro
1205 1210 1215
Cys Ala Arg Asp Asn Gly Gly Cys Ser His Ile Cys Ile Ala Lys Gly
1220 1225 1230
Asp Gly Thr Pro Arg Cys Ser Cys Pro Val His Leu Val Leu Leu Gln
1235 1240 1245
Asn Leu Leu Thr Cys Gly Glu Pro Pro Thr Cys Ser Pro Asp Gln Phe
1250 1255 1260
Ala Cys Ala Thr Gly Glu Ile Asp Cys Ile Pro Gly Ala Trp Arg Cys
1265 1270 1275 1280
Asp Gly Phe Pro Glu Cys Asp Asp Gln Ser Asp Glu Glu Gly Cys Pro
1285 1290 1295
Val Cys Ser Ala Ala Gln Phe Pro Cys Ala Arg Gly Gln Cys Val Asp
1300 1305 1310
Leu Arg Leu Arg Cys Asp Gly Glu Ala Asp Cys Gln Asp Arg Ser Asp
1315 1320 1325
Glu Val Asp Cys Asp Ala Ile Cys Leu Pro Asn Gln Phe Arg Cys Ala
1330 1335 1340
Ser Gly Gln Cys Val Leu Ile Lys Gln Gln Cys Asp Ser Phe Pro Asp
1345 1350 1355 1360
Cys Ile Asp Gly Ser Asp Glu Leu Met Cys Glu Ile Thr Lys Pro Pro
1365 1370 1375

```
Ser Asp Asp Ser Pro Ala His Ser Ser Ala Ile Gly Pro Val Ile Gly
            1380                1385                1390
Ile Ile Leu Ser Leu Phe Val Met Gly Gly Val Tyr Phe Val Cys Gln
        1395                1400                1405
Arg Val Val Cys Gln Arg Tyr Ala Gly Ala Asn Gly Pro Phe Pro His
    1410                1415                1420
Glu Tyr Val Ser Gly Thr Pro His Val Pro Leu Asn Phe Ile Ala Pro
1425                1430                1435                1440
Gly Gly Ser Gln His Gly Pro Phe Thr Gly Ile Ala Cys Gly Lys Ser
                1445                1450                1455
Met Met Ser Ser Val Ser Leu Met Gly Gly Arg Gly Gly Val Pro Leu
            1460                1465                1470
Tyr Asp Arg Asn His Val Thr Gly Ala Ser Ser Ser Ser Ser Ser Ser
        1475                1480                1485
Thr Lys Ala Thr Leu Tyr Pro Pro Ile Leu Asn Pro Pro Pro Ser Pro
    1490                1495                1500
Ala Thr Asp Pro Ser Leu Tyr Asn Met Asp Met Phe Tyr Ser Ser Asn
1505                1510                1515                1520
Ile Pro Ala Thr Ala Arg Pro Tyr Arg Pro Tyr Ile Ile Arg Gly Met
                1525                1530                1535
Ala Pro Pro Thr Thr Pro Cys Ser Thr Asp Val Cys Asp Ser Asp Tyr
            1540                1545                1550
Ser Ala Ser Arg Trp Lys Ala Ser Lys Tyr Tyr Leu Asp Leu Asn Ser
        1555                1560                1565
Asp Ser Asp Pro Tyr Pro Pro Pro Pro Thr Pro His Ser Gln Tyr Leu
    1570                1575                1580
Ser Ala Glu Asp Ser Cys Pro Pro Ser Pro Ala Thr Glu Arg Ser Tyr
1585                1590                1595                1600
Phe His Leu Phe Pro Pro Pro Pro Ser Pro Cys Thr Asp Ser Ser
                1605                1610                1615
```

<210> 5
<211> 3096
<212> DNA
<213> Homo sapiens

<400> 5

```
catcttctca cacgatctct cgcttcgcac tccttccttt gattggtttt caccatttac    60
tcagacgacg gtccttcttc gatctttgca cattcttcta tcatctacta ccttcatacc    120
cagctccgtc ccctaatatt catgcgcgga tggcccattc cgtggtgaaa attcccttct    180
actctgctaa tctgctgttc tctctccctc ccgtcgggtt ctgctcctgc cacgttctcc    240
cctctcccca ccaaaggctg ggttttcttt gtcagggctc ctttcccctt tggaagaagg    300
ggggctgtat ggccttggtg cgaggccctc cagtgacagg atcccccatc acccagagtt    360
ccacaggccc tggtagggag gaggggggagc agaagaggag gtgccatctt tgcctgctgg    420
ggaagggcag gggccaccca cacagagctc tcccatttgc tgtggaccct ggggccactg    480
cccagttcct tccaaaggaa agccagctcc ccaggtggtg ggagagtgat atggcttcct    540
cttaaactta gggaattgag tgtgtggttg cttctaagtg ccttagaagc cgggagcggc    600
tcctggaaag agcctgcctg ccacagcggg ccttaccctg gctgtgccca cagatgtccc    660
tggggcctgc cgctcctgcc cggctctcct ggcctccccc ggtgtgggtt gggaaaagca    720
```

cagcaaatta aaaaacacct ccatctctgg cctttgaaga atgcatctga acagccgaga 780
gtgtaaaccg tggtgaaatg tggtctttcc agtttgggga gaagcagggc agagctgggg 840
cttttgtacc cagggtttcc aagagctcct gcctcctcg gctgggctgg ccagggcccc 900
ccgctgggac ctccagctgt aatagggaag gttttactgg gttgctggcc actgtggact 960
gcccctaagg gcaggtatgc ctgcctttac ccgggttccc ctcctgcctg gaagatacag 1020
cccatgggag gcctgttgtc tgtgggatcc tccagcatca gagacactgg ggccagcgtc 1080
tgcctggtga ggtgcaggcc tggcaggccc ggtcccccac ctgcttgagc acccacggtg 1140
gtgggggctc gctgcctccc gagacaatct atgtcattgt tgtccaagga agctaattta 1200
gagtagaaag ttccgtgtcc agtcccactc tgtgcgtgtg ttagcagggg actctcgggc 1260
cggagctggg tccaccctgg taggggact tcatggggcc tgggcgacag cactgtgtat 1320
ttgtgtgtgt gtgtgtttgt gtgtgtgtgt gtctgaggag gtggaccagt ttctcaaaag 1380
gcctgtgacc ccaagaacca aggaatttca gcctgggtgg atcacacctt cactggtgag 1440
tgggacaagc tgggggccct cgccacagga gcagccaggg catggggcac agttggcctc 1500
attcacaaaa tgggagtata agtgatccct gctctggcgg ccaggacgat gagtgggaac 1560
acaccgtgtg ggggctgcct ggcctgggtg tgccgcgggt gtccttgttg gtgatggttc 1620
cacctgcttg tgccaccagt gccctctggg tctcacacac aactctcttc ccagcgaagg 1680
cccctcctgc cctcaggcct cagtgctgct tccgtctcgg aaggccccag gagctcctgc 1740
atcctgggcg tgattcctgt gtgcctgcag acccccctcgc ggctgccatc tcatcctttg 1800
gtgcacctgt tggccagacc tcctggtagc gggtgctgca ctcccctgaa tgtgccgggg 1860
cctgggggca gggacctggg ctcctccctc actgagtgga gggaactcag tgtcttggag 1920
ttggggtgcc tgcaggctgg gtggtgcagg tgaaatgcag acctctcagc tggtgttcca 1980
gagcagctgc cttcccccgc ccgagggact tcacccgcag cccagtcagg ggtggcgcct 2040
gggtgcatcg cccgcaggct gggtaggggt ggagcctggg tggccctgcc tgtgagctgc 2100
atagttgtcg cctttgaccc tgagttttct tcgttatctg tttggacctg tttggggcag 2160
gcaggggatg agatctgaag ataaatgcct tagctgtgac catctccttt tgtgagaggt 2220
caatgtccag ttccgctgca gttataacat cccatttttt gatttctttt tatttttccc 2280
tttttctttt tgagatggag tctcgctctg tcacccaggc tggagtgcaa tggggtgacc 2340
tcagctcact gcaacctcca cttctcgggt tcaagtgatt ctcctgcctc agcctcctga 2400
ctagcagggg ttacaggcgt gagccaccac gcccagctaa tttttgtatt tttagtagag 2460
gcaaggtttc gtcatgttgg ccaggctggt ctcaaactcc tggccttaag tgatctgccc 2520
gcctcggcct cccaaagtgc tgagatgaca ggtgtgagcc accgtgcccg gcccagaact 2580
ctttaattcc cacctgaaac ttgccgcctt aagcaggtcc ccagtctccc tcccctagtc 2640
cctggtccca ccattctgct ttctgtctca atgaatttgc ctaccgtaag tacctcatat 2700
aaattgaatc ataaagtatt tgtctttttа tatctggctt atttcactta gcataacatt 2760
cttaagtttc atccatgttg tagcatgtgt cagaatctct ctctttttt ttttttttt 2820
ttttttttt ttttgcagac agagtctcgc tctgtcatct agactggagt tcagtggcac 2880
gatctcggtt cactgcaaca tctgcctcct gggtccaagc aattctcctg cctcagcctc 2940
cttagcagct ggaactacag gcgcgtgcca ccatgccttg ctaatttttg tatttttatg 3000
tggaggcagg gtttcaccat cttggccagg ctggtctcga attcctggtc ttcaccacgg 3060
gggcccgaag gacccgggca aagcgtggag gggagg 3096

<210> 6
<211> 26928
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (12044),(12489),(26433),(26434),(26435),(26436),(26439),(26441)
<223> Identity of nucleotide sequences at the above locations are unknown.

<400> 6
gaagaccaag ggcacacagc gaggcagttt cagggcgggc agcctggggc cccacggggc 60
ggccccggac acttgttctc acctgtggag ggcagagaag ggaacaggga gagaagtggc 120
cggctgggag tggaggtggg tttgaggttt tactgtaaac taaatgtgta ccctctacct 180
tagttatgaa ttatgagaca cgaagactgc gaaacagaca cactcctcta aaagtgcctc 240
taggctgaca gggagaaagt cccgccaggc tcccagacgc caccttgag tccttcaaca 300
agcccgccag ggcctcttgc ccaccggtgt cagctcagcc actgaaccct ccaggaagaa 360
gacgtgctgg taggagaaga atctcaccca ggcacagcct ggaaggggca cagaaggggc 420
tccggaacca gcaagcccaa gttggaactc ccagtctgct actttctaga acgactgtgc 480
ccttggcggg tctaagtaga acctctccgc gcactctttc ctcctttgta aagtggggac 540
agcaatggcc accttgcagg ttcagagagg gcttgcagta cctcacagaa ctgagtgccc 600
gtgaacgtgt gtgttcctcc agatttgtga cagctttgcc aggctggagt caggctgaac 660
gcctctgccc tcatggggtt tatattctag gaagaccaac aaaaacaaga agacggaaaa 720
ttaaaacaac aaaagcccca ttgacaggcc gtgaagaatg ccatgaaaaa tgaatggcgt 780
tgtgctgcag tctttgggga aacgggctta cggaaagaag gacacttgag ctgctaccaa 840
tgagcagccg tccggtggga gggcagttca ggaagagcag acatccactg aggaggcgct 900
ggggcagagg gcagcctggt cgctggattc gggggaggaa ccacatcagg ccatgagctg 960
gagctggtgg tagaatgtac aggagaggcc agccagggcc agctcatgtc agacctcaag 1020
cggggaagat gaatcgagaa tgcaccccac gagcaatggg aagccagtct acgatttaag 1080
cagcaaaaat attttccctt cttccaccct gcatccagct ctaccagcac agcctggggt 1140
tctattttca agatagaata gacccagact cccagctctt cttacacttc tactactgcc 1200
acctgtcacc cactcatgcg tccccacttg cagcctcgac ccccttccac ctgatctcat 1260
ggcagccagg gaagctccag ggctcgtgag ggctgccatc tcaggaaaga agcaaaagcc 1320
ttcggcacct gcagggcctg ctccaaccac acttcttcct tgacctctca gcttccttag 1380
ccactccctt cccacatctc accctgctcc agccacagtg gtgtctctgt gggttctcaa 1440
acacaccagg tgcactcctg cctcagggcc tttgtgcttg ctgttctctg ctgggactct 1500
ttttttttct ttttttttg agacagggtc tcactctgtg gcccaggctg gagtgtagtg 1560
gtgtgatcgt agctcattgc aacctcaaac tcctgggctc aagcaatcct cccacctcag 1620
cctctcaagt agttagcttt tgttgttttg ttttgagatg ggatctcact ctgttgccca 1680
ggctggagtg cagtggggca atcttggctc accacaacct ctgcctccca ggctcaagca 1740
attctcctgc ctcagcctcc caagtagctg ggattacagg catgtgccac cacgcccagc 1800
ttatttttgt atttttagta gagacagggt ttcaccatgt tggtctggct ggtcttgaac 1860
tcctggcctc agatgatcca cctgcctcgg cctcccaaag tgctgggatg acaggcatga 1920
gcctgtctct agtagttagg actacagaga ggggccatca tgcctggtga tcctcccacc 1980
ttttctgctc caactctttc accccactta gcctcgtggc tcactctctt acctcttcag 2040
ctcctcagtc aggcctgagg acccctgttg aaaattgcaa accacacccc ccaccaccac 2100
cacccactat tgccagcact ttctactcca tttctctgct ttacttttct cctttgtact 2160
catcaccacc tgactcatta catgtttacg tatctttctt ctctccacta gcatggaagc 2220
tccaggagag cagagagtgt agttttattc cctgatgtgt ttcctgtgcc cgtaccaggg 2280
cctagcacac agtaggtgct cagtaaatgt gtgttggatg aacaaataca gtgaaaggat 2340
ctgatctaca tttataaaga aggcactctg gctgctgagt ggggatgaga ctgtcaggag 2400
gaaagaggcc cctgtggggg cctggccagc aggtgggtac aatggtagca gccaggagag 2460
agggcctctt ggactcaagt ggatggggcc tgctcagggc tccggccaca ggaacaaagg 2520
gaaggggccc caggatggcc tgtcatagag gacacattac aactggccca aagttcaagt 2580
caggtttcta aatttgggaa gggatacaga aaaactaaag actctactgg acagtcagtt 2640
attgaaatga ttacatagaa aatgtaccaa gaattaaaaa aaaaaaaaaa aagcattatg 2700
aaggggccac cagagactcc cagagaggaa agggactatg ggctggatgc ggtgactcac 2760
acctataatc ccagcacttt gggaggccga ggagggtgga tcacgaggtc aggagttcaa 2820
aaccagccta ggcaacatgg taaaaccccc gtttctacta aaaatacaaa aaattagctg 2880
ggcatggcag catgtgcctg taatcccagc tactcgggag gctgaggcag gagagttgct 2940

agaacccagg aggcagaggt tgcagtgagc cgagattgag ccactatgct ccagcttggg 3000
cgacagagca agactccgtc tctaaaaaaa agaaaaaaaa ggccagatga ggtggctcat 3060
gcctgtaatc ccagcacttt gggaggccga ggtgggtgga tcacgaggtc aggagatcga 3120
gaccatcctg gctaacatgg tgaaactcca tctctactta aaatacaaaa aattagccgg 3180
gcgtggtggc gggcacctgt agtcccagct acttgggagg ctgaggcagg agaatggcgt 3240
gaacctggga ggcggagctt gcagtgagcc gagattgcgc cactgcactc catccagcct 3300
gggcgacaga gttagactcc gtctcaaaaa aaaaaaaaaa aaaaaaatta gctgattagt 3360
tgggcttggt ggcgggcgcc tgtaatccca actactcggg aggctgaggc gggagaatca 3420
cttgaacccg ggaggcagag gttgcaatga gccgatatca cgccactaca ctccagcctg 3480
ggcgacagag caagactcca tctcaaaaaa gaaaaaaaaa aagaaagggg ctgtgctgtg 3540
gcctgggacc caaagcacac tactgcaagg tcccagggtg cctgactcca accggagcct 3600
tgagaacatt catttgcaaa gaatgaatta aaattcagca ctattttatt ctgcaggatt 3660
ccagcacccc aaggacagtc atttttagac ccttcagtaa cgtaataagt aaccggagga 3720
tgtgctgagc ttccacttcc ccagacggtt gcctgtcaca gctcatcagg ccaacaaact 3780
tttcttaggc ctcaaatttg gaaatgttca ctctcagttc gttccttaga tgcaagtcca 3840
tcccaatgaa gtaacagggg ctcagcacct gtccaatctc attgcttccg gggacagggg 3900
cccatgagga tgtcgtttca gcccggtgac acttgggcaa agtgcctttt ggtttccctc 3960
ccaggctgga acgtgctggc tctgtgaagt tacgctgggc acaagagccc cccccaaccc 4020
ggcaggactg actgctgtgg tcagaggcgc ccctggggct ttgggagcca cagaatcttc 4080
ctgagggcag cgccggagga ggccccagtg agagtgccca ctgccaggct cattcctcag 4140
gctgccgcag gcctctcccc aaaacaggca atgcttctca gcaacctgcc ccaggagcag 4200
gccagggaag gccgccatcg gcctacagtg ctgggctctg gagggcttgg ttggtaacag 4260
gccatggttt ctatgagcca gctggggtgt gaaggacaca ggctggattc acctctctgg 4320
gcctcagttt ctgcattcaa aaagtgggaa tcatgatatc tgctctattt cttatctctc 4380
agtgctgatg tgaacctcca ataagacttt taaaaatact ctttctacct tacttttatt 4440
tttcatttat tttaagataa tgtctagctg tctcacccag gctggagtgc agtggtgtga 4500
ttacggctca ctacagcctt aacctcccag gctcaagtga tcctcctacc acagcctccc 4560
aagtagctgg aactacaggc atgcaccacc gcacctggat aatttttttt tttgagacaa 4620
ggtttcactc tgttgcccag gctggagtgc agtggtgcac tcttggctca ctgcagcctc 4680
aacctccctg ggcttaggtg atcctcacac ttcagtctcc caagtagctg ggactacagg 4740
tatgtgccag tacacccagc taatattttt gaaggatggg gtttcactat attgcccagg 4800
ctggtcttga actccagggt ttaagcaatc taccttcctc agcctgccaa agtgctagga 4860
ttataggtat gagccacccc ccggcctata atcctaccac tttaaaaaag cctgtaattt 4920
tagcacttta aaaaatttttt ctaaattttt tatagagatg ggggacagct gtggtctcac 4980
tgtgttgccc aggctggtct tgaactccta ggatcaagcc atcctcctgg cctggcctcc 5040
caaagtgttg ggattataag cataagcctt accttacctt ttttttttga gttgcagttt 5100
tgttcttgtt gctcaggctg gagtgcaatg gcaagatctt ggctcactgc aacctccacc 5160
tcccgggttc aagcaattct cctgcctcag cctcccgagt agctgggatt acaggcatgc 5220
gccaccacac ccagctaatt ttgtattttt agtagagatg gggtttctct atatacctta 5280
attttaaagc actgcattca tgtaaattgt gattaacatg gattcaagag agggagtgag 5340
gatgaatgag ccaggcagtc acctcggctg tcaccctcca cttctctcct ccttctgaca 5400
gtcatcgtcc atccgtttct gcagctgttt gtttgactct cctgatcatt ttgcttgcca 5460
cataacttgc ctcctgggaa agaatgccct gggcaggccc acatgagtag tgaaaaataa 5520
tctgcagtga aaaataaaac taagtagtct ggtccacaga gcagtcttat tttttcactg 5580
cagatgaagg agttgacatt caggcttcat tctcatttat aagtgtttta aagacacata 5640
cagtggattg aacagtggcc ttcaaaaaga tgtatctaca tcctaatccc tgggacctgt 5700
gaatgttaac caagttagga aaagggtctt cccgggtgtc attaagttag agatcttgag 5760
atgaggagct catcgtggat tatccaggtg gaccctgcat ccaaggacaa atggtcctta 5820
gaaaagaaaa gcagaggctg ggcacagtgg ctcaagcctg taatcccagc actttgagag 5880
gccgaggtgg gtggatcacc taaggtcatg agttcgagag cagcctggcc aacatgatga 5940

aatcccatct ctactaaaaa tacaaaaatt agcaaggcat ggtggcgggt gcctataatc 6000
ccagctactc aggaagctga ggcaggagaa tggcttgcac ctgggaggcg gaggttgcag 6060
tgagccaaga tcgcgccact gcactccagc ctgagggaga aaagtgaaac tctgtctcat 6120
aaaagaaaag aaaagcagac agagatctga gacagaagag gagagtgaag gaaaaaaggc 6180
catgtgaaga tgaggcagag gttggagcca tgcagccaca agccaaggaa tacctggagc 6240
cccagaagtt gcaagaggta ggaagaagcc tcccctagag cctccagacg gagcacagcc 6300
ctgccaacac ctccacctca gacttctggc ctccagcact gtgagataat caactgctgt 6360
tgttttaagc caccagattt gtggtaattt gttatggcag ccacaggaaa ctaatacagt 6420
acctaatctt cacaaaccca tcttacagaa aaggaaactg aagtcagaga ggtagtggct 6480
tgtgcagtgt gttaggccat tcttgtatta ctataaagaa atacctgagg ccgggcatgg 6540
tggctcacgc ctgtaatccc agcactttgg gaggccaagg tgagtggatc acttgaggtc 6600
aggagttcaa gaccagcctg gacaacatgg tgaaacccca tttctactga aaatatgaaa 6660
attagccagg catggtggcg tgcatctgta gtcccagcta ctcaggaggc tgaggcagga 6720
gaatcacttg cgcccgggag gaggaggttg tagtgagcca agattgtgcc actgcactcc 6780
agcctgggag acaagagaga aaccctgtct caaaataaat aaaaaacaaa taaacacctg 6840
agactgggta gtttataaag aaaggggtta actggctccc ggttctgcag gctgtacaag 6900
catggtgccg gcatctgctt ggttgctggg aaggcttcag ggagttttac tcatcgtgga 6960
aggcagagcc agagcaggtg catcacacag caaaagcagg agcgagagag agagagagca 7020
gggaggtgtg cacactttta aatgagcaga tctcacgaga actcaccatt gcaaggacag 7080
caccaagcca cgaggggtct gcccccatga cccaaacctc ccactaggcc ccacccccaa 7140
cattgggaat tacagttcaa catgaggttt gggggggacaa atatccaaac tatatcattc 7200
caccctggc cccccagatc tcatgttctt ctcacattgc aaaatatagt catgccttcc 7260
cagtagcccc ccaaagtctt aactcatccc agcattaact caaaaatccc attcccaagt 7320
ccaacgtctc atctgaagat gagttccttt cacctacaag actgtaaaaa tgaaaacagt 7380
tatttactgc tgagatacaa tgggggcata ggcattaggt aaacattcct gttccaaaag 7440
ggagaaatcg gtcaaaagaa aggggctata ggccccaagc aagtccaaaa cccagcagag 7500
caatcattca atcttaaagc tccaaaataa cctccttaaa ctccatgtcc catagccagg 7560
gcacactggt gcaaggggca ggctcccaag gccttgggca gctctattcc tgcggctttg 7620
cagaattcag tccccatggc tgctcttaca gattggagat gagggcctgc ggcttttcca 7680
ggtgcagggt gcaagctgct ggtgatctac cattctgggg tgtggatggt ggcggccccg 7740
tcccgcagct ccactaggca ttgtcccagt ggggactcta tgtggggcct ccaaccccac 7800
atttcccctc caatgggaag gctctgcccc tgcagcagcc ttcttcctgg gctcccaggc 7860
tttctcatac atcctctgac atctaggtgg atggtgtcaa gcttccttca ctcttgcact 7920
ctgcacacct acaggcttaa caccacatgg aagctgccaa ggtgtatggc tggaaccctc 7980
tgaagcagca gcctgagctg tgactatggc cctttgagcc aaggctggag ctggaacagt 8040
ctagatgcag gcagggagca gtgtcctgag gctgtgcaga gcagcagggc cctgtgcctg 8100
gacaatgaaa ccattctttc ctcctcatcc tctgggcctg tgatgggagg gttgtggaag 8160
atctctgaaa tgcctttgag gccttttttgc ctctgaggcc tatttcctat tgtctcagtt 8220
attggcagtc ggctcctttt tagttatgca aatcctctag caagaggtta ctccactgcc 8280
ggcttgaact cctctcctga aaaagctttt tctttctttg tcacatggcc aggctgcaaa 8340
ttttccaaac ttttatgctc tgttttacct ttaaatataa cttctaactt taattcattt 8400
atttgctcct gcatttgagc atagggaatt caaagaagct gggccacatc ttgaatgctt 8460
tgctgcttca aaatttatgg ccacgcttgg tggctcacac ctgtaatccc agcactttgg 8520
gaggcctagg tgggcagatc acgagatcag gagatcgaga ccatcctggt caacatggtg 8580
aaacccatct ctactaaaaa tacaaaaaaa ttagcttggt gtggtggcgc agacctgtag 8640
tcccagctac tggagaggct gaggcaggag aattacttga acctgggagg cagaggttgc 8700
agtgagccca gatcatgcca ctgcactcca gcctggtgac agaataagat ttgatctcga 8760
aaggaaggaa ggaaggagga agggaagaaa tgtcttcccc ccagatgtcc tgggtcatcc 8820
ctcttatgtt caaacttcaa cagatcccta gggcatgaaa ataatacagc caaattattt 8880
gctaaggcat aacgaaagtg acctttgctc cagttcccaa taagttcctc atttccatct 8940

gagactcatc accctggcct tggcttgtcc atatcactgt cagcattttg gtcacaatca 9000
tttaaccagc taatcgggag gctgaggcaa gaggatcact tgaacccagg aggttgaggc 9060
tgcagtgagc tgtgatcaca tcactgcagt ccagcttggg caacagagca agatcctgtc 9120
tcaataaata aataaataaa tacataaata acttaagttt atttaaagct gcatctttgc 9180
caccatggag aaaggccagg ccagctcctt ctctctttct gcacgtgttc ctcccacctc 9240
agctgcctct gctcctcaag gaggaacaga gggagtagga aaggccatcc caggaggccc 9300
agcaccccat gacctggctc tggggccttg tgggtttatg gattcccagt gctgagtcat 9360
ccctcacagg ctcttgtggg caccttggac attggtcaga agcatgtggt ccccgggaac 9420
acacctttc ctgatcatct gggaagggca gcttgtgcca gcgaggccac ctgttcagcg 9480
ccacggcccg ccagacagct gcagccacag ccttgccttt gatcagagca aacaccagac 9540
atgtgtgtca tgccccccaac ccatctccag gggacacatg tcctttcttg ccaggcctga 9600
gatgaacaag agagggacaa gtccccaagc ctctctctcc ttcctgcctc acccactccg 9660
ctgttagatt ctcaaggtgg atggtgggct aactagggca accgaccatc ctggtttacc 9720
tagaactgag ggggcatttt caggaataaa actgcaaaag tctggagcaa acaggagcaa 9780
gttggtcact ctggggctgg tggagtcagg tttccttctg caggcccccc ccccgcaagc 9840
atgggtggaa cccaggacag gaacacagag caggccccag gaccgggctt gtcacttaca 9900
agtctttttt ttttttttttt ttttgagatg gagtcttgct ctgtcatcag ggctggagta 9960
cagtggtgcc atcttagctc actgcaacct ctgccttctg ggttcaagtg atccccctgc 10020
ctcagcctcc tgagtagctg ggactacagg tggcaccacc acgcccagct aatttttttgt 10080
atttctagta gagatgagat ggccaggctg gtcttgaact cctgacctca agtgatctgc 10140
ccgccttggc ctcccaaagt gctgggatta caggtgtgag ccactgtgcc tggccccact 10200
cacaagtctt aaaccatgcc tcagcacatc aatgccattt acaaaaaggt agagggattt 10260
tccaggcaaa aatagatgaa agacatagga tgattgatca tgtcctgctt aaacataggt 10320
ctgatgctat taagaattga gggctgggag cggtggctca cgcctgtaat cccagcactt 10380
tgggaggccg aggcgggcgg atcacgaggt caggagatcg agaccatcct ggctaacacg 10440
gtgaaacccc atctctacta aaaatacaaa aaatggccgc gcgcggtgac tcacgcctgt 10500
aatcccagca ctttgggagg ccaaggcggg cggatcacga ggtcaggaga tcgagaccat 10560
cctggctaac acagtgaagc cccgtctcta ctaaaaaata caaaaaaaat tagccaggca 10620
tggtggcggg cgcctgtagt cccagcaact tgggaggctg aggcaggaga agaatggtgt 10680
gaacctggga ggtggagctt ccagtgagcc gagatcacac cactgcactc cagcctgggc 10740
gacagagtga aactccatct caaaaaaaaa ataaataaat aaataagaat tgttagtatt 10800
ttgcaggtgt gacaaatgat tctgtttctg tggcagaatg ttctcaggag atctcttttg 10860
aactctcatg gaaagcatca tgctgttggc aacatcacat ttatttttat ttatttatta 10920
ttttttagag acagggtctt gctctgttgc ccaggctgga gtgcagtggc acaatcacag 10980
ctcactgcag cctcaacctc ctgggctcaa gcaatcctcc tgcctcagcc tcccaaagta 11040
gctgggacca caggcgtgag ccactgcact cagcccaatg taccttcaat atttacattt 11100
ctggcaaagg tagcaaaacc ttaacaaatt ttgaatctag ataataaaat tatgaggctg 11160
ggtgcagtgg ccctgacagg gatggctcac atctgtaatc tcaacatttt gggaggccaa 11220
ggtaggcgga tcacctgagg ccaggagttt gagaccagcc tggccaacat ggtgtaaccc 11280
tgtctctaac aaaaatacaa aaaaattagc cagacgtggt ggtgcacgtc tgtcatccca 11340
gctactaggg aggctgaggc aggagaattg cttgaacccg agaggcagag gttgtgatga 11400
gccgagatcg cgtcattgca ctccagcctg ggcaaaagca agagcgaaac tctctctcca 11460
aaaaataaaa aaaaaataaa ttaatgaatt aattaaaata aaataaaata atggatagtc 11520
actgtaaaga aaaaataaat gtatatatca gccaacaagt gatggaatag agcaccccat 11580
ctccctggct ggacagatac atcccacaac acctggaagg cggctccatg tagaactttc 11640
tggactgctt gaggtgctgt gctggagcac ggtgacagag gagctggacc atggacctcc 11700
cccggccccc accaagggcg aggtcccccct gtggtgggtc tgagggaggc atccgtatgg 11760
cctctgcggc ttgggcaggg aatttggggt ccaagtactt ggtgcaaagc ctggaaagag 11820
ggtttgggtg ctgagggcat atccccctggg ccacatgggg gcagaagtgg ggccccctga 11880
agcttggagt cctgggcagg ggcatctatt ttgctgtctg aggccttcag tacttgaagc 11940

aaaatggagg cagaatgtcc caccttaatg cccctgattc ctccaaacca attccagaga 12000
cagcaagggc cagaacaggg atggccctgc ccagggtcat gcancgagga agtggccagg 12060
ctgggatctg aacccaggct aatcccctcc cttgtcctcc tccaggccct caccctgca 12120
tagagccctc cagctcactc atcctcggcc agctccatct cctcagcttg taaacccccc 12180
cgggattttc ctttcttaaa aaacaaaggc ttggccaggc acggtggctc acgcctgtac 12240
tttgggggtg gctcccagca ctttgggagg ccaaggtggg cggatcatga ggtcaagaga 12300
ttgagaccat tctggccagc atggtgaaac cctgtattta ctaaaaaaaa aaaaattaac 12360
tgggcatggt ggctagctac ttaggaggct gaggcaggag aatcgcttga acctgggaga 12420
aagaggttgc agtgagccaa gatcgcgcca ctccacttta acctggcaac agaacaagat 12480
tccgtttcna aaaacaaaca aacaaacaaa taaacaaaaa aaggcggagc gcgatggctc 12540
gcgcctgcaa tcccagcact ttgggaggct gaggcgggcg gatcacttga ggttaggagt 12600
ttgagaccag cttggccaac atggtgaaac cccatttcca ctaaaagtac aaaaatcagc 12660
caggtgtggt ggtgggtgcc tgtaatccca gctactcagg aggctgaggc aggagaatcg 12720
cttgaaccca tgacctggag gctacagtga gctgagattg cgccactgta ctccagcttg 12780
ggcaacaaga tttgtttctc taaaaaaaaa aaaaaaaaga ctggcccttc cccttcagct 12840
cttcctcagg gtccctgagc actctacacc cccgtctaca ctgagcactc caccctgctg 12900
tctacactga gcactccacc ctgccatcta cactgaggac tccaccccac tgtctacact 12960
ggctgcctcc cgccctcacc tcctgctaag gccattcccc gctgcatctg tcttctagat 13020
tctgcagcct tcagcacgct gggccccctcc tttgtccccc tgagccacct ccagcctccc 13080
cctgagctgc tactcctctc ccagcagcct ccacccaagc ccctccagtc cccaagctgt 13140
cccttgcatc cagcactgcc cttccacgtg ccccttcccc ccagcttcac agcagggtgg 13200
ggcctccagg ccctgcccac tgtgcccatc cacaagttgt ggtgggagct ccgaggggag 13260
gcaggggtgt gcatggactt gggacgtcca agtctgggac caggggcagc tggttggtgg 13320
agtgtggagg gggataggga ctttcaggta gagaggctgt aggggcaaga tcgggacggc 13380
ggatgtccct aaggagggct ctgacctggg aaatattgtg cagcttcctc tttgccattc 13440
ctggagctca gacactggcc ggctctcacc ccgcccttcc tgcaggacac agctccatcc 13500
cagtgagttc ctagtgtaga catctccagc agcacggatg ggaaaggaag tcatcaaagg 13560
tgcccaggac cggaggcttt ttctggaggt ggcagaggag ggtgtgggtc tcagggctct 13620
ggctgagggc aagcgtggga ggtcttaggt ctgcaccagc cccgtgaagg cccctcctgc 13680
tccctggtgg agtcctagag ggaacagcag cccctaggct ctagcaggag tgggtagggg 13740
cttttctggc ttcctactgt gccagcagga tagctgggcc tggcactgag cccaaagatc 13800
acatgccggg gcattggcgc agtgaggaac agacccttgc caaagctggc aaagaagacc 13860
ccatggggtg cagctggtga agctgagagc tcaatgtttg ggggagcctg gcaaaagggg 13920
tcctcccctc cctctgcagg ccaggatcgc aggttttccc tacatgttgg taattctcaa 13980
acaatcccat ggccactgga gcaaagatca cagtgggcgg cggcctcggg agcagtggac 14040
agggcacgca gtgcctttga tgccagagcc ctcgccccaa agtcaacaaa ctctgcagcg 14100
gactttgcac ccggactttg ttttcaccat acaaggaaag ggacagatca caggccctct 14160
cgctgccctc gctgagccgg aagctgcagc gtgagctctc tcaagcccca tttctaggtt 14220
ccccaggcgc acccctgagc ccctactcgc ctattaagtt ctcctaatag cccttcaagg 14280
tcttaatgta tgtccattag acagagggga aaactgaggc gagggcaagt gacttgaccg 14340
aggttcctcg gcgagcaggg cgtggagctg agaacctcgt tattactgct ccccacacaa 14400
ccctctggcc gttcttggaa gaaggctgag ccccgggggg gccagagtga cccaaacacc 14460
atgggccgcc tgcggtaaca cgtgcggcca cgaaggggca gcagtttccc gcccggccgg 14520
gctctctccg gcgctcagta tccgtcccag gccaagaaga agaaactcgg ggaggagggc 14580
ggaggggggct gcgtgggagg gcgtggaaga tggacgtggc caggggagtg gcagctgcac 14640
acagtggatg ctgttaagat gaagggaaag aacgtgggct ccgagatcac tggacacggt 14700
tccacctttc ttcccgctca ctgcatggcc ctgggcgggt tgttgaaccc ttggaaacct 14760
gtttttcctt ttttcctttt tttttgagac agggtcttgc tctgtggccc agactggagt 14820
gccgtggcac gatcttggct cactgctgcc tcccaggttc aagtgatcct cccagctcag 14880
cctcctgcgt agctgggacc ccaggtatgt gtcaccacag ccggctaatt tttgtatttt 14940

tttgtagaga cgggatttcg ccgtattgcc caggctggtc tcaaactcct gagttcaccg 15000
gatcttcctg cctcagcctc ccaaagtgct gggattactg gcatgagcca ccgcacccag 15060
cagagacctc agttttctaa cctgtgccag caggaataat gatagctgcc tagcttggct 15120
gtgctgggaa ttaagtaaga tgaccgggta gcaaatatga agtattactg gacacagagg 15180
gccccaggct gggttagcag cggtggtcag ggctgctgct tcctggcctg agctcgaagg 15240
agggccctca ttaccacctg ggtgagtcct cgtccaagcc tggcactgct gcgtgggaat 15300
aacttctgcc acccaagttg gcagattgtg tgcaaagtta agtcctgact ctgtggggtg 15360
gacttcgagg cctcttcatc ggacctgctt ccggtgactg cattcgcacc tcctcctgtt 15420
cctggtttaa cacagcccag ctttcctcct gctgagccct ccctgggcct gctgtcaccc 15480
tcgtgccgct gtgcctcgca gtgccactcc ctgtaccctg aatactttgc cctgcctctc 15540
cacccagctg agagtcaggg cccctgtgag gctctgccca gcccgtcctc cgggtttctg 15600
cctctgctga gcacttccct gcatgattgc ttctgagagt ccccccagcc tgtgagcttc 15660
tcaggactgg gacagcttct caggaccgag gcttcctggt ctgcttgcaa ttttacaggc 15720
gggcacattt tcccttggcc aacatcagag actggacatc tgcagatctg tgctagccac 15780
tgagcaccca ggcaccccag caggtagctc tgtaaccaac ccattctgta aagctgaggc 15840
tcagagaggt gaagcgcctg gcctggggcc acagcctgcg tcagctgcag agccaggagc 15900
tgagatatgc acctgcggct ctgctcacag ggtcctgcac agactgctgc tggagccacc 15960
tatgtagagt caagagagtt catgttaact ccctctcaca tccctcagcc agggtggggg 16020
ctgacgatag acactcaggg atggcctacc ctccccaaca acccccgtca ggtttgccgg 16080
atctccttgg aagaaaagtt ctgggcagaa ttccaccgtt ggcctggcct acactctcct 16140
tagtggctta ggaccctcag cggtggataa gttgtgggca gaagagatgc aatcaggatt 16200
ctcacccact caccccttgc cagccccaat aagctcaata agctgggctc ggtctgagga 16260
agtgtccagg aaatgtgcaa atggcctggg acagccctgt gttcctttca gtaaggttgc 16320
tgaaggtgag gctgaaagtt ggagaaacag aagccagtgc ttatggtttt aattaagata 16380
atggaatgta tgtatgtatg tatgtatgta tgtatgtatt tatgtattta tctttagaga 16440
tagagtctca ctctgttgcc caggctggaa tgcggtgaca caatcatagc tccttgcagc 16500
ctcgacttcc tatgcccaaa tgatcctcct acctcagcct cctgagtagc tgggactaca 16560
gacacacgcc aactatgcct agctaatttt tatttctatt ttttgtggag actgggttct 16620
cactttgttg cccaggctgg tcttgaaccc ctagcttcaa gcaatcctcc tgcctcagcc 16680
tcccaaagtg gagggattac aggtgtgagc caccacacct ggcctggaat ttatttgtat 16740
tctgcttata aaattaatac attcttattg cagaaaagtt tgaaaataaa agaaaggaca 16800
aagaacaaaa agcgtatata atttcacagc tcagatctca ctgctattaa catttttatt 16860
tactttcagg ctttttttctt tctaggtaca tatgcagaga ttattttatt ttatttattt 16920
tattttatat tttatttttat attttttatt tcattatttt attttatttt attttattat 16980
ttttagagac agggcctcac tctgtcaccc aggctggagt acaatggagt gatcatagct 17040
cactgcagcc tcaaacacct gggctcaagc aatccccccca ctcagccttc tgagtagttg 17100
ggactaaagt gtgagtctgg ctaatttttt ttacttttttg tattgacaga ggtctcacta 17160
tgttgcccag gctgatctca aactcctggg ttcaagcgat cctcccacct tggactccca 17220
aagtgctggg attacaggca tgagccacca tgcctggcct aaaatgccac tttttgtcat 17280
ttactaaaat cccatggaca ctttgacatg tctgtattct atgctattga tctgactgtt 17340
ggcatctaca tcattatggc catctatcat ctatcataat ccattttaac attaaaattg 17400
tgctgctgct tagatttttc tggcctgtct cctatttgta ttcttccaga taaattttag 17460
aatcatttta tcaaattccc cttgcagaaa aagccctatt ggattttggt tgaaaaatac 17520
tgaatttttа cattaactta ggaaagggct gggcacggtg gctcacgcct gtaatcccta 17580
cactttcga ggccaaggca ggtggatcac ttgaggttgg gagtttgaga ccagcctggc 17640
caacatggtg aaactcggtc tttactaaaa atacaaaaat tgccaggcgc attggctcac 17700
ctgtaatccc agcactttgg gaggccgagg tgggtggatc acgaggtcag gagatagaga 17760
ccatcctggc taacacggtg caaccccgtc tctcctaaaa atacaaaaaa ttagccaggc 17820
gtggtggtgg gcgcctgtgg tctcagctac ttaggaggct gaggcaggag aatggtgtga 17880
acccaggagg cggagcttgc agtgagccaa gatcgcgcca ctgcactcca gcctgggcga 17940

cagagtgaga ctccatctca aaaaaaaata ataataataa tacaaaaatt agccgggggt 18000
cgtggcgtgc acctataatc ccagttactt gggaggctga ggcaggagaa tcgcttgaat 18060
ccaggaggtg gaggttgcaa tgagcagaga tcgtgccact gtactccagc ctgggtgaca 18120
gagtgacact ctgtgaaaaa aaaaaaaaaa ttctgaagga ttgagactct tagactctta 18180
ggtcttccta tccaagagca caatatagct tttcatgtat tcaagccttt ttcaatgcat 18240
caacagaatt ttacagtttt tttcatgata tcctgctatt tcttataaaa tgtattccta 18300
gatattctgc atgttttccg gttgtttgtt aataaatatt tttcatttgt cattatttcc 18360
taattggctg ttatttgtat atatgacatc tgttgaattt tttgattact ttgaaaatgg 18420
ccattctttt gtgttttttt ttaactttct attttgagat aattttgact tacagaagat 18480
ttgcaaaaat agtacagaga gttcctgttt ccccccttatg ttaacccagt ttctccttat 18540
gttaacatct tacataacta cagaacaatt gtcaaatcta agaatcaacc tgggcacaat 18600
gctattaact aaactgcaga agctgttcag atctcaccag ttcttctact gctccccttt 18660
tctcttccag tgttcaatcc ggaatcctac attatattta gttgtcattt ctctttggtg 18720
tcttccaatc tgtgacagtt cctcagtctt tctttgtctt tcatgacttt catttttta 18780
tacttttgaa aaatactggc cggttgtttt gtagaacgcc ctcagtttgg gtttgcctga 18840
agtttttttgt gattagatcg aggtcatgca ttattggaga gggtgccacc gcctcgatgt 18900
gcaagctcaa tgcatcatat cagagggttt gtaatgtcag tttataccgc cggagaccct 18960
aacctggagc atttcgtgaa ggtgctgtct gccaggattc tccactagaa agttactatt 19020
tttccctttt taattactga atgtctgagg ggaaatactt tgagactatg caaatatcct 19080
gtttctgctt taacttcggc tcactaagtt tagcattcat ctatggatct cgcttatagc 19140
aagtattact gtggagttct aatggtaatt ttctgtttct ctcattcctt caacctttat 19200
taatatgctt cttcctcact tattcatttt gtttcagttg tttataccaa catggatttg 19260
tggatattgg ttttattctt tgggttgcaa ttgaatccta tcattatttt gttagtcagt 19320
tgttccatcc gaccttggtc attaggagcc cttgaaattt ggctcccatg cctttttttt 19380
ttttttgag accgagtctc actctgtcac ccaggtttga gtgcagtggc atgatcttgg 19440
cttcctgcaa cctccgcctc ccaggttcaa gcaattctcc tgcctcagcc tcctgagtag 19500
ctggtattat aggcgctcca ccaccttgcc cggctaattt tttgtatttt tagtagagat 19560
ggggttttat tatgttggcc aggctggtct caaactcctg acctcaggtg atctgcccgc 19620
ctcggcctcc caaagtgctg ggactacagg cgtgagccac cacacctggc ctcctatgcc 19680
attttaacat gcccgtcttt tctttttctt tcctactttc tgtgactgta agaagctcca 19740
ggatacattt ttgctgccct agacttagcc tcaatcagtt ctcagaaaag ctctggttct 19800
ttttatggga tacttagaaa actagctctg tatggcctgg cgcggtggct cacgcctgta 19860
atcccagtac tttgggaggc cgaggtgggc agatcacaga tcacgaagtc aggagatcaa 19920
gaccatcctg gctaacatgg tgaaactctg tctctactaa acatacaaaa aattagtcca 19980
ggcgcggtgg cgggcgcctg tagtcccagc tactcaggag gctgaggcag gagaacggca 20040
tgaacccggg aggcggagct tgcagtgagc cgagatcggc agccactgca ctccagcctg 20100
ggccacagag cgagactccg tctcaaaaaa aaaaaaagga aaaagaaaaa agaaaactag 20160
ctctgtatgc tagttttttt tttaagacag ggtctctctt gccccagctg gagtgtagca 20220
gcacgatcac agctcactgt agcctcaacc ttctgggctc aagcaatcct cctgcctcag 20280
tctcctaagt agctgggtct acaggcatgc accaccgtac gtggcaattt ttaaaaactg 20340
tttgtagaga tggagtctcc ctatgttgcc tggtctggaa ctcctggcct caagtgatcc 20400
tcctgcctcg gcctcccaaa gtgctgagat tacaggcatg agccactgta cctggcctgg 20460
ccaaggtctg tctttttttа aaagaagttg ttgtatagtt gtttttttttt ttatttttttt 20520
ttctgagacg gagtctcgct ctgtcgccca ggctggagtg cagtggtgcg atctcggctc 20580
actgcaagct ccgcctccca ggttcacgcc attctcctgc ctcagcctcc cgagtagctg 20640
ggcctacagg cgcccgctac cacgcccggc taatttttttg catttttagt agagacgggg 20700
tttcaccgtg ttagccagga tggtctcgat ctcctgacct cgtgatccgc ccgcctcggc 20760
ctcccaaagt gctgggatta caggcgtgag ccaccgcgcc cggcctgttg tatagttttt 20820
atctcgagtt ttctagcgat ttaatcatat tggttacaaa aaaggatgat tttactacct 20880
cctttccaat gtttctacat atttttttcat tttatctaac tgcattttaa aataaacttt 20940

taattttaga atggtttcat atttacagaa aatgtgcaaa gatagtacag agagttcctg 21000
tgtactccac acccggtttc cttattatta tcttaacgtg atacacaatt aataaaccag 21060
taacattatt attcactgaa gtccacactt tctttttttt tttttctgag acggagtcta 21120
cttctgtcac ccaggctgga gtgcagtggc gcaatctcgg ctcactgcaa cctccacctc 21180
ctgggttcag gcaattctgt ggctcagcat cccaagtagc tgggaataca ggtgcccgcc 21240
accacgcccg gctaattttt tgtattttta gtagagatgg ggtttcacca tgttagccag 21300
gatggtcttg aactcctgac ctcgtgatct gcctgcctca gcctcccaaa gtgctgggat 21360
tacaggcgtg agccaccgcg cccggcgtcc atactttctt tagatatcct tccttttac 21420
ctaacgtcct tcttctggtt caggatccca tccagaaagc aacattaccc ctcgccatca 21480
cgtcttcaca ggctcccctt gacgggaaga gttcctcaga ctttccttgt ttttgttgac 21540
cttgacagtt ttgaggagga ctggtatctt agtctgtttt gtgctgctat cacagactag 21600
ctgagaccga tacatgatac atgaaaaaaa atgtattctt acagttgtgg aggctgggaa 21660
gttcaagacg aagttgctgg ttggtttggt ctctggtttc aagatggcgc cttgctgctg 21720
catcctctgg agaagaagaa tgcggtgtcc tctcactgca gaagatggaa gcgctaaaag 21780
gaatgaactc cctttgccaa gccatttttat aatgggcatt aatccacaaa ggatgaaacc 21840
ctgagaaaca tcaagcttta aagcactggt tctcaacctt tttggtctca ggagcccttt 21900
atactcttaa aacgttttga ggatcccaaa aaaaggcttc tacaggttcc atcttttaat 21960
atttaccata tcaaaaatta aactgaaaaa attttaaatt atttattcat ttaaaataac 22020
aaggataaac ccattacatg ctaacataaa tcatgtattt tatgaaaaat agctatattt 22080
atcaaaacaa aaattagtga gaagagtggc atgtataatt ttttttgttt atttttttgtt 22140
tttagatgga atcttattct gtcgcccagg ctggagtgca gtggtgtgat ctcggctcac 22200
tgcaagctct gcctcccagg ttcacaccat tctcctgcct cagcctcctg agtagctggg 22260
actgcaggtg cctgccacca cgcccggcta atttttttgta tttttagtag agatggagtt 22320
tcaccgtgtt agccaggatg gtcttgatct cctgaccttg tgatccaccc gcctcagcct 22380
cccaaagtgc tgggattaca ggcttgagcc actgcgtctg gcctaaattt ttgtgaatgt 22440
ctttaatgcc tgccttctca tatttgtttc tgcattcaag ttattgcaaa atgttgtgtt 22500
ggttgaagtt tgtaaagaaa atgtggcctc atacagttgt gtagttggaa aggcaagagt 22560
attttgattc tctcttcaaa caactatgga caacctgctg ttacaaaacc agaatgcaaa 22620
aagttgtagt aaatacaggt taggtgtagt gtggaatctg aaagcatgtg aatgaacttt 22680
ctgagttttg taacattaaa gtccagttgc gttaagctac tgtgatagca tatagcattg 22740
tcctaatact ggaattagta tcagaagtgg ggtgctactg ttaataaata aaaagaaata 22800
aataaatcat gtgatactgg ctcagaagtc aggcagtagg ctgtgtggaa cctgacatca 22860
cgccatgtaa tacattggca accatttgat ccagctgtct gtcatgatga cttggaaagt 22920
caaccacata cttacagagc ctgtagacat aggggaaaat agtataaaac agaatactaa 22980
cagtggacct tggttcttgc cagttgcatt tagccaaata ttaaacaaaa gagatattct 23040
tgggcagcaa ctggaccatc ttcaagtaaa agtgaaaggt aataaacaga gtccagacat 23100
ttgtgcccat gcgggttaag aaaaatccag ttgcttctag acaccgtata tgaaaacaac 23160
gctgaaaaca agcctttgag tggtaaaggc cgattaacac tcagcgcggt aacaaagacc 23220
aggtgggcta acccgaaatg aaatgagaag cctgtggtga tgaggaggca gagaagtaaa 23280
atcaagtttg agcatttcgt ttaggagagt ttgggctctg attacttgca catgcaaacg 23340
aactggaaac aaacagatca gatgtctacc acttcttcga gggaattgca ttgccaaaga 23400
agtcatgaaa gcagactcta tactgattag gcattaaaac aaaaacaatc tttaggcccc 23460
taaacttgca tgggcaggaa gtgggctgtc aaagctgttc atcctctaag gtggacctag 23520
ttcctagtcc ccagtataca cttcagatgt ggccctggag gacactggac atggaggacc 23580
tcccagagga tgaggctagg gcttcatttc tccaatgacc tcagctgcct ctatttcccc 23640
ttcttcctct ggaagtccta tcatcgttat tattattatt atcatcattt ttattttgag 23700
ataaggtctc gctctgttgc ccaggctgga gtgcagtgac atgatcatgg ctcactgcag 23760
ccctcccagg ctcaagtgat cctcctgcct cagcctcctg agtagctggg agtacaggca 23820
catgccacca tgcttggcta ttttttttttt cagtagagat agggctctca ctatgttgcc 23880
agggctgatc tcaacctcct gggttcaaga gatcctccta cctcagctcc tgagtagctg 23940

ggattcgggt gcacaccacc atgccaacta atttttaatt ttttttttgta tggacaggat 24000
gtacagtgtt agaaatggat tgcttgcaga ggcaggagga tcacttgagc ccaggagttt 24060
gatcacactg tgaaccatga tcgcacccct gcactccaat ctgggcaaca gagtgagacc 24120
ttgtctcaaa aaaaaaaaaa aagagagaga gagagagact caaagatagg caaaaaagtg 24180
ggaaagcttt atagtggaca aaaaggaacg ctctaagtct gccctattgg catggtgctg 24240
aaggtgggct aactagagat agggggtact atgtggttga ctatgggtgc atctttggct 24300
ttcccctgggt gatcctaagt tggaagcagg gacaaaaatt agggaagctg ttagttattc 24360
atcacgttct ggcagtagtg gactggttgt gatagaagtt attgttttgg ccaggtgcgg 24420
tggctcatgc ctgtaatcct agccctttca gagttcaacg tgggtggatc aggaaggagg 24480
gaggatttgg gaggtcagga gttagcctgg ctaacctggc gaaatcccat ctctactaaa 24540
aatacaaaaa ttagctgggc gtggtggtgc atgcctataa tcccagctac tcgggacgct 24600
gaggcaggag aatcagttga acctggggag gcggaggttg cagtgagcca agatcgtgcc 24660
caatttcatc tcaaaaaaaa aaaaaaagtt atcgtttagc ttcctcgatt gttactggac 24720
gtagtaatct ggcttcctgc aagtctaact ttcagcagac tggctacatg ggctgtgtac 24780
tgtagataag gcagtaagta aagcaaaaat tgatagagca tcaaggataa atagaaaatc 24840
cgtaatcaag cagaagattt gaacacttca ctttcagtaa ctgataaaac aagtagacaa 24900
aaaaaatcag taaggatgta gaagatttga acaacgtaat taacaaactt gacttgattt 24960
acacgtctag aaccctgcag aacacacact ttttcaagca tactcagaac atttatataa 25020
agtgaccata tggtggacca taaagcagtt tcaacaaatc tcacaggagt aaaataacag 25080
accgtgtttt ctgaccgtaa gtacagttaa cctagaaatt gaaaacaaaa agctagaaaa 25140
accccatgta tctggaaatt ttaatataca ctttgaaata acaaatggat cagagattaa 25200
ttcaaatagg aatttagaaa taccttgaac tgaaaaataa tgagaatact ataccccaaa 25260
actgtggggt gcagctgaac agtatataga cgaaaagtat actcatatgt gcatacctta 25320
aggagcgggg aggattgaaa gttaatggga ggcaaaagca ggtggatcac ttgaggttag 25380
gagttcaaga tcagcctggc taacagggtg aaaccccatc tctactaaaa atacaaaaaa 25440
ttatccaggc gtagtgaggc tgaggcaaga gaatcgttgg aacccaggag gcagaggttg 25500
cagtgagccg cgattgcgcc actgcacccc agcctgggag acagagcgag actccatctc 25560
aagaaagaaa aaaaaaaaag aaaaggccag gcgcggtggc tcatgcctgt aatcccagca 25620
ttttgggagg ccgaggtggg cggatcacga ggtcaggaga tcgagactat cctggctagc 25680
acggtgaaac cccgcctcta ctaaaaatac aaaaaaatta gccaggcgtg gtggcgggtg 25740
cctgtagtcc cagctactca ggaggctgag gcaggagaat gtcatgaacc caggaggcag 25800
agcttgcagt gagccgagat cgcgccactg tactccagcc tgggcaacag agagagactc 25860
tgtctcaaaa aaaaaaaaaa gttaatggga taaacatcca tctcaagaag ttagaaagga 25920
atgacaaata aaccaaaaaa aaaaaaatca aaagaagaaa atcataaggt caagactata 25980
aagagagtgg ctgggtgcag tggctcaggc ctgtaatctc agcattttgg gaagcagagg 26040
tgggcagatc acttgagccc aggagttcaa gaccagcctg agtaacatag agagacctca 26100
tctttgctga aaataaaaat aaaaaattag ccaggcatgg tggtactgag gtgggaggat 26160
cacttgagcc taggaggttg aggctgcagt aagccatgat tgtgccactg cacttcagcc 26220
tgggtgacag agtgggaccc tgtctctaaa aaactaaaat aaggctgggc gcggtggctc 26280
aaatctgtaa tcccaccact ttgggaggcc aaggctgagg tcagcagttt gagaacagct 26340
tggccaacaa gatgaaacct catctctact aaaaatacaa aaaattagtt gggtgtggtg 26400
gcatgtgcct gtaatcccag ctacttagga ggnnnnctnt ngattatatt ttctccttcc 26460
tacgtcgtta ttggactgaa ttcagaatga tgactctcat tggagctctt cctgtctcct 26520
aactacagtg gcttccgacc ccactctggt tttcacttca ccccctctgct gctcatacga 26580
gtagatactt ccttccttct ttctcacttg ttgctcttcc tcaacccccc ccgttggtgt 26640
cccctcctct ttatcttttt ctcgcgacac ctgcgttctc ttgccctctt atcatccctt 26700
tctcgaggcg gtcctttcct ttatccagct taaatacctt ctcctctgtt tatttggggg 26760
ttgggttttt atctctcacc ctcccctctaa tttctttcct ctttccgcac ccatcaagcc 26820
tctcgtggtt tctcttcctc tactctcggg tccccccccct ctccccttct ttttttcttc 26880
acccccccaa gcgctttgcc ttttttttct ttgcccttta ttcccccc 26928

<210> 7
<211> 29430
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (4336),(4345),(4349),(4392),(4447),(4490)
<223> Identity of nucleotide sequences at the above locations are unknown.

<400> 7
aggggaaggg ccggctccgt agctcacacc tataatccca gcactttccg aggagagagg 60
atcatctcag gccaggagtt caagaccagc ctgggcaaca cagcaagacc gcatctctac 120
aaaaacttct tttaaagctt aaaaaaaaaa aaaaaagcaa agaggacagt tcaggagaaa 180
agcctgtaga ggcagcacac taaggaggag acgcagccca ggcaccagga ggggctggcc 240
atgggcactc actcctccag caggcgagtg cccagcacca gctggcccac ccagacaccc 300
aggacacggc ctgaatggct ccgtattcac gtgggtggta ataaacaagc aatacacata 360
gccaataagg acaccttagt aatgttacat cataaacgct gcagatcagg gaaatggtgc 420
agggtgaagt gggttggggg gctgcatgct acatgagaag tgggtcgggg ggctgcatgc 480
tacctgagac agagcaggcc ttgctgggaa agaaggagcc ggcaggcctg ggcaaaggtc 540
ctggggtggg agcacactgg agcagagtgt gggggtagca tggcgggtgc tggtcctctg 600
ggcgccttcc caccacgtca tgtgcccatg tgcccaaggt ctctcgtttc acagcccccc 660
gaagctcagg ggtcacagct acacagcccc cagatacctt ggcctgcccc aggtcattcc 720
atccagtgat ggacctgctg acctctagcc tgacctctgg gcagcgtaat ttgagaagga 780
ggagaaggga gggcaacaga cctggggcga tgagggatgc acagggtggc agacacctga 840
ggctgcacct tggagcctca gttctgggtg tgggtggggg atggacaggc tgagggctga 900
agcagctggg cccggccacc atcacacccc aggacccacc agatcaccat gaaaaaccga 960
atgtcaactg gcagcccaga gtgcagaaca aacctttcag aaacacggtg gtgactgccg 1020
catcatgaac ataaaataat tacgccctct ccccagggat caccctgca ggagtttgtc 1080
ccaagaaaca ccagaaagaa ggaaaacgtc tgagtcacaa tatttgctga ggccttattt 1140
gtaatagcaa aaaaaaaaaa aaaaaaagaa caatctccag cggcaggggt aactagacta 1200
ttgtctccgt ggaaaggtag caccaattaa ctagtaacaa aatgactgcg gtaacaacaa 1260
aacgttcgac atgtcaacac caaaaaccac acacccagca taaccgtgaa ccatgatttc 1320
tactagaatg aatggcagtt atgagaaagc accagcggag acaaagattg aaaaagtaaa 1380
ggtggcctca ttagggagac aagtctctgg gtaatatatt gtaatactgg taaatatata 1440
gtttttaata tatttttaa ttccaaattc catatatgtt cctatgaagc tatttctgca 1500
aatatttttt tcaggaccgt acatcacaaa ggcaaaaggg ccaggtcagc tctccagctg 1560
agagtgacca cttcagagca gacggcagac tccagggtta gcaagcctgg ctgagacctg 1620
gcccatgaca atcactcaac ccctctgacc tcaacatcct gtctgtgaaa tggggataat 1680
tactgcacct ccacatcaca gagtgcgagg cttaaacagg atgcttcata gaaaagcgct 1740
caagaggtaa cagccgggag ggggtagtgg ttttcattaa ttaaatgttg ccttcatcca 1800
gccctgggcc agctccaaca caaagcacac accatccact cagactcagt tgcctggatt 1860
caaagcccgg cctggcctcc agctgtgaga ttccgggcag gatttcccat ctcccagagc 1920
ctcagtttcc tcattcatga aacaggaagt gatcattcct tttatttta tttttatttt 1980
tattttgaga cggagtttca ctctagttgc ccaggctgga gtatgatggc gcaatctcag 2040
ctcactgcaa cctcggcctc ccagtttcaa gcgattctcc cacctcagtc tcctgagtag 2100
ctgggattac aggcacacgc caccacgccc agctaatttt gtatttttag tagagacggg 2160
gttttgccat gttggtcagg ctggtctcga actcctgacc tcaggtgatc cgcccgcctt 2220
ggcatcccaa agtgctggga ttacaggtgt gagccaccaa gcccagttga caactgcttt 2280
taaagacacc tctggctgct gtggaaaaca gcctggtagt gcctcaaaaa gttacacata 2340

gaatgatcct atgaccagta attccactcc tacatatata cccaaaagaa ctgaacccct 2400
ctactcatgt atgtacacat acaggtacac gcatgttaac agcagtgttc acaaagccaa 2460
aacatggaaa cagctcaaat gtccataacc gatgaacgga taaatgaaac gtagtctatt 2520
caccacctga cggaggtgag aggggccata aaaaggaatg atgcataaaa acgaatatta 2580
tggccaggta tggtggctca cgcctgtaat cccaggactt tgggaggctg aggcgggcgg 2640
atcacgaggt aaggagttcg agaccagcct ggccaacacg gtgaaacccc atctctacta 2700
aaaatacaca aattagctgg gcatggtgga gggcgcctgt aataccagct actccggagg 2760
ctgaggcaag agaatccctt gaacctggga aacagaggtt gcagtgagct gagattgcac 2820
cactgcactc cagcctgggc gacagaccaa aactccgttt cggaaaaaaa agaaaaaatt 2880
agccaggtgt ggtggcgggt gggtccctgt aatcccagct ctacttggga tactgaggca 2940
ggagaaccac ttgaacccgg gaggtggagg tagcggtgag ctgagattgt gccactgcgc 3000
tccagcctgt gtgacagaag gagactctgt ctctaaaaaa caaaaacaaa aaaggcccga 3060
cgcggtgtct tacacctgta atgccaacac tttgggaagc caaggcaggc agatcatctg 3120
aggtcaggag tttgagagca gcctgggcaa cacggtgaaa ccccatctct actaaaaata 3180
cagaaattag ccaggtgtgg tggcacatgc ctgtaatccc agctactcgg gaggctgagg 3240
caggagaatc gcttgaaccc aggaagcgga ggttgcagtg agccgacatt gcaccattat 3300
actccagcct gggtgacaga gtgagattct gtctcaaaaa aaaaaaaaaa aaaaaaaaaa 3360
ctaaacaaaa gcaaaaaaac caatgagtaa tgttgtcaag tgaacttcat cccaatggga 3420
atgcagataa tttgtttaaa aggcaccatg cacactgggc aggctggctt cccctgggaa 3480
cgtcttcttt tgcctggatt cccagttggt ttaatcgggc gtagaacact ttcttcaatc 3540
cgggattcag gcacccctgc tcagcacaaa ctcagtacac cccgcactct gctgtgggtt 3600
cttggcacta ttaggagaat gtgagggggt gattcagatc tatctctagt gggtgcatgt 3660
ctgccactcc caggaacgcc cacttctggc aagtcagtgt cagagaaagg ccagctcgtg 3720
gcccctcctg ccttgagtcc caggacccgt gatcagtcct acccggagca gaatcaggag 3780
tttgaaaacc caagtgccaa caatctcatt ttaacccatg taagcatatc caatatttat 3840
atatagaatt cataacagat gtctgggctt ccattccaat agcctatatt ttacactgtt 3900
tatttacatg gttacaccaa acaagactca attcaaggta acccaatcct ttgctactat 3960
accaaaataa gcaacatttt cagtccatgc cttatatata ttcaccaagc attacactag 4020
gcctccaact gctcatcgga gcaagctgca gcctggacac aagctagaga ttaatcagtc 4080
aggaatgatc ctgcgtccag tgccagcatg atggaagaga cagagaaaca gaagacatca 4140
gggctccaga gtcaaggagc ctgcaggtta gttgggcagg atatacacac atacacacac 4200
acacgcacac acaaaaccac ccaagaagaa aaggtgggat gaatgcatgg acaggtaatg 4260
cctggagcct ggggatggat aagctgactg caggtggccc aggcaggctt cctggaggaa 4320
gaagacctgg ctgtangtgg ggtangcang ctttctaaat ggggaaaatc tggctgtggg 4380
tggagttggc angtttccga aaagaagaaa agctgactat gggtacacct ggctgttggt 4440
ggaacangca ggcttcttgg aagaagaaaa tctggctgtg ggtggatcan gcaagcttct 4500
tggaagaagt aaacctgact atgggtggac caggcaggct tcctagagga agaagaccgg 4560
ctgtgggtga accaggcagg cttcctagac agaggaagat ctggctgcgg ttagagtggg 4620
caggcttcta agaagaggaa gggctgactg tgggtagacc tggctgtggg tagactgggc 4680
aggcttcctg gaggaggaag agctggagca ttgaaaaaca aacatgactt ggtgaatgtt 4740
gagcatgccc aggcctgatc cccagaggca attacgcact caagttactt aattctactc 4800
acaatgcctc acaaacaact tctctgacac ctaacacagc tctgggcacc ttctagcttc 4860
agctcctcaa agcagttatt cacgctacta ccctgcacac ctcctcacac cccaacccca 4920
gggacaggag ttctgccaga tgccaaagct cctgatgcca aagcctgggt ctgcttccgg 4980
gctcctcttg gtctaactgt ccaccccgca tcggcatgat gtgcaaaaac aaggctttgc 5040
aatctgccct gatgcctggc ggagcgagtc cctcccgatt cgtctccttc agaaacacct 5100
gggctgccct ggtcctgtta tacccccaac acattctaca gtcagctccg caagttccac 5160
aaagatcaac gctggcgttt ttatggcatt ttatttacag tttttacaat ataaaaaagg 5220
aaggatgcca cagctcagcc agcaggacag acagagatct atgatgcttc tgctgcacca 5280
ttgtttgtgg tcaagaaagt ctgttttcaa tgatttatta aattgtggtg ggagatggat 5340

ggtggcagtg gttaccagca acatgaatgt tcttaatgcc actgaacttc acacttacaa 5400
atggttacga cgataagtgt tatatgtatt ttaccacaat taaaaacagg taaatgcagg 5460
ccgggcacgg tggctcacga ctgtaatctc agcactttgg gaggccaagg caggcagatc 5520
acctgaggtc aggggttcga gaccagtctc gccaacacgg tgaaactctg tctctattaa 5580
aaatacaaaa attagccaga tgtggtggtg catgcctgta atcccagctt ctcaggaggc 5640
tgaggcagga aaatagcttg aaaccgggag gcagaggttg ccatgagctg agattgtacc 5700
attgcactcc agcctgggtg acaaaagcaa aactctgtct caaaaaaata aaataaaata 5760
aaaataggta aatgcaaaca tatggtatag taatattatg ggctattatg agctacaaaa 5820
aagaatgact tgggactaca gttacagccc tcattcagga atttgtttta aatgtgggtt 5880
ggtcgctaag gcatgtacac aacattttga cgttcaaata ttcctagatt tggacagtga 5940
gcacccctct aagctggctc ttctgtccca gaggtcccca ccagtcctcc agaacttctt 6000
tgctttctta cacaataaga tgccccatgc tcggcttgta cctttccttg ccccagccct 6060
agaaccagct tcttcgtgga caagctctga ctcctttggg tggagaatgg tattcagaaa 6120
cccagacctg ggctctggtg tgctcactgc tacttggggt cattgcttct aggcctctct 6180
gctgatggag gtaggatata cacgtacagt cttccctctt cccagattcc gtacttgagc 6240
tcgcctactt gctaacattt atttatatcc cccaaattaa acctcacagc acttctgcaa 6300
tcactcactg acttgcagag tgtgaaaaaa ctgagtcacc atcacacgtt ccaaactgag 6360
gtcaactgag gccacaacgc cccatcttct tgctccggct gtcgagatgt aagcaagtgt 6420
ccttctctcg gtctagctag tgccatgctt tccacatcac tgtgcttttt gtgggcaatt 6480
ttgctgtata aaatgtcccc tgcacatatg ctgctgtgta gtgctcctag gtgcatgagg 6540
ctgccccacg ccttacagag agaatatgca tgagaggctt tattcaggta tgagttatag 6600
cgtagttggc catgaattca atgttaatga atcaacaata tacagtaaat aaggtgcttt 6660
ttagagacag ggtctcactc tgtcacccag gctttagagt ccagtggtgt gaccttggct 6720
cactgccgcc tcaacctcct gggctcaagt gatcctccca cctcagcctc ccaaactgtt 6780
gggattacag gcgtgagcta ctgcactcag cctaaataag gtgtcttaga aacacacata 6840
agacaaggtt atgggctgag tgcggtggct catgcctgta atcccaacac tttgggaggc 6900
caaggtggga ggttcacttg aggccagaag tttgagacta gcctgggcaa catggcaaga 6960
cctcatctgt atatttttt aaatcagaca ggtgtggtgg tgcatgccta tagtcccagc 7020
tactggagag gctgaggcag gaaaatggcc tgagcccagg aggtcaaggc tgcagtgacc 7080
catgattgta ccactgcatt ccagcctggg gtgacacagc aagacgctgt cttaaaaaaa 7140
aaaaaaaaaa aagccaggtc aggtatcgaa cagttggcaa aaacgttgtg acctgaggct 7200
cacaggaacc tagcccgatg tttcccctag gagcaatggt tcagtattca ataattcagg 7260
gttcccagtg actttatgga gcataacttt caagaataac aagaaccaac tgtacgtgtg 7320
tatgtatact cacactttta ttttatttta ttttatttt tgagacagag tctcactctg 7380
tcacccaggc tggagtaaaa tggcgtgatc tcgactcact gcaacctccg cctcccaggt 7440
tcaagtgatt ctcagcctcc caagtagctg ggattacagg tgtgccccca caaccggcta 7500
atttctgtat ttttagtaga gacggagttt cgccacattg gccacgctgg tctcaaactc 7560
ctaacctcaa gtgatccacc cacctcagcc tcccaaagtg ctggaattac aggcatgagc 7620
tgccgtgcct agcctacata cacttttata cacacatgca tctatgacta tttctctatt 7680
tctgtgcatg tgtgcgtggc agtacctaca gtttcagcta tgtgtctggg tactgtctcg 7740
tccaagtttg taagcacctt ctccaaagtg caaagcctgg cttgtgttac tatccatatg 7800
tttacttatt tgctcaatca atttacttat tagctccata accagcttcc catctgctcc 7860
agtagcctct gctgtcagtc acctctgcac cctaccccac cttgcttccg gatgctggat 7920
gccaatcacc cccgacacct ctacatagca ccaccctcga catgctgctt ctttatttct 7980
tatttatttg tttgagatgg agtcttactc tgttgcccag gctggagtgc agtggcacga 8040
tccaggctca ctgcaacgtc cgcctcctgg gttcaagtga ttctcctgcc tcagcttctc 8100
aaatagctgg gattacaggt gcccaccacc acgcccagct aatttttgta tttttagtag 8160
agatggggtt tcaccatgtt ggccaggctg gtctcgaact cctgacctca agtgatccac 8220
cttggcctct caaagtgctg ggattacagg tgtgagccac cgcgcctggt ctgcttcttt 8280
aaatgccagg caccaacatt tgtgcaatgg ggtgggagga aagaacaggg aggagagcac 8340

actgccggcc cctgcactga atccactgat caatctgggg gcaactgcca tctccatctc 8400
ctgtcttcct atccgtgaac atctactgca gtcctctcca atgtccttct gtaaagttgt 8460
attatgtttt gcatacaggc cttgcatatt agttctcaga tataatccat atactttata 8520
taaaattcaa accacattta aaaaaataaa actagcatga ctataacgga gtctgcaaca 8580
ttctcacaga ctttatgata aaacatgaaa cttcaaagat acttagggtg gggcagggac 8640
aatgtttaag gctgcctgga agcctcccca tccctgagcc agaaagtcct atctcccctt 8700
caaggggaaa tgcttgaaaa agcactgatc aggctaaaat gacagggatc agggagtaat 8760
caaagtacaa gtgagctggt ctcctccatt ctgagcacag caaagttcag tctctccaag 8820
tccaagaatc atacacctgt ttgccaagaa tgaagttcag gtgtctacaa gtggctgaaa 8880
atattcattg ctgggccatt aacaacattc ttggcaaaac cataccttag cttctcgtgg 8940
aaatttctta aggtagaaga aacaggaaac acccaggctc gcttttatgt agacagttcc 9000
atgaagccag ggaccttccc cacatccacg tttcaattac ctgcacgcag ctcacagtgt 9060
attcaacatc tacgcgtctc tcctactggg gtggcggtgg ccactcaaac cctcatgcag 9120
ctacgatgac cgcaattttg gcaacataat ttcatgtttt tccttgggct tttacccaag 9180
tcagtgacac aattctgcag ttgtctaaag attcaaaatg agggacttga catttacaac 9240
aataataaaa tcttgggttt cctttaacca agcacatgtt ctgcctttta gagaaagctc 9300
tgcaaactca agctggagtg ggatacttgc tgacatcttc aagcacccca ggaatagctc 9360
tactccccca tttccacctt ggctgaacca tctatatccc accaattccc ccaacatccc 9420
tccatccgtc catccatcca cccaaggacc tgctaagcca ggaggtctct cccatctacc 9480
ccacagcctg gcctcagccc acaagggctc tctctacatg aatcccaccg caccagagta 9540
gaccaagtct cccgtagact ccaccctgac cacctccatg cctccagcca ttcccacccc 9600
taaaaaccct ccctggtctc tacacccagc tgatgaatac ttggctgaat gtgacctggc 9660
ctcctggacc caggtgaagc ccacgtcctc cgtaagcccg ccagctcacc ctgcctctgc 9720
accttcactg gagagagccc gcacttcacc tcctcagggc aggcatggct gatgccaccc 9780
agtggaatct ggtgcaaagc agggcccggt gcagagcagg gctgcctgca gagcaaggcc 9840
ctggtgctgg ggccgagcac ctccaatgct ggccgtggaa ccatccctcc cattccaggt 9900

gctgtctcca tcaagaatga gcgagctgct gacatttgca tgacaataat gaataaatac 9960
catattttgc ttcaaatcca gaatagatgt ggccagggtt ggcatatgac tgttgggaaa 10020
ggacagtttg cctcttccca aaccaacttg gattataaaa agcttttctt aacgaccaca 10080
agagcggagg agctcagggg cagacaaaag gaaggctggc tgcagaaggc gggagagtgg 10140
ggccttcagg ggcgggtggg gagagagaaa gcctggagct gcacccccaa ggtctgtgta 10200
catcaggtgc tacagaataa caccacctct tccagcttgg cccccacctg ccctctccca 10260
gcccagtcac ccagacagca ccccactccc cacacacacc tcacatctgc ccgcctcaca 10320
ctcaccagct tcggctctca atgcaacctg gaacctgccc ttggcctctc agctcagcca 10380
cccccattcc tgttggcccc tggcccccca tcgaattctc tctaatccta atgcacacac 10440
ttgcacactc aaacacacac acacacacac acacacacag cccagaggaa aaccataatt 10500
gactgaggtc caggcaagtt tcccgagcag ggaccacatt tcaaaggtca gggaagcagg 10560
cgaacaggaa acatacaggg ggcacgtttg ggggtggagc aggaaataag aaatcacttg 10620
caaaagataa aaagaaaatg aggtagctgg tttcagacac ctcggagcac acagaacagg 10680
acaggcgcct ccgggtcttc cctcaacagg gagatgggcc aggcaggtcc ctgctgctcc 10740
accgcagagc tgggggctat ggccctgaca ccaaggccct ggggcaggcg gggaggcagc 10800
tgttctcctg cctgtgctcc cgggcagggc ctggccccac aagggaactg gccgaaggct 10860
ctgcttggct actccggaaa gtcctgggag acaagcaaag gacttgctag gtcactccaa 10920
acggcccaga tgtgacaact gtgaagaagc cacaccaaag caaggtgaca gaacaatgtt 10980

ggtgacgtca ggttatcagc ttacgctcaa ctccacttac ccggactcac ccgtaacctg 11040
ccgtctcttc ccaaccagta aaggatgcct aggtagaggg gcacaaggcc tggagcataa 11100
ttaccatttt aaaggctctg agaagtcctg cggtgaggaa gcctagttca ctttctctcc 11160
cctaggattt cccaactgcg cctgatcaca gaacattttt tcatttccac tcaggaaaca 11220
tattttgaaa aacactggcc tagaggcaga agtgaaatgg aaaacacaaa agtaaaactg 11280
aacaggaggc actgggcaga gaacggtcag aggcgccctg aatcctggac cggtggagat 11340

ccccagcttg gcatgctccc ctccctgggc ccagaccgcc tccccccatt tcctggataa 11400
gaaggctaat gcgcatcagg gtgaagggct tgcctgggct acacccccag gctcgcccca 11460
caccaatcgc gctcctgcga gagccagtga ctttcttgat ttggctactg tggaattgtt 11520
tgcaactaac caccccagat acagatacaa atgacaggat gatcagatgt aaaggaccca 11580
caggtctctg tgatacggct tcatgcagcc agcatggcta gtgccgtgca gaatgagaat 11640
gaccccaggc aagtccttgc ctcccagacc cagaacccca tggagcccac cagggctggt 11700
tcacaagcac tgtctgggtc gggcagagat tccagcaaga ggagggaaca tccatgcacc 11760
ggagccagtt accagaagca aatcgcctct tccaaaaccc aggctattaa tggagtccac 11820
tgttgagtgg agctggggtc tagctatgga atactgcaca gcagagatct tcctgagaga 11880
aagcagtttt ccctgaaagc catgtgtcct ccactaactg tgttttaatt gggcgaacgt 11940
ctgtatctca ttgcagtggc cgcgcatgtg ctgacaaggg gctgggggcg gggtggggag 12000
cagaagctca ggggcctggg agggaaggaa acaggccacc agggctcccc agaaggcatg 12060
tatctctctc acaaacacac gcatgcacac acacgtgcac acatactctg caagccctga 12120
gttagcaact gtggaatgtg accagctcag tgatcccagg acaagctgct agggaatatg 12180
acatttgatt gatgtctgca aatgtgcgtt ttcactaatt agaaggttta gggcagagca 12240
gagaaaaata tgtatttcag agtcccagtt tgacctgcca gaaaccagcc cattactaac 12300
attcttattt tcaacaaaat atagcattct gattacatac catcttggtt ccacgcctcc 12360
tgccttgcca agccccccgga agcggcccaa ggccatggca aatagtgaga gaaacagttc 12420
cagggtggag actgactcag gggtgtcagt cagtggggcg ctgatggccg gtgggaggcc 12480
agcagtcatc accctctcct tgggacagtt gagtagctct cccccagggt catgtggcca 12540
ctcaggttca tatgggaggc gagaggagtg gcagagtcca ggagagtggc tccgaagtca 12600
ctgttccctc caggcctcag tgtcttcatc cattaaatgg gtaggctgag gtctgggatg 12660
acaaggaggg cttgcactta ctgaaaccca tgggaggctg ttcgccgatt tcttttattg 12720
atggaagaaa acactcgtat aattcaagta ccaattaaaa ggcaggcact ggaaccaccg 12780
tctgccaatt cctagttttg cctataccaa atttgagcaa gttaattgac ctctcccagc 12840
ctcagtttct tcgtctgtaa aatgagggta gggatggccc ccagcccaca gggcagctgg 12900
aaggattaaa gaaatcaaac atctcttaga gcccacctgg cacactgtga tacacaacaa 12960
atgttagcta tttttgtcta tgaagtctag attttatatc ttgggtgttc taaagcagga 13020
tacatttatt taaaaacaag gattttcatt aaacacgtac cccacagaca gcaacccccat 13080
ggagactgct cttaattcag gccagtatcg aaacgactct aactacaagc tttatacagg 13140
tctcttggct gtccttcaaa tccaactaag gtggtacttc tgaagcactg tgcacatgtg 13200
tgtgtgcatg cacacgtgtg ggaagggcgg gctcacggat ccctcaggta ccccacccac 13260
gcagtctcaa gtcacaaagc gacagagcag ccgaggaagg tctgtgcccc actggaccct 13320
cgtgaagcca ccaactctac ctctgcgccg tgtcctgcag actgggctac cctttgggtg 13380
gggaccagca tttgatgcaa gaaaggcaga cagaaaagga aaagggcaag ttcgactcca 13440
gataacacag acagtaccaa gccccagggt ccataaatgc cacgcagatg gaagcattta 13500
ctgcgaggcc acacagcaaa cgcacggatc cagggacgga ggtgcagact gcggtgcccc 13560
tgagccatga ccctgcaaat taccaccatg ggaaaggagg ctgccaaacc ccccgacagt 13620
cggctgggct ggcacagact cgtggtttcc atcgaggtgg gaggaggtgg gacgtcccag 13680
ccccctcccc atgcccactg cagagggaag cggccgtttc ccctgtgtgg ttacaaaggt 13740
ctcattgttc ttcctcacag ggaggaaact ggaggaccga gctcagaacg cattttagaa 13800
ctggcagaaa agaacatctg gggaaggaaa cacatttcag aaacaaacat acctttgtac 13860
cagcttttat tttctttaag tgttgaaaaa ataataataa taaagacatg ccaaatttat 13920
catcgctcta caaaatccct ttattgagca aaacgtggca gctctacttt caaatgatta 13980
ctgttcctgg aaaattgcag caacgtggat gccaaggccc gaaggccgcc atcagcagcc 14040
aaacaaaaga tgccacctcg ggctccgcga cactgtacca tgccagggaa ctggacagat 14100
ttggggaatg ccacggtttg cctttaaccc cttgcctcct ggtctcctga tgcatctcag 14160
aggctaacat tctttgagga actggcattt cttagttgta aatatgcatg tgggtttggg 14220
agctgcctgc aaagtccagt gttgacgatc agctttgatt tccttggaat caagtttacg 14280
tgtcgagtct ggaagttaag aagaatttgg agaagctgag cactatggtg ttgcaggccc 14340

tgggtgaact cttccaccaa gcattcattg tggactgaca gcgtgcgagg ggctctgcag 14400
gcaggtgcac aggacgaaac acattccgtc cgggggaaac ctgcaggaaa gctccctctt 14460
cttcctaagg tgccgggcct agcttcatgg gtccctaccc tccacgcctg tcacactttc 14520
tgagtctcat gtgggagctg cttctggttc ctgacttcac tcagtcctca taggaggtgg 14580
aactactgtc accccatttt acagatgggg agactgggca caaggggacc aagaaaccaa 14640
tgcaaagtca cacttgtggg atcagtgaca ggggagatca attcccaggt tctttctgca 14700
agagttaaat tgttttcatg ctgcctaagg gggggcaact gaaagaccac tgcatatctt 14760
tgccaaaagg gtcaagcaca ggagccgcag ccagtgggtc agatccgcag aggcgctggg 14820
gtgaccctcc ccatacctgg agggatgctt gtcccctcct ggccttcact gggtcccctc 14880
atgaccgtgg cctcccagga cctcagcaca atcccggtcc tgtgctccag gacaagccct 14940
ccgtccccaa gactgtgagg aaatggaacg aagaggggct cgctgcagcc cagcacccac 15000
actgcccctt ctcaggggca agaaccgtcc tggaggactt ggctttggag ggggagcctg 15060
ggaggccagt aagtcaacaa gcctctactg ctcatgggtg ggatcccacc gcaggccccc 15120
acctgctggg gcgggcaggg acgggcggca cagcttggcc agggcagata acccccacct 15180
tggccagggc gaaggcagga cacgtgggct ccagcctggc cccaccatcc ctgcacaaca 15240
ctgggcaaag tccacgtttt cctcaactgg gtgttgacat ctgcaggaca ggggcatgga 15300
ggtacagagc gctgaagcca cacagcaacc taggagcgag actccatgcc tccccgggga 15360
cccctcccca ccatgaggac catgaaggct tcccatgtgc cgcaaggact ctggtgtgga 15420
gacacacgtc tcctacacag ccaggcctaa cgctcttgta actgggtggt cccacctggg 15480
ctcacagctg gagggccagg agctcaaggc ttcgcagggt ctgctctcat cccagaggcg 15540
atggggagcc acagcaggct gcaggagaga gggtgggccc cctccacttc agaggcccca 15600
tctggcccac agactggaga gcacatctct cagcaaccac ggagcgccaa ctgcgcacag 15660
ggcctggtcg tcagagcggg gcaaaggcac tgaccgtcac ggccagggcg agggaagacg 15720
ggtgggcagg gaccttgggc agaggggaa gaacctggtg cccaggctgg ccctgccttc 15780
agcagtgaag ctgagtgggg aggcgctgat gcaggggcc agaaagggct gctggtcagc 15840
cgggaggagc cccccacaga ggaagcagcc agcccagacg cagatggcag ggtcccctca 15900
acaatgtcct ctgaaaagga gaggcgggga ctgctctggt gacacctaca aatagatagt 15960
cagccctcag ccccctgcca tacttctgac aaagcagagg cccccagggg aggcgcaccc 16020
gaaggtacct gcacctgtcc cccagactcc tagagcccac ctgaccccat cccaccaggg 16080
ctccagctac aaaataaatg ccgaggccag ctaggcaagg acgcacactc ggtaccgact 16140
gaataggctc cacgttgtca tgagcgcaac ccacaggcca ccaggccaca ctatgcagag 16200
ctgagatggt ttcggccaag cagcctctca gctgagctga acaagtccag agtccccggg 16260
gggtcgtcac tatggagtaa caattgcgat gcgatggtaa ccctaacagc taaccgtcac 16320
tgagccaggc cctgagctag gtacttttca acgctgcctc tctgcagcct caggacgagc 16380
ctgtgggagc ataaagatca ttccctatca cggatgggga aactgagctc tgaagcagtt 16440
aacgtgcttg tcccagaccg cagagctagg agcaggacac aacagcaggt caggcaggaa 16500
cgggtgaggg gggcctgcat gggcttctct ggaggctgcg catacacgca acccccagga 16560
ccccgaccct gcacctgcag ctcgctactg ccccctcagt gactccagca aacctcgggg 16620
taggggaagg aggctgggaa tacctcgggt gtccgaaaca gcagcttctg cttggaggcc 16680
actgctgcat aatggttgct gcccagcaca ccccaagcca cctgtgccac ctgtggtgac 16740
cttccagcat gccttggtga ccaagctggc cttaggtgct gtgggcagcc aagaatagaa 16800
cagggcccac ccctcctctt cacactaaca caaagcaaga ggcgggcact tcgactgagt 16860
gcatccctct agctcaaggg cctcacggat cacaggggtc agggcaagat cccaattctg 16920
cattcccgtc tgcctttcat cctgctctgc caacaacagc cagtgaggct ggggacatcc 16980
ctgaacctgt ttctcacctg aaacacatca taccattgga ccccagccct ccgggagagg 17040
ccctaatccc tgactgtggt gagatcagat cactggttaa gtacccagaa gggccttggt 17100
caggggctcc aggggtgggg ggtgatgggc gtggtggtat cccgctctgg gctatagtcc 17160
accctgatgg aggaggtctg tggtcagaac cgggctgtgc agggcacagg agcccagagg 17220
gacccccaga gctcacctgg tggtctctga gcagggctcc ctcaaccctc agagaaaagc 17280
acagcaagga ggccgcccag agcccagcgc ctagcaccca gtggcgtgcc agacctgcct 17340

ggatcctgga gatctctcat caccctccaa gtcagtcatg cccaacccag ggacccacag 17400
cccacggggc cgtgaaggtg tgctgagtcc aagaaggcct tcgacactgg gaagccaagt 17460
ggcacctcct ggtgtggagc aggcggaatc ccaccagcct ctgctctgcc agtgggcaca 17520
gctggacgat gagcagaagg ggctgttgct taataaacgt catttcctta agaggataaa 17580
acctttcaaa acagatggaa attttttttt aattaaaact ggtggccaaa gagatggaaa 17640
gcacccccttg tgcctccctc ccatcgtgac ccatcctctg cacacctcaa gctgttcgct 17700
gcccaggtgt ctcctgaggc actgggggcg ggtgagaatc cgtgagccct cggccagccg 17760
tggctctctg gagctctgcc ccaggccatc agggcacacg ccgggcaccc tgggggccac 17820
acagggcaga gcccagctgg gtcagcacac agggccacac tgggcacaca agtctctgag 17880
cctcccctgt ggacgcagct ctcactatcc caccccacta ggtcccgggg atctgtccca 17940
cagggtgata tgctgtcaca gaccactacc agagccatgg cctgctgttc cgcccgcagc 18000
caggtagtca cttgctccac agggacaggc aacgccgcac ttgggggctg ctctgcggca 18060
ggactagagc tccagcagct cagccctcct gagaaggaga actccatgct ctaagaggca 18120
gacgcagcgg acggcaccaa agccaccaca agcccacggg gccctgcatg gcaggtcagg 18180
agtccctgac cactcgctct ttgtaaccag agctgcagtg gagtctacga ggcaaggact 18240
gtgggcggca gtggccacag caaatgaatg agtgtcccaa gggagcaggc ggctgcgggg 18300
aggcacagcc gggacccagg agtcctccgg cactgcagca aactccctgg gccccctgag 18360
cagcgaccag gtggcaagtg catgaactcc cgggggcata acctgggagg gtgacactct 18420
cttcgtgttc aaattcttga gaacgcatta aaaatatcac tcagtcacct actctatagt 18480
tttaactcaa aagtaccaaa gtagccaggc gcggtggctc acgcctataa tcccagtact 18540
ttgggaagct gaggcaagag gatcacttaa gcccaggagt tccaaatgaa cctgggcaac 18600
atggagggac cccatttcta caaaaaaagt gttttaaaaa attacctggg cctggtggtg 18660
tgtgcctgta gtcccagcta ctcaggaggc tgaggcggga gaaccacatg aacccagggg 18720
aggtagaggc tgcagtaggc tgtgatggca ccactgcact ccagcctggg taacagagtc 18780
agactctatc tcaaaataaa tttaaaaagc accaagccag gcttggtggc tcacacctgt 18840
aatcccagca ctcagggagg ctgaggcaag tggatcacct gagtcagaag ttcgagacca 18900
gcccagccaa catggtgaaa ctccatctcc actaaaaata caaaaattac ccaggcgtgg 18960
tggcgggtgc ctgtaatccc agctactcag gaagctgagg caggagaact gcttgaaccc 19020
aggaggcaga ggttgcagtg agccaagact gtgctactgc actcaagcct gggagacaga 19080
acgagactcc atctcaaaaa ataaataaat caatcaaaac caccaagact ttttaatata 19140
aacatttatt attccataat tccttttttg catgattaaa aatgtttata taaagtttcc 19200
tgaaaatggt aagaatgcca agtgaaggct gcaaatgccc aagcccccac cgtggcatct 19260
cacggagtct gggccctagg aggctggtgg gtaccacgtg gacccgagac ttcacagtca 19320
agtccctttg gggtacactg ggtttcccac accccagaaa tatgggctct tactgcagga 19380
ccatgggggt cctcacactt ggcccagaag ctgtcacata gccagacagg tgttctacaa 19440
cctaggctag agggagctca tgctccagca gaattcgagc cagaggaggt aaaagatggg 19500
taagatctgc tccctggaca gatgaggcct tggcctcaga acagttactg atcatctacc 19560
agacatcaca ctagaggcag aggggcgcag acgaagacag cccctgtcct caaggccctc 19620
ccaggttggg tggaccatgg aaggttccag acagatctgg caagagaagt gcccacacca 19680
ggggcagaag atgggcaggt ctgctcaggg cggcacggcc tgccaggcca aaaagttcca 19740
acttcagatg ctggagaatg ggcacgactg tctgagaaag ggaaggatgt gatgaaaact 19800
acttggagaa aaattaatct ggccagagca taagataaat gggcaaaggg gaggttccag 19860
aaagcaagga gaccaagtaa aagctgatgt cattggctct gaatctaggc tttcactgaa 19920
tatgcaccgc agggcctgta ggtaaagcct cagagcccag ggagtctgag tggaggagag 19980
ggcaggggac agagctgggg cctgtgtcta cagtgctcag gaggaatagg catggacgtc 20040
agctcggagg ctccagctga agtgaggagg cggccagggc agcacggcca cgcccggatc 20100
cagactcctt ttgggaagca agttcgctct gggggaaagt ttggagaaat ggcctttacc 20160
cgcagaagca agccccagaa catatcttgc tccaaaacta tctcgtacag tgaggacgtt 20220
aagcttcagg tcccctagag gagacagtct gctccttcct ggggcagaac ccaaggtggc 20280
cagagcctgg aaggcaccca gcacccaggc tggtgtgttc cagcccaggc cacacgctca 20340

gatagctatt aatgccccgt tgagcaattt cctgagagct ttgccaggca ggtaccgcct 20400
ccccatctga actaatacag gggtacatcc caaggaagaa atgaaaggtg cccacatttt 20460
gctctgggat taactaggga ggggagtgat aattaactca gtaattatat ttgccatcgg 20520
gctaatgcta aaattagtgt gcattagaat ttctttcctg agcagacacc ggagtgagtt 20580
gggcagcagg agtggctcgg gcaagtcggc acaaagggca cctccagagc cttccacaaa 20640
tgtcagcaaa acccacaaat gtcaaggccg gctccactgc acccagcaga tgaattcact 20700
tccacagcct gagaccgcca gctcatcgga ggccatttaa aatccagccc tctgacacct 20760
gctggatatc accatttacc gtccccagat caagagatca aagggtggaa cctgatagga 20820
cggctctgaa gttcaccaca aaagcataaa cgtgcaagca gagccaatac gtcttttgaa 20880
aaggacaatg aggtgggaat ttacataact gatcttaaaa tatgttctga tgcttcagag 20940
atggagacag cagcattccg gtacacaaag acactcacag gcagtggagc acagtgaagg 21000
gtctggaatc aggacccagg tgtctgtgga cactacacat aaaagagcag catttacaat 21060
gaatggatag gatggaccat cccaccaagg tgttggacaa ctccctattc actggccaga 21120
cccctacctc ataccatata caaaaaaaaa aaaaaaaaaa aaacccagac agaataatgt 21180
ctgaatgtaa aacataaaac agtaacagtc ctggaagaaa ataatggagg atatatttat 21240
aatctggaga tggagtaaca agggatagga aaaaagccat agggaaaaag tagagttatg 21300
attatatgaa gcttcttaat atctttatga taatgtacca ccagaaacaa ggatgaagga 21360
ctagctacag accagcagtg aaacctgaaa caaacagaac aaagaattaa agtccatacc 21420
aaataaagac ctcccacaaa tctataagaa aaagataaac aggctggcac cgtggcttat 21480
gtctgtaatc ccagcacttt gggaggcgga gatgggtagg tcacttgagg tcaggagttc 21540
gagaccagcc tggccaacat ggtgaaaccc tgtctctacc aaaaatacaa aaattagcca 21600
ggcgtggtgg cgcatgcctg tagtcccagc tacttgggag gctgagccag gagaacagct 21660
ggaacccggg aggcagaggt tgcagtgaac caagatggca atcgcgccac tgcactccag 21720
cctggaggac acagcgagac tctgtctcaa aaaaaaaaaa aaaagaagaa gaagaaaaaa 21780
gaaaagaaaa agacaacaga aaaatgggcc aaggataagt gtaggcaatt tgcagaaaag 21840
taaataccaa taaaccagaa atgagggttg tgcaaatcaa aaggtgttat aatttttaac 21900
caaactggac caaagaaaac accaaaaacc aaaatcttgt aattgccagc atcagagagg 21960
atataggaaa gtgtgtgttc tcgtagatgc ttgcaggtat gaactgctac agccttttag 22020
gagttatgta tgtatgtatg cttgtatgta tgtatttgag acagggtctc gctctgttgc 22080
ccaggctaga tctgttgcag tgctgtgatc atggcttact gcagccttga cctcctgagc 22140
tcaatagatt ttcccacctc agcctttcaa gtagctgaga ctacaggagt gtgcaatcat 22200
actcagctaa ttttttaaat tttttgtaga catggggggt ctcccaattt tgcccaggct 22260
ggtctcgaac tcctggactc aagtgatcct cctgcctcaa cctcccaaag tgctgggatt 22320
acctggatga gccactgtgc ccggcctcaa tatctttaaa aacagaaatg gacacactct 22380
ttgactagga atgtatccta taaaaacact tatacacatg cagagacaca cgagcaagca 22440
tgctttgtaa tagcaatgaa ggctggaaaa actcctcaat caggtaaatg ctgtcaagtg 22500
cacctgtgta ctatgaaatg gcacttggct tttaacaaga gcaaagacag aaaagcaaaa 22560
gtacaaagta gggtgtgatg gcacatgcct gcagtcccag ctactcagga ggctgaggca 22620
ggaagatcct ttgagcccag gagttggagg ccaggagctg ggcaatagtg agaaaaaata 22680
aaattaaata ataataataa taaaataggc tgggcacagc ggctcatgcc tgtaatccca 22740
acactttggg aggctgaggt gggaggatcg cttgatccca ggagttcaag gccagcctgg 22800
gcagcaaagc aagacaccca tctcaacgac aaattttaaa aaatcagcca ggcaggctgg 22860
gcatggtggc tcacgcctgt aatcccagca ctttgggagg ccgaggcagg cagatcactt 22920
gaggtcagga gttcgagacc agcctggcca acgtggcaaa accctgtctc tactaaaaat 22980
acaaaaatta gctgggcatg gtggcagatg cctgtagtcc cagctactga ggcacaagaa 23040
tcgcttgaac cagggtggca gaagttacag tgagccgaga tcgtgccacc gcactccatc 23100
ctgggcgtga gtgagactcc tgtctcaaaa aaaaaaaaaa aaaaaaaaca aggagccagg 23160
cacggtgggg tgagggaggg cacagaagca gcgcctcttc tgggggcacc cccaatctct 23220
agcgatccag aggcctcagg atcctgaagg gagaaaaaac gtgaagctcc gtgctagaag 23280
agaccataga gattggaatc agctggttct attttacaaa aaaaggaaac tgaggccctc 23340

agaaggtgag tgcctctcaa tgccccacag ggaggcaggg agagggctct gagccctgca 23400
gggccctgga ttcttgcaat ggggtggagt ggagcctgtg ccgcccccac caggcacctt 23460
ctcaggagag gagccgttgt catatccttg aaggggtcct tgagcccctc aaaaggctaa 23520
aaaccacttt cctccttgag tgaaccttca cctcagttta accacaagaa aaactacatt 23580
aaggcccagc gcagtggctc atgtctgtaa tcccagcact ttgggaggct gaggtgggtg 23640
gatcgcttga gcccaggagt tcaagaccag cctgggcaac atagtgaaac cctgtctcta 23700
caaaaaacaa caaaatcagc tgggcgtggt ggtgcacacc tgaggtccca actacttgcg 23760
ggctgaggtg agaggattgc ttcagcccag gaggtagagg ctgcagtaag cggtgactga 23820
atcactgcac tccagcctca gcaacagagc aagactcaaa aaaaaaaaaa aaagcaggcc 23880
gggtgtggtg gctcacgcct gtaatcccag caccttggga ggccgagcgg gaggatcagg 23940
agatggagac catcctggct aacacggtga aaccccgtct ctactaaaaa tgcaaaaaat 24000
tagccgggcg tggtggcggg tgcctgtagt tccagctact caggaggctg aggcaggaga 24060
aaggcgtgac cctgggaggt ggagcttgca gtgagctgag atcacaccgc tgcactccag 24120
cctgggcgac agagcaagac tccatctcaa aaaaaaaaaa attaaatctc aaaaaaaatt 24180
acattaaggc aaactaaaag atgtttaaaa tatatatatt aaattaaata cactccaata 24240
gagcaaatac gaaaataccc agaaaacaca atccccgcac ccccaggaca acctcccagg 24300
gggtccacag caagagaccc caagcacgag agacagagaa cagtgtccct gtggcggaac 24360
ctctggccca tcaggctcta ttagaaaata aggctcttgc cactgagaga aagaggcaca 24420
gtcgcccagc agccacgggc tctggcacac cacgagtcag gccagcaaag tgtcaactgc 24480
cccctacaag gtgacaaact aggacaaact ggaaaccaga ggctggacct ggagcacagg 24540
gaccaccaca tggggctggg gaatgggcag ggacctcaga gcgccaccca catgcctaag 24600
agcagcgcgt atgcgcatgc ctctgcatgg cttagggaca cagggagctc cccccacccc 24660
caacccagga aggcagcccc cactacccag gtagggaacg gataggacca gcaccccgtt 24720
ctgctcgtaa ctcagggctc caggcccccc cgggggcaac cagcacagag ctcagacccc 24780
aaatatcttc acccacctcc tggtccccat ctggacaagg gtgctgggga ctggctctca 24840
gtcacaccct cggggtactc ttcaaaggac agctggatgc cccagggcag gagcttttgg 24900
cccccagctc cctcacccca gacaccagct cttgggaccc caccagcatg ggcaaggtgg 24960
acaccatcgt cccgattttg cagatgagga aactgaggct gagggctggc acacggctct 25020
ccagagctga agagaatgca gagagcagcc ggagccagcc ggtgggtccc tgaggccggc 25080
tcgtagcaag ccacagctgc ctccgcccat cacacttgga cctcactggc cccaggacag 25140
ccctccaggg cggcctggca cagagcccac accctgctgc ttcctgaaca aataagtgaa 25200
caaggccacc aagccgagga cctggatgta gccccggctc ccgccagggc ctccccaaca 25260
gactccccat ttggagagcg cattaagtgt ttccaaagcc tcacaaacca cagatgtccg 25320
gctgtctcac ggcttctgta acctgaactt ggccctcact ctgccctccc agcactcctc 25380
tcagggccca ggccccctcct ctgagatgcc agcactgact ccccaacttg tccccatcac 25440
ctggctcgtt cctgaacctc ggcaggagag tctcaggcca gatcctccca ccagccacct 25500
ccaccaggat gcaggaggca tgagacctgc tcgtgccggc tgggagatgc aaccaaccaa 25560
gatcaatcca atcagcggat gaactgacaa atataatgtg gtccctccac acaatggaat 25620
attattcagc cacaaaaagg gctgaaatag gccgggcgtg atggctcaca cctgtaatcc 25680
cagcactttg ggaggccgag gccggcagct cacttgaggt caggagttca agaccagcct 25740
ggccaacatg gtgaaatccc gtctctacta aaaatacaaa aattagctgg gcgtggtggc 25800
gggcacctgt aatgcaagct acttgggagc ctgaggcagg agaatcactt aaacccagga 25860
ggcagaagtt gcagtgagcc aagatcgcac caccgcactc caacctgggc aacagagcaa 25920
gactccattt caaaaaaaaa ataaaaggct gaaacaccca tacgtggtac tacttggatg 25980
actcctgaaa acgttacagt aaccaaggaa gtcagccacg aagacgcatt gtaagattcc 26040
cttcatgcaa aatgcccaga acaggcagaa ccacagaggc agaaagtcga ctggtgttca 26100
ccaggggatc cggggagagg gaacgggaag tcaccgtgta atgggtatgg gttttatttt 26160
ggggtgatgg aaatctctta taacttgata gaagagaggg ttgtaaacac tgtgaatgta 26220
ccaaatgcct gccttctata ctttaatatt ttatattata taagtttcac ctcaatttaa 26280
aaaaaaaaca actcgacacc tttcacctag gaaagatctg gctttagctt gcatttcctg 26340

taactcctgc ctaaagcctt ccagaagctt ccgctgcctt gtggatcaca accagactcc 26400
acaccatgat ctggcctcta agggcctctc gcaggacacc ccgagggtga aggagcaccc 26460
gtgggcccac ctctgcatag ctgcaaagct tctttccctg tcctcccctc tacatgggaa 26520
gctctgcccg caggggcggg gccttatctg ccattctatc gcactcaacc ctagcacttc 26580
actcggtagc agacaccaaa gcaaaacagc aacagcatta taccgggcca ggtgcacgtt 26640
aactcactga attcatggta ggaaggattc tattcccatt ttacaggtga gaaaactgag 26700
gcacacaaag gtagcatcag cttcctaagc ctcccagcac aggaagcggc caggctggaa 26760
tcagaccctg ggcgcagggg ctctgtccac agtgctaact aactactcct gcccccgagg 26820
gctgcagcgg tgagtgagtg agtttgtcag tggactggat gtccaaggtc atacaggaaa 26880
aatccagact attgtaataa cagcctctag accggctggg gccagaaaga tcgaggacgc 26940
tgacacacaa ctgcgctcac tgcagctctg ccagggatgg ggctaaaggt ctcacacagg 27000
gcagttaggg ctccccatag cctgggagag gaacggggtg agataacaga aactaggtat 27060
ggtgcccgaa gtcaaacagc cactgagcat gtaaacccag gtgggtctga ccccaaaccc 27120
ctccaccccc atcagccctg caacccgtcg ctgcaaggga gaaagcaact cagaggcctc 27180
acctgcctac atcccccacc cgtgtgtgtg agttctacta aatgcctgag cagtgacaca 27240
gcacggctga aattaaacgg gttccaaaaa cgacaggaag cacgaagtga atctccccag 27300
gaaagtgctg aacaaatgct ggatcgggtt caccggcgaa tttcttggaa ctgaagaggg 27360
gagctaaaca cacggggccc tgctttggag gggactctct cagggtgctc cacacagcac 27420
ttggttaacc ccactcagcc cttctgggct ctcccagagg gcccggcctt ggccttgggc 27480
atctacagga ggaacctcca gggggagagg gggtgcctgg acaggccggc cctggaacaa 27540
gcacttgggc cccgaggaga gaggactagg gcttgggagc tggggaagtt ctcagcactg 27600
ggaccactag aacaaagcca tttccgtgcg ttcacagctt ccaattgcaa caggaagcaa 27660
tcaggaaaaa taattagcgg cccacttact ggcttcgctg aggtccgagg catgtatttc 27720
acacagtaaa accagggata taacatcaaa accgttctgc agaaagattc ctccctttcc 27780
ttccatttta ggcctggatc accacattca ctggggctcc caggccttgc tgcctaatgt 27840
taaaataatc aactctattt ttgcctcaca cacaactgaa ctctacagct ataattcttt 27900
ctcctcaggg gctcgaacca catggacgac aggcatttga ctccagcaac atcaccccaa 27960
aacgtgcaca aaacccaaaa ctgcaatgag gtgaaaggca acgcggtcgg cctagaaacc 28020
ccccctttaa aacaaacagt ttccccaaaa ccccttttgc ctccttgacc caggcatttc 28080
cggaaaaagg agcggcgctg gcctgtactc cccagatact gtcgctgttt tgtcttcacc 28140
ttgttttgct agctccagac aaggccccac aatgtaaaca cgctcctgaa agaggcagat 28200
ttggggtgaa actgtccata gaatctctag gcttgggtca gaggcaggag gacgtgaaac 28260
aaactccaag ctcctcctgt tccccgctgt cccccacacc tccaagcaga ggctgcagcc 28320
tgggggatct gactacaggg ccaccccgct gcaccattca cactggaaat attcagggag 28380
acagctgttt gccttaagga ggcccagaca aaggggcccg aggtcctccc cgctaaactg 28440
ccacaaacag aacaggagcc gcggcgtgca caggcacttg cggccgtgcc acttggccag 28500
ccatactcca gaaaaacaaa acacgcacat ccgaagagaa tgatttaggt agcaagaggc 28560
ttgcttgaaa aaccacatgg caatctccaa attaaaagaa catgtgtagc gtttcacgac 28620
tgcttaagtt tcctgagtcc tcctgacctc aactccaccc cctgggaaac accaaaagtt 28680
ggatgagaaa gttcccccgc cctacctctc cccacgggag tgtacaactg aggcacaagc 28740
ctgcctcccc cactgccccg cgatctggga ccacgtctcc tccgcgtagc cgacccgggg 28800
atggacacta tctggggacc cggcggccac acggggcatt cgggtcgccc gggcacctgg 28860
caggtgtcag tccgcttgga aacccacagc cacgcggctc acaggagcag cgccaccggc 28920
taggccgccc cgcgcccggg ctcagaactt tctcgctgcc acttcagccc gtcctcggag 28980
cacgcggggc ggccgcgcgg ccgctggaaa caggcttgcg aaccggctcc ccgggccagg 29040
cccgcctccg cgccccaagt ccccgctcgg tgcccggccc gggccacacg ggcccagcgc 29100
gggctcggct cggctcccgg cttcccgcgg gctcgggcag gtgaggaccc gcccgcgccg 29160
cacctggcgg agcgggcgcc ctcctcgcca gcccgggacg cagcgtcccc ggggagggcc 29220
cgggtgggga gacaaagggc ccgcgcgtgg cggggacgcc ggggacggca gggggatccc 29280
gggcgcgcgc cccaactcgc tcccaactcg ccaagtcgct tccgagacgg cggcggcgcc 29340

cgcgcacttg gccgcggggc cgcccgggcc attgtccgag caacccgcgg cccgtcttac 29400
acgccgggcg cgggaaggta tcgaatcagg 29430

<210> 8
<211> 33769
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (33739),(33749),(33758)
<223> Identity of nucleotide sequences at the above locations are unknown.

<400> 8
cttcccctta cactggtcct tcgacccgcc tcggatgaaa actgaatggg tttagcctta 60
gaggctctcg gtctctaagg gaggtgggtc aggatgccgg ggacagggtc ctcttcctgg 120
ggcaacgtgg gggaacgagc cacctacccc tccactgaat tgccctgggg tgtgggtacc 180
gacggctcat tcggtgtcca gggtctgaga tgtgttgaca ggaagaatga aaggggatgg 240
gagggatggg gcgaaagaag ccacctgcag ccccaggaac tatctggcca gcacaccgtc 300
acccagcggc ctgagccacc cctgccagag ccaggaggag accctgccaa tgggtcacca 360
gtgtgcagga actcagaagg tcatcacagt taataccctc catgccccaa tgtgggaaaa 420
caggtttttt cacaacaaac aagataattt ttgttatttt ggcaaaagga ggcagggcag 480
ccccggacac ctccatccca cctcatcacc cagccgcagg gccccggcca tccctgcaga 540
cagagtggat gtcacaacct ccctgcaccg aaccaagtgc agctcccagg ccacaggcca 600
cccaggaaag gtccagtggc ccccggaggc tcccaccgca ggcctcccac cacagccggc 660
accaacccag gatagctgtg ttctcctggc ttcttttcac acgggtagca gaaagctgag 720
atccggggaa agctgagatc cagggaaagc tgagaatcgg cctctgctgc ccggacgccc 780
acccccagct ctgctcccag ctccagggcc tccttctcag gtgcccttac aggaggcaga 840
gggcttgagc cacctcctgg gcctggggca cgcaggatga acggggtcac ggtgcaggcc 900
actgtccact gcgcagatcc caaggccata aacagcctgg ccacagtggc ttcccagctg 960
gcaggcggcc agattatttt tgttgtttag caattgatta agtttctccg ctgcccccag 1020
gggtaagtgg tggggcaaat gccgcaaccg cagcatttga cccgggatcc tgtgccaagt 1080
gaccataggg tcacaaagca caagggaagt ggctgggccc gatgctggct ctgctggaac 1140
ctgaggccgg ccactgtcac ctgcacggtg cctgggacct tccagcaagc acagagaagc 1200
tatggccctc caggagcagc tggcaggcac cttggcctgc agtcaggggc tctgtctgct 1260
cagctctaaa acaggaaagt cgctgctctg cctggggtca gggcagccag agagtgacca 1320
agtcagtgcc ggcctcagga agggacctgc aggcgggtcc cttcctctcc catccctcgg 1380
tgccagccag cccctcctgt ggccccccac tgcctgcctc tgcccccatg ccccaccaca 1440
acctcaggcc catggctgca tggccactcc ccaggcaggc agtggggatg ggatttcacc 1500
atgttggcca ggctggtctc gaactcctga cctcaggtga ggagttccta aagtgctggg 1560
attacaggcg tgagccaccg cgccagccct ccctgtggta ctaaacactc acacccccct 1620
gctggggacc ctggtgaggg aacacagcct cacaagtgaa gtgtggtttt gttgagcaaa 1680
tgacgcctgg gcagccctct catctttgcc taaaactgaa gaatttaggg gcgtggatgt 1740
ataaaacagt tggtgactta aatgaaaaag aaggccacac tccccccttt aggcaggcgg 1800
cctaattctt taaaagccag cacagggtgc ctttctgaac ccaggcacac agtaggtgtt 1860
caatggacag cagcggttac ttgtactgct catgacaccc tgtctgtggc ctctgcagct 1920
ggctccagcc tgacgcatgg ctgcgccccc ccgcaaggcc accccggtat acatggaaac 1980
tctgtggaga aggccttggg ggccggccag gacgccaggc ccagatccca tctgcgccct 2040
tcctccatag acctcagcga gctctcggca ccatgtgcct caggcccatt taagaagtag 2100
ggccggccag gcatggtggc tcatgcctgt aatcccagca ctttgggagg cccaaggtgg 2160

gtggatcacg agatggtcag gagatcgaga ccatcctggc taacacggtg aaaccccatc 2220
tctactaaaa atacaaaaaa taagccgagt gtggtggcgg gtgcctatag tccaagctac 2280
tcgggaggct gaggcaggat aatcgcttga gctcagcagg cagaggttgc agttagcgga 2340
gatcgcgcca ttgcactcca gcctaggtga cagagagaga ctctgtctca attaaaaaaa 2400
aaaaaaataa aaaaaagaag cagggccagc cacggacgac ccctcacaca gctcccagga 2460
cgcgtgcctg ggtatagggc tcaggaccat gaccgctgca gtggccccca agaaacgtta 2520
cttttgtcac ccaccccgcc tcagtggcag tagccaaaat aacggattag aatggaacca 2580
tgtgacaatg ccactgcccc aactgacaga agatggctat cagcagttca cgcggcccca 2640
cctatcacaa gtgcagggca ctctacaact tatgcatcct tccccagaca ccgtcctttc 2700
gaccctccca ggtcagcaag gcacacaggg cctacatttc acagccacac agcagagggc 2760
tgaggctgga actcggatgc tctgatttcc gttcaatcac atccccagag gtggcacaga 2820
gacgggggc ttctcttgac aaagtcaaga aagtcactgc cagctccact gaagaccaaa 2880
gaacctcagc tctcaaaccc tcttgaaggt gttaccgaac tctcccagcc tgtttcctgg 2940
gtcccgatgt tggtcccgtg ggacacagga agaggaagaa gctccctaga gcagagcctg 3000
gtgcacctgc cacactctca gagggctgcg cacgggcgga ggagccgtgt gcaggagtgg 3060
ggtctggatg gaggggcgct gtggccgggg gcaggggggca ggggaagggt gctccaggtg 3120
gtgggcacag cacgagcagg ggcagggagg tccacactca gatgtgcaca gggagaaaca 3180
aatcgtgcat ttccattgga ataggcggta aaaggtagaa aaacagagtg ggggccagga 3240
agggagtcgg agccttctag tgtctctctg caggtgagcg gcagcccgag gtgtcagctc 3300
agcagacttg gggtccaggg gccgtgtctt ctatcactga ccccagggca cacggaactg 3360
gggagggaga gcagaggcac agggcacggt cagtgaaacg aaacaaggag tcatcaccaa 3420
atgcggaaag ggcaaggagt gcccgcagcc gcacaagggt tctgtctggg caacgtgggc 3480
gtcccaccag gccccgcacc ctgcaagcgc aaagctcgcc actgaagata aagggaagct 3540
gttggagctg cggagctggt ctggggtccg catggagctg ggcttatgct gcagtcacaa 3600
gggggacatg gaagaggctg caggggacaa aaccagtgac cacagtctaa ctctgagcct 3660
gtggaaaggc gcccacagca ttcacccatc ccagagatgc cattcccccc gtgcccccgc 3720
tccacggtga cagcgttctc caggaatatg atgcgccccc ctcctcttgc atcagccctg 3780
acagtgagta ttcaggccaa aaagcagaag agcacagctg cgtggttcca tttccatgta 3840
gttctggaac aggcaacgct aatccaaggt gatagaagtc aggagagtgg tggaggggggc 3900
gggggttgag gatggcaaag gggcaccggg aactttccca gtggtagaaa tgttctctgt 3960
ctggaccgtg tggtagttat gcagacatat gcagctgtca aagttaatcc aaatgtacac 4020
gttaaaatgt gtgcgtttta ttgcctgcaa gttatacctc aattaaaaaa ataaagttag 4080
cactcaggct tcttccacaa cttcctgaac cgtgtgagct gattttcttg ctattaaaaa 4140
ttcacggtcc atggctgaga acagcagctg ccttctgttt gcaaagtcaa cgccaatcac 4200
tgcccggccg cggcagactc ggccccacag gacctccttt ctttttccc tttgacctac 4260
ttccctgata agtgacaaga cagccagact ctgggaacaa acgcccgtta ttcggccccg 4320
agctgagcgg gccctgcttc ctgagctaat ccgcccggac agacggaggg acgtgagggg 4380
ctttgccgtc ggctccagct gtcagtctgc ccgtcagact cgacagtggc cccctctgtt 4440
cctcccgctg cccccactcc atccccgact tctttttgtt tcctgtccct gacagacgaa 4500
catctgttaa aactctgtct gggtgagctg tggccagcgg cccacaaatc cccaagccgc 4560
accccagcct catctgggcg ctgccgggag cactgcctgg ccacccctctg gacatagctc 4620
tgagagccac cggccagggc acgtgtggcc cgagtggcat ggtgcacgcc gctaagccca 4680
ctgcccaaag gcccccaagc aggagggatg tgcaggagac aaaagtcaaa agaacagggg 4740
cacgttccac agaggatggg gctggagggg tggcagtgag gaacagcagc ttccgaggat 4800
ggcggtggca actcccaaat aaggcctcac tcctgctgtt tttagctcat tccacataat 4860
tggaaaaaca tggcagaaac cgaagccagc tgctgccttg gtcctggggc tgtgtggagg 4920
gggtggggag gccggaggcc caggctctgc actcgactgc tggggatgag agtgactctg 4980
agctgcagag agcagcatcg cagccgccat ggtcccattg agccccggcc acgctgggcg 5040
gcagaggctc gtgggatata cctgccctgt ctcatggggg tcacttcagg aggggcgggg 5100
gagccaggac acagcccagg gctagcggtc accctgcagc tcaggggcca cgtaaatagt 5160

gccaccttga aggcacacag cagtgcgggg ccccccccgc caccaacgca tccctacctc 5220
taggaggccg cctgtgtgcc cctgggaacg ctgctccctg tcccttgggg tcctggtgtg 5280
accaccctct cagccccttc cttggggaag gcacctgact ccctacaccc agctggcttt 5340
catttgctca aaatcaggaa aaagcagaat tcaagacatc acagaaatgt cttcgcctgt 5400
aactccatga aagataaacg gtcagacacc caggagggag tcccagggac ccttgagtct 5460
cacctgaggc tctggcttca aacctcgaga tgtttccagc catgctagcg ccgccccca 5520
caacctgccc cacacagtcc tcccttggga actcacagat ttggccccca cctgccccgt 5580
ttcttctggt ggagtgggtg cgttgggttg gggtggggct ggggactctg gatgtgtctt 5640
aagagtctga gtgattctga cacagccagg ccctgccccc ctcctgacct tcgccccaca 5700
ggaaagggag ccacacgcct gaagcgccca gcacacccca ctccgtcctc cccaggtcac 5760
ccgctggccg tgtgagccgt gctccccact gcccccttcac ccaccccagc tcctcctggc 5820
agcacccagc cttggaagct acttctgatt acaaccgccg aaggaagact cgctccctcg 5880
gcactgaccc agacagcctg caccatcacg ctgctcagca caacccacac agccttcctc 5940
caaaccccat ggagcgggga gtataatcac ccccttttcta ccaacggaca aactgaagca 6000
cagagaggtt aagtcacttt cctaagctcc caacacgatg acaaaaaata gaaggtcagc 6060
ccgcaagtgg aactaggtgc tccaagtccc cggtctgcct gacactgcac ctcctcgccg 6120
ccacggtccc gggtccgcct gacactgcac ctcctcgccg ccacggtccc gggtccgcct 6180
gacactgcac ctcctcgccg ccacggtccc gggtccgcct gacactgcac ctcctcgccg 6240
ccacggtccc gggtccgcct gacactgcac ctcctcgccg ccacggtccc gggtccgcct 6300
gacactgcac ctcctcgccg ccacggtccc gggtccgcct gacactgcac ctcctcgccg 6360
ccacggtccc gggtctgcct gacactgcac ctcctcaaca ccaccacggt cccgggtctg 6420
cctgacactg cacctcctca ccaccaccac agtcccgggt ctgcctgaca ctgcatttcc 6480
tcatcaccac agtcccgggt ctgcctgaca ctgcatttcc tcatcaccac ggtcccgggt 6540
ctgcctgaca ctgcacctcc tcaccgccac ggtcccgggt ctgcctgaca ctgcactttc 6600
tcaacaccac tccttggccg gctcccaact acaaaccaag ccatgtcttc catcctgaat 6660
cctcttggcc taaacatcac tcacaatgcc tccctcggga acaggcacaa gtcccaccag 6720
cacagcctcc ttcgttacct gcgtttccgc tagcccaggg ccagctccag agccctcacc 6780
acagagcctc tatccttcac ccccggacac tggacctcac caacccatag cctggaggag 6840
atccctgtgt gaccccaggg cctcctctgc ccgactctga atttcactgc ccaacgtgac 6900
acctcggaag gctctctggg cactggcagc cctccatggg caccgctcct tctggccagc 6960
tctgacatcc cggctggtga ggtgccctgc acgaggcctc tgcccactgg gacctcacag 7020
ccgtgctgtc agctgcaaca agcgacagaa tttcacgttt tcttcacgtt gcccctgggt 7080
gagcagctcc aggtagtttt cagtcgaggc gaggcgtccc gtcagcagcc aggcggcaca 7140
gctaattcat gcccgccggg cgcacggccg caataccaat gggcacctgc agcctggaaa 7200
gccacagagg aaccgagaac agcgactgtg ctcaggtgac aggactgtgg tcttttaaca 7260
aaacattttc ctttaacgtg atattttacg gcaaggaatg aaacctggag ggcaggacat 7320
ttggatacta aagccccagg ctgccgcgtg gtctgctttg tgaagtctga agcccgcgcc 7380
ccattctggc cccgctcaca ggtccggctc tgactcacca gcttcaatgc taggccgtgc 7440
ctgtcctcca accagaacat gacttcctta aggacaaagc cgtttctcgc ccatccccat 7500
ctccctctgg attaagaaat atgggaagat cttctagaac cacctcaaat ttgcagagag 7560
ccatcctggt gacaaaccct tgaaatgctt ctaagaagag tttaggtttc ttctcaactc 7620
taaaacctct agaaaactct atttccacac cagctgcccc tggaacactt cagcttcaaa 7680
agggcccagg gcagggagac ggaggagcca gcatccacac cgagcaccag cctgttaatt 7740
aacgggaagc gggtggggcc catctccagg cagctctgag gtcagactgg ggaaccatgc 7800
ttacaaaaaa aagtgaactg aaacgctcac gtcctcatgc aaaaccagac tcccagttgc 7860
atctttctgt ctcattgagg agctttttcc tcccttgac agaacaccct acacacggca 7920
tctggaacca aagcagaaag attcaggctc agagtaaaac agtccccaca ctggctgcat 7980
gtggacgttc ccggcccaga gtctcgccca agcagggcct ataaatgaca caaaatgttt 8040
ttctcctgcg tgccagtcat gctccaactg agttatgtgt aaaagtgcct ctcacggctg 8100
agggcaaaaa cagttcccac aagactagag aaaggtgacc cctgacggct gagtctctag 8160

ggagcgtgga gctgcgtgct cagccctgcg gccctgacgg ctctggaatg gaaaagctat 8220
ccaactggaa gggcagggct cgctgctagt ccagcggtcc aaccccacag gtgtctgtgg 8280
tgtcagctcc atgccacaga gcccagggct ggggccagag ccaccaggcc ccctgccagc 8340
ctgcaggggc ctcctcctct gggtagccta accacccccc gtgagcgcag gcagcctcct 8400
ctaatcacca cagggcctgt ccccccctct cccccgcttg caggaaaatg agccctgagg 8460
actccccagg gctgctctgg gcctggacat ggagactggg aattacattt gcagaaggag 8520
cgcaatgccc ttgaagggct cagccacgag cagccagtcc ccagggctca gaaggcccag 8580
ctgttagaac cctgggagcc agcaaagagc caggggctcc acctaagtct atagcccctg 8640
cctcttctgg ttgggaaaga aatcaacgcc cctttactgg ctcccactga cagcccactc 8700
ccccaggtat gggaggattc tgggacgatg caggcaaacc tggaccctga gtgaacctgc 8760
cccagctctc acgggcctgg caccagccac agcacctaag gcgccggtca tggtgacaac 8820
atgaaggtga taagggcatg gacagtggac atggcagctg gacactgggc acccactgga 8880
tgccaggcac ccagcacggc tccgtcaccc ctggatgagc agtggccctt tgcaagccag 8940
ggtagcctgg gcaagttatt tgggggtctc caagcttgtc cagctgtgcg acttcactga 9000
gccatgagtc tgggatttta tcagggccca cacccgttcc tggaactctg atacgtgagg 9060
gagccacaca gggacccttа acaaaagctc ccagggcaac atgttctctt gcctcagtct 9120
cccaaatagc tgggattaca ggcgcacgac taccgcccgg ctaatttttg tatttttagt 9180
agagacaggg tttcaccatg ttggccaggc tggtcttgaa cccctgacct caaatgatcc 9240
ttccactgtt agggcaaggc acctgacagg cacgactgca cgatctgctt gttgggggct 9300
gtgtccattc cccactcctt cgacaaatgt ccacacccag ccttgctttg acaccccaag 9360
aacagagatg gtgacacctg cttcctacat gcccattgct ctcccaaggc agacatcccc 9420
agcagatgca acacagtgtt taggcagaca tcaccaatcg atggtggcaa cagacaccag 9480
gccctgctcc ctctaactcc agtggccagg ccccaagcca gctctcacct gcccactccc 9540
aacccacagc agcaagactc agaaatggca aaaacacaaa gagaacagaa acgccccata 9600
gcgggaggat gactaaaaga catgtcttga taagatattg ttcaggcata ggccaggcac 9660
agtggctcat gcctgtgatc ctagaacttt aggaggctga ggtaggtgga tcacctgagg 9720
ttaggagttc aagaccagcc tagccaacat ggtgaaaccc catctctact aaacatacaa 9780
aaattagcca gacatagtag cgggcgcctg taatcccagc tgcttgggag gctgaggcag 9840
gagaattgct tgaacctggg aggtggaagc tgctgtgagc cactgtactc caacctggac 9900
aacagagcaa gactctgtct caaaaaaaaa aaaaaaaaaa gatatccttc actaaaactc 9960
atgtctttga tacatattta cctcctgcaa tcgcaaatgc ttctgcagtg cataaagtga 10020
aataaatagc aggaagcctt acggttcgat cacccacaca gacacacagt cacatacagg 10080
aaaaacgcag ggagggctgg ggaacaaaaa aacagaagat aaaatgtgga gacagacaca 10140
ccaagagagt aagagaccac ctccagacct cccttcagct tctcaaacac acgagccggg 10200
cccgttacag aatttgcggg gaccgctgca aaatggaagt gcagacagcc ccttactcaa 10260
aaggtaggaa tttcaggtca acaacagagc tcacctcata tgactacaca ggtcacacag 10320
cccgtgaagt cggtcccaac accagcatgc tcctgcctca aagccgctgc acgtgctgtt 10380
ccttctcgcc tttccctctt ttagtccttc agatctcagg cctcctgaga gagacctctg 10440
acctgccggc tcaggcggcc acacccccag tacaggagtc tccggctcag cccctgctgt 10500
gttccgtacc cgatccaggt ctgtcctatg tccatctgtg tgccggcttg cttcctgaca 10560
tggcccccac cacacgtgtg cctcggggca ggggaacagg cccgtctcat taactgcttt 10620
cttctcagat attttctgga atatttgtgg atattgggca acatatatgc tccacctttt 10680
tcagactagc caggacgagc tgcatttttt ttttttttt tttgagacag ggtctcactc 10740
tgttgcccag gctggagtat agcggcatga tcttggctca gtgcaacctc cgcctcctag 10800
gctcaagcaa ttctcctgcc tcagtctccc aagtagctgg gattacaggc ccgtgccact 10860
actgcccagc taatttttat atttttagta gagatggagt ttcaccatgt tggccaggct 10920
ggtcttgaac tcctgacctc aaatgatcca cctgccttgg actcccaaat tgttgggatt 10980
acaggcgtga gccactgcgc ccggcccgag ctgcctgttt tacacctttg ccatattccg 11040
gtgattctct ctcccctccg tcccccggcc ctgactgtgg tggccactcc ctgccgtcat 11100
gagcccgtat gtcctcactc tttccctttc cgccaggact tcaaccaaca ctgcagagcg 11160

cagggtccag ctccagcact gagttcagcc tcttctcacc aacagacagg caggaaagaa 11220
aacaaactct gagaaggcca aggttcccgg gcagccagca agccaagcat ccttctccgc 11280
tgaggcttgt gcagccgagg caccccctcc tccagggagc aggcagcgtc ctggggcagt 11340
ctgcgaggga gaccagggcc cttgctccac cagggcccca ggtatggggg cagcagcaaa 11400
ctcatggctc tgggagccag accccacctg ctagaaccta ctatgccacc tgctgtgggc 11460
aaccccaggc tggtgacttg ccctggcctc ctctgtaaac aaagggctca tccaacctgg 11520
tcaaaccact cctccccttc aagggtctat aatcctccct taacctgctt ggtccaaacc 11580
cctggtgtcg ccaggtcact caggaggcag ctcatctgga ctccttccct gggtccagtt 11640
tctctctcaa cattgccttt gaggccgagg tgaacggtca acagcgaagg gccccagagg 11700
tgatggagga gcgggtgtcc aagacactca ccctttctaa tgcactgact ccctcgtgga 11760
ctcacttgtg ccgtctcccc cacccaccca gccccagagc ccagagtgcg agcgccagag 11820
gcccgggatt ctgtctgcac cgcggggtcc ccagtgcctc ggagcaatgc cagcacccgg 11880
caagtgttcg acaaatgcct gctgaatgag caaatggatg gatgaacgaa tgaatgagca 11940
agcagatgaa tgaatggggt gctgtccaga gccgtgagga ctaggccgcc caagtcccca 12000
tttctcaaat tctccttctc ccgacttggg aaacaagatg cttggtcggg gaggctctcc 12060
aaccatcccc tgcagcagcc ggcacagcgg acagaccctt tgatgtaaca gccatgtctt 12120
cattaaagat gccctgctct cagaaagaga aagacaaata caaacctgga aaatcctcac 12180
caaacgcagg acccctgcca gggagcagag aaaagaccca cacgccacgg gcgccacgac 12240
cacacacaca ccccagccgc tgcacacaaa cacagaccct agccagcaag aacaggggga 12300
ccaggaaact gttcctaaag tcaggacccc catgtgctca gacagcagtg agagcaagga 12360
cacttctcca tccaccggat gccaggagag tccttttagg gggccccaca ccgagactct 12420
gcccttagga ctgttcctga gtgtggaagc cagcccactt ggaagccccc tgccctcccg 12480
agtgggacac cggcacagga agcaggccct gtcccccacc actttctgca agctgggccc 12540
catcacgcta cagaaacggg gaggactggt cccagggatg gcgctttcct gacacctctc 12600
gttacccccc cgcttgccag gccccagggt cagccccaga ggccagactg gctatcccag 12660
gcccgggagc atccccgaag gcgagctgca tcctgaacgt gtgtgatttc ccgaagggcc 12720
cgccccgaac cgacacctgg aaagaaagat cctcagccgg tgccccagag gagaagagcc 12780
atgcctcact gcaacacagt cccaggaagc accaagtgcc tgaggaccaa ggcggagagt 12840
aaaaaagtgg aaaatatctg gggcaaaaat aaaacaaaac aaaacaggat tgacctcctg 12900
ggctcaagca atcctcccaa ctcagcttcc cgagtagctg ggaccacaga cttgaatcac 12960
cacacccgcc aagtggatca tttcgaacgg gtttgccgag gttccttctg gggcaccccc 13020
ggcggccgca acccattccc gccaggcccc gccccgcccg cccgccccgt cccgtcccac 13080
cgcctcacct gccttacacg tcctgccgtt gtcctgcagc tgcacacccg tggggcaggc 13140
gcatgtgtag aaaggctcgc ttggggacag caggcacagg tgggagcagc cgccattgtc 13200
ctcctcacag cgagtgtgga ctgagaaaac caggacagac tgagagaagg ttccagaaga 13260
ggaccgtcac ttgtttctga atgagtcaca tcctgcctcg tcccccgtga cagcctccag 13320
tgtgtccctc tgcccaaaca tcggcctcaa gtggcatcag ggacctcccc gcgggcacca 13380
ttccacctgc ctcatcgctg gccccgtcca catggggccc tcagcctggc cagacggcct 13440
gcaatttccc caaaaccagc cgtgaccttc ctggccaccc tcacacccag atgtgacctg 13500
cccatggagt gacatcctcc ccatctgctt cctcccacca agctcctatg actagaacac 13560
cctccccagc tcctcggagc ccccaaagga cacccctctg caaaggctgc cccccacgct 13620
ccaatggccg gggtcaggac ctgcctgtgt ggtagtgacg ggaaccccag agacaatggg 13680
ctcctgggca aaaggcttgt cttgtctttg tgctatgtgt ggacccagca gcttccatag 13740
gaacactgtc cttcttgctg ggatggccaa gcttgtcact ctcccaagcc ctcctatgac 13800
caacagcaat tgaacggaac tcgataaatg cttccagcac ctcattcaaa ccaggggaaa 13860
gctgggtgta gcagccccaa aatacggata taactggaac aacaaactca tcaaaatgaa 13920
cctctccctc cctcatgctg ccccaagtgt agatgggttt tgtgaccacg actttctcac 13980
caggaaacag ctccagagag ccccaccctc ctgtgtcctg ctctgggaac agctggcacc 14040
cctaggcccc acatttcaat tcaaagtcca aaccttccat aatggcctgg ccagaaatct 14100
ccatccctgg tccctgtggg agtgggccac tgtccccaga gccgcagccc cactgtcaca 14160

gaagctggtg catttcccca tcagggacct ctgtcacaac ccagcgtggc ccccaggctg 14220
agaactgctg attctgggca gattattcat tgataaatac gcgacttgca gggccaagca 14280
tggtggctca tacctgtgac cccagcactt tgggaagtca aggtgtgagg atcactggag 14340
cccacgagtt tgagacaagc ctgggcaacg tggcaaaatc tctcatctct attaaaaata 14400
catacacaca cacacacaca cacacacaca cacatatata tgtatatata aataaccata 14460
tatatatata cacacatacg tgtatgtgta tataaataca tatacacaca cacacagaca 14520
acttcttctg ggccttgaaa acgaggcaac cttccttgga aatccccttg ccactgctga 14580
gcctgaaata gcccccatga gctctgcaga ggggtcctct gcaggcccgt gtcccccagc 14640
cagccacaca cctccctcca ttgcagcagg tacccccttta gagagggggc cccccagagc 14700
atgggcttct gcagggaggg gtcacctgcc cccccacccc acccacgccc gcgcaccccc 14760
acgccccccgc atcctcccac tcccctgccc cgcgcccccg ctcccccccag ccccctcacc 14820
ctctcccccg tgccccaacc ggcactcaca aaaaggctgc cgctcctggc tcagcacctg 14880
gatgtccatg ggtgagtata gggcactcag gatctccttc ctcttccccc cagtgcgctt 14940
gttgcaggca tggatggagc gggtctgcca gtctgtccag tacagagtgt ccccggagag 15000
cgtcagggcg aaggggtgcg tcaggctgcc ctccaccacc ttctgcctgc agtcagggaa 15060
gcggggtgga ggagccatca ggagggtccc ccgacagtca ttgctgctga cccaattaat 15120
ttctttttttt ttttttgaga tggagtctcg gtctgtcgcc caggctggag tgcagtgatg 15180
taatctcagc tcactgcaac ctccgcctcc cgggttcaag caattatcct gcctcagcct 15240
cccgagtagc tgggatcact gatgcccacc actacgccca gatgattttt gtatttttag 15300
tagagacagg gtttcatcat gttggcaagg ctggtctcga actcctgacc tcaggtgatc 15360
cacccacctc agcctctcaa agcgctggga ttacaggcgt gcgccaccat gccaggcttc 15420
ccatttgctt tcaaccagac aagtgaggcc aggtcaagag ccccaggagc tggcgccctc 15480
gtacatttct cccggcgtgc acagggcacc tcccaaacac agcctgtgat ggtgacacac 15540
gggctccccc aggtcaagtg gcaaagtctc ccccagggaa gaaaggagga agccatgcct 15600
ggcaaaaagc acacctctcc tgcccaacgc tttaacctct gtatacaaat caggccatgt 15660
gcactcgctc cttcttacaa tgctcataat ttatactttc agagtaaatg aaacttggca 15720
tcaacccgag aaacagctat tcttttctag atgcttacag tgcccagcaa atgaggactc 15780
gggtgtaatg agattatgga cactggaaac aggatcataa tgtgacgtgg tcggtaatgt 15840
gcagttttat ttgcttaatg accctcgccc cgtgacaggc tccctgaggg tgggcctggg 15900
ggcagaggtc cccgccacgt ccccagcccc cagcacagtt gccaggagag ggtgacactc 15960
atgaagtggc acagggaaga tgggagctgt gggctctgca gatccaccac ctcttctgtt 16020
catttttgtt gatgctgttt tttaagaaaa ttattgaagt aaaattcaca ggacatacgt 16080
ttacttttttt ttttttttttt ggagatgggg tctcactctg tcacccaggt tggagtgcag 16140
tggtgtgatc tcagctcact gcaacctctg cctcccaggt tcaagcgatt ctcccacctc 16200
cgcctccaga gtagctggga ccacaggcgt gcaccaccac acccagctaa tttttgggggg 16260
gtatcttttt ggtagagaca gggtttcgcc atgttgccca aggctggtct tgaagccctg 16320
agctcaggcg atccacccgc cttggcctct caaagtgctg ggattacagg cataagccac 16380
tgcacccagc ctaaatttac cactttaaag tgaatagtgt tacctagtgc attcgcaagg 16440
cggtgcagcc tccacttctg tctagttcca aagcacttcc attgccccac aggcaaaccc 16500
cacacccggc agcagtcatg ccccagtccc cgcccccagc cccggcaaac acttttgatg 16560
gacttaacta cacacattct caacatctca tataaacgga atcacaatat acagcctctg 16620
atgtctgtct tctttgactt ggcaccatgt tttcgaggtt catccaggct gtagcatgtc 16680
agtgcttcat cccgttttag gggtgaacca tattccagtg tgcagacaga aaccaatctg 16740
tgcatccatt cacccactgg gggacctttg tgtcatttcc accctcggct gttgtgcaca 16800
gtgctgctac ggacattact gtccattcac attttgtgtg aagacctgtt ttcgattctt 16860
aagagtatac agctaggagc ggaattgctg ggtcatacgt aaatcaatgt ttacgtctca 16920
aggaatcaac aaactgtttt ccacaatgtt gtctttttttg tttgttttct gagacagggt 16980
cttgctctgt cacccaggct ggagtgcggt ggtgtgatca tggctcactg cagcctcaat 17040
ctcctaagct caatccatcc tcctgcctca gcctcctgag tagctgggaa cacaggtatg 17100
taccaccatg gccagctaat tttctaattt tatttttttt tgtttttgtt ttttgagac 17160

agagtctcgc tctgtcgccc aggctggagt gcagtggtgc catctcagct cactgcaagc 17220
tctgcctccc gggttcacac cattctcctg cctcagcctc ccgagtggct gggactatag 17280
tcaccggcca ccacgcctgg ctaatttttt tgtatttta gtagagatgg ggtttcaccg 17340
tgttacccag gatggtctcg atctcctaac ttcatgatcc acctgccttg gcctcccaaa 17400
gttctgggat tacaggcgtg agccaccacg cccgacctta cttttaattt tttaatttta 17460
ttattttatt ttatttttt ttttttgag acagagtctc gctctgtagc ccaggctgga 17520
gtgcagtggc gggatctcag ctcactgcaa gctccacctc ccaggttcac gccattctcc 17580
tgcctcagcc tcccgagtag ctgggactac aggtgcccac cacgatgccc ggctaatttt 17640
ttgtattttt agtagagaca gggtttcact gtgttagcca ggatgatctc aatctcctga 17700
cctcgtgatc cgcccgtctc agcctcccaa agtgctggga ttacaggcgt gagccaccgc 17760
gcccagcctt ttttttttt ttttttttt ttttgagata gagtcttgct ctgtcgccca 17820
ggctggagtg cagtggcggg atctcagctc actgcaagct ccgcctccca ggttcacgcc 17880
attctcctgc ctcagcctcc cgagtagctg ggactacagg cacccaccac cacacctggc 17940
taatgttttg tatttttagt agagacgagg tttcaccgtg ttagccagga tggtctcgat 18000
ctcctgacct cgtaatccgc ccgcctcggc ctcccaaagt gctgggatta cacgcgtaag 18060
ccatggcgcc cagcccatgt ggccattttt cagtgagaga agccagaggc ccatcactct 18120
cggttgctcc ctgggccatg ctctgcctca gccagaagca ctgagggaag gtcagcctcg 18180
gcccttgccc cagccacagt cacagataaa ggggcctgca caggtctgtg tggctccaga 18240
gctcgtcacc caacacacga cgcttccatg tgaatagccc caggtgcatc atgaagagcg 18300
atggccgctg cagaggcaga agaatcccgc ggggaagcag gtgggagaga ggctgagaac 18360
agaccagacc ctggagctac agaccctatg ttccaaccct ggctgggact agctgtgtgg 18420
ctctgggcaa attcacatgc ttctctgtgc acaggggatc aaaatagcaa acacaggcta 18480
ggcacagtgg ttcacaccta taatcccagt gctttgagag gccgaggtgg acacatggct 18540
taagctcagg agtttgagac cagcctgggc aacatggtga aacctcgtct ctacaaaaaa 18600
aataccaaat aaattagcca ggcgtggtgg tacgtgcctg tggtctcagc tacttggaag 18660
gctgaggcgg gaggaacact tgagcccaag aagtcaaggc tgtggccgcg tgtggtggct 18720
cacgcctgta atcccagcac tttgagaggc tcaggtgggt ggatcacttg tgatcaggag 18780
ttcaagacca gcctggccaa catggtgaaa ccccgtccct actaaaaaaa tacaacaatt 18840
tgccaggcgt ggtggcgggc acctgtaatc ccagctactt gggaggctga ggcaggagaa 18900
tagttagaac ttgggaggtg gaggttgtag ttagccaaga tggtgccgct gcactccagc 18960
caggggggaca gagcaagact ccatcccaaa aaaaaaaaaa acaaacaaac aaacaaaaaa 19020
agaggtcaag gctgcagtga accatgattg tgccaatgca ctccagcctg ggtgacaaag 19080
tgagaccctg cctcaaaaca ataaaaatat aaataaaaat aaaacataat agcaaacgtt 19140
tcatagaggt ggtatgagca ttaaatgaac tgataaacgt ccctggaaaa cagtaagtgc 19200
tatggaagga ttcgctgccg ccaccgccac caccattagc atgtttcaac ctccatcacc 19260
ctcactgtcc cctgtcacca tcctttgacc agggcactcc cagctgcagc ctttctatcc 19320
tcttgtccac ccttcataac tgtaagatca ctcagctccc aagaaccaca gtctacaggg 19380
taaccacatt tccaaatctc aaaccagacc cgctggtctg cacttccagg gacaacagga 19440
tattttcaaa ccagcccaaa agagatgtgt ggctcagcat aagaggaaca ggagaaactg 19500
aggcctcttg ccctgagaat gagcttggaa gtggatgtcc cggcctcact caaaccttca 19560
gatgactgag gcccagccag gagcttgagt gtaccctcag gtcataccct gagccagaag 19620
cacccagcta atccactcct catcactgac tccctcccca taaaaaacct gtttgctgtt 19680
tcaggctgtt aagttgtggg ctgttttgtt acacagcaat ggataactaa cacacgaggc 19740
ctggcaagtg tggagcaaag ctgcccaagc cctcaagtct gttcatgtgg gtgttggcct 19800
gtgtttgcag aaatccagcc actgagtcct cccatgcagt cactactgcc ctctgcacag 19860
acacctgcca catccctgcc tgggccagga gctccactag tgcaggaatg gggtctgccg 19920
tcccaggagg atccctgaca cctagcacag ggctagcagc aggcagcact tggttagtga 19980
ataaactgcc cttcacctgt acacagaagg gatgtttcta taaggggtaa ttaagtacag 20040
agctgggaag ctatgctgac cagaaggctc taaaagcaat taaccaacga ggggaaaacc 20100
cttcctactc attctcggcc cattttattg agcactgacc atgtggaagg ccccctggtg 20160

agactgggga atgcaccaat aactgagaca gcttccggct gttgccctca ggatgcctga 20220
gctgggatag ggccagggtg ggggtggtgc gtgtgacagg gttactgttc acaaccctgc 20280
cgggccataa gccctcccca acaattccaa aatccaaaac gctctgaaga tggaaagctt 20340
ttgttgctca tctggtgaca aaacctcatt tggtgcatgg gccgggtgcg gtggctcacg 20400
cctgtaatcc cagcactctg ggagccgagg ggaaggatcc cttgagctta ggagtttgag 20460
accagcctga gcaacatgtg agaccccgtc tctaccaaaa atacaaaaat tagccaggtg 20520
tggtggcgca ctcctgtagt cccagctact cgggaggctg aggcgggagg atcgcttgag 20580
cctgggaggt gggggctgca gtgagctgag attatgacat tgcactccag cctgggtgaa 20640
agagtgagac tctgtctcaa aaaaacaaag ttaaaaaaaa aaaaactgtg catgggtgtg 20700
ggctacagat agtcttttct gccctactta gaatgaacgt gccacatttg ctatagaaat 20760
attcaagggc tggtggcaaa tgccacacag accctgacgc tgttccaagt tctgagaagt 20820
cctgcattcc tcagggcccc agagtttcag agaagagtct gtaggcctga gttaagaagg 20880
aacgccttca aaagccctgg ggacaaaggg gaaaggggtg ccccaggact gcgtgggtac 20940
ctaccggaac gagccgtcca ggttggcacg gtggatgaag ctgagcttgg cgtcagccca 21000
gtagagcttc tgctcctcca ggtcgatggt cagtccattg ggccagtaaa tgtccgagtc 21060
cacaatgatc ttccgggtgc tgccatccat ccctgcccgc tcaatccggg gcgtctcacc 21120
ccagtctgtc cagtacatgt acctgtgacg ggggcagggc aagagaagca gctaacacag 21180
atctgttttt tgtttttgtc tgcatagatg cagacatgaa acaacagaca gtgaacttgc 21240
cctaaaatct cacccatcgg aaataaccaa caggtatggt ttcaggtatt cctgccttaa 21300
gctgggcaat caaaatatac tatttccaac ttgttctcag ttaacagtaa attctgggca 21360
ccttcccttc ttgtggatag aaagattcct tgttcttttg atgattgcct agtgtactct 21420
gctgtaagtt ttttaaagaa cttcaggtta tttctgattt ttttgctacc atgaaaatgc 21480
tgtaaatgaa cctctaaaag gcaattcaaa acactcagga tggaatatta tttagtggta 21540
taaagaaatg agctatcggc tgggcccagt ggctcacacc tctaatccca gcactttggg 21600
aggccaaggc gggtggatca cgaggtcggg agatcaagac catcctggct aacacagtga 21660
aaccccgtcc ctactaaaaa tacaaaaacat tagccaggcg tggtagtgag cacctgtagt 21720
cccagctact taggaggctg aggcaggaga atcatttgaa cccgggaggg ggaggttgca 21780
gtgagcagaa atcgcaccat tgcactccat cctgggcgac agagcgagac tccatctcaa 21840
aaaaaaaaaa aagaaaagaa aagaaatgat ctatcaagcc atgaaaagac atggaggaaa 21900
cttaaatgca tgttagtagg tgaaagagcc aatctgtatg agtccagttc taaacactct 21960
ggaaaaagca aatacacaga gacagtaaag catcagtggt tgccaggagt tggagaggag 22020
agggatgaat gagtggagca cagaaaatca gggcagtgga actatcctgt atgacatgga 22080
atggtgggtg catgtcctta ctcatctgtc taaaccaaga atgtacaaat caagggcgaa 22140
ccctcgtgta aacgtggatt ttgggtgatg gtgcgtcagc cagctttcat cagttgtaac 22200
aaatgtacca ccctgcacag gatgctgaca gttgggaagg ctgtgtgggt gtgaggacag 22260
ggatgtatag gaactcagta cctgctgctc atcaattttg ctgtgaacct acaactgttt 22320
gaaaaaatta agtctattta aaaacaacaa aacatggcca ggcacgatgg cttgcacctg 22380
taattccagt acttcgggag gctgaggtgg gtgggtcact tgagccaccc tgggcaacat 22440
ggcaaaatcc cacctctaca aaaaataaaa attaaaaaaa agttagctgg gcatggtggc 22500
acactcttgt agtcccagct acttgggagg ctgacgtggg aggatccctt cagccctggg 22560
aggtcgaggc tgcagtgagc tgtgactgta ccactgcact ccagcctgga tgacagagtg 22620
agaccctgcc taaaaaaaaa aaaaaaaagg ctgggtgcgg tggctcatgc ctgtaattcc 22680
agcgctttgg gaggccgaga tgggcggatc acgaggtcag gagatcgaga ccatcctggc 22740
taacacggtg aaaccccgtc tctactaaaa gtacaaaaaa aaaaattagc cgggcatggt 22800
ggcggacacc tgtagtcaca gctactcggg aggctgaggc aggagaatgg cgtgaacccg 22860
ggaggcggag cttgcagtga gccaagatca caccactgca ctctcagcct gggagacagc 22920
aacactccgt ctcaaaaaaa aaagaataaa acccatggct gggatggacc ctgaacctgc 22980
agctgcagct gttcctgggt aggtctgtgg gcgacgtggc tttgcttctc catgttccca 23040
agagacaagc atcacccatc catgagaaac aagcacatcc tcagggcgcc cttacgtgat 23100
ctctggccaa tgaaccaaga caaagtgagc agacaccagg tctgggatgg caggtcccac 23160

ccccaccagt gcccagtgtg ccctgtttgg aggtgaccac agggtgtgtg cccagaggct 23220
gggcgtgact ctcagcggag accagagggg aaccacacca gcttggagga ctcagttccc 23280
atcccagcca gctgggatga gccacaggac acaagggctg gcagacctat tgtgttttgt 23340
ccacccttca cagcagagaa aggggacagt gcccagaatg tcctctgagg agcctcctcc 23400
cactcttggt ccttgtaaaa tggtgctgac tcccttgctc ccttcttcct ggggtgggcg 23460
gcaaacccca ttcccctcag ccttagcaag tgatttagaa acaggcagct cgcccaagcc 23520
aggcatgaga gtgatcccgg gacacaggga gaacaagccc cgctttgccc tctgggggtc 23580
tccattcagc agaagaggca aatgacagac acacagccgc ctcctccccc accatggtgc 23640
tctgcagcct caggagcctc aggtgcacca agggccaccc catccagggg gccatgcttc 23700
cttgagtggt atcgttcctg agcgagtacc atctccacct tccagagggg ctgtgacaag 23760
atcaacaaga atgagggcat aggagcctcg aaccaaacat gccctcttcc ctgcagaggc 23820
tgactgcgcc cagctgctat caccaagccc ctgctcctcc ggccccgtgg ggacagggta 23880
agaggggtgt cacatggaac agctctccaa acagtccctc tcaagctgct gtctcctgtg 23940
catctagtga gaacccaacc aacaaaggga aggtgggaat tgctattccc attaggcaga 24000
tgagaaaact gaggccccga aaggctggcc tgttccaggt tacaggcgct gagcggctgc 24060
tctgggaaca cacttggtgt ctgctgaggg cccgagcccg gccatcatat gactcaccct 24120
tcgccagcaa agcccgggtg tgggtgaact tttcctggca gcctgggact ccaaggtgct 24180
ggcagccagc ccagggaagg ctcccgcgtg cctgcggcag acgccttgct ttacctgcac 24240
gtccccaccc ctaggagcct ggacagagcc cagaccctcc gccacctcct gagaaggtat 24300
caggggcatc agtctggact tgggggggaa tccacacagg ccttccccaa atgctccacc 24360
gtggcccatg gaaaaggctg gaaaacgtgc aggagcagga gcctccgcat ggagcataat 24420
tcacattcct tccccgagtt tcataacaga ggcctgctgg tttccttaaa tggggaattt 24480
gcgagccagt cggtgaccag agactggttg gcgtggacgt gctcttgcag agtctcaaac 24540
gctaccacaa gcccagccaa attccacgga ggaaaatcga cttccgaaga aaagagctgc 24600
agcatggcct tcgtgcagag ccagctgcgg ttgtggttgt gtgttatttt agggaagggc 24660
cattttgcat tttaaagagg gggttgggtt tcaccctggc tttaatttga gacccggggg 24720
ccactgcagc cccttgtcag gctggtacag gccggggact cctcccatgc taagccagtg 24780
tctttctggc cccagatcct caggggccag agggtcatcc ccagagcccg ctctgccacc 24840
cacatgggta ccctgggcct gggagggatg tgccttccct caaccctgcc tggatgtccg 24900
cacggggcca cctgcattgc tgaaactgca acgaagtcga gtctcaggag gggccccccc 24960
ggctgcaggg ctcttgatcc ttttggccac gtgcacactg aggtggacgc tcggacccag 25020
agacccccct catgatgatg gccggggcag gaacccccctc ctctgaggaa ggaccctggt 25080
gggggacagc actgcaggag ggcacaggag atgacggggg ctctagcagg gccgggagga 25140
aggccaagat gctcctcgca accgtgtgcc tgtggccagg acagaggaca aacccaccct 25200
ccactgtccc cactctcagg acagcagtcc tgccccagga ctcagcgccc acacttatgc 25260
ctgaggacca ctattcaagt cagtatttgg cgagcagggg ttgctgccgc gggcgctgtg 25320
acaggctgga atcctctccc tctccctctc cctctccgga gacatggagc ctacagggac 25380
agagtcagca cctcagggta ggaccatggc tggcgtcatc agcatcactg gatctgatga 25440
gtgggagccg gcatctcact gttttcactc tctcattcaa atgactggag caaagggaag 25500
gtgtggggag aggcccagga atcaacacta aggtcaactt tgcccccagg ggcaggggtg 25560
ggagtgaaca gccacaggtg tgatcctggg gagggcttct gggagagaat tcagaggcaa 25620
gcatgtagag gaaccatttc aaatagttaa gaaaagccag agccaaacag ggacagttgg 25680
ctcgcagaga tgatgcaggc aaagccagct cagatctgag catgggaaag actactccca 25740
accaagggcc cagcatctcc caaccaagca ccaagtacct cccaaccaaa tgccaagcac 25800
ctcccaatca aatacctccc aaccaagcac ctagcacctc tcaactggac accaactact 25860
cccaaccagg caccaagtac ctcccaacca agtgccaagc acctcccaac caagtaccaa 25920
ttacctccca accaagcgcc tagcacctcc caactgagca tcatgcacct cccaacagag 25980
catctagcac ctcccaactg atcacctccc aacctagcac cgagcacctc ccaaccaagt 26040
gcagagcacc tcccaaccaa gtgccaagca cctcccaatc aaatacctcc caaccaagca 26100
cctagcacct ctcaactgga caccaacaac tcccaaccaa gcgccaagca cctcctaaca 26160

aagtaccaat caccttccaa ccgagcacct agcacctccc aactgagcat catgcacctc 26220
ccaacaaatc acctagcacc tcccgactga tcacctccca acctagcact gagcacttcc 26280
caaccaacat agcaaaagcc ataaagaagt aaaaagacaa aaccacgtag gcatggagac 26340
tggacttctg gtggcgagga aagggcattt ttattataac gacagctaac atttgttgaa 26400
ctcacaaact gttcttggtg ttttcctcat gacatgcagc atggtcacgc ctctgtacag 26460
acaaggatac tgaggcacag agtggcaccg tgccaacctt gtctcatctt tttatcgaac 26520
ctacatgcag agtgccagca aatccagctg tcttttctct tcagaacaga tcccaaatct 26580
cgccactcct tacccccaca agtgaggtgt ccccgctgct gctttctgtc gccaggatcc 26640
cggtaataac cgtggagagg gctcctgccc ccacgccacc caccccacag ctcactctcg 26700
ctccagccac caggggatgc cttccagcac gagtcagagc tggcacctcc tctgctcgag 26760
acctcatgtg tcctctcctc acaccttggg ccctgtttcc ctacattctg ctacagcccc 26820
tcaaacaggc cccgccccaa accagcccag ggcctttgca ctggctgatc cctctgcctg 26880
gaccgcgctg cccccagaca gccacacggt tctcagcctc atctgcttcc agtctcgact 26940
caaaagtcac caagaggcct tcccagcacc tgagctccga cggaagcccc tcgccacagc 27000
acccaagcac tgctttatcc ccctacgcac acgtcccttt caaatactat tcatttacca 27060
tctcctccca ctcactgaaa gggccagaga ctgggctata cccgctgcgt ggggagcagg 27120
accaggcgca agggctcaca aatgcagtgg atgcctggtt gggaggtgag ggagctgcag 27180
cgacccacgc tgggagggaa cgcaatgaca ggaggagcgc aggtcctggc gacacgatgg 27240
ccatggcagc cgctggtgag caaccgcagg ccggccctgg gagagggctt ctagcaagct 27300
gctatcttca gcctctccga ctactgcaga tgccccctcc tagccagaga cactgctaca 27360
ccagccgacc cttccaaaaa gaaggtcagt aaccccgcga ctcctggagc cacagtgcag 27420
ggggagaggg ctgagagggc aacagttcac caagcggaac agaggctgcc ccggaggtca 27480
gctggctccc cggcagctgc aggggtggct agcccactcg gagggcagcg agggcatacg 27540
aggggctcca gggatgagtg gttgcccagc acagcacccc tgggaggccg ggggcacttc 27600
tcaggtagtg ggggcacgag gctgctctgg cctgacctca gggactcaaa atactttggc 27660
gataaattcc accgtgtccc acccctgctg gtaccccata cttacacaca gactggttca 27720
gatgcagaca ctctcgcgca catactcgct cacacgggca catacacgtg cacacacagt 27780
cacatgcgca cactcataca cacacaaata tccactcaca cgcatgcatg cacacacacg 27840
gacacacaca ggctcacacg tatgcacgca tatgcgtgca cacgcacaca cacacacaca 27900
cgctcacatc ctcccactcc cacactcagt tgctcagaca cacacacgcc tggctctcac 27960
acaaacctgt tgggctctga aaggctccag cccttcccat gctcgtcaga agccagtcaa 28020
tggcttccta agtcaccaca cagatcaaag aggtgaactt ggccacatgg cactctgctt 28080
cctgagctcc caaacaccag ccttggtgag gacagaccct cacccacac cctcattccc 28140
actaccctgg gcaggcccag aggaggggca tctgcaggat ctggcaacca gcccctcccg 28200
cccggctcct gcagccggca ccatgggagt caggggggagg tcactgcaaa gggcaacagc 28260
aagttggtgg ccccaggact agagcccagg ggtcttcagt cctactccag agcttggaca 28320
ctgtcccaca gggcatggcc aagggaaggg cttccagagc cctgacttca gggaggaggg 28380
caggcgggct cctgtggcag gcctggatgc atggccgccc actcctggga ctttctaacc 28440
tagaatatct aggtcaggct gggtgcagtg gctcacgcct gcaatcccaa cactttggga 28500
ggccgaggag ggtggatcac ttgaggttag gagtttgaga ccagcctggc caacatggcg 28560
aaaccctgtg tctactaaaa atacaaaacc tagccaggtg tggtagtgca cgcctgtaat 28620
cacagctact caggaggctg aggcaggaga atcacttgaa ctcgggaggt ggaggttgca 28680
gtgagctgag atcgtgccat tgcgcaaaga agatctaggc cggcccctca accggtgagg 28740
tccaggctgg gagtgctgag agactgtggt gacactgaat gaactaacag gcaaagggct 28800
tccaactgag cctgggggtg gtgggaaatg gctcttgtgt tctagtcaag acctctgcca 28860
accagttctg acactgaccc agcacagaac ctgacaggtc agcaagggcc agggcttagc 28920
acagcccagg taagggtgtg tgtacggccc ccagagtcac tcccaggctg caagaaaagg 28980
gacaaaggag ggacaagggg tggccaagca aactgttccc tctgctcggg agtctggggt 29040
gacctggcct agctggccag tggagctggg ccacctcccc ttaaactctc caccccggac 29100
ttcgactcca aagctttcct gccacccacg ctctccccac ctgggatcac ggccaggccc 29160

tgagccttca agggcccagg tgaactcagc cagactagga gctgaggagg acacagggca 29220
gcttccagaa cggacccgag aaccactccc agcaggttct gcttccagac aaggagctgc 29280
actttttcag ccaatgcaat tagaaagcca ggagaaggtg caaattccac ctgcctgagc 29340
gtccgcactt cccaggccgc ccaccataca cacagcaaag atgtgtttaa ccattcaaac 29400
ccatggccaa ccacatcggt tgcctcagac atgcaagttt taaaaaggaa cataactatg 29460
ggccaggcac ggtggttcac gtctgtaatc ccagcacttt gggaggccga ggtgggtgga 29520
tcacctgagg tcaggagttc gagaccagcc tagacaccat ggtgaaaccc catctgtacc 29580
aaaactacaa aaattagctg ggcgtggtgg tgggcgcctg taatcccagc tacttgggaa 29640
gctgaggcag gagaatcact tgaacccggg aggcgaaggt tgcagtgagc cgagattgtg 29700
ccactgcact ccagcctggg caacaaggga gactccatct caattaaaaa aaaaaaaaaa 29760
aaaaaggaac ataactatgg agtctcaagg ggaagtaatt ccttcaacaa taacaaatct 29820
tgaaagctga gctctttttt ttttttgaga caggatctcc tcactttgtc gcccaggctg 29880
gagtgcagtg gtgggatcac agctcactgc agcctcgatc tcccaggctc aaatgatcct 29940
cctacctcag cctcccaaga agctgggatt acaggtgcat accatcacac ccgattcatt 30000
tttgtatact ttgaagagat ggggtctcac catgttgccc agtgtggtct tgaattcctg 30060
gactcaggtg atctgcccgc cttggcctcc cagagtgctg ggattacagg cctgagccaa 30120
caccccacg ggttcatttt cagagtcgca ccgagtgctg gggttacagg cctgagccaa 30180
ccccccacg ggttcatttt aagagtgaca ccgagtgctg gggttacagg cctgaaccaa 30240
ccccccacg agttcatttt cagagtcgca ccgagtgctg gggttacagg cctgagccaa 30300
ccccccacg ggttcatttt aagagtgaca ccgagtgctg gggttacagg cctgagccaa 30360
caccccacg ggttcatttt cagagtcaca ccgagtgctg gggttacagg cctgagccaa 30420
ccccccacg ggttcatttt cagagtcaca ccctttttct gaaaaacaac ttgggctcat 30480
gcaaattcga gagagagatg gtgacactcc ccgccccctg gacccaggtg gagtcgcagc 30540
agggtttacc cgtgagcggg gtccaaggcg atggccctcg gctggtcaag gtcctgccag 30600
aagagcacct tccgggatgt gccattgagg ttggccacct cgatgcggtt ggtctctgag 30660
tccgtccagt acagcttctt gcccacccag tcgcaggcga ggccgtcggg agagaccagg 30720
ccggagatga ccacgttctg cacggcggcc cccgtctggt tcaggtaggt ctgcttgatg 30780
gcctcctcgc tcacgtctgt ccagtacacg gctcccttgg aaaactggaa gtccactgcg 30840
gccgcatcct ccaggccgct gaccacgatg gtggactcca gcttgactcc gccggcgtcc 30900
accagccgta cgtcccggcg gttggcaaat agcaggagcg gcgaggctgt ggggcagaag 30960
caaaccgtga gggccactgg ctaagccagc aagatacaca gccctgggat ggagcactat 31020
gcccagagca ctcctggtac tgccctgccc atgcccaaga cctccagttc cttcctccca 31080
cccctaaggc gttgtcagga agttgcctgg gcagccccgg cccgcatcat tcagaggctc 31140
ctgcagcgca gcaaacagcc ttcttcccac attcggtgac agcacctgtt tgtttaccaa 31200
ctgttacgtc tgttccccca gatatgggtg acccttcctg ccatgcccaa aacctcccac 31260
atcgtcctcc agaggctaca ggggccctgt cctgttctgc agagaagcca catccccttt 31320
gttggcctga cacaggggat ggggacatgc aggcacagca ctggccatgc tgctcgctac 31380
agacccagcc acagggccac atttttgag gggttcagag cccaggccag acagagcctc 31440
aagattccct tacaagtctt tgaccactgt ccaagctcag gcccgtttcc ttggccgtgg 31500
catcagcttc ccatccaccc ctgtattcca tgtttctccc accctgcttc tggacattcc 31560
tacatttaaa gggtcactct ggaatgccac cccttggctc agacaccttc cacagctccc 31620
tgtgccagtg ccatgcagaa caaggtcaga ccccctagcc tggcctccaa ggccttggcc 31680
tctggcctca cctacacttc tctccaccac cccaccccaa gcattcctga tctgcctgcg 31740
gccaggctgg ctccctcacc tccctgtgca ccgcagccct cagccccttc tgcctgtgca 31800
agaagcctca tctcacagac aacggtctca ttcccacaac gggctcaatg agaaatcagg 31860
agaggccttc agaccatcac cccaccagac acctcagacg tcggaccagg agggtccagc 31920
aacccccaac acagactcag agggactaag aagccacatg aggagtgaac acaagatgtg 31980
gacaggagga ggttaagggc ctccagggag ctccatcagt ccgtgttctg ctgtcagcag 32040
ggttaggctg ggctggccac aaacaccccc aaaaaacatc tgaagccttg gcttgaaaca 32100
gctgacattc ctcatgaaaa ctgcagaccc ctgggtcctc ctgcgcagat gggggagccc 32160

agccaacccc acactcccac cttcaccaag aaagagaaag ccaaaacaaa ctcaactcag 32220
ccaatgacaa tcacagaact gaatcctgta gttagttcag ttggtttcat ttcagcaggg 32280
gaaagatttg cagcctctat gagggtagct gggaacacaa agggccagag catggcccag 32340
gagaccccag cgcagtgggg tagatggttc cgagcacagg cctccctgcc aagacaagca 32400
ctggctcaaa tcctggcccc tcccattccc aggagacatg ctccacagga tgggaggaca 32460
cacagaggac ctgaggccag gaaaatgaca gcggcgcctc cgccgcccca cccgtgctgt 32520
catcatctta ggtctacagt tctttgtggc aacgagggac actgtgaaag tcaaacaaca 32580
ggaaggcata ggccacaaat aaagacaaac gggacttcat gggaagctaa agattttgtg 32640
catcaaaaga cactatcgag agagtaaaaa ggcaacccac agaatgagag aaaatatttc 32700
caaatcatag atctactaag agattaatat ccatgaaata cagagaactc ctaaaactca 32760
acaatgagaa aacaactaag ccaactcaaa aatgggcaaa caacttgaac agacatttct 32820
ccaaagatga catataaatg gccaataaac acatcaaaac aggcttaata tatccctaat 32880
catcagggaa atgcaaatca aaactacaat aagataccat cttgcaccaa ttaggacggc 32940
tactatcaaa aaaacaaaat agcaagtgtt ggtgaggatc tggagcaact ggaacccttg 33000
tgcaccactg gcaaaaatgt gaaatggtgc agctactatg gaaaacagca tggcagttcc 33060
ccaaaaactt aaacacagaa ttaccatatg acccagcaat ttcgctttgg gttatatacc 33120
caaaagaact gaaaacaggg acacaatcag atatgcatac accttggatc acagcagcat 33180
ccttcccaac agctaaaaca tggaggcagc caggcatggt ggctcacgcc tgtaatccca 33240
gcactttggg aggctgaggc gggtggatca cctgaggtca ggagttcgag accagcctgg 33300
ccaacatggt gaaaccccgt ctctactaaa atacaaaaat tagctgggcg tagtgacggg 33360
cacctgtaat cccagctact cacaagtctg aggcaggaga atcacttgaa ccctggaagt 33420
ggacgttgca gtgagccaag attgcgccac tgcattccag cctgggtgac acagcgagac 33480
tctgtctcaa aaaacagcaa aacaaaaaca aaaaaacaaa caaacatgga agcaacccaa 33540
gcgtccctct actgagggat gaatagcggg gcaaaatctg ctccatccac acaatggagt 33600
actattcagt ctcaaaaagg aaaaagattc tggtcaggca cggtggctca tgcctgtaat 33660
cccagcactt ggggaggctg aggcgggtgg atcacctgaa gtcaggaatt caaggcccgc 33720
ctggccaaga ctggcaccna gctacacana aagtatangg ccccggaaa 33769

<210> 9
<211> 72049
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (8356),(8385),(38585)
<223> Identity of nucleotide sequences at the above locations are unknown.

<400> 9
tataccttgc gcggaccttc ggctcctgtg gtgaagacaa tatgaagaaa atagaaatta 60
cccataattt tgccacacag acttagttgt gtccatgtat cttgtgcacc ttttttctgt 120
ttacggatca aaatcgactt ttagggtcag gcgcggtggc tcacacctgt aatcccaaca 180
ctttgggagg ctggagttgg ggttgggggg tggatcactg aagatcagga gtttgagacc 240
agcctggcca acatggcgaa actccatctc tactaaaaat aaaagattag ccaggcgtgg 300
tggtgggtgc ctctaatccc agctactccg gaggctgagg caggagaatc gcttgaaccc 360
aggagacaga ggttgcagtg agccaggatc acgccactgc actccagcct ggcaacagag 420
cgagactctg tctcaaaaaa aaaaataaaa ataaaataaa taaatacata aattgacttt 480
taggagattg gttcaaacaa tgtgtgtaat gttgtgtctg agtgtttttc atttatcgtt 540
catgcaaatt ccgacatcat tcactcttct ccagagtgtg ctgttttcct gcctgtgtca 600
tcacccgtca ccttgaatgc cctcgtttag gtaaaataag tacattttat tcaaaaatat 660

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 260

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 260

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

What is claimed is:

1. A transgenic animal having somatic and/or germ cells comprising a nucleic acid which comprises a promoter region capable of directing protein expression in animal and/or human cells that is operably linked to a sequence comprising at least 15 contiguous nucleotides of SEQ ID NO: 2 including at least the thymine at position 582 of SEQ ID NO: 2.

2. A transgenic animal having somatic and/or germ cells comprising a nucleic acid which comprises a sequence which encodes SEQ ID NO: 4 and which includes at least a codon for the valine corresponding to the valine at position 171 of SEQ ID NO: 4, and wherein the nucleic acid further comprises an operably linked promoter region capable of directing protein expression in animal and/or human cells.

3. The transgenic animal of claims 1 or 2, wherein the nucleic acid comprises SEQ ID NO: 2.

4. A transgenic animal for the study of bone density modulation having somatic and/or germ cells comprising a nucleic acid which comprises a promoter region that directs protein expression in animal and/or human cells operably linked to a sequence comprising at least 15 contiguous nucleotides of SEQ ID NO: 1, wherein bone mass is modulated relative to non-transgenic animals of the same species in more than one parameter selected from among bone density, bone strength, trabecular number, bone size, and bone tissue connectivity.

5. The transgenic animal of any of claims 1 to 4 wherein the promoter region is CMV, RSV, SV40, and EF-1a, CMVβActin, histone, type I collagen,

TGFβ1, SX2, cfos/cjun, Cbfa1, Fra/Jun, Dlx5, osteocalcin, osteopontin, bone sialoprotein, or collagenase promoter regions.

6. The transgenic animal of any one of claims 1 to 4, wherein the promoter region is a bone specific promoter region.

7. The transgenic animal of any one of claims 1 to 4, wherein the promoter region is a CMVβActin promoter region.

8. The transgenic animal of any one of claims 1 to 4 wherein the promoter region is a type I collagen promoter region.

9. An embryo of the transgenic animal of any one of claims 1-8.

10. The transgenic animal of any of claims 1 to 3, wherein the transgenic animal express a human HBM protein.

11. The transgenic animal of claim 10, wherein the human HBM protein is expressed greatest in bone tissue.

12. The transgenic animal of any of claims 1 to 3 which exhibits a HBM phenotype.

13. The transgenic animal of any of claims 1 to 3, wherein bone mass is modulated relative to a non-transgenic animal of the same species in more than one parameter selected from among bone density, bone strength, trabecular number, bone size, and bone tissue connectivity.

14. The transgenic animal of any one of claims 1 to 3, wherein a human HBM protein is expressed and wherein the transgenic animal is fertile and passes the human *HBM* gene to its offspring.

15. The transgenic animal of claim 4, wherein a human LRP5 protein is expressed and wherein the transgenic animal is fertile and passes the human *LRP5* gene to its offspring.

16. A transgenic animal produced from the transgenic animal of any one of claims 1 to 15 or its offspring.

17. A transgenic mouse having a genome comprising an alteration of the gene encoding LRP5 wherein the alteration is caused by the introduction of a nucleic acid for gene targeting by homologous recombination into embryonic stem cells or pluripotent cells comprising a first section homologous to mouse *LRP5* gene and a second section homologous to another section of mouse *LRP5* gene, and between the first and the second section a middle section comprising an engineered deletion of a portion of the *LRP5* gene, a nucleic acid sequence change, or a nucleic acid insertion, and wherein the nucleic acid is capable of homologous recombination with the endogenous gene.

18. The transgenic mouse of claim 17, wherein the middle section of the nucleic acid for gene targeting comprises an engineered deletion of the ATG start codon, an engineered frame-shift mutation, an engineered stop codon, a neomycin resistance sequence, a loxP recombination site, or a synthetic transcriptional pause sequence.

19. The transgenic mouse of claim 17, wherein the nucleic acid for gene targeting further comprises both intron and exon sequences of the mouse *LRP5* gene.

20. The transgenic mouse of claim 17, wherein the nucleic acid for gene targeting further comprises a codon encoding a glycine to valine change at position 170 of the amino acid sequence of the mouse *LRP5* gene, and wherein the altered gene encodes a HBM protein.

21. The transgenic mouse of claim 17, wherein the alteration is a disruption of the *LRP5* gene such that it is not expressed.

22. A transgenic mouse having a genome comprising an alteration of the gene encoding *LRP6,* wherein the alteration is caused by the introduction into embryonic stem cells or pluripotent cells of a nucleic acid for gene targeting by homologous recombination comprising a first section homologous to mouse *LRP6* gene and a second section homologous to another section of mouse *LRP6* gene, and between the first and the second section a middle section comprising an engineered deletion of a portion of the *LRP6* gene, a nucleic acid sequence change, or a nucleic acid insertion, and wherein the nucleic acid is capable of homologous recombination with the endogenous gene wherein the transgenic animal has modulated Wnt activity, Dkk activity, lipid levels or bone mass.

23. The transgenic mouse of claim 22, wherein the alteration is a disruption of the *LRP6* gene such that it is not expressed.

24. The transgenic mouse of claim 22 or its offspring, wherein the mouse is fertile and transmits the altered gene to its offspring.

25. The transgenic mouse of any one of claims 17 to 24, wherein the transgenic mouse is fertile and transmits the altered gene to its offspring.

26. The transgenic mouse of any one of claims 17 to 24, wherein the transgenic mouse is fertile and transmits the altered gene to its offspring wherein the offspring exhibits a phenotype of modulated bone mass as indicated by at least three parameters selected from among bone density, bone strength, trabecular number, bone size, and bone tissue connectivity as compared to wild-type mice.

27. The transgenic mouse of any one of claims 17 to 26, wherein the transgenic mouse is produced by the introduction of a mouse embryonic stem cell into a mouse blastocyst.

28. A transgenic mouse produced from the transgenic mouse of any one of claims 17-27.

29. An animal model for the study of bone density modulation comprising a first group of animals composed of the transgenic animal of any one of claims 1 to 28 and a second group of control animals.

30. The animal model of claim 29, wherein the group of control animals comprises transgenic animals having cells which comprise a nucleic acid encoding human LRP5.

31. A method for studying bone mass determinants comprising the steps of:

(a) providing a first group of transgenic animals according to any one of claims 1 to 28; and

(b) measuring at least one parameter of bone development in the transgenic animals.

32. A method for studying modulators of bone mass comprising the steps of:

(a) providing a first group of transgenic animals according to any one of claims 1 to 28;

(b) administering a test compound; and

(c) measuring at least one parameter of development in the transgenic animals administered a test compound.

33. A method for studying bone mass comprising the steps of:

(a) providing a first group of transgenic animals according to any one of claims 1 to 28;

(b) administering an experimental procedure; and

(c) measuring at least one parameter of development in the animals administered an experimental procedure.

34. The method of claim 32, wherein the test compound is administered by injection, orally, by suppositories, in an implant, or topically.

35. A method according to claim 33, wherein the experimental procedure is chosen from among an ovariectomy, restricted bone loading, and increased bone loading.

36. A method according to claim 32, wherein a group of transgenic animals are expressing HBM and wherein bone mass density is modulated in the transgenic animals that are expressing HBM and are administered the test compound.

37. A method according to claim 36, wherein a group of transgenic mice are expressing HBM and wherein bone mass density is increased in the transgenic mice that express HBM and are administered the test compound.

38. The method of claim 32, wherein the test compound comprises a hormone, a growth factor, a peptide, RNA, DNA, a mineral, a vitamin, a natural product, or a synthetic organic compound.

39. The method of claim 33, wherein the experimental procedure comprises a surgical procedure, a gene therapy procedure, a drug therapy procedure, a dietary regimen, or physical exercise.

40. A method for studying an effect of HBM on bone disorders comprising the steps of:

(a) providing embryos of animals with a bone disorder phenotype;

(b) introducing the nucleic acid any one of claims 1 to 4 into a first group of the embryos;

(c) transferring the embryos to pseudopregnant mice; and

(d) measuring at least one parameter of development in the resultant mice.

41. A method for identifying surrogate markers of bone formation/resorbtion comprising the steps of:

(a) providing an animal model of bone development according to claim 29;

(b) measuring quantitatively a candidate surrogate marker in the animals; and,

(c) comparing the measurements of a group of transgenic animals to measurements of a control group of animals.

42. A method for studying effects of HBM on cardiac disorders comprising the steps of:

(a) providing a first group of transgenic animals according to any one of claims 1 to 28; and

(b) measuring at least one parameter of cardiac health in the transgenic animals administered a test compound.

43. A method for studying effects of HBM on cardiac disorders comprising the steps of:

(a) providing a first group of transgenic animals according to any one of claims 1 to 28;

(b) administering a test compound; and

(c) measuring at least one parameter of cardiac health in the transgenic animals administered a test compound;

wherein the test compound comprises a hormone, a growth factor, a peptide, RNA, DNA, a mineral, a vitamin, a natural product, or a synthetic organic compound.

44. The method of claim 43, wherein the parameter of cardiac health is blood serum lipid concentration.

45. A method of screening of cardio-protective treatments for bone mass modulation effects comprising the method of claim 31, and further comprising the step of administering a cardio-protective treatment to a subgroup of the first group the first group of animals.

46. An isolated cell derived from the transgenic animal of any one of claims 1 to 28 or its progeny.

47. The isolated cell of claim 46, wherein the cell is an osteoblast or an osteoclast cell.

48. A transgenic animal wherein the expression of endogenous LRP5 is modulated by an altered gene control sequence introduced by homologous or non-homologous recombination.

49. A method of modulating bone density comprising the step of,

administering a compound which modulates the endogenous expression of LRP5, HBM, or LRP6.

50. A method of screening of cardio-protective treatments for bone mass modulation effects comprising,

providing a first group of animals according to any of claims 1 to 28;

administering a cardio-protective treatment to a subgroup of the first group the first group of animals; and,

measuring at least one parameter of bone modulation in at least the treated mice.

51. The transgenic animal of any one of claims 1 to 16, wherein the nucleic acid further encodes for at least an alanine to valine substitution at position 1330 of SEQ ID NO: 4.

52. A nucleic acid for gene targeting by homologous recombination comprising a first section homologous to mouse *LRP5* gene and a second section homologous to another section of mouse *LRP5* gene, and between the first and the second section a middle section comprising an engineered deletion of a portion of the *LRP5* gene, a nucleic acid sequence change, or a nucleic acid insertion, and

wherein the nucleic acid is capable of homologous recombination with the endogenous gene.

53. The nucleic acid of claim 52, wherein the middle section comprises an engineered deletion of the ATG start codon, an engineered frame-shift mutation, an engineered stop codon, a neomycin resistance sequence, a loxP recombination site, or a synthetic transcriptional pause sequence.

54. The nucleic acid of claim 52, further comprising both intron and exon sequences of the mouse *LRP5* gene.

55. The isolated nucleic acid of claim 52, further comprising a codon encoding a glycine to valine change at position 170 of the amino acid sequence of the mouse *LRP5* gene.

56. A method of producing a transgenic mouse whose genome comprises an alteration of the gene encoding LRP5, the method comprising:

(a) providing the nucleic acid of claim 52;

(b) introducing the nucleic acid into mouse embryonic stem cells;

(c) selecting those embryonic stem cells that comprise the nucleic acid;

(d) introducing an embryonic stem cells of step (c) into a mouse blastocyst;

(e) transferring the blastocyst of step (d) to a pseudopregnant mouse; and

(f) allowing the transferred blastocyst to develop into a mouse chimeric for the nucleic acid.

57. The method of claim 56, wherein the introduction of the embryonic stem cell is by microinjection.

58. The method of claim 57, further comprising:

(a) breeding the chimeric mouse to a wild-type mouse to obtain mice heterozygous for the alteration; and

(b) breeding the heterozygous mice to generate mice homozygous for the alteration.

59. A method for identification of genes associated with bone mass comprising the steps of:

(a) providing an animal model of bone development according to claim 29;

(b) measuring a profile of gene expression in the animals; and,

(c) comparing the measurements of a the first group of animals to measurements of the control group of animals.

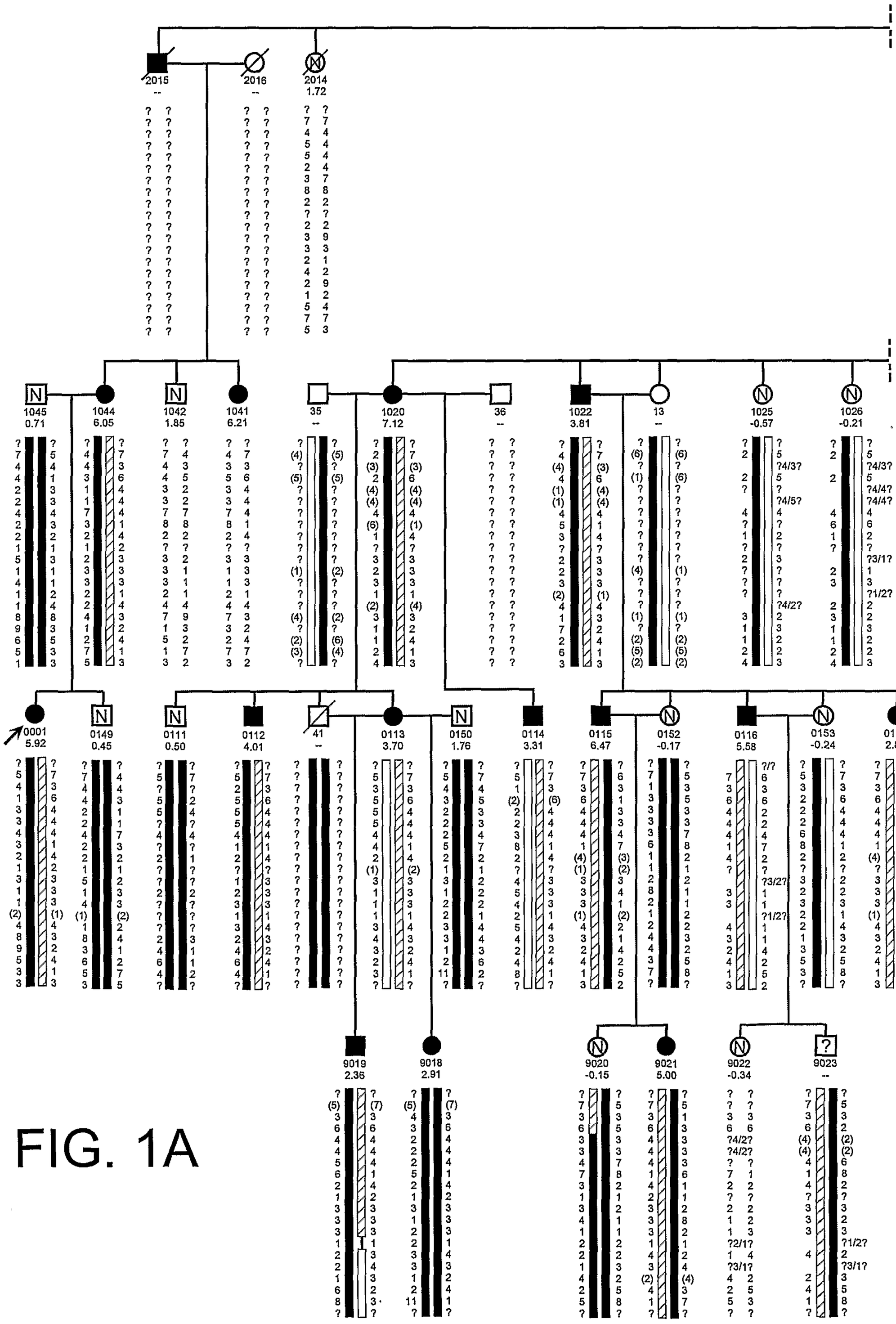

2015 --
2016 --
2014 1.72
1045 0.71
1044 6.05
1042 1.85
1041 6.21
35 --
1020 7.12
36 --
1022 3.81
13 --
1025 -0.57
1026 -0.21
0001 5.92
0149 0.45
0111 0.50
0112 4.01
41 --
0113 3.70
0150 1.76
0114 3.31
0115 6.47
0152 -0.17
0116 5.58
0153 -0.24
0117 2.85
9019 2.36
9018 2.91
9020 -0.15
9021 5.00
9022 -0.34
9023 --

FIG. 1A

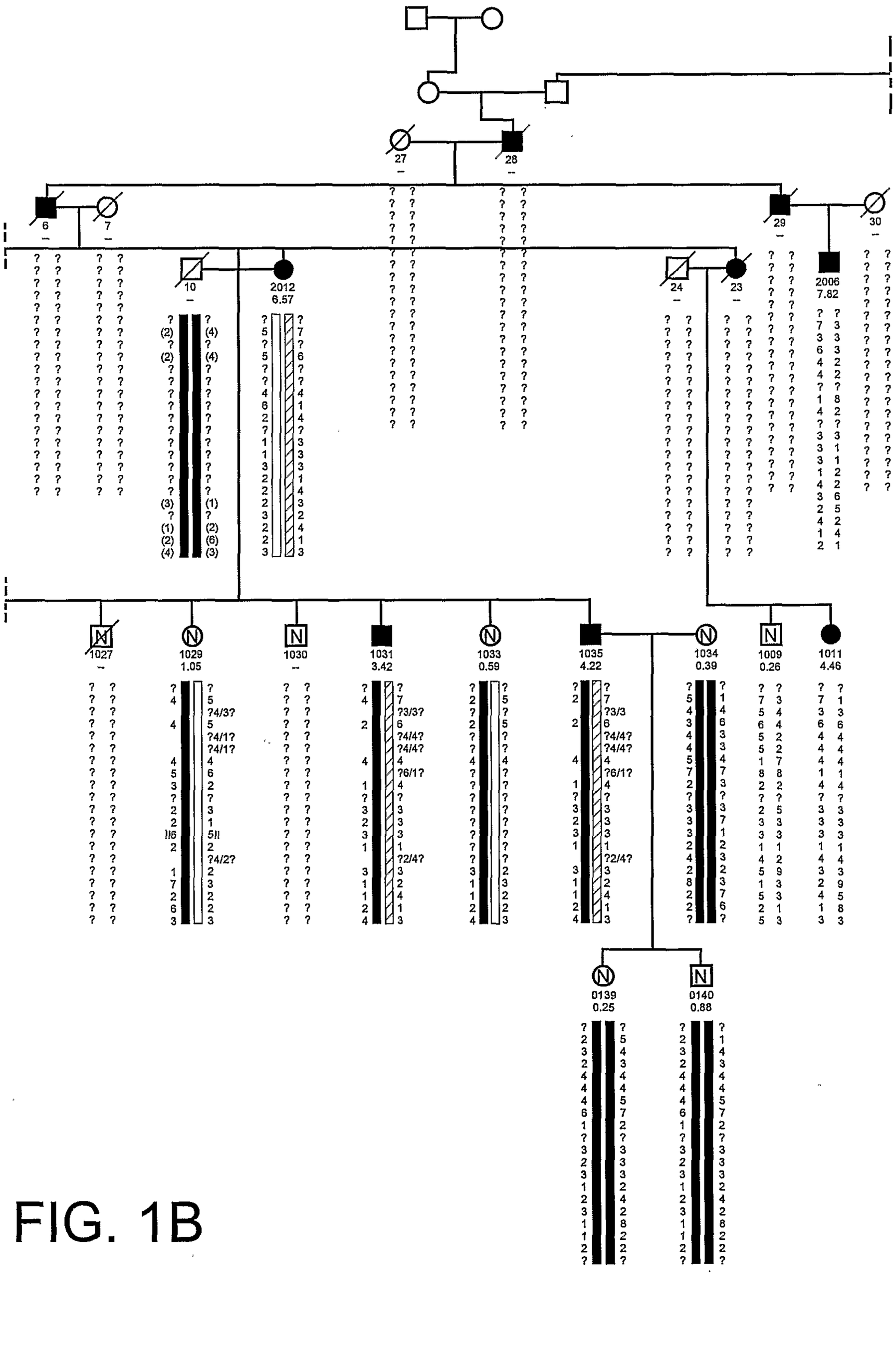

27
28
6
7
10
2012
6.57
29
30
24
23
2006
7.82
1027
1029
1.05
1030
1031
3.42
1033
0.59
1035
4.22
1034
0.39
1009
0.26
1011
4.46
0139
0.25
0140
0.88

FIG. 1B

FIG. 2A

BAC/STS Map of the HBM Region

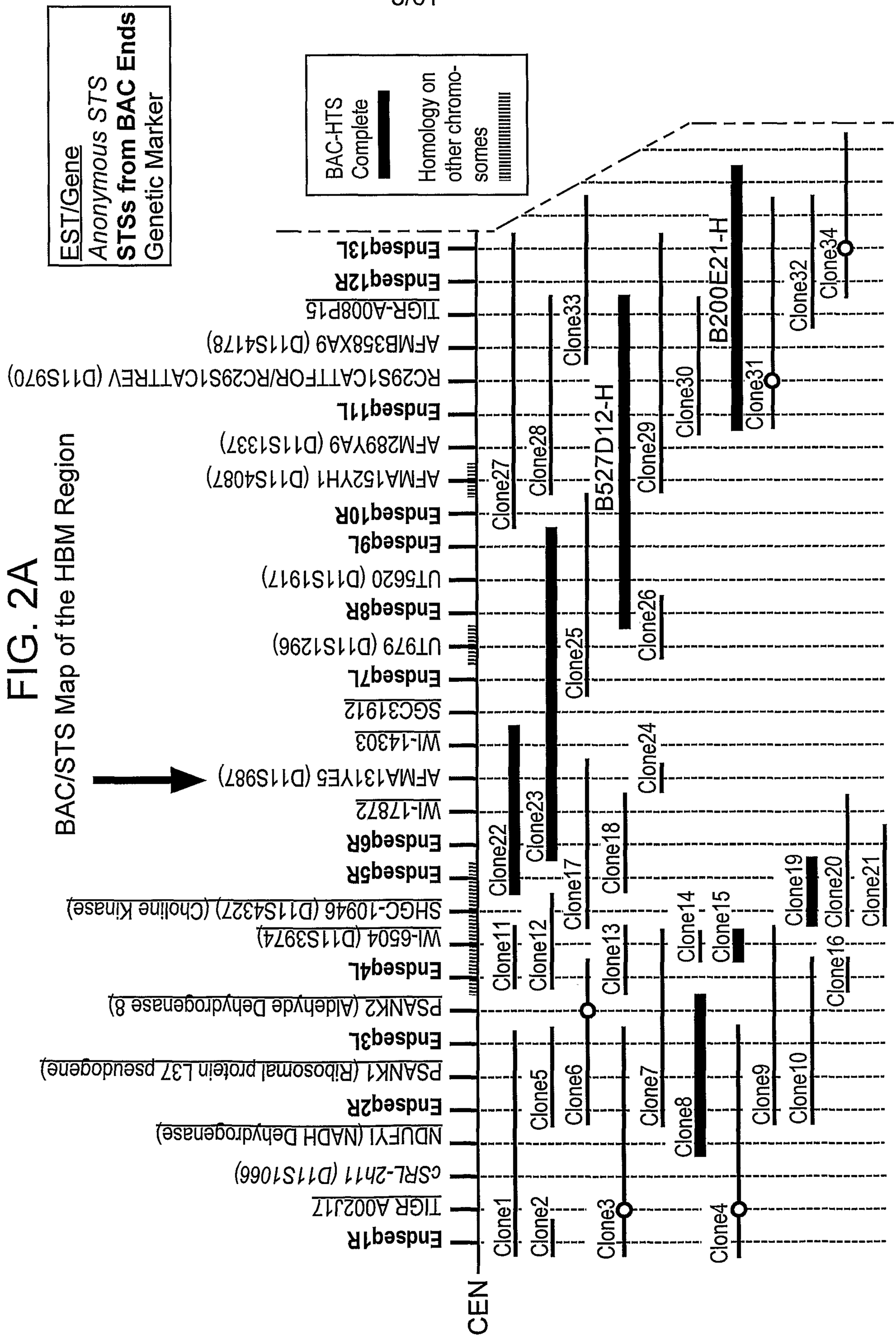
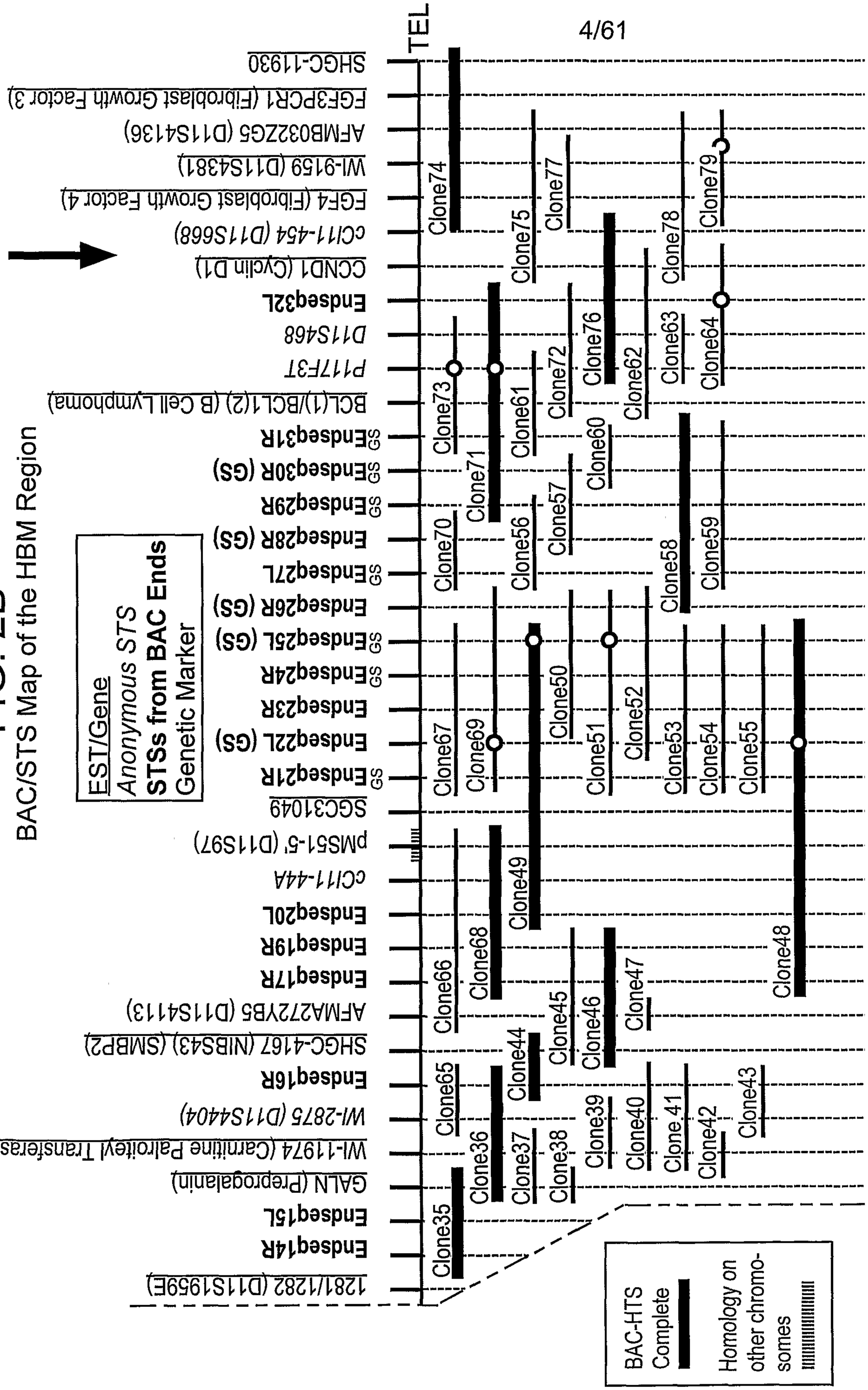


FIG. 2B

BAC/STS Map of the HBM Region



Exon 1
ACTAAAGCGCCGCCGCCGCGCCATGGAGCCCGAGTGAGCGCGGCGCGG
GCCCGTCCGGCCGCCGGACAACATGGAGGCAGCGCCGCCCCGGGCCGCC
GTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGGCGCTGTGCGGCTGC
CCGGCCCCCGCCGCGGCC

Exon 2 Coordinates: 527d12_Contig308G 30944-30549
gccccacag**CCTCGCCGCTCCTGCTATTTGCCAACCGCCGGGACGTACGGCT**
GGTGGACGCCGGCGGAGTCAAGCTGGAGTCCACCATCGTGGTCAGCGG
CCTGGAGGATGCGGCCGCAGTGGACTTCCAGTTTTCCAAGGGAGCCGTG
TACTGGACAGACGTGAGCGAGGAGGCCATCAAGCAGACCTACCTGAACC
AGACGGGGGCCGCCGTGCAGAACGTGGTCATCTCCGGCCTGGTCTCTCC
CGACGGCCTCGCCTGCGACTGGGTGGGCAAGAAGCTGTACTGGACGGA
CTCAGAGACCAACCGCATCGAGGTGGCCAACCTCAATGGCACATCCCGG
AAGGTGCTCTTCTGGCAGGACCTTGACCAGCCGAGGGCCATCGCCTTGG
ACCCCGCTCACGGgtaaaccctgctg

... 9408 nt ...

Exon 3 Coordinates: 527d12_Contig308G 21141-20945
ccccgtcacag**GTACATGTACTGGACAGACTGGGGTGAGACGCCCCGGATTGA**
GCGGGCAGGGATGGATGGCAGCACCCGGAAGATCATTGTGGACTCGGA
CATTTACTGGCCCAATGGACTGACCATCGACCTGGAGGAGCAGAAGCTC
TACTGGGCTGACGCCAAGCTCAGCTTCATCCACCGTGCCAACCTGGACG
GCTCGTTCCGgtaggtacccac

... 6094 nt ...

Exon 4 Coordinates: 527d12_Contig308G 15047-14850
tccctgactgcag**GCAGAAGGTGGTGGAGGGCAGCCTGACGCACCCCTTCGCCC**
TGACGCTCTCCGGGGACACTCTGTACTGGACAGACTGGCAGACCCGCTC
CATCCATGCCTGCAACAAGCGCACTGGGGGGAAGAGGAAGGAGATCCTG
AGTGCCCTATACTCACCCATGGACATCCAGGTGCTGAGCCAGGAGCGGC
AGCCTTTCTgtgagtgccgg

... 1827 nt ...

Exon 5 Coordinates: 527d12_Contig308G 13220-13088
tttctcag**TCCACACTCGCTGTGAGGAGGACAATGGCGGCTGCTCCCACCTGT**
GCCTGCTGTCCCCAAGCGAGCCTTTCTACACATGCGCCTGCCCCACGGG
TGTGCAGCTGCAGGACAACGGCAGGACGTGTAAGGCAGgtgaggcggtgggacg

FIG. 3A

... 20923 nt ...

Exon 6 Coordinates: 527d12_Contig309G 7705-8100
ctccacag**GAGCCGAGGAGGTGCTGCTGCTGGCCCGGCGGACGGACCTACG**
GAGGATCTCGCTGGACACGCCGGACTTCACCGACATCGTGCTGCAGGTG
GACGACATCCGGCACGCCATTGCCATCGACTACGACCCGCTAGAGGGCT
ATGTCTACTGGACAGATGACGAGGTGCGGGCCATCCGCAGGGCGTACCT
GGACGGGTCTGGGGCGCAGACGCTGGTCAACACCGAGATCAACGACCC
CGATGGCATCGCGGTCGACTGGGTGGCCCGAAACCTCTACTGGACCGAC
ACGGGCACGGACCGCATCGAGGTGACGCGCCTCAACGGCACCTCCCGCA
AGATCCTGGTGTCGGAGGACCTGGACGAGCCCCGAGCCATCGCACTGCA
CCCCGTGATGGGgtaagacgggc

….. 3211 nt …..

Exon 7 Coordinates: 527d12_Contig309G 11311-11482
ttcttctccag**CCTCATGTACTGGACAGACTGGGGAGAGAACCCTAAAATCGAG**
TGTGCCAACTTGGATGGGCAGGAGCGGCGTGTGCTGGTCAATGCCTCCC
TCGGGTGGCCCAACGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCTCT
ACTGGGGAGACGCCAAGACAGACAAGATCGAGgtgaggctcctgtgg

…… 13445 nt …..

Exon 8 Coordinates: 527d12_Contig309G 24927-25143
ccgtcctgcag**GTGATCAATGTTGATGGGACGAAGAGGCGGACCCTCCTGGAG**
GACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTTCATCT
ACTGGACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCAA
GGCCAGCCGGGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGCTC
AAAGCTGTGAATGTGGCCAAGGTCGTCGgtgagtccgggggggtc

….2826 nt ……

Exon 9 Coordinates: 527d12_Contig309G 27969-28256
gttcgcttccag**GAACCAACCCGTGTGCGGACAGGAACGGGGGGGTGCAGCCACC**
TGTGCTTCTTCACACCCCACGCAACCCGGTGTGGCTGCCCCATCGGCCT
GGAGCTGCTGAGTGACATGAAGACCTGCATCGTGCCTGAGGCCTTCTTG
GTCTTCACCAGCAGAGCCGCCATCCACAGGATCTCCCTCGAGACCAATA
ACAACGACGTGGCCATCCCGCTCACGGGCGTCAAGGAGGCCTCAGCCCT
GGACTTTGATGTGTCCAACAACCACATCTACTGGACAGACGTCAGCCTG
AAGgtagcgtgggc

……3102…..

FIG. 3B

Exon 10 Coordinates: 527d12_Contig309G 31358-31582
cctgctgccag**ACCATCAGCCGCGCCTTCATGAACGGGAGCTCGGTGGAGCAC GTGGTGGAGTTTGGCCTTGACTACCCCGAGGGCATGGCCGTTGACTGGA TGGGCAAGAACCTCTACTGGGCCGACACTGGGACCAACAGAATCGAAGT GGCGCGGCTGGACGGGCAGTTCCGGCAAGTCCTCGTGTGGAGGGACTT GGACAACCCGAGGTCGCTGGCCCTGGATCCCACCAAGGG**gtaagtgtttgcctgtc

……1297 nt……

Exon 11 Coordinates: 527d12_Contig309G 32879-33064
gtgccttccag**CTACATCTACTGGACCGAGTGGGGCGGCAAGCCGAGGATCGT GCGGGCCTTCATGGACGGGACCAACTGCATGACGCTGGTGGACAAGGTG GGCCGGGCCAACGACCTCACCATTGACTACGCTGACCAGCGCCTCTACT GGACCGACCTGGACACCAACATGATCGAGTCGTCCAACATGCTGG**gtgaggg ccgggct

…….2069 nt…..

Exon 12 Coordinates: 527d12_Contig309G 35133-35454
gtgttcatgcag**GTCAGGAGCGGGTCGTGATTGCCGACGATCTCCCGCACCCGT TCGGTCTGACGCAGTACAGCGATTATATCTACTGGACAGACTGGAATCT GCACAGCATTGAGCGGGCCGACAAGACTAGCGGCCGGAACCGCACCCTC ATCCAGGGCCACCTGGACTTCGTGATGGACATCCTGGTGTTCCACTCCT CCCGCCAGGATGGCCTCAATGACTGTATGCACAACAACGGGCAGTGTGG GCAGCTGTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGCTGCGCCTCA CACTACACCCTGGACCCCAGCAGCCGCAACTGCAGCC**gtaagtgcctcatggt

…….2006 nt……

Exon 13 Coordinates: 527d12_Contig309G 37460-37659
gcctcctcta**CGCCCACCACCTTCTTGCTGTTCAGCCAGAAATCTGCCATCAGT CGGATGATCCCGGACGACCAGCACAGCCCGGATCTCATCCTGCCCCTGC ATGGACTGAGGAACGTCAAAGCCATCGACTATGACCCACTGGACAAGTT CATCTACTGGGTGGATGGGCGCCAGAACATCAAGCGAGCCAAGGACGAC GGGACCCAG**gcaggtgccctgtgg

……6965 nt……

FIG. 3C

Exon 14 Coordinates: 527d12_Contig309G 44624-44832
ctttgtcttacag**CCCTTTGTTTTGACCTCTCTGAGCCAAGGCCAAAACCCAGACA GGCAGCCCCACGACCTCAGCATCGACATCTACAGCCGGACACTGTTCTG GACGTGCGAGGCCACCAATACCATCAACGTCCACAGGCTGAGCGGGGAA GCCATGGGGGTGGTGCTGCGTGGGGACCGCGACAAGCCCAGGGCCATC GTCGTCAACGCGGAGCGAGG**gtaggaggccaac

……1404 nt…..

Exon 15 Coordinates: 527d12_Contig309G 46236-46427
ccaccctcccgcag**GTACCTGTACTTCACCAACATGCAGGACCGGGCAGCCAAGA TCGAACGCGCAGCCCTGGACGGCACCGAGCGCGAGGTCCTCTTCACCAC CGGCCTCATCCGCCCTGTGGCCCTGGTGGTGGACAACACACTGGGCAAG CTGTTCTGGGTGGACGCGGACCTGAAGCGCATTGAGAGCTGTGACCTGT CAG**gtacgcgccccgg

…..686 nt…..

Exon 16 Coordinates: 527d12_Contig309G 47113-47322
ggctgcttgcag**GGGCCAACCGCCTGACCCTGGAGGACGCCAACATCGTGCAGC CTCTGGGCCTGACCATCCTTGGCAAGCATCTCTACTGGATCGACCGCCA GCAGCAGATGATCGAGCGTGTGGAGAAGACCACCGGGGACAAGCGGAC TCGCATCCAGGGCCGTGTCGCCCACCTCACTGGCATCCATGCAGTGGAG GAAGTCAGCCTGGAGGAGTTCT**gtacgtgggggc

…..3884 nt……

Exon 17 Coordinates: 527d12_Contig309G 51206-51331
ttgtctttgcag**CAGCCCACCCATGTGCCCGTGACAATGGTGGCTGCTCCCACAT CTGTATTGCCAAGGGTGATGGGACACCACGGTGCTCATGCCCAGTCCAC CTCGTGCTCCTGCAGAACCTGCTGACCTGTGGAG**gtaggtgtgacctaggtgc

….3905 nt…….

Exon 18 Coordinates: 527d12_Contig309G 55236-55472
gttctcctctgtccctcccccag**AGCCGCCCACCTGCTCCCCGGACCAGTTTGCATGTG CCACAGGGGAGATCGACTGTATCCCCGGGGCCTGGCGCTGTGACGGCTT TCCCGAGTGCGATGACCAGAGCGACGAGGAGGGCTGCCCCGTGTGCTCC GCCGCCCAGTTCCCCTGCGCGCGGGGTCAGTGTGTGGACCTGCGCCTGC GCTGCGACGGCGAGGCAGACTGTCAGGACCGCTCAGACGAGGTGGACT GTGACG**gtgaggccctcc

…….3052 nt…..

FIG. 3D

Exon 19 Coordinates: 527d12_Contig309G 58524-58634
tctccttgcag**CCATCTGCCTGCCCAACCAGTTCCGGTGTGCGAGCGGCCAGTG**
TGTCCTCATCAAACAGCAGTGCGACTCCTTCCCCGACTGTATCGACGGCT
CCGACGAGCTCATGTGTGgtgagccagctt

……..1448 nt……

Exon 20 Coordinates: 527d12_Contig309G 60082-60319
gtttgtctctggcag**AAATCACCAAGCCGCCCTCAGACGACAGCCCGGCCCACAGC**
AGTGCCATCGGGCCCGTCATTGGCATCATCCTCTCTCTCTTCGTCATGGG
TGGTGTCTATTTTGTGTGCCAGCGCGTGGTGTGCCAGCGCTATGCGGGG
GCCAACGGGCCCTTCCCGCACGAGTATGTCAGCGGGACCCCGCACGTGC
CCCTCAATTTCATAGCCCCGGGCGGTTCCCAGCATGGCCCCTTCACAGgta
aggagcctgagatatggaa

….1095 nt…..

Exon 21 Coordinates: 527d12_Contig309G 61414-61552
cttccctgccag**GCATCGCATGCGGAAAGTCCATGATGAGCTCCGTGAGCCTGA**
TGGGGGGCCGGGGCGGGGTGCCCCTCTACGACCGGAACCACGTCACAG
GGGCCTCGTCCAGCAGCTCGTCCAGCACGAAGGCCACGCTGTACCCGCC
Ggtgaggggcggg

……6513 nt……

Exon 22 Coordinates: 527d12_Contig309G 68065-68162
ttggctctcctcag**ATCCTGAACCCGCCGCCCTCCCCGGCCACGGACCCCTCCCT**
GTACAACATGGACATGTTCTACTCTTCAAACATTCCGGCCACTGCGAGAC
CGTACAGgtaggacatcccctgcag

…….2273 nt…..

FIG. 3E

Exon 23 Coordinates: 527d12_Contig309G 70435-70901

tcaaacattccggccactgcgagaccgtacag**GCCCTACATCATTCGAGGAATGGCGCCCCC**
GACGACGCCCTGCAGCACCGACGTGTGTGACAGCGACTACAGCGCCAGC
CGCTGGAAGGCCAGCAAGTACTACCTGGATTTGAACTCGGACTCAGACC
CCTATCCACCCCCACCCACGCCCCACAGCCAGTACCTGTCGGCGGAGGA
CAGCTGCCCCGCCCTCGCCCGCCACCGAGAGGAGCTACTTCCATCTCTTC
CCGCCCCCTCCGTCCCCCTGCACGGACTCATCCTGACCTCGGCCGGGCCA
CTCTGGCTTCTCTGTGCCCCTGTAAATAGTTTTAAATATGAACAAAGAAAAAA
ATATATTTTATGATTTAAAAAATAAATATAATTGGGATTTTAAAAACATGAGA
AATGTGAACTGTGATGGGGTGGGCAGGGCTGGGAGAACTTTGTACAGTGGAG
AAATATTTATAAACTTAATTTTGTAAAACA

FIG. 3F

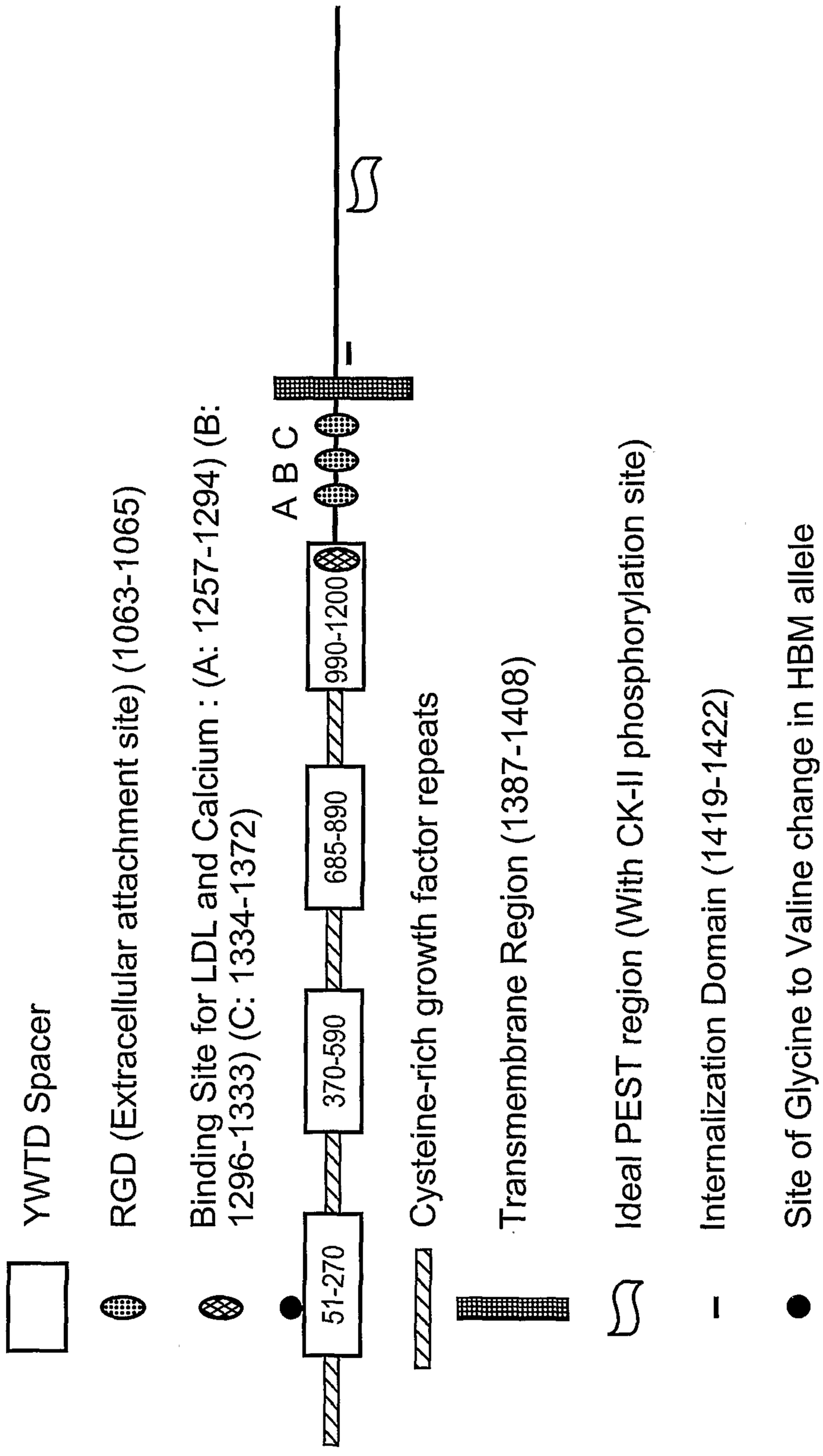

Model for a LDL Receptor-Related protein, Zmax1
YWTD Spacer
RGD (Extracellular attachment site) (1063-1065)
Binding Site for LDL and Calcium : (A: 1257-1294) (B: 1296-1333) (C: 1334-1372)
A B C
Cysteine-rich growth factor repeats
Transmembrane Region (1387-1408)
Ideal PEST region (With CK-II phosphorylation site)
Internalization Domain (1419-1422)
Site of Glycine to Valine change in HBM allele
200
370-590
685-890
990-1200
51-270

FIG. 4

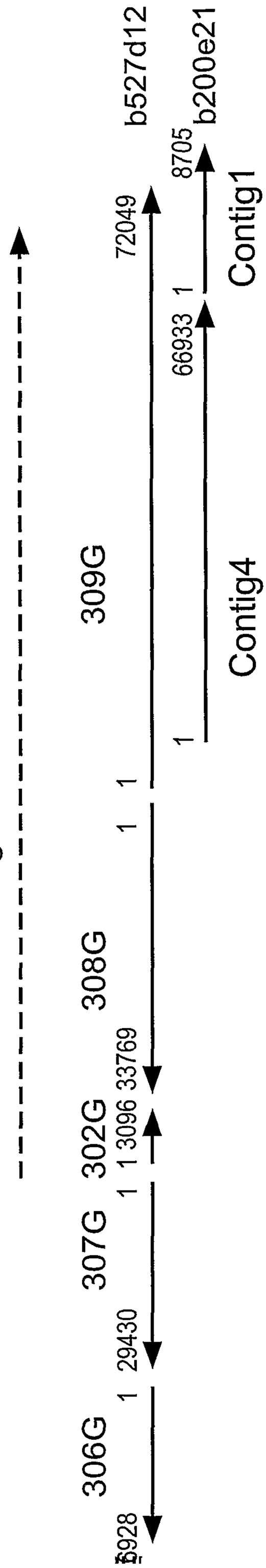

b527d12
b200e21
8705
72049
1
Contig1
66933 1
Contig4
1
309G
1 1
High Bone Mass Gene
302G
308G
1 3096 33769
1
307G
1 29430
306G
1 5928

FIG. 5

FIG. 6A

```
   1  ACTAAAGCGCCGCCGCCGCGCCATGGAGCCCGAGTGAGCGCGGCGCGGGCCCGTCCGGCC   60
  61  GCCGGACAACATGGAGGCAGCGCCGCCCGGGCCGCCGTGGCCGCTGCTGCTGCTGCTGCT  120
   1            M  E  A  A  P  P  G  P  P  W  P  L  L  L  L  L  L   17

 121  GCTGCTGCTGGCGCTGTGCGGCTGCCCGGCCCCCGCCGCGGCCTCGCCGCTCCTGCTATT  180
  18   L  L  L  A  L  C  G  C  P  A  P  A  A  A  S  P  L  L  L  F   37

 181  TGCCAACCGCCGGGACGTACGGCTGGTGGACGCCGGCGGAGTCAAGCTGGAGTCCACCAT  240
  38   A  N  R  R  D  V  R  L  V  D  A  G  G  V  K  L  E  S  T  I   57

 241  CGTGGTCAGCGGCCTGGAGGATGCGGCCGCAGTGGACTTCCAGTTTTCCAAGGGAGCCGT  300
  58   V  V  S  G  L  E  D  A  A  A  V  D  F  Q  F  S  K  G  A  V   77

 301  GTACTGGACAGACGTGAGCGAGGAGGCCATCAAGCAGACCTACCTGAACCAGACGGGGGC  360
  78   Y  W  T  D  V  S  E  E  A  I  K  Q  T  Y  L  N  Q  T  G  A   97

 361  CGCCGTGCAGAACGTGGTCATCTCCGGCCTGGTCTCTCCCGACGGCCTCGCCTGCGACTG  420
  98   A  V  Q  N  V  V  I  S  G  L  V  S  P  D  G  L  A  C  D  W  117

 421  GGTGGGCAAGAAGCTGTACTGGACGGACTCAGAGACCAACCGCATCGAGGTGGCCAACCT  480
 118   V  G  K  K  L  Y  W  T  D  S  E  T  N  R  I  E  V  A  N  L  137

 481  CAATGGCACATCCCGGAAGGTGCTCTTCTGGCAGGACCTTGACCAGCCGAGGGCCATCGC  540
 138   N  G  T  S  R  K  V  L  F  W  Q  D  L  D  Q  P  R  A  I  A  157

 541  CTTGGACCCCGCTCACGGGTACATGTACTGGACAGACTGGGGTGAGACGCCCCGGATTGA  600
 158   L  D  P  A  H  G  Y  M  Y  W  T  D  W  G  E  T  P  R  I  E  177
```

SUBSTITUTE SHEET (RULE 26)

FIG. 6B

```
601  GCGGGCAGGGATGGATGGCAGCACCCGGAAGATCATTGTGGACTCGGACATTTACTGGCC  660
178   R  A  G  M  D  G  S  T  R  K  I  I  V  D  S  D  I  Y  W  P   197

661  CAATGGACTGACCATCGACCTGGAGGAGCAGAAGCTCTACTGGGCTGACGCCAAGCTCAG  720
198   N  G  L  T  I  D  L  E  E  Q  K  L  Y  W  A  D  A  K  L  S   217

721  CTTCATCCACCGTGCCAACCTGGACGGCTCGTTCCGGCAGAAGGTGGTGGAGGGCAGCCT  780
218   F  I  H  R  A  N  L  D  G  S  F  R  Q  K  V  V  E  G  S  L   237

781  GACGCACCCCTTCGCCCTGACGCTCTCCGGGGACACTCTGTACTGGACAGACTGGCAGAC  840
238   T  H  P  F  A  L  T  L  S  G  D  T  L  Y  W  T  D  W  Q  T   257

841  CCGCTCCATCCATGCCTGCAACAAGCGCACTGGGGGGAAGAGGAAGGAGATCCTGAGTGC  900
258   R  S  I  H  A  C  N  K  R  T  G  G  K  R  K  E  I  L  S  A   277

901  CCTCTACTCACCCATGGACATCCAGGTGCTGAGCCAGGAGCGGCAGCCTTTCTTCCACAC  960
278   L  Y  S  P  M  D  I  Q  V  L  S  Q  E  R  Q  P  F  F  H  T   297

961  TCGCTGTGAGGAGGACAATGGCGGCTGCTCCCACCTGTGCCTGCTGTCCCCAAGCGAGCC  1020
298   R  C  E  E  D  N  G  G  C  S  H  L  C  L  L  S  P  S  E  P   317

1021 TTTCTACACATGCGCCTGCCCCACGGGTGTGCAGCTGCAGGACAACGGCAGGACGTGTAA  1080
318   F  Y  T  C  A  C  P  T  G  V  Q  L  Q  D  N  G  R  T  C  K   337

1081 GGCAGGAGCCGAGGAGGTGCTGCTGCTGGCCCGGCGGACGGACCTACGGAGGATCTCGCT  1140
338   A  G  A  E  E  V  L  L  L  A  R  R  T  D  L  R  R  I  S  L   357
```

SUBSTITUTE SHEET (RULE 26)

FIG. 6C

```
1141  GGACACGCCGGACTTCACCGACATCGTGCTGCAGGTGGACGACATCCGGCACGCCATTGC  1200
 358    D   T   P   D   F   T   D   I   V   L   Q   V   D   D   I   R   H   A   I   A    377

1201  CATCGACTACGACCCGCTAGAGGGCTATGTCTACTGGACAGATGACGAGGTGCGGGCCAT  1260
 378    I   D   Y   D   P   L   E   G   Y   V   Y   W   T   D   D   E   V   R   A   I    397

1261  CCGCAGGGCGTACCTGGACGGGTCTGGGGCGCAGACGCTGGTCAACACCGAGATCAACGA  1320
 398    R   R   A   Y   L   D   G   S   G   A   Q   T   L   V   N   T   E   I   N   D    417

1321  CCCCGATGGCATCGCGGTCGACTGGGTGGCCCGAAACCTCTACTGGACCGACACGGGCAC  1380
 418    P   D   G   I   A   V   D   W   V   A   R   N   L   Y   W   T   D   T   G   T    437

1381  GGACCGCATCGAGGTGACGCGCCTCAACGGCACCTCCCGCAAGATCCTGGTGTCGGAGGA  1440
 438    D   R   I   E   V   T   R   L   N   G   T   S   R   K   I   L   V   S   E   D    457

1441  CCTGGACGAGCCCCGAGCCATCGCACTGCACCCCGTGATGGGCCTCATGTACTGGACAGA  1500
 458    L   D   E   P   R   A   I   A   L   H   P   V   M   G   L   M   Y   W   T   D    477

1501  CTGGGGAGAGAACCCTAAAATCGAGTGTGCCAACTTGGATGGGCAGGAGCGGCGTGTGCT  1560
 478    W   G   E   N   P   K   I   E   C   A   N   L   D   G   Q   E   R   R   V   L    497

1561  GGTCAATGCCTCCCTCGGGTGGCCCAACGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCT  1620
 498    V   N   A   S   L   G   W   P   N   G   L   A   L   D   L   Q   E   G   K   L    517

1621  CTACTGGGGAGACGCCAAGACAGACAAGATCGAGGTGATCAATGTTGATGGGACGAAGAG  1680
 518    Y   W   G   D   A   K   T   D   K   I   E   V   I   N   V   D   G   T   K   R    537
```

FIG. 6D

```
1681  GCGGACCCTCCTGGAGGACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTT  1740
 538   R  T  L  L  E  D  K  L  P  H  I  F  G  F  T  L  L  G  D  F   557

1741  CATCTACTGGACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCAAGGCCAG  1800
 558   I  Y  W  T  D  W  Q  R  R  S  I  E  R  V  H  K  V  K  A  S   577

1801  CCGGGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGCTCAAAGCTGTGAATGTGGC  1860
 578   R  D  V  I  I  D  Q  L  P  D  L  M  G  L  K  A  V  N  V  A   597

1861  CAAGGTCGTCGGAACCAACCCGTGTGCGGACAGGAACGGGGGGTGCAGCCACCTGTGCTT  1920
 598   K  V  V  G  T  N  P  C  A  D  R  N  G  G  C  S  H  L  C  F   617

1921  CTTCACACCCCACGCAACCCGGTGTGGCTGCCCCATCGGCCTGGAGCTGCTGAGTGACAT  1980
 618   F  T  P  H  A  T  R  C  G  C  P  I  G  L  E  L  L  S  D  M   637

1981  GAAGACCTGCATCGTGCCTGAGGCCTTCTTGGTCTTCACCAGCAGAGCCGCCATCCACAG  2040
 638   K  T  C  I  V  P  E  A  F  L  V  F  T  S  R  A  A  I  H  R   657

2041  GATCTCCCTCGAGACCAATAACAACGACGTGGCCATCCCGCTCACGGGCGTCAAGGAGGC  2100
 658   I  S  L  E  T  N  N  N  D  V  A  I  P  L  T  G  V  K  E  A   677

2101  CTCAGCCCTGGACTTTGATGTGTCCAACAACCACATCTACTGGACAGACGTCAGCCTGAA  2160
 678   S  A  L  D  F  D  V  S  N  N  H  I  Y  W  T  D  V  S  L  K   697

2161  GACCATCAGCCGCGCCTTCATGAACGGGAGCTCGGTGGAGCACGTGGTGGAGTTTGGCCT  2220
 698   T  I  S  R  A  F  M  N  G  S  S  V  E  H  V  V  E  F  G  L   717
```

SUBSTITUTE SHEET (RULE 26)

FIG. 6E

```
2221  TGACTACCCCGAGGGCATGGCCGTTGACTGGATGGGCAAGAACCTCTACTGGGCCGACAC  2280
 718    D   Y   P   E   G   M   A   V   D   W   M   G   K   N   L   Y   W   A   D   T    737

2281  TGGGACCAACAGAATCGAAGTGGCGCGGCTGGACGGGCAGTTCCGGCAAGTCCTCGTGTG  2340
 738    G   T   N   R   I   E   V   A   R   L   D   G   Q   F   R   Q   V   L   V   W    757

2341  GAGGGACTTGGACAACCCGAGGTCGCTGGCCCTGGATCCCACCAAGGGCTACATCTACTG  2400
 758    R   D   L   D   N   P   R   S   L   A   L   D   P   T   K   G   Y   I   Y   W    777

2401  GACCGAGTGGGGCGGCAAGCCGAGGATCGTGCGGGCCTTCATGGACGGGACCAACTGCAT  2460
 778    T   E   W   G   G   K   P   R   I   V   R   A   F   M   D   G   T   N   C   M    797

2461  GACGCTGGTGGACAAGGTGGGCCGGGCCAACGACCTCACCATTGACTACGCTGACCAGCG  2520
 798    T   L   V   D   K   V   G   R   A   N   D   L   T   I   D   Y   A   D   Q   R    817

2521  CCTCTACTGGACCGACCTGGACACCAACATGATCGAGTCGTCCAACATGCTGGGTCAGGA  2580
 818    L   Y   W   T   D   L   D   T   N   M   I   E   S   S   N   M   L   G   Q   E    837

2581  GCGGGTCGTGATTGCCGACGATCTCCCGCACCCGTTCGGTCTGACGCAGTACAGCGATTA  2640
 838    R   V   V   I   A   D   D   L   P   H   P   F   G   L   T   Q   Y   S   D   Y    857

2641  TATCTACTGGACAGACTGGAATCTGCACAGCATTGAGCGGGCCGACAAGACTAGCGGCCG  2700
 858    I   Y   W   T   D   W   N   L   H   S   I   E   R   A   D   K   T   S   G   R    877

2701  GAACCGCACCCTCATCCAGGGCCACCTGGACTTCGTGATGGACATCCTGGTGTTCCACTC  2760
 878    N   R   T   L   I   Q   G   H   L   D   F   V   M   D   I   L   V   F   H   S    897
```

FIG. 6F

```
2761  CTCCCGCCAGGATGGCCTCAATGACTGTATGCACAACAACGGGCAGTGTGGGCAGCTGTG  2820
 898   S  R  Q  D  G  L  N  D  C  M  H  N  N  G  Q  C  G  Q  L  C   917

2821  CCTTGCCATCCCCGGCGGCCACCGCTGCGGCTGCGCCTCACACTACACCCTGGACCCCAG  2880
 918   L  A  I  P  G  G  H  R  C  G  C  A  S  H  Y  T  L  D  P  S   937

2881  CAGCCGCAACTGCAGCCCGCCCACCACCTTCTTGCTGTTCAGCCAGAAATCTGCCATCAG  2940
 938   S  R  N  C  S  P  P  T  T  F  L  L  F  S  Q  K  S  A  I  S   957

2941  TCGGATGATCCCGGACGACCAGCACAGCCCGGATCTCATCCTGCCCCTGCATGGACTGAG  3000
 958   R  M  I  P  D  D  Q  H  S  P  D  L  I  L  P  L  H  G  L  R   977

3001  GAACGTCAAAGCCATCGACTATGACCCACTGGACAAGTTCATCTACTGGGTGGATGGGCG  3060
 978   N  V  K  A  I  D  Y  D  P  L  D  K  F  I  Y  W  V  D  G  R   997

3061  CCAGAACATCAAGCGAGCCAAGGACGACGGGACCCAGCCCTTTGTTTTGACCTCTCTGAG  3120
 998   Q  N  I  K  R  A  K  D  D  G  T  Q  P  F  V  L  T  S  L  S  1017

3121  CCAAGGCCAAAACCCAGACAGGCAGCCCCACGACCTCAGCATCGACATCTACAGCCGGAC  3180
1018   Q  G  Q  N  P  D  R  Q  P  H  D  L  S  I  D  I  Y  S  R  T  1037

3181  ACTGTTCTGGACGTGCGAGGCCACCAATACCATCAACGTCCACAGGCTGAGCGGGGAAGC  3240
1038   L  F  W  T  C  E  A  T  N  T  I  N  V  H  R  L  S  G  E  A  1057

3241  CATGGGGGTGGTGCTGCGTGGGGACCGCGACAAGCCCAGGGCCATCGTCGTCAACGCGGA  3300
1058   M  G  V  V  L  R  G  D  R  D  K  P  R  A  I  V  V  N  A  E  1077
```

FIG. 6G

```
3301  GCGAGGGTACCTGTACTTCACCAACATGCAGGACCGGGCAGCCAAGATCGAACGCGCAGC   3360
1078   R  G  Y  L  Y  F  T  N  M  Q  D  R  A  A  K  I  E  R  A  A   1097

3361  CCTGGACGGCACCGAGCGCGAGGTCCTCTTCACCACCGGCCTCATCCGCCCTGTGGCCCT   3420
1098   L  D  G  T  E  R  E  V  L  F  T  T  G  L  I  R  P  V  A  L   1117

3421  GGTGGTGGACAACACACTGGGCAAGCTGTTCTGGGTGGACGCGGACCTGAAGCGCATTGA   3480
1118   V  V  D  N  T  L  G  K  L  F  W  V  D  A  D  L  K  R  I  E   1137

3481  GAGCTGTGACCTGTCAGGGGCCAACCGCCTGACCCTGGAGGACGCCAACATCGTGCAGCC   3540
1138   S  C  D  L  S  G  A  N  R  L  T  L  E  D  A  N  I  V  Q  P   1157

3541  TCTGGGCCTGACCATCCTTGGCAAGCATCTCTACTGGATCGACCGCCAGCAGCAGATGAT   3600
1158   L  G  L  T  I  L  G  K  H  L  Y  W  I  D  R  Q  Q  Q  M  I   1177

3601  CGAGCGTGTGGAGAAGACCACCGGGGACAAGCGGACTCGCATCCAGGGCCGTGTCGCCCA   3660
1178   E  R  V  E  K  T  T  G  D  K  R  T  R  I  Q  G  R  V  A  H   1197

3661  CCTCACTGGCATCCATGCAGTGGAGGAAGTCAGCCTGGAGGAGTTCTCAGCCCACCCATG   3720
1198   L  T  G  I  H  A  V  E  E  V  S  L  E  E  F  S  A  H  P  C   1217

3721  TGCCCGTGACAATGGTGGCTGCTCCCACATCTGTATTGCCAAGGGTGATGGGACACCACG   3780
1218   A  R  D  N  G  G  C  S  H  I  C  I  A  K  G  D  G  T  P  R   1237

3781  GTGCTCATGCCCAGTCCACCTCGTGCTCCTGCAGAACCTGCTGACCTGTGGAGAGCCGCC   3840
1238   C  S  C  P  V  H  L  V  L  L  Q  N  L  L  T  C  G  E  P  P   1257
```

SUBSTITUTE SHEET (RULE 26)

FIG. 6H

```
3841  CACCTGCTCCCCGGACCAGTTTGCATGTGCCACAGGGGAGATCGACTGTATCCCCGGGGC  3900
1258   T  C  S  P  D  Q  F  A  C  A  T  G  E  I  D  C  I  P  G  A   1277

3901  CTGGCGCTGTGACGGCTTTCCCGAGTGCGATGACCAGAGCGACGAGGAGGGCTGCCCCGT  3960
1278   W  R  C  D  G  F  P  E  C  D  D  Q  S  D  E  E  G  C  P  V   1297

3961  GTGCTCCGCCGCCCAGTTCCCCTGCGCGCGGGGTCAGTGTGTGGACCTGCGCCTGCGCTG  4020
1298   C  S  A  A  Q  F  P  C  A  R  G  Q  C  V  D  L  R  L  R  C   1317

4021  CGACGGCGAGGCAGACTGTCAGGACCGCTCAGACGAGGTGGACTGTGACGCCATCTGCCT  4080
1318   D  G  E  A  D  C  Q  D  R  S  D  E  V  D  C  D  A  I  C  L   1337

4081  GCCCAACCAGTTCCGGTGTGCGAGCGGCCAGTGTGTCCTCATCAAACAGCAGTGCGACTC  4140
1338   P  N  Q  F  R  C  A  S  G  Q  C  V  L  I  K  Q  Q  C  D  S   1357

4141  CTTCCCCGACTGTATCGACGGCTCCGACGAGCTCATGTGTGAAATCACCAAGCCGCCCTC  4200
1358   F  P  D  C  I  D  G  S  D  E  L  M  C  E  I  T  K  P  P  S   1377

4201  AGACGACAGCCCGGCCCACAGCAGTGCCATCGGGCCCGTCATTGGCATCATCCTCTCTCT  4260
1378   D  D  S  P  A  H  S  S  A  I  G  P  V  I  G  I  I  L  S  L   1397

4261  CTTCGTCATGGGTGGTGTCTATTTTGTGTGCCAGCGCGTGGTGTGCCAGCGCTATGCGGG  4320
1398   F  V  M  G  G  V  Y  F  V  C  Q  R  V  V  C  Q  R  Y  A  G   1417

4321  GGCCAACGGGCCCTTCCCGCACGAGTATGTCAGCGGGACCCCGCACGTGCCCCTCAATTT  4380
1418   A  N  G  P  F  P  H  E  Y  V  S  G  T  P  H  V  P  L  N  F   1437
```

SUBSTITUTE SHEET (RULE 26)

FIG. 6I

```
4381  CATAGCCCCGGGCGGTTCCCAGCATGGCCCCTTCACAGGCATCGCATGCGGAAAGTCCAT  4440
1438   I  A  P  G  G  S  Q  H  G  P  F  T  G  I  A  C  G  K  S  M   1457

4441  GATGAGCTCCGTGAGCCTGATGGGGGGCCGGGGCGGGGTGCCCCTCTACGACCGGAACCA  4500
1458   M  S  S  V  S  L  M  G  G  R  G  G  V  P  L  Y  D  R  N  H   1477

4501  CGTCACAGGGGCCTCGTCCAGCAGCTCGTCCAGCACGAAGGCCACGCTGTACCCGCCGAT  4560
1478   V  T  G  A  S  S  S  S  S  S  S  T  K  A  T  L  Y  P  P  I   1497

4561  CCTGAACCCGCCGCCCTCCCCGGCCACGGACCCCTCCCTGTACAACATGGACATGTTCTA  4620
1498   L  N  P  P  P  S  P  A  T  D  P  S  L  Y  N  M  D  M  F  Y   1517

4621  CTCTTCAAACATTCCGGCCACTGCGAGACCGTACAGGCCCTACATCATTCGAGGAATGGC  4680
1518   S  S  N  I  P  A  T  A  R  P  Y  R  P  Y  I  I  R  G  M  A   1537

4681  GCCCCCGACGACGCCCTGCAGCACCGACGTGTGTGACAGCGACTACAGCGCCAGCCGCTG  4740
1538   P  P  T  T  P  C  S  T  D  V  C  D  S  D  Y  S  A  S  R  W   1557

4741  GAAGGCCAGCAAGTACTACCTGGATTTGAACTCGGACTCAGACCCCTATCCACCCCCACC  4800
1558   K  A  S  K  Y  Y  L  D  L  N  S  D  S  D  P  Y  P  P  P  P   1577

4801  CACGCCCCACAGCCAGTACCTGTCGGCGGAGGACAGCTGCCCGCCCTCGCCCGCCACCGA  4860
1578   T  P  H  S  Q  Y  L  S  A  E  D  S  C  P  P  S  P  A  T  E   1597

4861  GAGGAGCTACTTCCATCTCTTCCCGCCCCCTCCGTCCCCCTGCACGGACTCATCCTGACC  4920
1598   R  S  Y  F  H  L  F  P  P  P  P  S  P  C  T  D  S  S        1615
```

SUBSTITUTE SHEET (RULE 26)

FIG. 6J

```
                                                                  4980
                                                                  5040
                                                                  5100
4921  TCGGCCCGGGCCCACTCTCCTGTAAATAGTTTTTAAATGAACAAAGA
4981  AAAAAATATATATTTTATATAATTTGGGATTTTTAAAACATGAGAA
5041  TGTGAACTGTGATGGGGTGGGCAGGGCTGGGAGAACTTTGTACAGTGGAGAAATATTTAT
5101  AAACTTAATTTGTAAAACA  5120
```

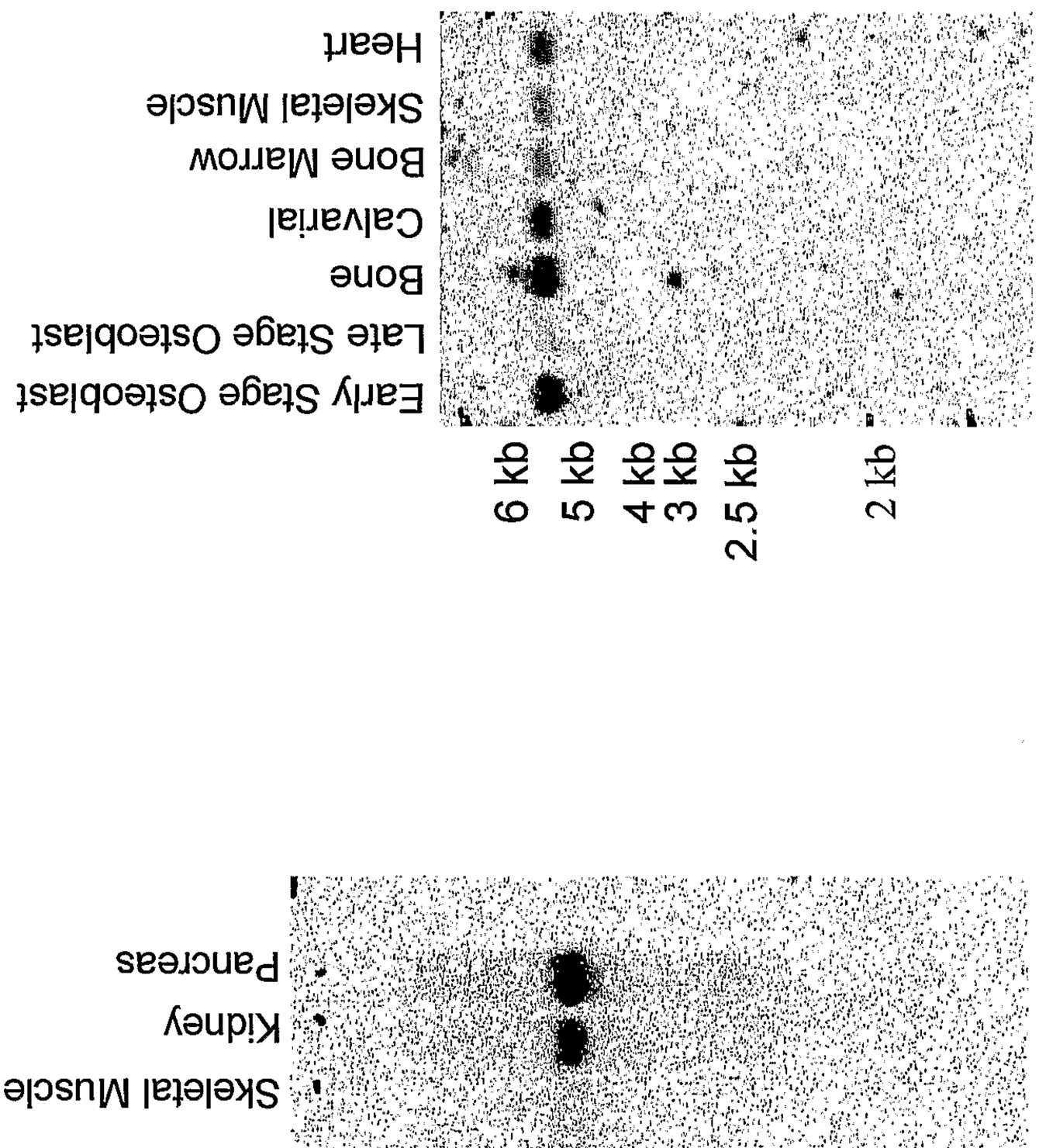

Northern Blot Analysis - Zmax 1
2 kb
1.5 kb
3 kb
4 kb
5 kb
6 kb
Early Stage Osteoblast
Late Stage Osteoblast
Bone
Calvarial
Bone Marrow
Skeletal Muscle
Heart

FIG. 7B

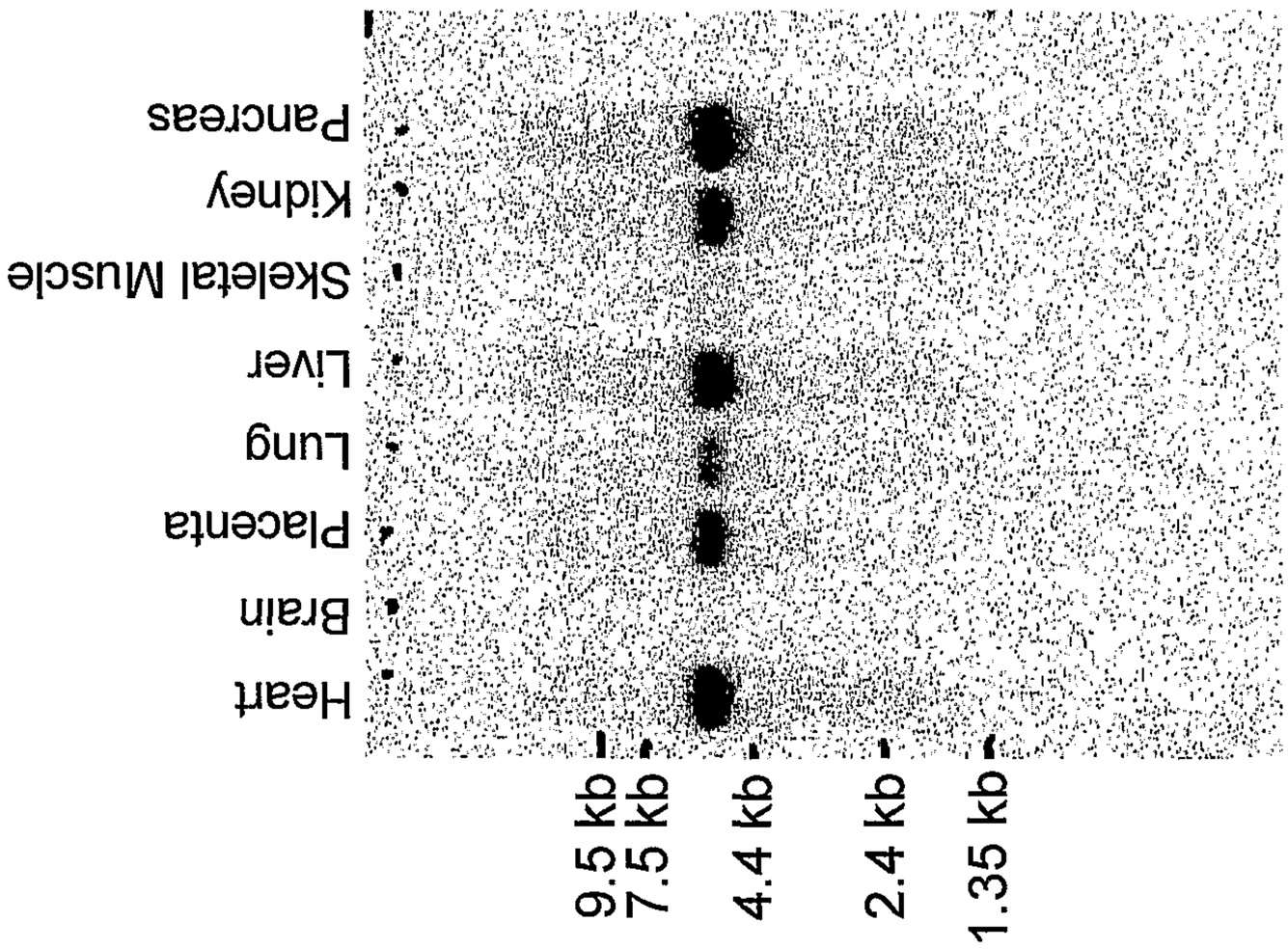

Northern Blot Analysis - Zmax 1
1.35 kb
2.4 kb
4.4 kb
7.5 kb
9.5 kb
Heart
Brain
Placenta
Lung
Liver
Skeletal Muscle
Kidney
Pancreas

FIG. 7A

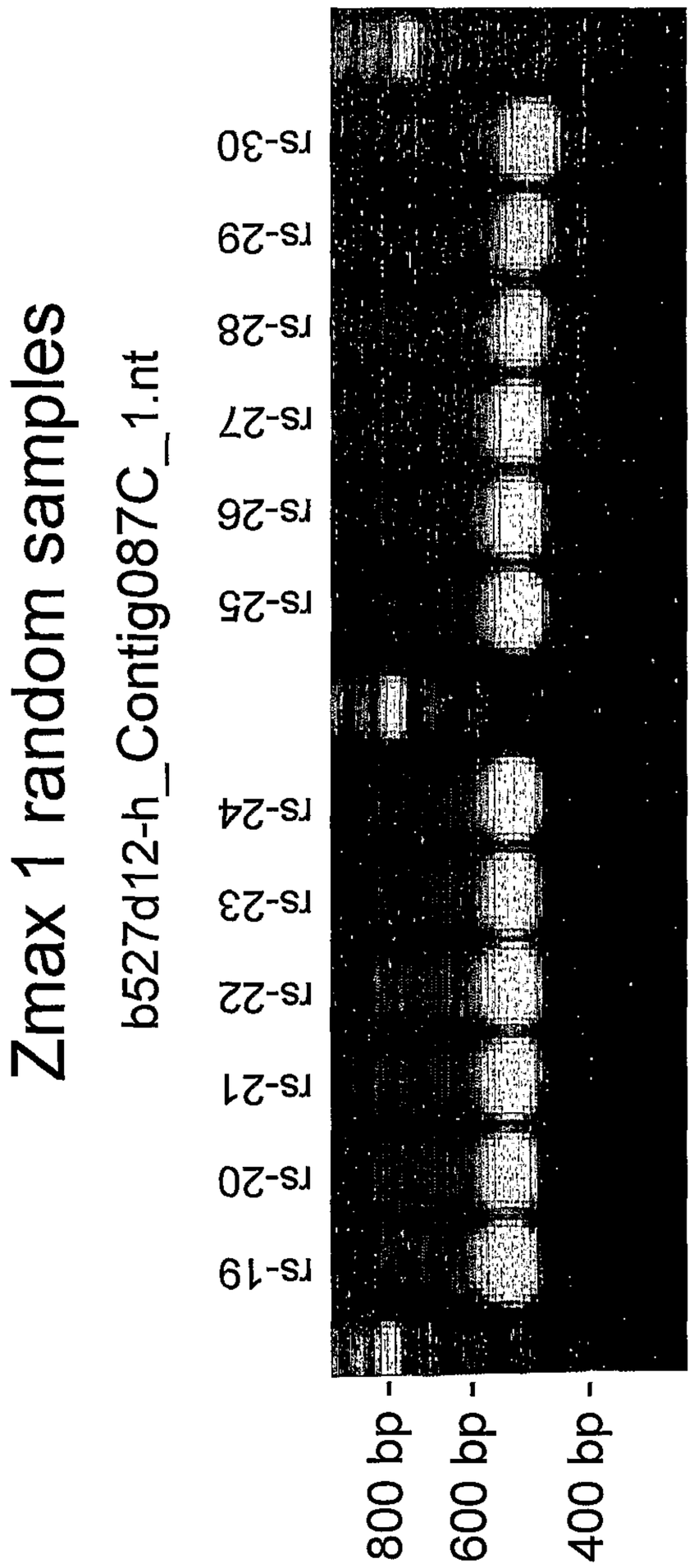

Zmax 1 random samples
b527d12-h_Contig087C_1.nt
rs-30
rs-29
rs-28
rs-27
rs-26
rs-25
rs-24
rs-23
rs-22
rs-21
rs-20
rs-19
800 bp -
600 bp -
400 bp -

FIG. 8

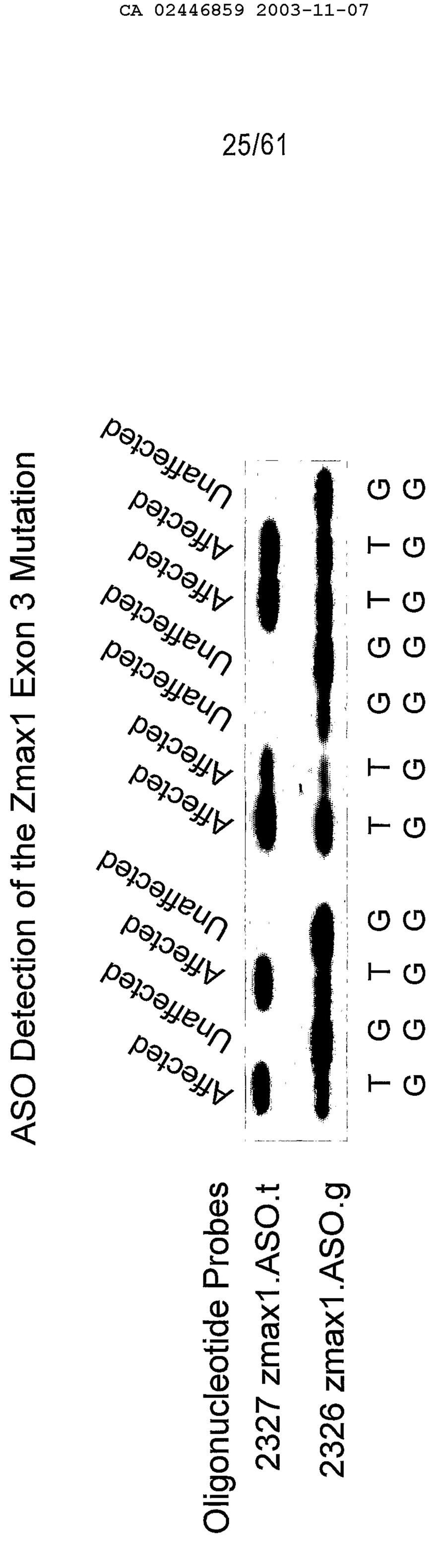

ASO Detection of the Zmax1 Exon 3 Mutation
Affected
Unaffected
Affected
Affected
Unaffected
Unaffected
Affected
Unaffected
Affected
Unaffected
Affected
Unaffected
Probes
Oligonucleotide
2327 zmax1.ASO.t
2326 zmax1.ASO.g
T T G G T T G
G G G G G G G
Genotype

FIG. 9

Mouse Zmax1 In situ hybridization
100X Magnification

Antisense probe



FIG. 10A

Mouse Zmax1 In situ hybridization
100X Magnification

Sense probe



FIG. 10B

Mouse Zmax1 In situ hybridization
400X Magnification
Antisense probe

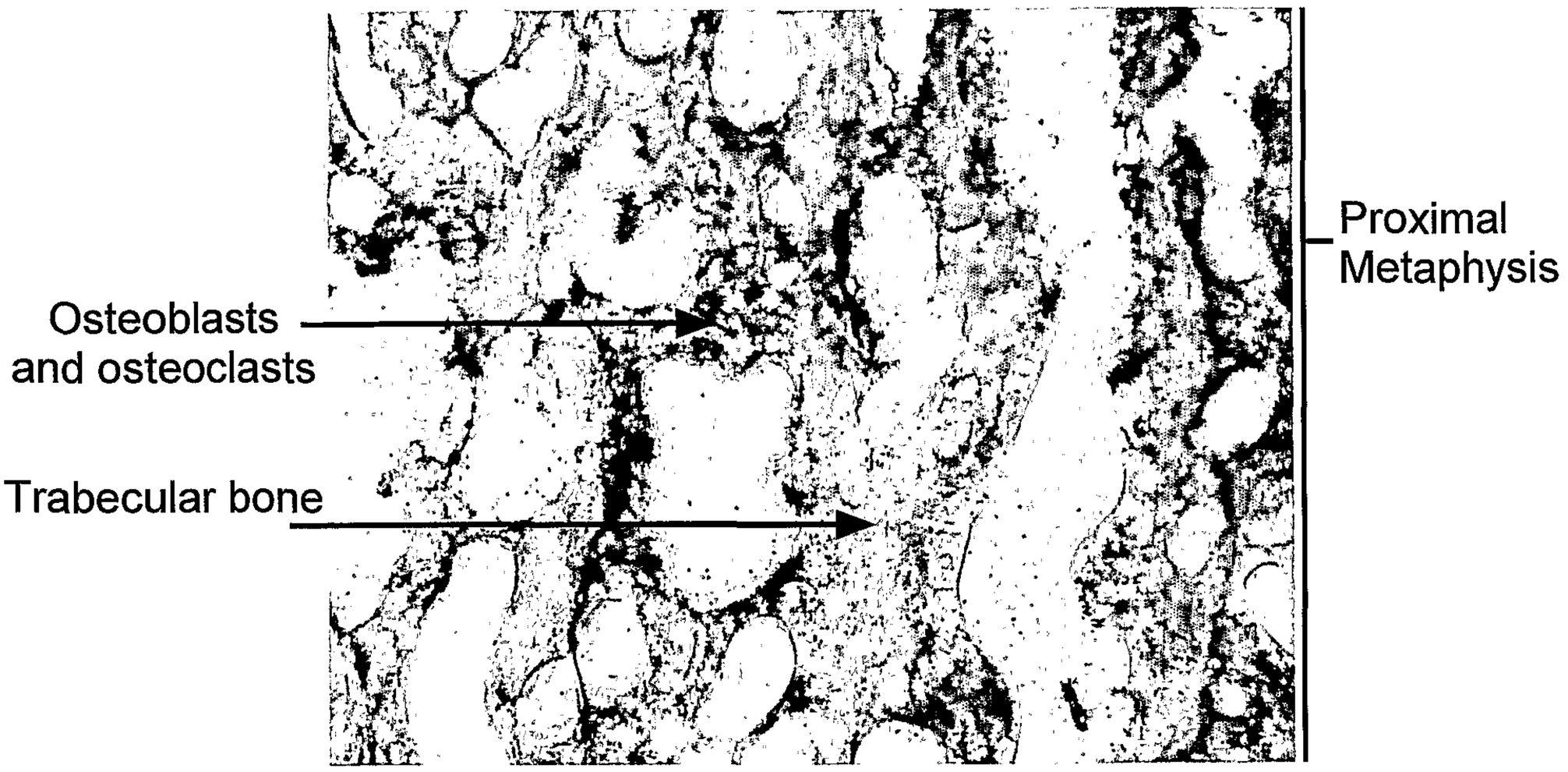


FIG. 11A

Mouse Zmax1 In situ hybridization
400X Magnification
Sense probe

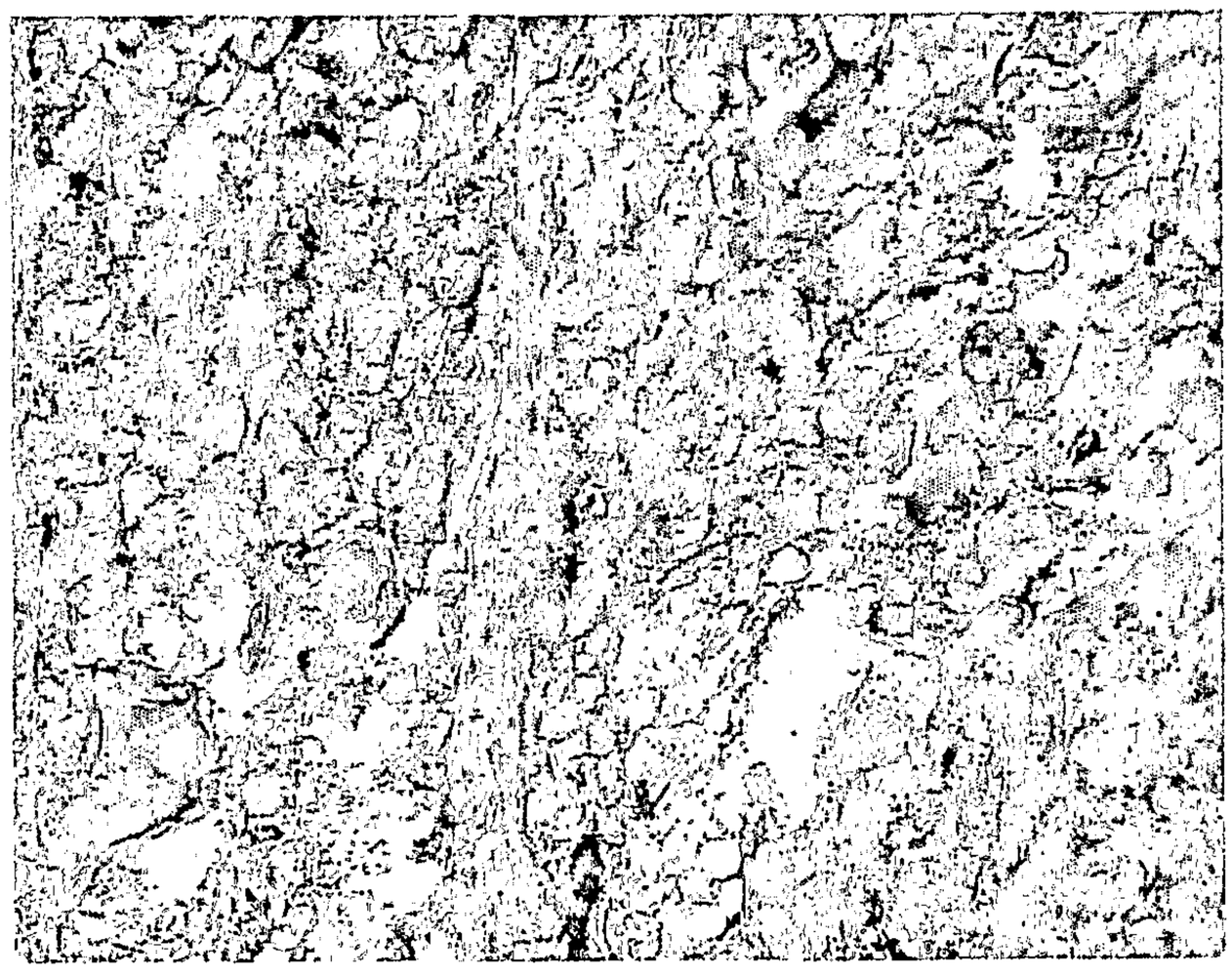


FIG. 11B

Mouse Zmax1 In situ hybridization
400X Magnification
Antisense probe

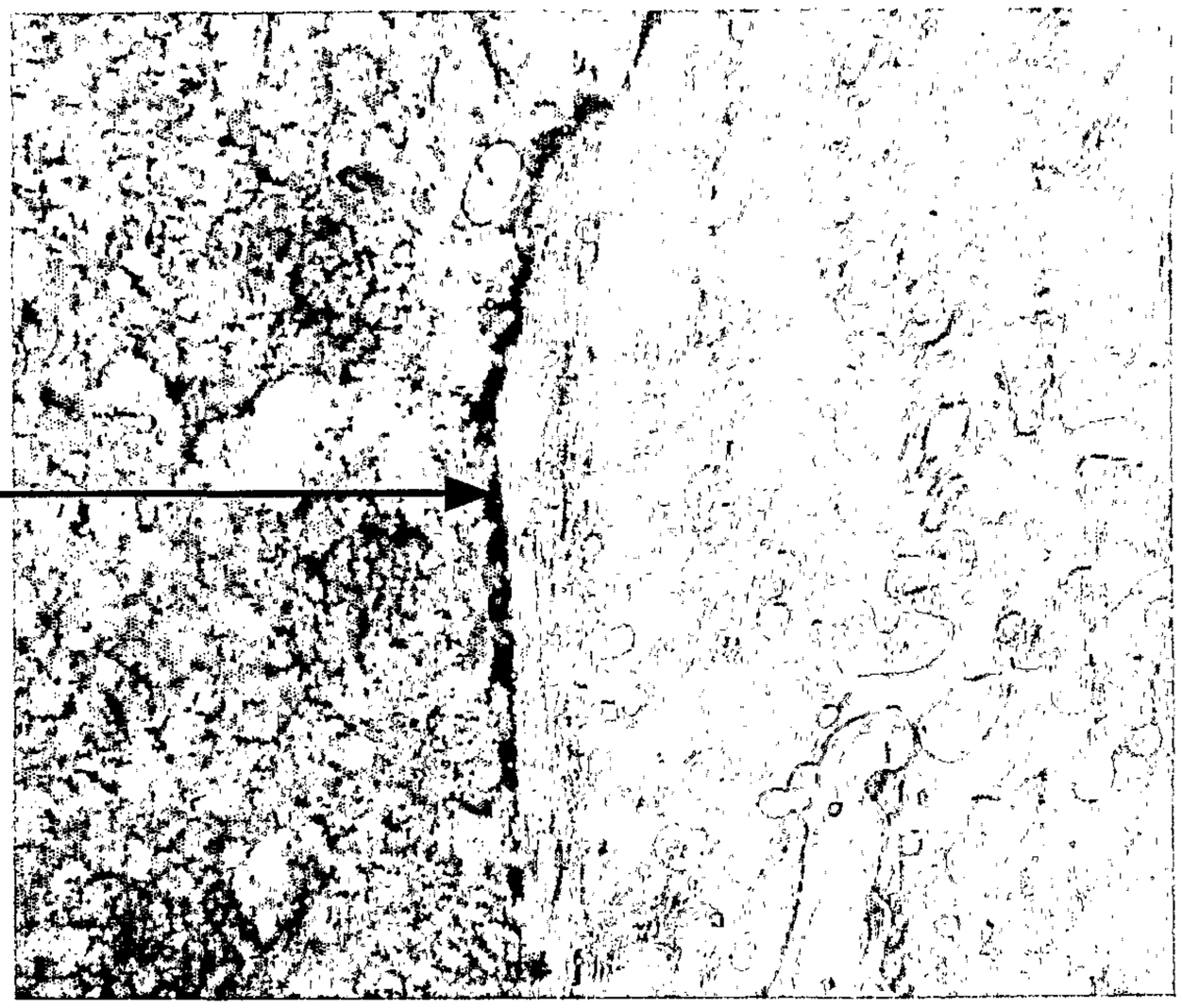


FIG. 12A

Mouse Zmax1 In situ hybridization
400X Magnification
Sense probe

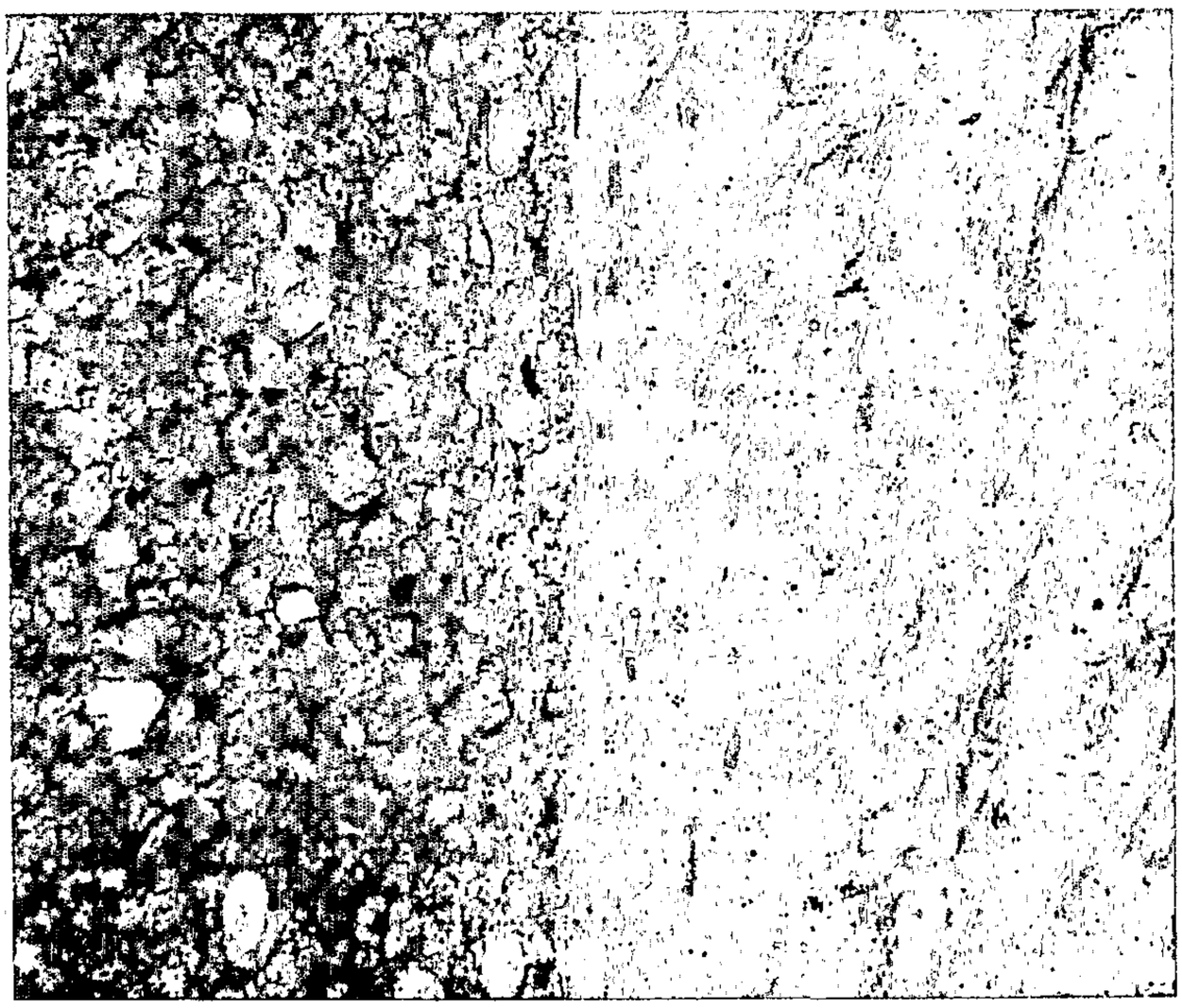
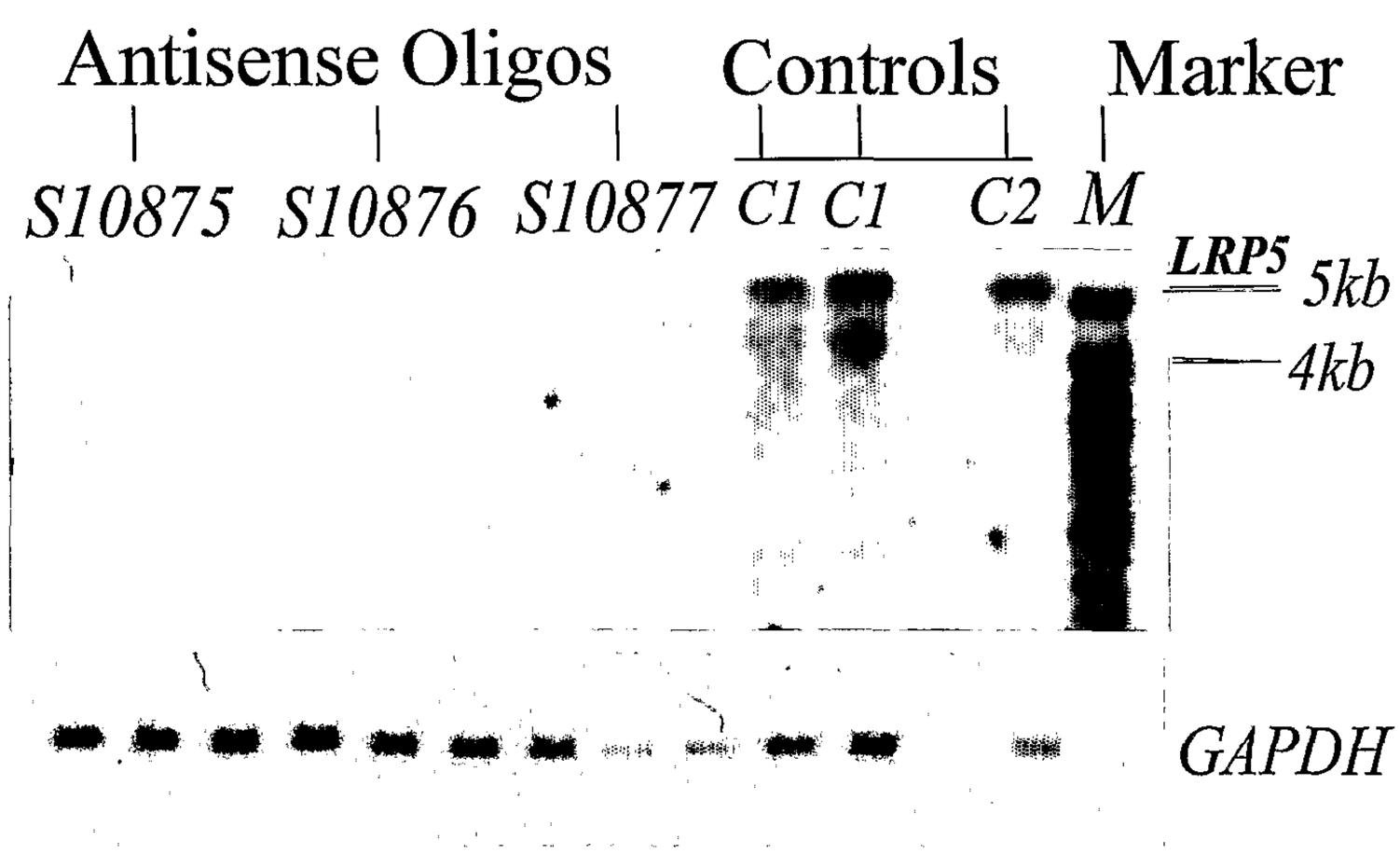


FIG. 12B

Antisense Inhibition of Zmax1 Expression



MC-3T3 cells

FIG. 13

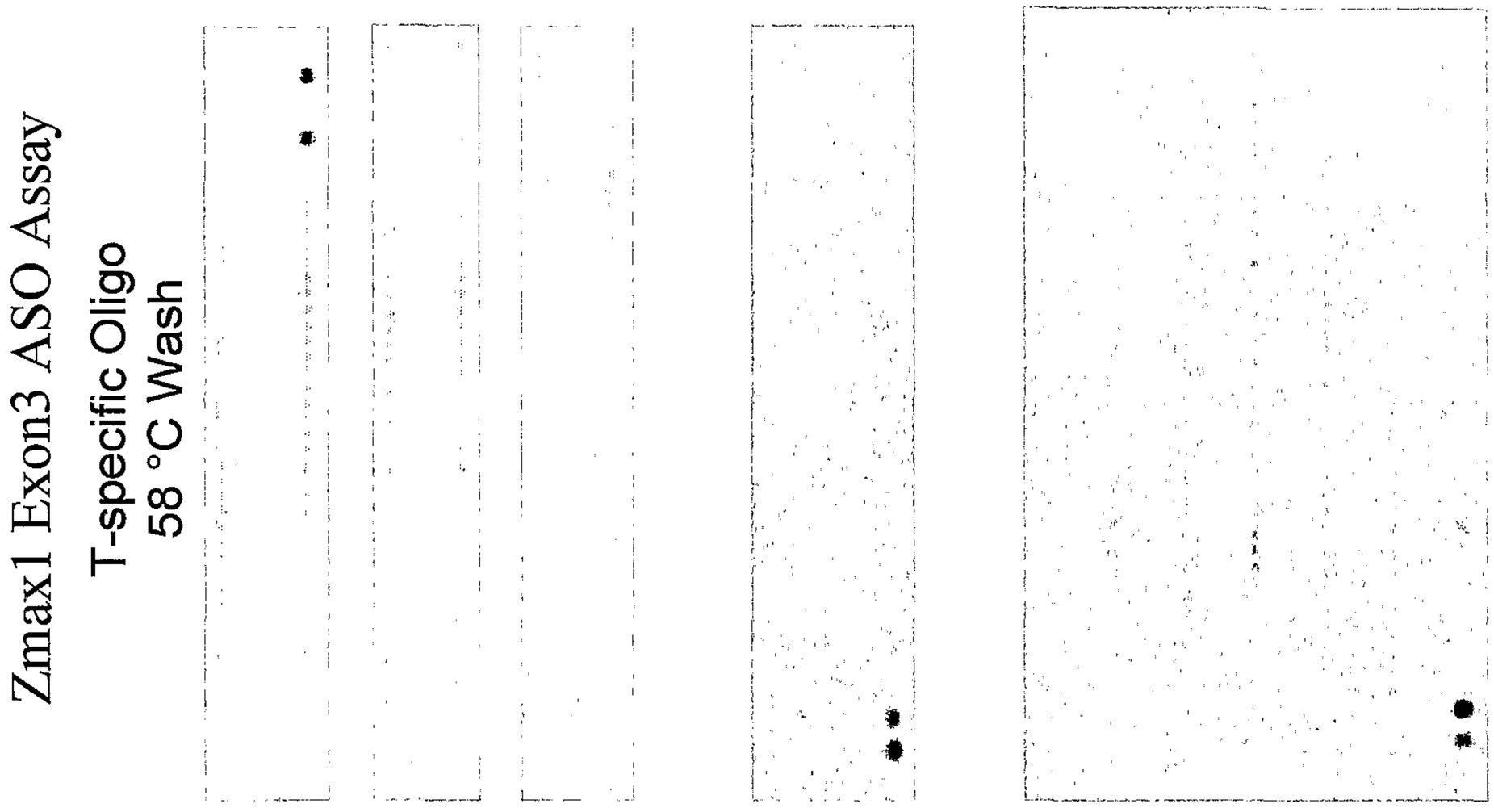


FIG. 14B

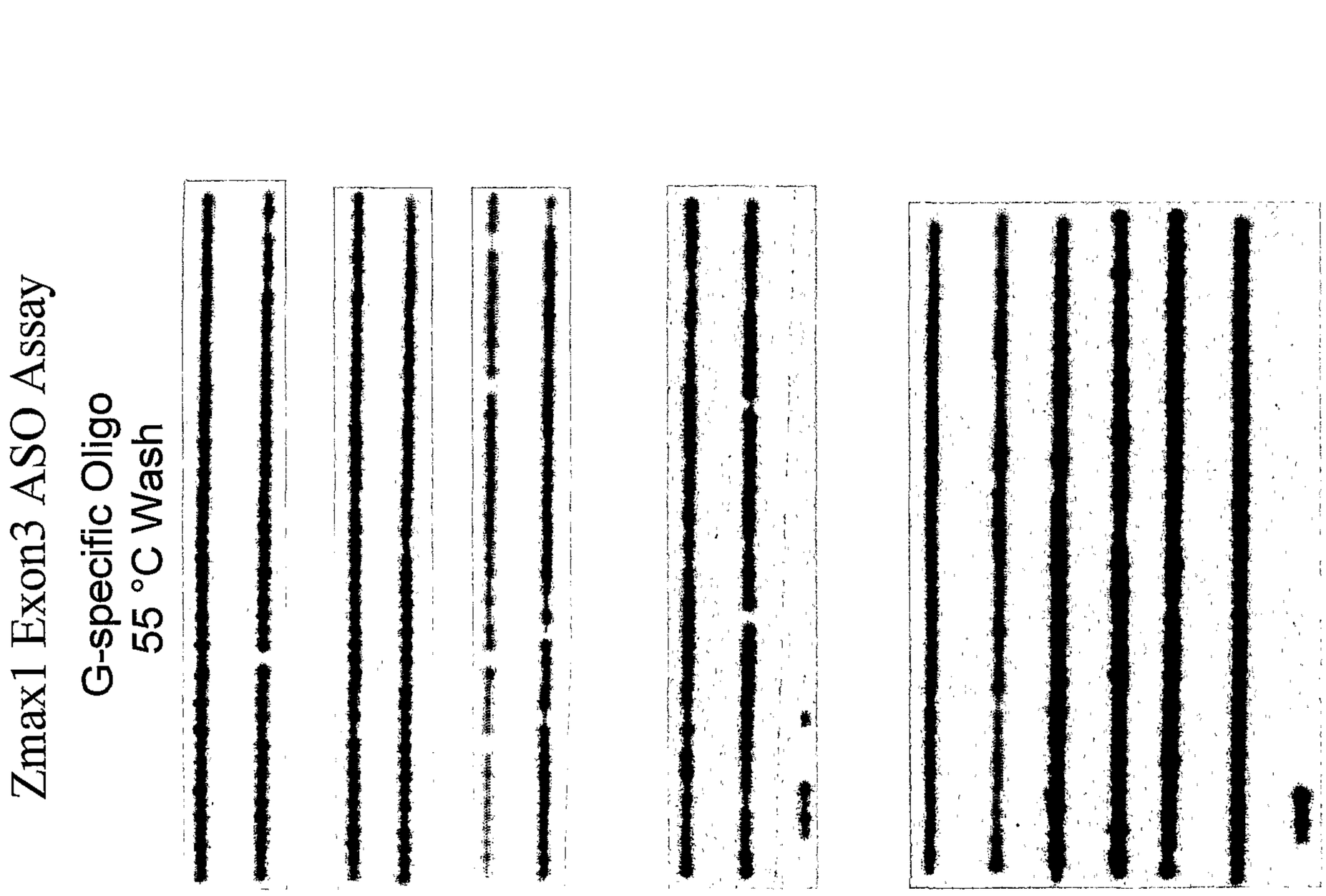


FIG. 14A

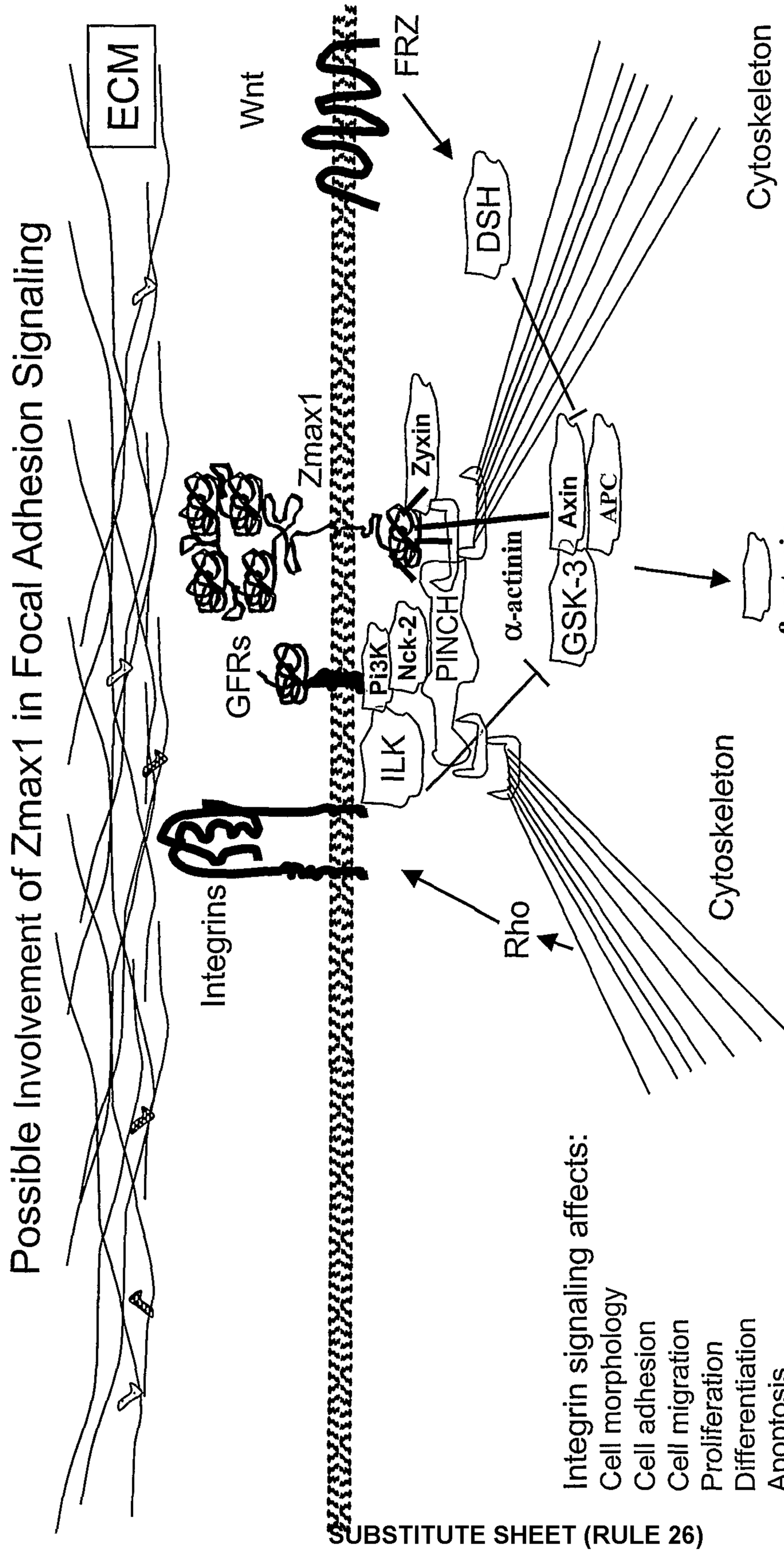

Possible Involvement of Zmax1 in Focal Adhesion Signaling
ECM
Integrins
GFRs
Zmax1
Wnt
FRZ
Pi3K
Nck-2
PINCH
Zyxin
ILK
α-actinin
DSH
Axin
APC
GSK-3
ß-catenin
Rho
Cytoskeleton
Cytoskeleton
Nucleus
Integrin signaling affects:
Cell morphology
Cell adhesion
Cell migration
Proliferation
Differentiation
Apoptosis

FIG. 15

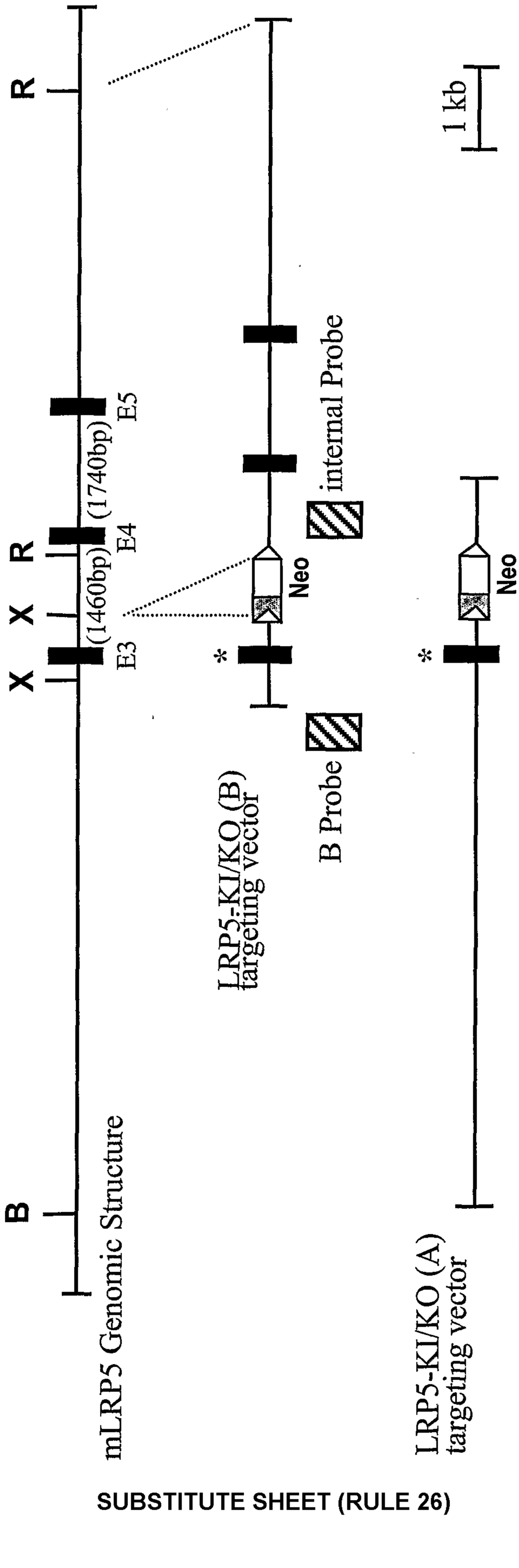

ZMAX Gene Targeting Knock-in / Knock-out vectors
B
mLRP5 Genomic Structure
X X R
(1460bp)
(1740bp)
E3
E4
E5
LRP5-KI/KO (B)
targeting vector
Neo
internal Probe
B Probe
R
1 kb
LRP5-KI/KO (A)
targeting vector
Neo
A Probe

FIG. 16

HBM: Constructs for transgenic mice

Confirmation of expression

*Transient transfection into HOB-02-02 cells**

CMV βActin		Type I collagen	
HBM X 1,000	Zmax1 X 1,000	HBM X 10	Zmax1 X 10

**Fold increase compared to Zmax1*
in cells transfected with empty vector

FIG. 17

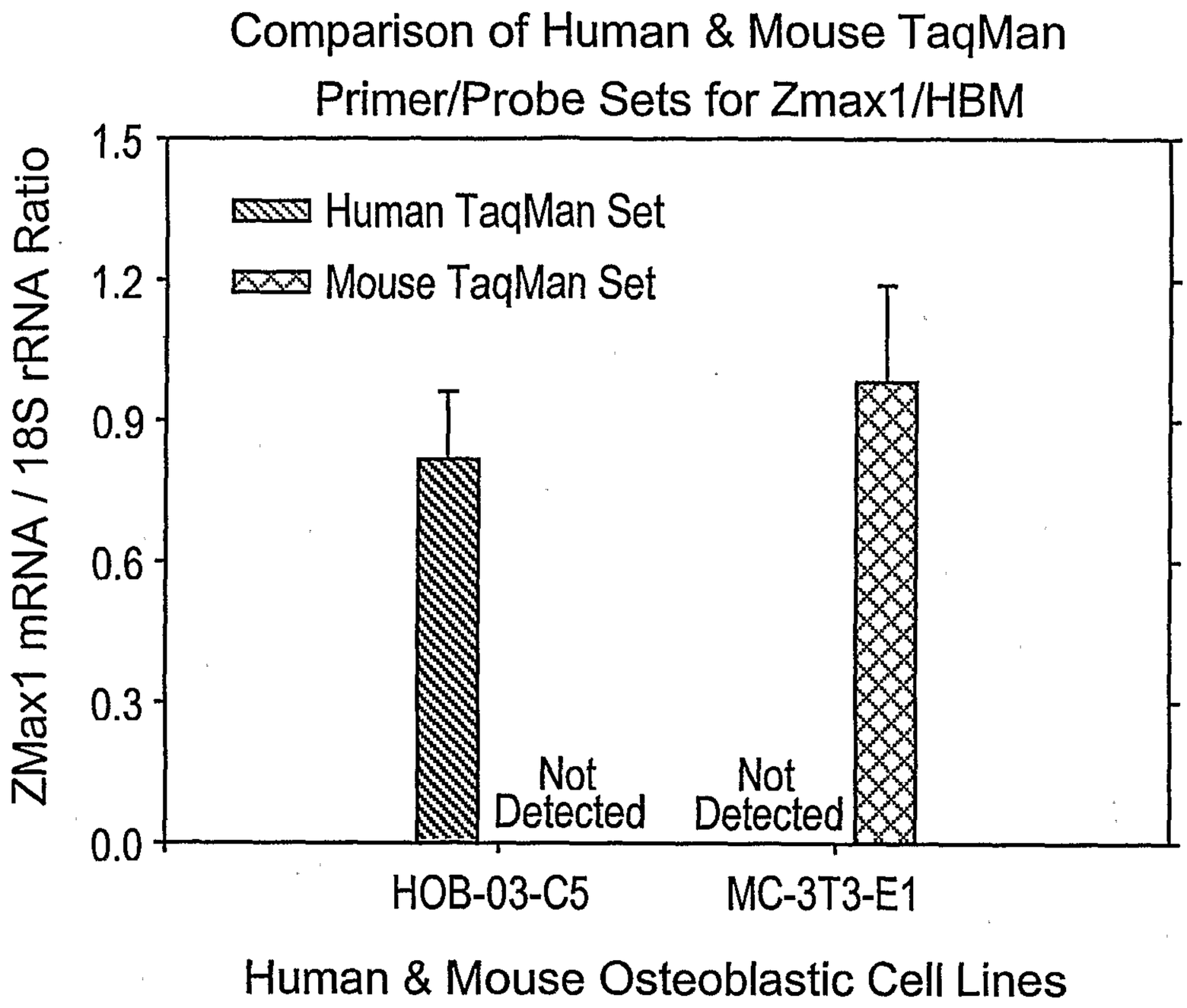

Comparison of Human & Mouse TaqMan
Primer/Probe Sets for Zmax1/HBM
ZMax1 mRNA / 18S rRNA Ratio
1.5
1.2
0.9
0.6
0.3
0.0
Human TaqMan Set
Mouse TaqMan Set
Not Detected
Not Detected
HOB-03-C5
MC-3T3-E1
Human & Mouse Osteoblastic Cell Lines

FIG. 18

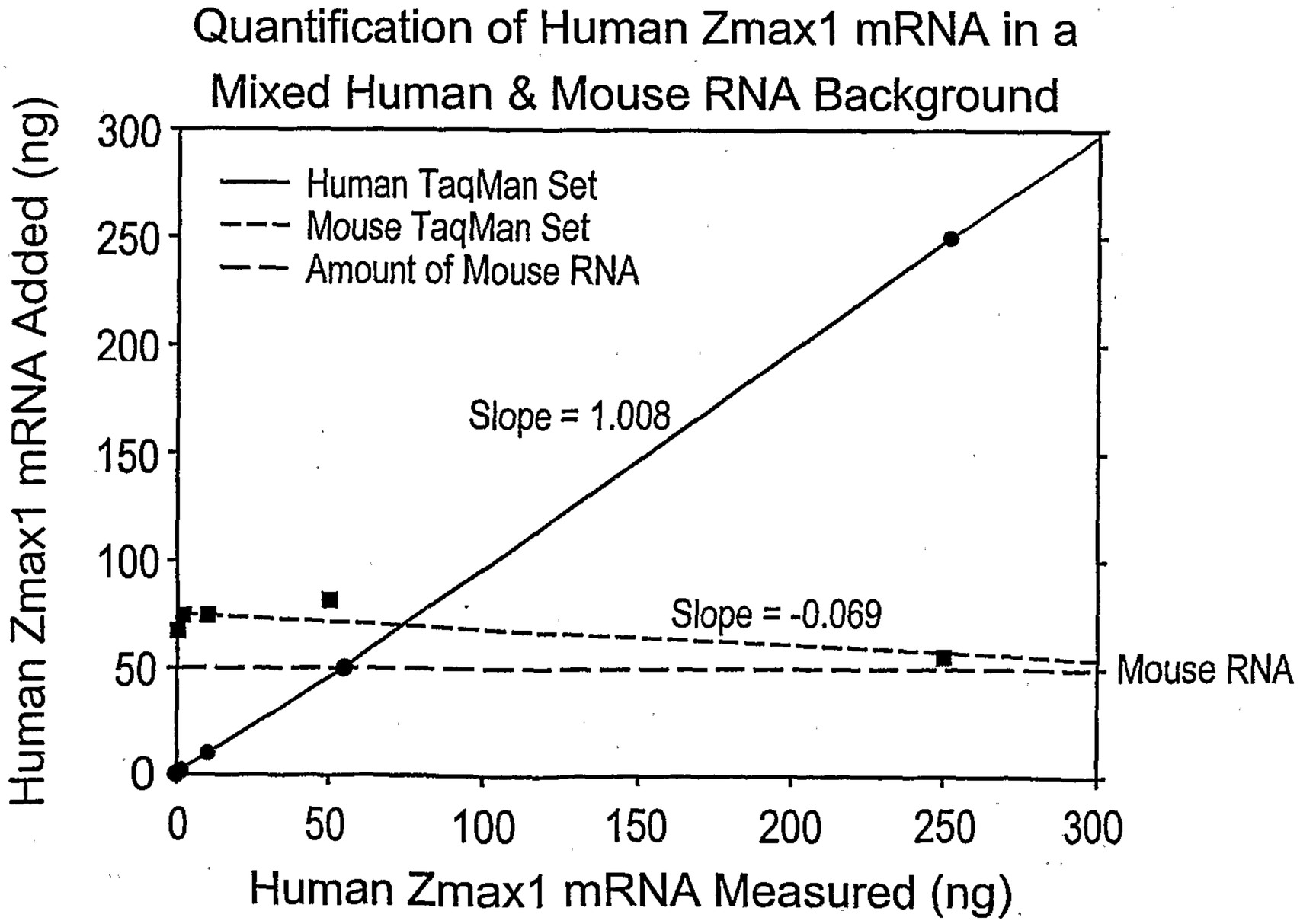

Quantification of Human Zmax1 mRNA in a
Mixed Human & Mouse RNA Background
Human Zmax1 mRNA Added (ng)
300
250
200
150
100
50
0
Human TaqMan Set
Mouse TaqMan Set
Amount of Mouse RNA
Slope = 1.008
Slope = -0.069
Mouse RNA
0
50
100
150
200
250
300
Human Zmax1 mRNA Measured (ng)

FIG. 19

HBM: Transgenic mice

*mRNA expression by Taqman analysis**

	Tissue						
Line	Tibia	Femur	Heart	Gonad	Brain	Kidney	Liver
HBMMCBA							
2	7-10		20-90	2-30	6-11	5-9	<1
13	1-2		6-7	3-4	5-6	<1	<1
18	10-11						
HBMMTIC							
19	7-8	19-20	1				1
35	1	1	0				0

* relative to Zmax1 in HOB-03-C5

FIG. 20

HBM: In vivo pDXA - transgenic mice

HBMMCBA construct

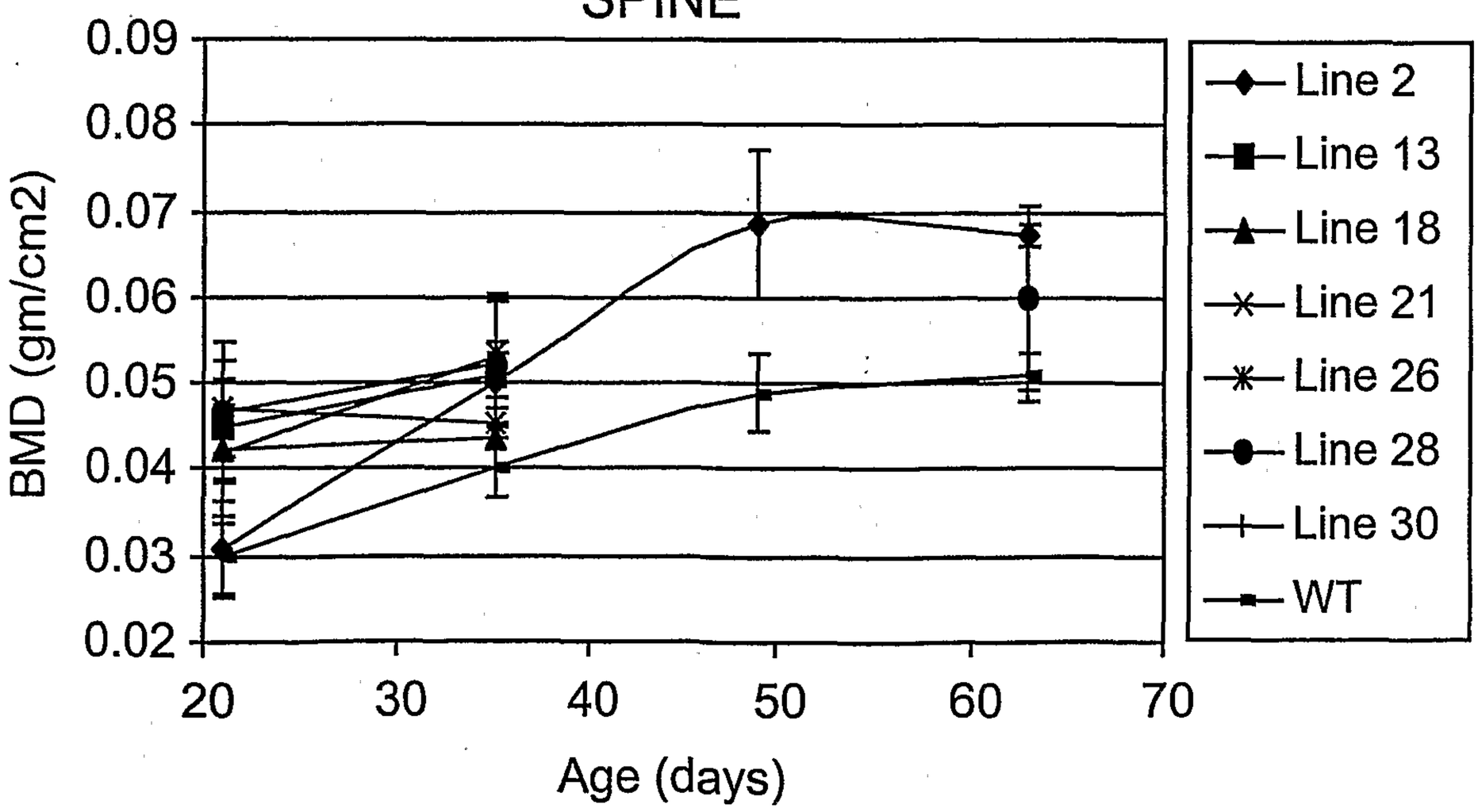


FIG. 21A

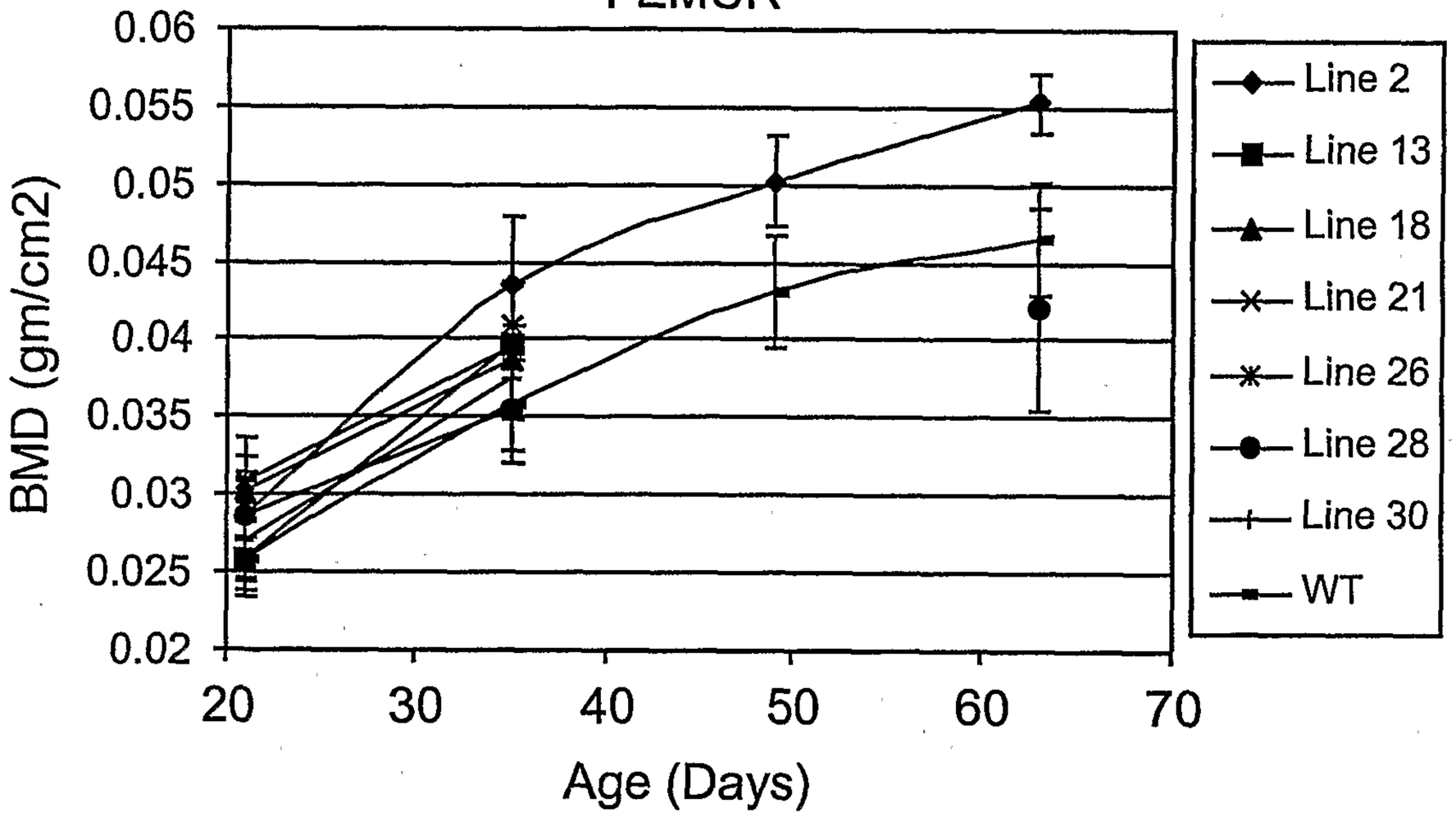


FIG. 21B

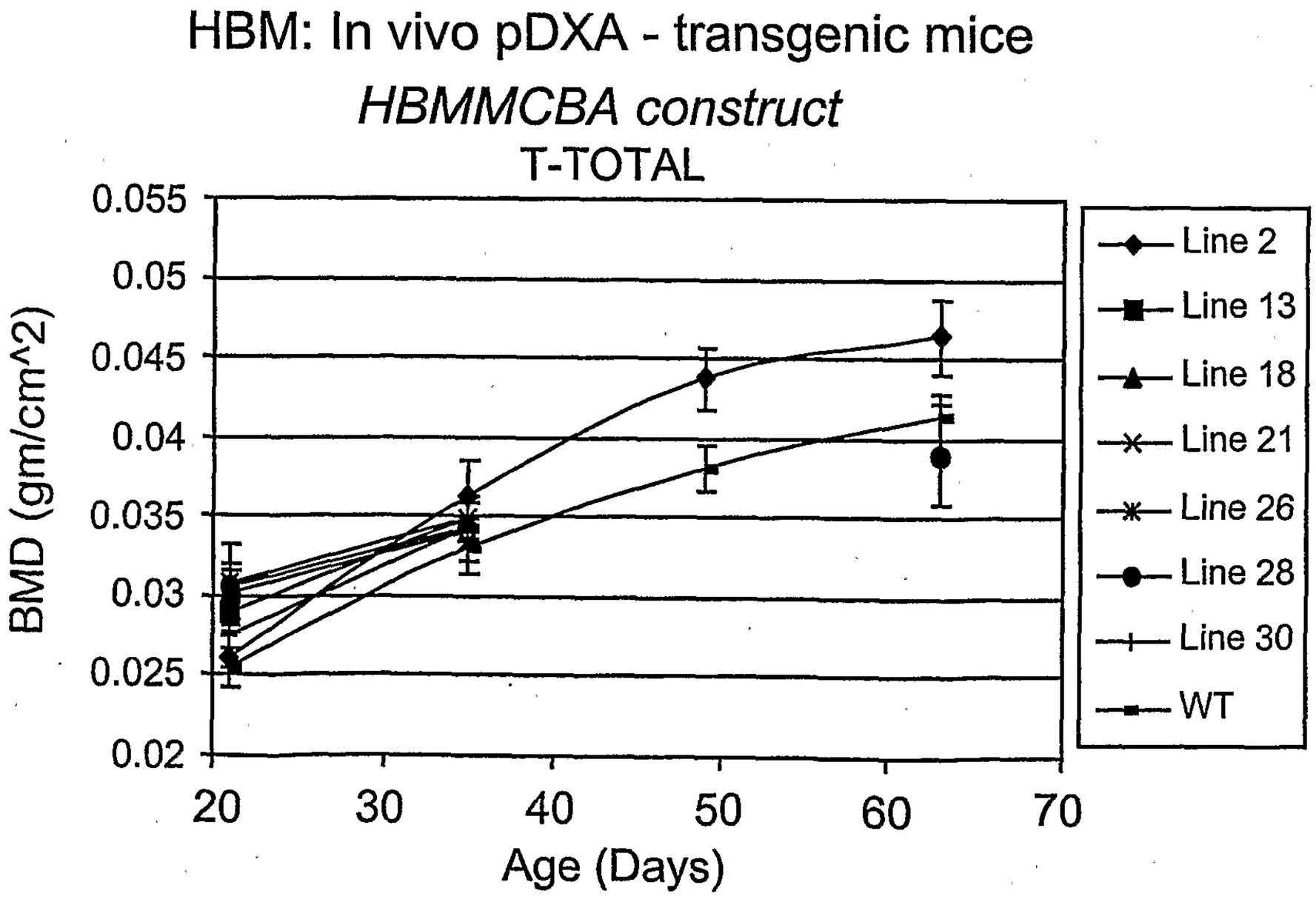


FIG. 21C

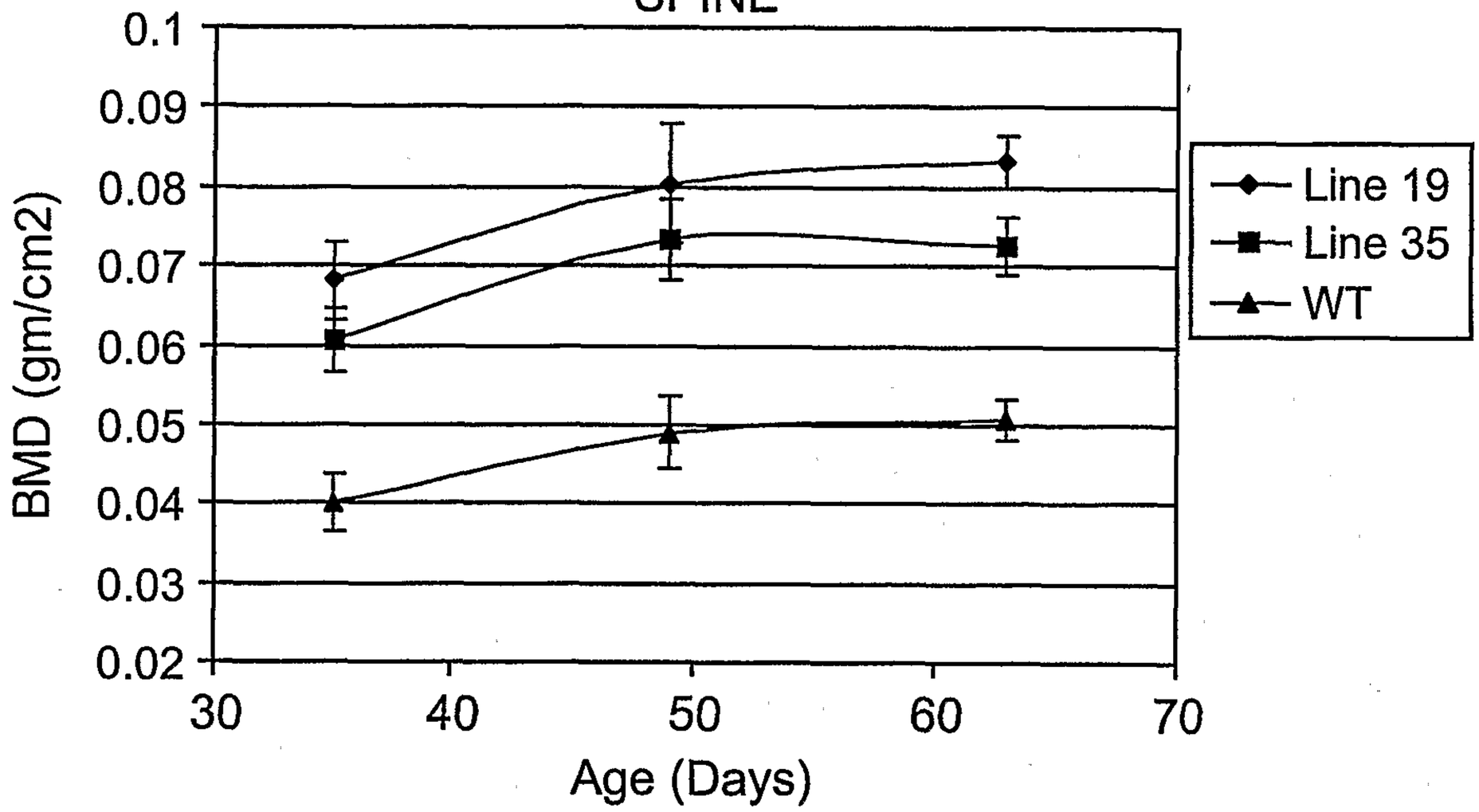


FIG. 21D

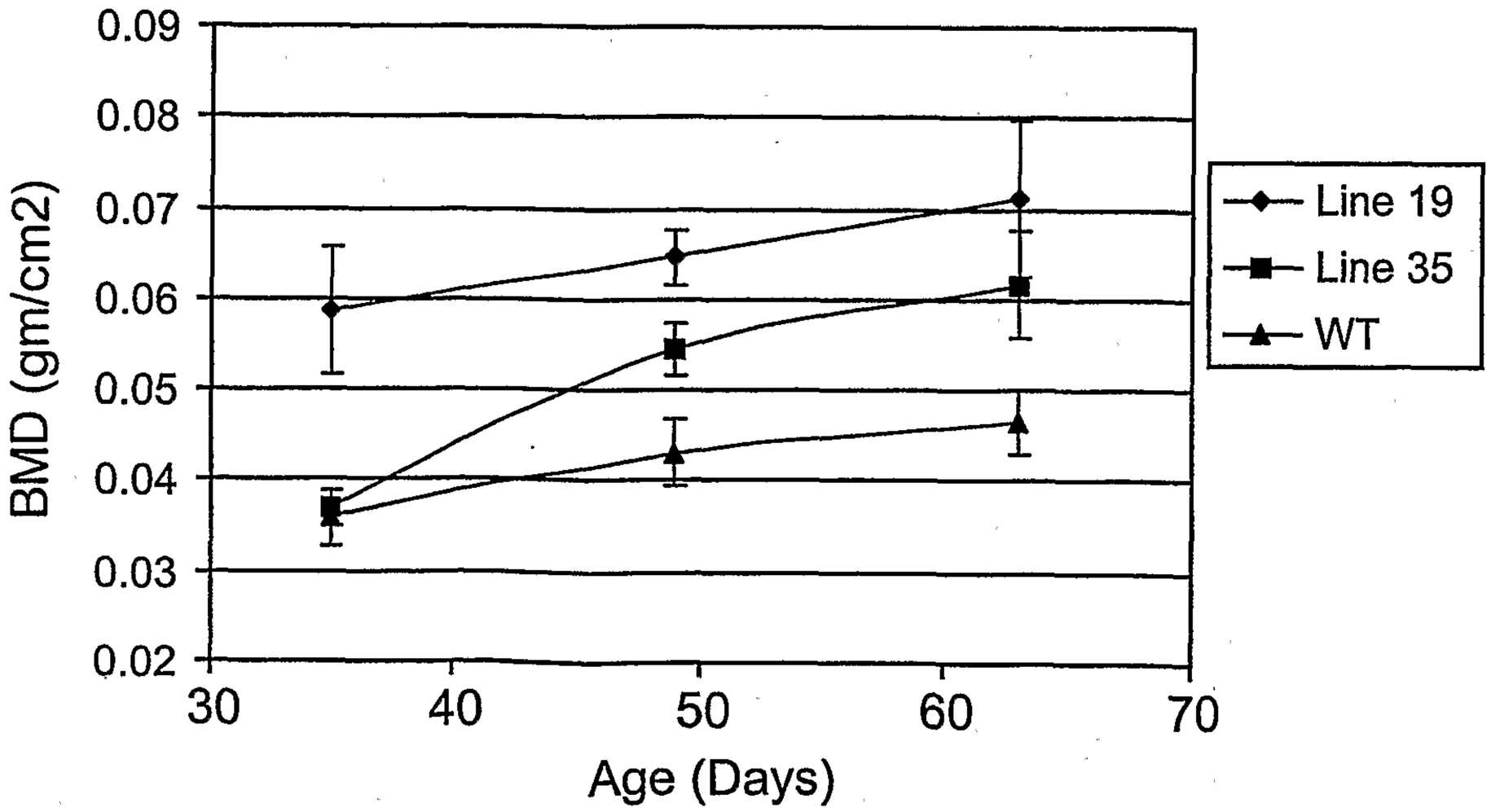


FIG. 21E

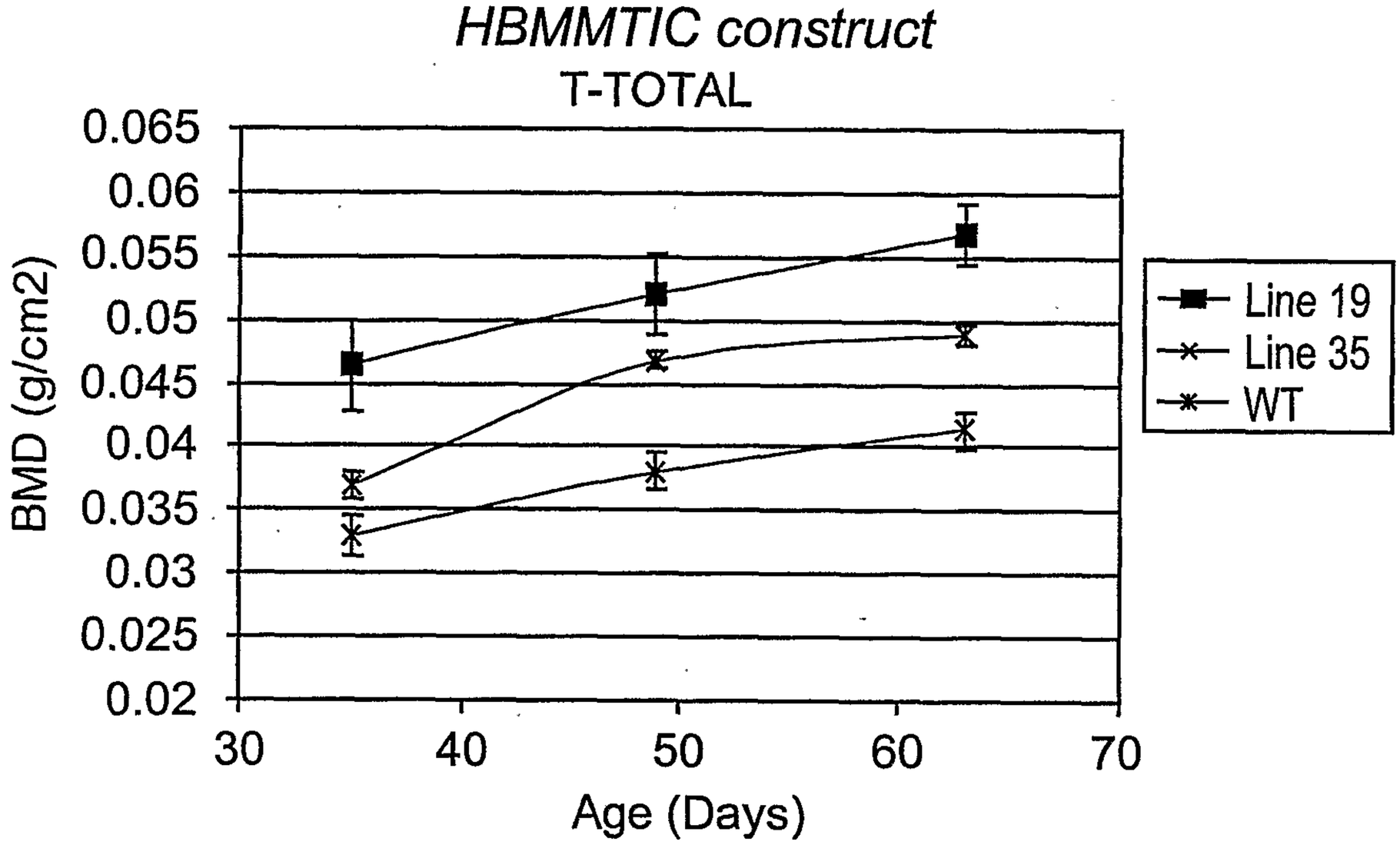

HBM: In vivo pDXA - transgenic mice
HBMMTIC construct
T-TOTAL
BMD (g/cm2)
0.065
0.06
0.055
0.05
0.045
0.04
0.035
0.03
0.025
0.02
30
40
50
60
70
Age (Days)
Line 19
Line 35
WT

FIG. 21F

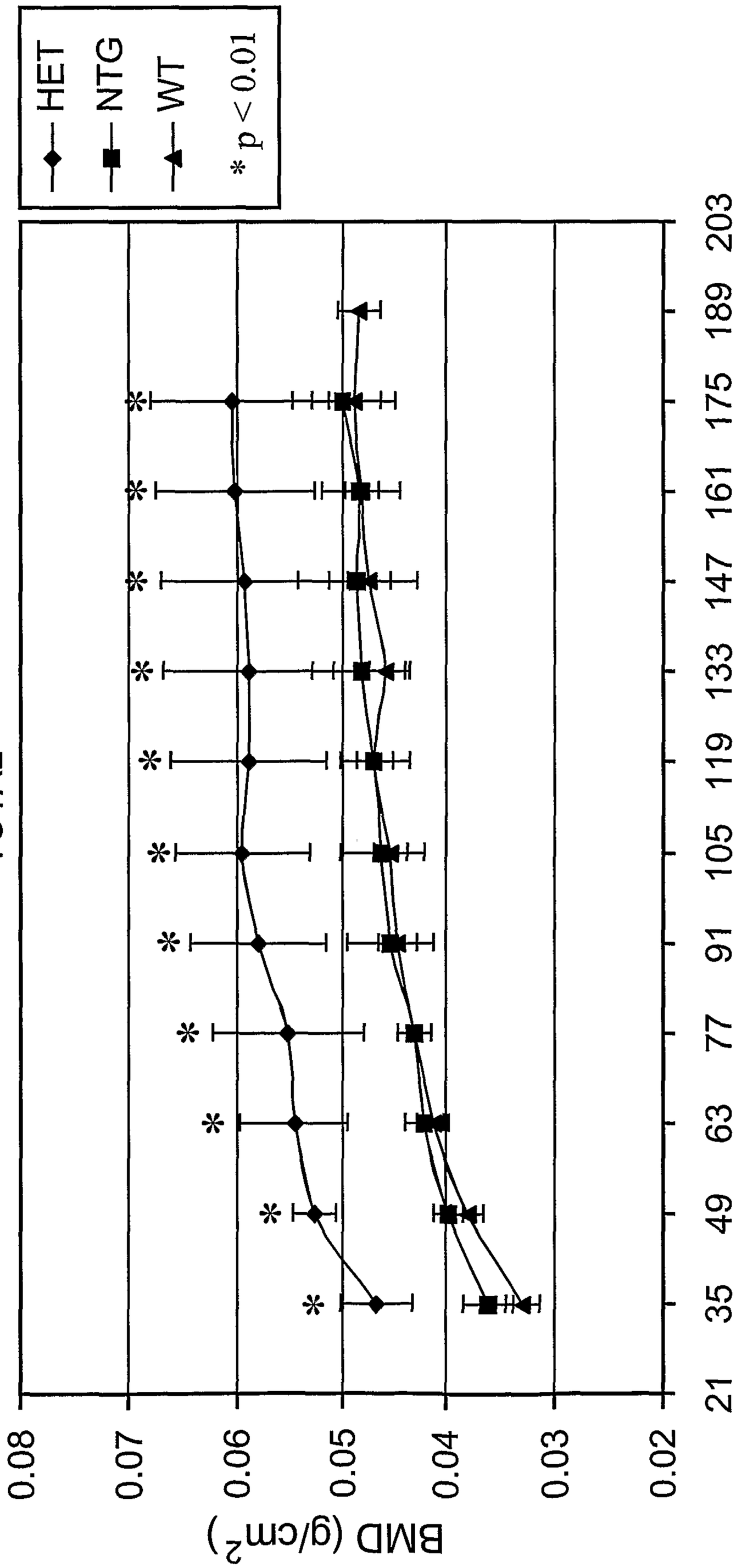

HET
NTG
WT
* p < 0.01
HBM: In vivo pDXA - transgenic mice
HBMMTIC line 19
TOTAL
BMD (g/cm²)
0.08
0.07
0.06
0.05
0.04
0.03
0.02
Age (days)
21
35
49
63
77
91
105
119
133
147
161
175
189
203

FIG. 21G

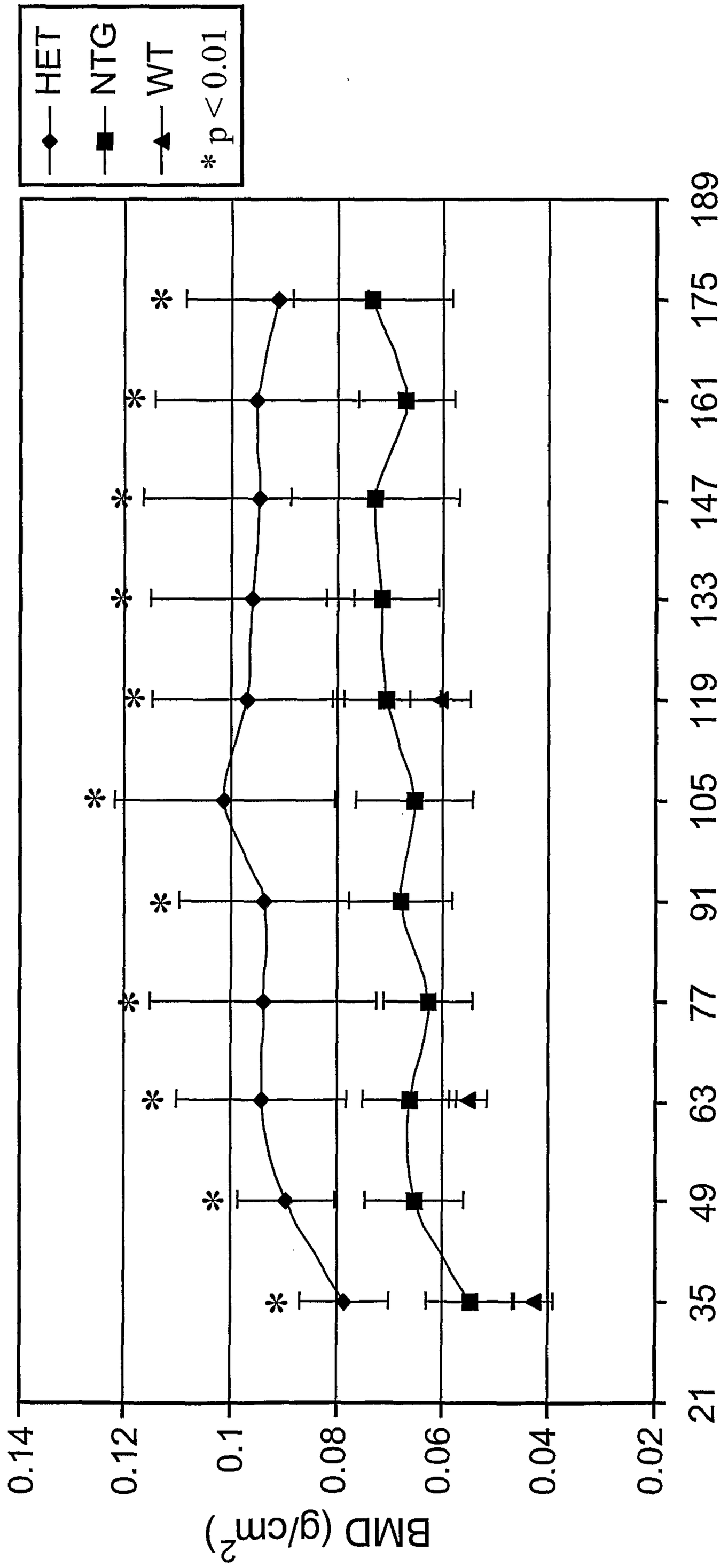

HET
NTG
WT
* p < 0.01
HBM: In vivo pDXA - transgenic mice
HBMMTIC line 19
SPINE
BMD (g/cm²)
0.14
0.12
0.1
0.08
0.06
0.04
0.02
Age (days)
21
35
49
63
77
91
105
119
133
147
161
175
189

FIG. 21H

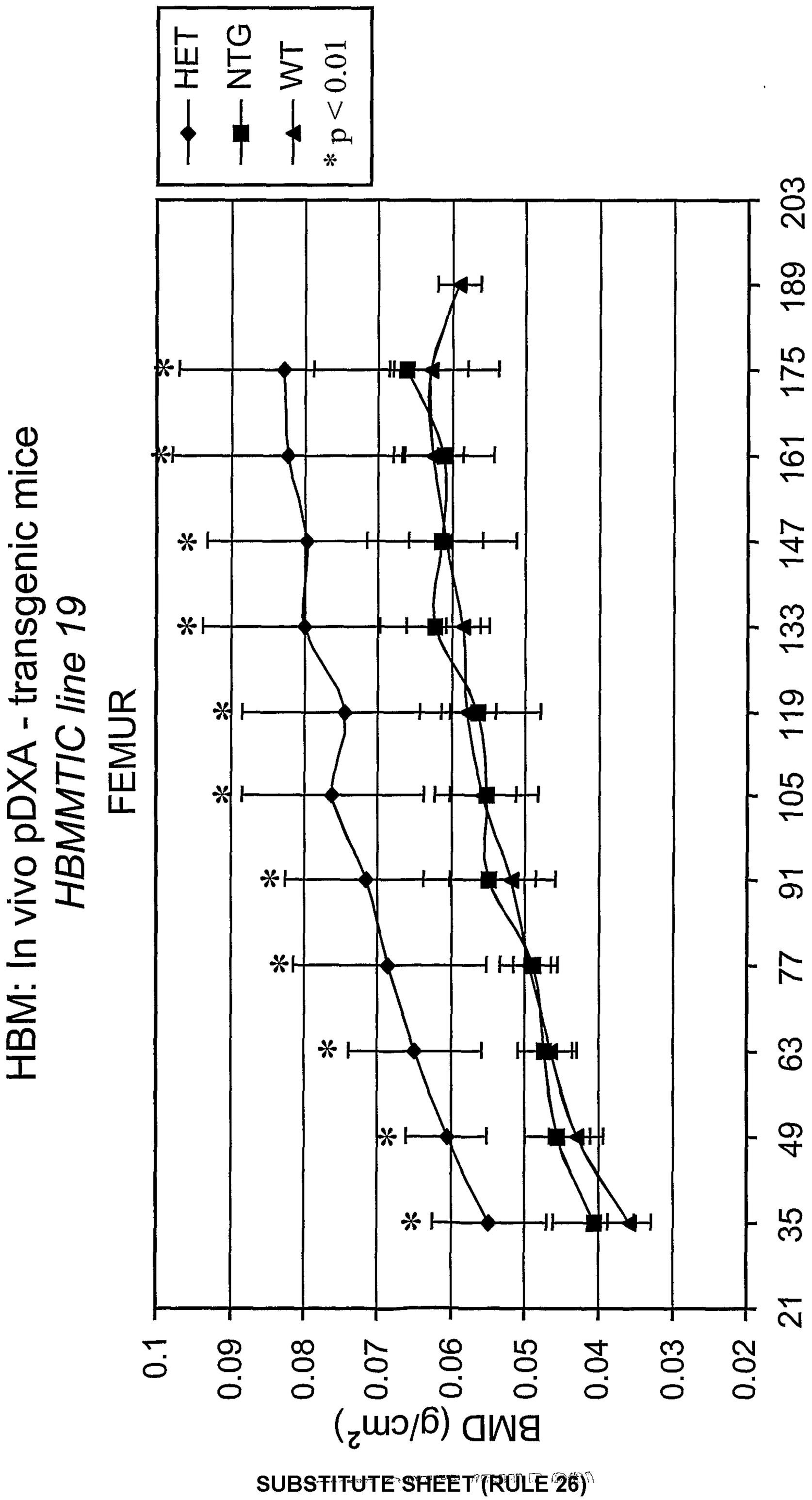

HET
NTG
WT
* p < 0.01
HBM: In vivo pDXA - transgenic mice
HBMMTIC line 19
FEMUR
BMD (g/cm²)
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
Age (days)
21
35
49
63
77
91
105
119
133
147
161
175
189
203

FIG. 21I

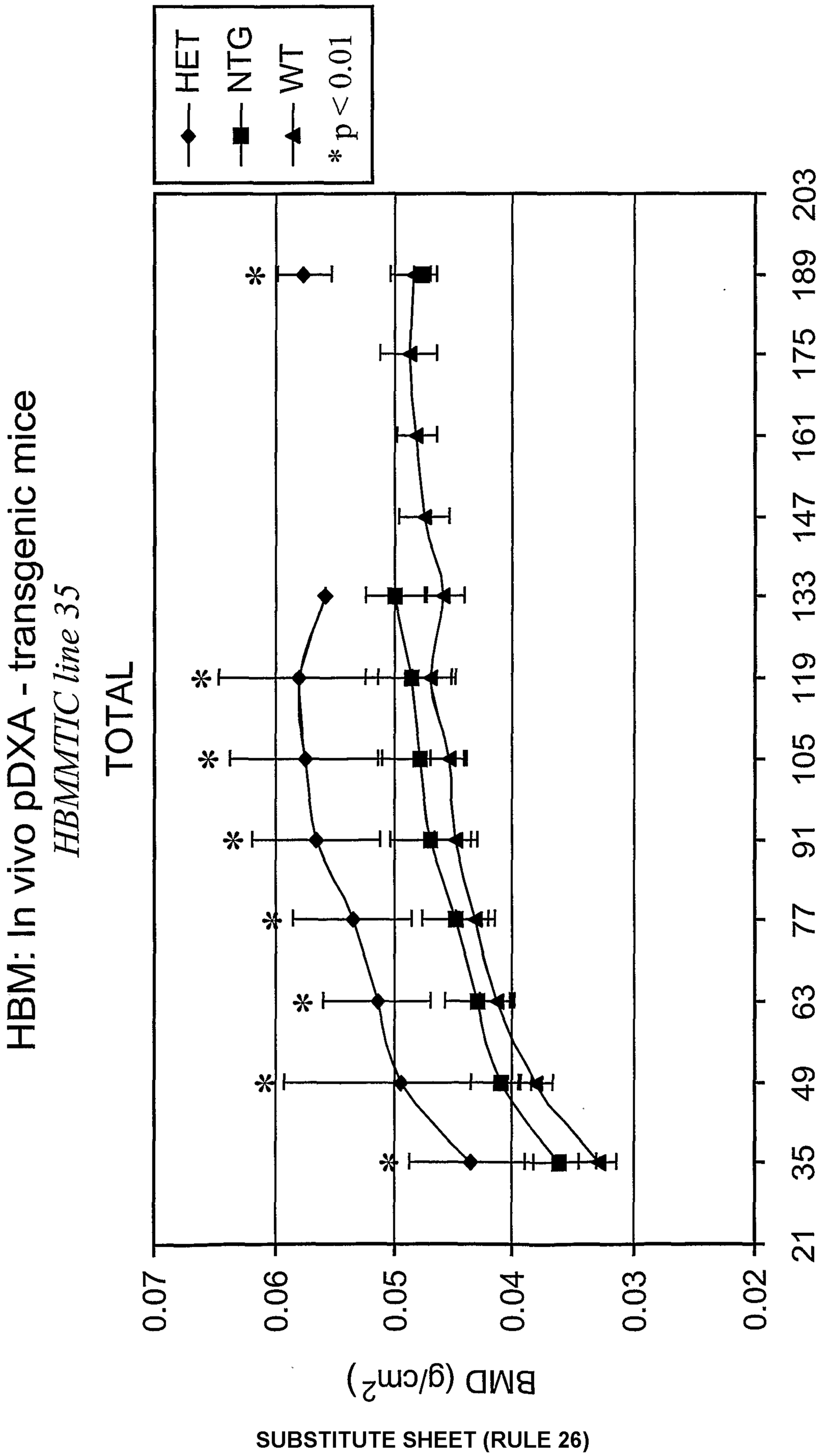

HET
NTG
WT
* p < 0.01
HBM: In vivo pDXA - transgenic mice
HBMMTIC line 35
TOTAL
BMD (g/cm²)
0.07
0.06
0.05
0.04
0.03
0.02
Age (days)
21
35
49
63
77
91
105
119
133
147
161
175
189
203

FIG. 21J

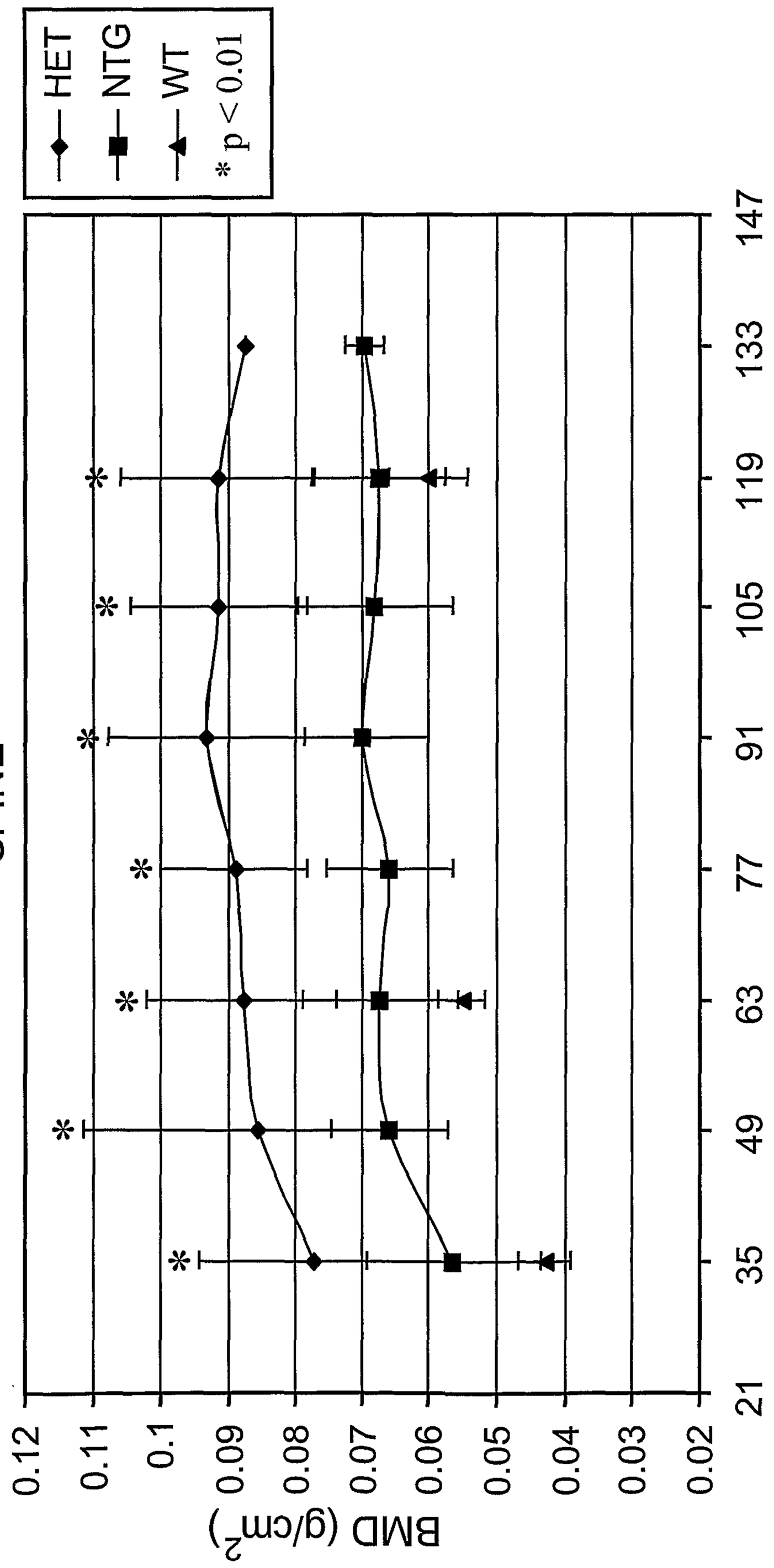

HET
NTG
WT
* p < 0.01
HBM1 transgenic mice - pDXA
HBMMTIC line 35
SPINE
BMD (g/cm²)
0.12
0.11
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
Age (days)
7
21
35
49
63
77
91
105
119
133
147

FIG. 21K

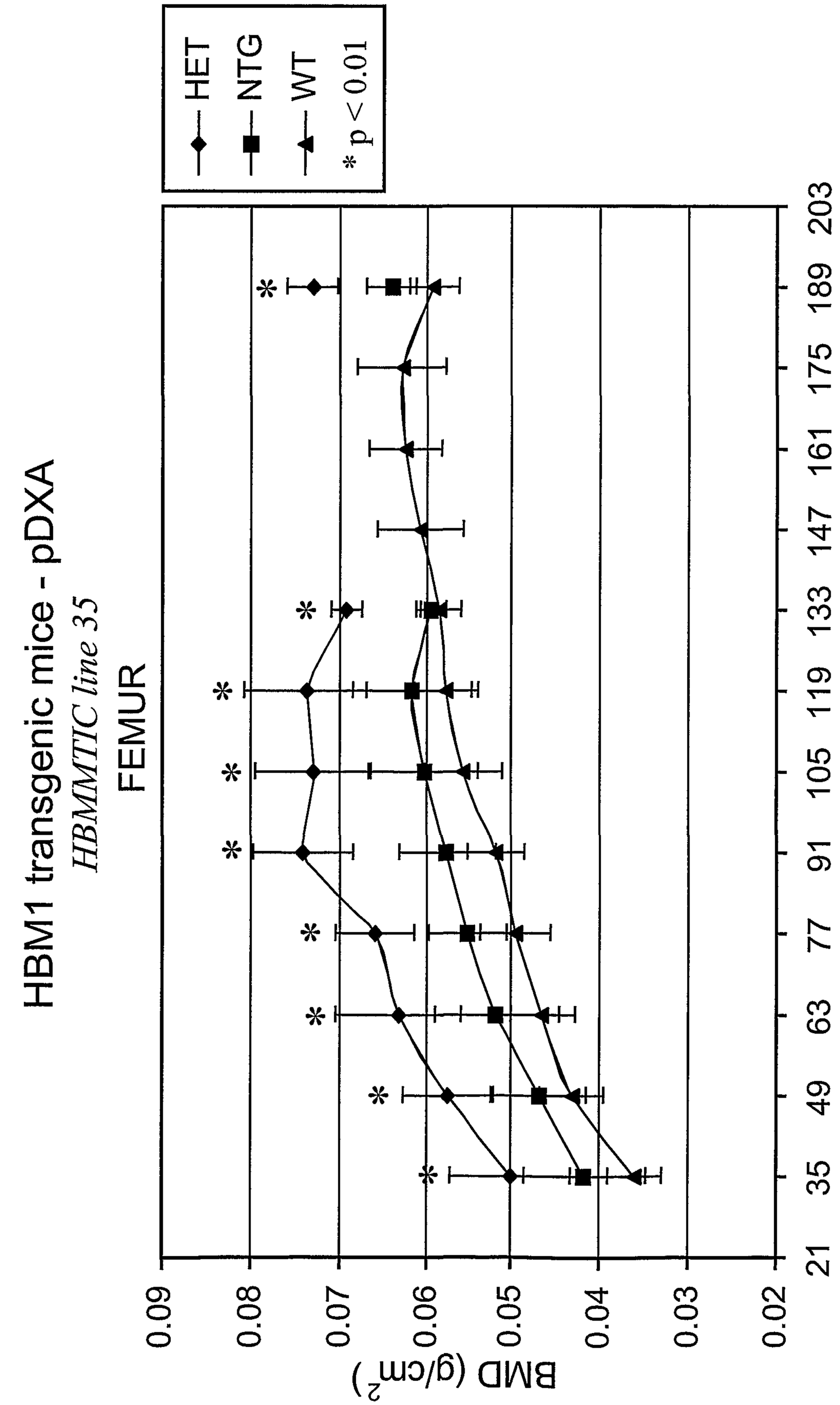

HET
NTG
WT
* p < 0.01
HBM1 transgenic mice - pDXA
HBMMTIC line 35
FEMUR
BMD (g/cm²)
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
Age (days)
21
35
49
63
77
91
105
119
133
147
161
175
189
203

FIG. 21L

HBM transgenic mice: in vivo pDXA*

BMD % changes vs WT in 5 week old animals

Line	n	Femur	Spine	Total	Tibia mRNA
HBMMCBA					
2	11	21	24	10	x7-10
13	4	11	27	5	x1-2
18	1	8	9	3	x10-11
21	1	10	12	6	0
28	15	0	30	4	x1
30	7	4	32	4	0
HBMMTIC					
19	5	63	70	41	x7-8
35	1	4	47	6	x1

** CV: Femur (2.7%); Spine (6.4%); Total (1.7%)*

FIG. 22

HBM transgenic mice: in vivo pDXA*

BMD % changes vs WT in 9 week old animals

Line	n	Femur	Spine	Total
HBMMCBA				
2	3	19	32	12
HBMMTIC				
19	2	52	64	37
35	3	32	43	18

** CV: Femur (2.7%); Spine (6.4%); Total (1.7%)*

FIG. 23

FIG. 24A

HBMGI_2AS

TCTAGACTCG AGGCGGCCGC CCATTGTGCA CTAAAGCGCC GCCGCCGCGC CATGGAGCCC
GAGTGAGCGC GGCGCGGGCC CGTCCGGCCG CCGGACAACA TGGAGGCAGC GCCGCCCGGG
CCGCCGTGGC CGCTGCTGCT GCTGCTGCTG CTGCTGCTGG CGCTGTGCGG CTGCCCGGCC
CCCGCCGCGG CCTCGCCGCT CCTGCTATTT GCCAACCGCC GGGACGTACG GCTGGTGGAC
GCCGGCGGAG TCAAGCTGGA GTCCACCATC GTGGTCAGCG GCCTGGAGGA TGCGGCCGCA
GTGGACTTCC AGTTTTCCAA GGGAGCCGTG TACTGGACAG ACGTGAGCGA GGAGGCCATC
AAGCAGACCT ACCTGAACCA GACGGGGGCC GCCGTGCAGA ACGTGGTCAT CTCCGGCCTG
GTCTCTCCCG ACGGCCTCGC CTGCGACTGG GTGGGCAAGA AGCTGTACTG GACGGACTCA
GAGACCAACC GCATCGAGGT GGCCAACCTC AATGGCACAT CCCGGAAGGT GCTCTTCTGG
CAGGACCTTG ACCAGCCGAG GGCCATCGCC TTGGACCCCG CTCACGGGTA CATGTACTGG
ACAGACTGGG TTGAGACGCC CCGGATTGAG CGGGCAGGGA TGGATGGCAG CACCCGGAAG
ATCATTGTGG ACTCGGACAT TTACTGGCCC AATGGACTGA CCATCGACCT GGAGGAGCAG
AAGCTCTACT GGGCTGACGC CAAGCTCAGC TTCATCCACC GTGCCAACCT GGACGGCTCG
TTCCGGCAGA AGGTGGTGGA GGGCAGCCTG ACGCACCCCT TCGCCCTGAC GCTCTCCGGG
GACACTCTGT ACTGGACAGA CTGGCAGACC CGCTCCATCC ATGCCTGCAA CAAGCGCACT
GGGGGGAAGA GGAAGGAGAT CCTGAGTGCC CTCTACTCAC CCATGGACAT CCAGGTGCTG
AGCCAGGAGC GGCAGCCTTT CTTCCACACT CGCTGTGAGG AGGACAATGG CGGCTGCTCC
CACCTGTGCC TGCTGTCCCC AAGCGAGCCT TTCTACACAT GCGCCTGCCC CACGGGTGTG
CAGCTGCAGG ACAACGGCAG GACGTGTAAG GCAGGAGCCG AGGAGGTGCT GCTGCTGGCC
CGGCGGACGG ACCTACGGAG GATCTCGCTG GACACGCCGG ACTTCACCGA CATCGTGCTG
CAGGTGGACG ACATCCGGCA CGCCATTGCC ATCGACTACG ACCCGCTAGA GGGCTATGTC
TACTGGACAG ATGACGAGGT GCGGGCCATC CGCAGGGCGT ACCTGGACGG GTCTGGGGCG
CAGACGCTGG TCAACACCGA GATCAACGAC CCCGATGGCA TCGCGGTCGA CTGGGTGGCC
CGAAACCTCT ACTGGACCGA CACGGGCACG GACCGCATCG AGGTGACGCG CCTCAACGGC

FIG. 24B

HBMGI 2AS

ACCTCCCGCA AGATCCTGGT GTCGGAGGAC CTGGACGAGC CCCGAGCCAT CGCACTGCAC
CCCGTGATGG GCCTCATGTA CTGGACAGAC TGGGGAGAGA ACCCTAAAAT CGAGTGTGCC
AACTTGGATG GGCAGGAGCG GCGTGTGCTG GTCAATGCCT CCCTCGGGTG GCCCAACGGC
CTGGCCCTGG ACCTGCAGGA GGGGAAGCTC TACTGGGGAG ACGCCAAGAC AGACAAGATC
GAGGTGATCA ATGTTGATGG GACGAAGAGG CGGACCCTCC TGGAGGACAA GCTCCCGCAC
ATTTTCGGGT TCACGCTGCT GGGGGACTTC ATCTACTGGA CTGACTGGCA GCGCCGCAGC
ATCGAGCGGG TGCACAAGGT CAAGGCCAGC CGGGACGTCA TCATTGACCA GCTGCCCGAC
CTGATGGGGC TCAAAGCTGT GAATGTGGCC AAGGTCGTCG GAACCAACCC GTGTGCGGAC
AGGAACGGGG GGTGCAGCCA CCTGTGCTTC TTCACACCCC ACGCAACCCG GTGTGGCTGC
CCCATCGGCC TGGAGCTGCT GAGTGACATG AAGACCTGCA TCGTGCCTGA GGCCTTCTTG
GTCTTCACCA GCAGAGCCGC CATCCACAGG ATCTCCCTCG AGACCAATAA CAACGACGTG
GCCATCCCGC TCACGGGCGT CAAGGAGGCC TCAGCCCTGG ACTTTGATGT GTCCAACAAC
CACATCTACT GGACAGACGT CAGCCTGAAG ACCATCAGCC GCGCCTTCAT GAACGGGAGC
TCGGTGGAGC ACGTGGTGGA GTTTGGCCTT GACTACCCCG AGGGCATGGC CGTTGACTGG
ATGGGCAAGA ACCTCTACTG GGCCGACACT GGGACCAACA GAATCGAAGT GGCGCGGCTG
GACGGGCAGT TCCGGCAAGT CCTCGTGTGG AGGGACTTGG ACAACCCGAG GTCGCTGGCC
CTGGATCCCA CCAAGGGCTA CATCTACTGG ACCGAGTGGG GCGGCAAGCC GAGGATCGTG
CGGGCCTTCA TGGACGGGAC CAACTGCATG ACGCTGGTGG ACAAGGTGGG CCGGGCCAAC
GACCTCACCA TTGACTACGC TGACCAGCGC CTCTACTGGA CCGACCTGGA CACCAACATG
ATCGAGTCGT CCAACATGCT GGGTCAGGAG CGGGTCGTGA TTGCCGACGA TCTCCCGCAC
CCGTTCGGTC TGACGCAGTA CAGCGATTAT ATCTACTGGA CAGACTGGAA TCTGCACAGC
ATTGAGCGGG CCGACAAGAC TAGCGGCCGG AACCGCACCC TCATCCAGGG CCACCTGGAC
TTCGTGATGG ACATCCTGGT GTTCCACTCC TCCCGCCAGG ATGGCCTCAA TGACTGTATG

FIG. 24C

HBMGI_2AS

CACAACAACG GGCAGTGTGG GCAGCTGTGC CTTGCCATCC CCGGCGGCCA CCGCTGCGGC
TGCGCCTCAC ACTACACCCT GGACCCCAGC AGCCGCAACT GCAGCCCGCC CACCACCTTC
TTGCTGTTCA GCCAGAAATC TGCCATCAGT CGGATGATCC CGGACGACCA GCACAGCCCG
GATCTCATCC TGCCCCTGCA TGGACTGAGG AACGTCAAAG CCATCGACTA TGACCCACTG
GACAAGTTCA TCTACTGGGT GGATGGGCGC CAGAACATCA AGCGAGCCAA GGACGACGGG
ACCCAGCCCT TTGTTTTGAC CTCTCTGAGC CAAGGCCAAA ACCCAGACAG GCAGCCCCAC
GACCTCAGCA TCGACATCTA CAGCCGGACA CTGTTCTGGA CGTGCGAGGC CACCAATACC
ATCAACGTCC ACAGGCTGAG CGGGGAAGCC ATGGGGGTGG TGCTGCGTGG GGACCGCGAC
AAGCCCAGGG CCATCGTCGT CAACGCGGAG CGAGGGTACC TGTACTTCAC CAACATGCAG
GACCGGGCAG CCAAGATCGA ACGCGCAGCC CTGGACGGCA CCGAGCGCGA GGTCCTCTTC
ACCACCGGCC TCATCCGCCC TGTGGCCCTG GTGGTAGACA ACACACTGGG CAAGCTGTTC
TGGGTGGACG CGGACCTGAA GCGCATTGAG AGCTGTGACC TGTCAGGGGC CAACCGCCTG
ACCCTGGAGG ACGCCAACAT CGTGCAGCCT CTGGGCCTGA CCATCCTTGG CAAGCATCTC
TACTGGATCG ACCGCCAGCA GCAGATGATC GAGCGTGTGG AGAAGACCAC CGGGGACAAG
CGGACTCGCA TCCAGGGCCG TGTCGCCCAC CTCACTGGCA TCCATGCAGT GGAGGAAGTC
AGCCTGGAGG AGTTCTCAGC CCACCCATGT GCCCGTGACA ATGGTGGCTG CTCCCACATC
TGTATTGCCA AGGGTGATGG GACACCACGG TGCTCATGCC CAGTCCACCT CGTGCTCCTG
CAGAACCTGC TGACCTGTGG AGAGCCGCCC ACCTGCTCCC CGGACCAGTT TGCATGTGCC
ACAGGGGAGA TCGACTGTAT CCCCGGGGCC TGGCGCTGTG ACGGCTTTCC CGAGTGCGAT
GACCAGAGCG ACGAGGAGGG CTGCCCCGTG TGCTCCGCCG CCCAGTTCCC CTGCGCGCGG
GGTCAGTGTG TGGACCTGCG CCTGCGCTGC GACGGCGAGG CAGACTGTCA GGACCGCTCA
GACGAGGCGG ACTGTGACGC CATCTGCCTG CCCAACCAGT TCCGGTGTGC GAGCGGCCAG
TGTGTCCTCA TCAAACAGCA GTGCGACTCC TTCCCCGACT GTATCGACGG CTCCGACGAG
CTCATGTGTG AAATCACCAA GCCGCCCTCA GACGACAGCC CGGCCCACAG CAGTGCCATC

FIG. 24D

HBMGI_2AS

GGGCCCGTCA TTGGCATCAT CCTCTCTCTC TTCGTCATGG GTGGTGTCTA TTTTGTGTGC
CAGCGCGTGG TGTGCCAGCG CTATGCGGGG GCCAACGGGC CCTTCCCGCA CGAGTATGTC
AGCGGGACCC CGCACGTGCC CCTCAATTTC ATAGCCCCGG GCGGTTCCCA GCATGGCCCC
TTCACAGGCA TCGCATGCGG AAAGTCCATG ATGAGCTCCG TGAGCCTGAT GGGGGGCCGG
GGCGGGGTGC CCCTCTACGA CCGGAACCAC GTCACAGGGG CCTCGTCCAG CAGCTCGTCC
AGCACGAAGG CCACGCTGTA CCCGCCGATC CTGAACCCGC CGCCCTCCCC GGCCACGGAC
CCCTCCCTGT ACAACATGGA CATGTTCTAC TCTTCAAACA TTCCGGCCAC TGCGAGACCG
TACAGGCCCT ACATCATTCG AGGAATGGCG CCCCCGACGA CGCCCTGCAG CACCGACGTG
TGTGACAGCG ACTACAGCGC CAGCCGCTGG AAGGCCAGCA AGTACTACCT GGATTTGAAC
TCGGACTCAG ACCCCTATCC ACCCCCACCC ACGCCCCACA GCCAGTACCT GTCGGCGGAG
GACAGCTGCC CGCCCTCGCC CGCCACCGAG AGGAGCTACT TCCATCTCTT CCCGCCCCCT
CCGTCCCCCT GCACGGACTC ATCCTGACCT CGGCCGGGCC ACTCTGGCTT CTCTGTGCCC
CTGTAAATAG TTTTAAATAT GAACAAAGAA AAAAATATAT TTTATGATTT AAAAAATAAA
TATAATTGGG ATTTTAAAAA CATGAGAAAT GTGAACTGTG ATGGGGTGGG CAGGGCTGGG
AGAACTTTGT ACAGTGGAGA AATATTTATA AACTTAATTT TGTAAAACAG AAAAAAAAAA
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA GCGGCCGC

FIG. 25A

ZMAXGI_3AS

TCTAGACTCG AGGCGGCCGC CCATTGTGCA CTAAAGCGCC GCCGCCGCGC CATGGAGCCC
GAGTGAGCGC GGCGCGGGCC CGTCCGGCCG CCGGACAACA TGGAGGCAGC GCCGCCCGGG
CCGCCGTGGC CGCTGCTGCT GCTGCTGCTG CTGCTGCTGG CGCTGTGCGG CTGCCCGGCC
CCCGCCGCGG CCTCGCCGCT CCTGCTATTT GCCAACCGCC GGGACGTACG GCTGGTGGAC
GCCGGCGGAG TCAAGCTGGA GTCCACCATC GTGGTCAGCG GCCTGGAGGA TGCGGCCGCA
GTGGACTTCC AGTTTTCCAA GGGAGCCGTG TACTGGACAG ACGTGAGCGA GGAGGCCATC
AAGCAGACCT ACCTGAACCA GACGGGGGCC GCCGTGCAGA ACGTGGTCAT CTCCGGCCTG
GTCTCTCCCG ACGGCCTCGC CTGCGACTGG GTGGGCAAGA AGCTGTACTG GACGGACTCA
GAGACCAACC GCATCGAGGT GGCCAACCTC AATGGCACAT CCCGGAAGGT GCTCTTCTGG
CAGGACCTTG ACCAGCCGAG GGCCATCGCC TTGGACCCCG CTCACGGGTA CATGTACTGG
ACAGACTGGG GTGAGACGCC CCGGATTGAG CGGGCAGGGA TGGATGGCAG CACCCGGAAG
ATCATTGTGG ACTCGGACAT TTACTGGCCC AATGGACTGA CCATCGACCT GGAGGAGCAG
AAGCTCTACT GGGCTGACGC CAAGCTCAGC TTCATCCACC GTGCCAACCT GGACGGCTCG
TTCCGGCAGA AGGTGGTGGA GGGCAGCCTG ACGCACCCCT TCGCCCTGAC GCTCTCCGGG
GACACTCTGT ACTGGACAGA CTGGCAGACC CGCTCCATCC ATGCCTGCAA CAAGCGCACT
GGGGGGAAGA GGAAGGAGAT CCTGAGTGCC CTCTACTCAC CCATGGACAT CCAGGTGCTG
AGCCAGGAGC GGCAGCCTTT CTTCCACACT CGCTGTGAGG AGGACAATGG CGGCTGCTCC
CACCTGTGCC TGCTGTCCCC AAGCGAGCCT TTCTACACAT GCGCCTGCCC CACGGGTGTG
CAGCTGCAGG ACAACGGCAG GACGTGTAAG GCAGGAGCCG AGGAGGTGCT GCTGCTGGCC
CGGCGGACGG ACCTACGGAG GATCTCGCTG GACACGCCGG ACTTCACCGA CATCGTGCTG
CAGGTGGACG ACATCCGGCA CGCCATTGCC ATCGACTACG ACCCGCTAGA GGGCTATGTC
TACTGGACAG ATGACGAGGT GCGGGCCATC CGCAGGGCGT ACCTGGACGG GTCTGGGGCG
CAGACGCTGG TCAACACCGA GATCAACGAC CCCGATGGCA TCGCGGTCGA CTGGGTGGCC
CGAAACCTCT ACTGGACCGA CACGGGCACG GACCGCATCG AGGTGACGCG CCTCAACGGC

FIG. 25B

ZMAXGI_3AS

ACCTCCCGCA AGATCCTGGT GTCGGAGGAC CTGGACGAGC CCCGAGCCAT CGCACTGCAC
CCCGTGATGG GCCTCATGTA CTGGACAGAC TGGGGAGAGA ACCCTAAAAT CGAGTGTGCC
AACTTGGATG GGCAGGAGCG GCGTGTGCTG GTCAATGCCT CCCTCGGGTG GCCCAACGGC
CTGGCCCTGG ACCTGCAGGA GGGGAAGCTC TACTGGGGAG ACGCCAAGAC AGACAAGATC
GAGGTGATCA ATGTTGATGG GACGAAGAGG CGGACCCTCC TGGAGGACAA GCTCCCGCAC
ATTTTCGGGT TCACGCTGCT GGGGGACTTC ATCTACTGGA CTGACTGGCA GCGCCGCAGC
ATCGAGCGGG TGCACAAGGT CAAGGCCAGC CGGGACGTCA TCATTGACCA GCTGCCCGAC
CTGATGGGGC TCAAAGCTGT GAATGTGGCC AAGGTCGTCG GAACCAACCC GTGTGCGGAC
AGGAACGGGG GGTGCAGCCA CCTGTGCTTC TTCACACCCC ACGCAACCCG GTGTGGCTGC
CCCATCGGCC TGGAGCTGCT GAGTGACATG AAGACCTGCA TCGTGCCTGA GGCCTTCTTG
GTCTTCACCA GCAGAGCCGC CATCCACAGG ATCTCCCTCG AGACCAATAA CAACGACGTG
GCCATCCCGC TCACGGGCGT CAAGGAGGCC TCAGCCCTGG ACTTTGATGT GTCCAACAAC
CACATCTACT GGACAGACGT CAGCCTGAAG ACCATCAGCC GCGCCTTCAT GAACGGGAGC
TCGGTGGAGC ACGTGGTGGA GTTTGGCCTT GACTACCCCG AGGGCATGGC CGTTGACTGG
ATGGGCAAGA ACCTCTACTG GGCCGACACT GGGACCAACA GAATCGAAGT GGCGCGGCTG
GACGGGCAGT TCCGGCAAGT CCTCGTGTGG AGGGACTTGG ACAACCCGAG GTCGCTGGCC
CTGGATCCCA CCAAGGGCTA CATCTACTGG ACCGAGTGGG GCGGCAAGCC GAGGATCGTG
CGGGCCTTCA TGGACGGGAC CAACTGCATG ACGCTGGTGG ACAAGGTGGG CCGGGCCAAC
GACCTCACCA TTGACTACGC TGACCAGCGC CTCTACTGGA CCGACCTGGA CACCAACATG
ATCGAGTCGT CCAACATGCT GGGTCAGGAG CGGGTCGTGA TTGCCGACGA TCTCCCGCAC
CCGTTCGGTC TGACGCAGTA CAGCGATTAT ATCTACTGGA CAGACTGGAA TCTGCACAGC
ATTGAGCGGG CCGACAAGAC TAGCGGCCGG AACCGCACCC TCATCCAGGG CCACCTGGAC
TTCGTGATGG ACATCCTGGT GTTCCACTCC TCCCGCCAGG ATGGCCTCAA TGACTGTATG

FIG. 25C

ZMAXGI_3AS

CACAACAACG GGCAGTGTGG GCAGCTGTGC CTTGCCATCC CCGGCGGCCA CCGCTGCGGC
TGCGCCTCAC ACTACACCCT GGACCCCAGC AGCCGCAACT GCAGCCCGCC CACCACCTTC
TTGCTGTTCA GCCAGAAATC TGCCATCAGT CGGATGATCC CGGACGACCA GCACAGCCCG
GATCTCATCC TGCCCCTGCA TGGACTGAGG AACGTCAAAG CCATCGACTA TGACCCACTG
GACAAGTTCA TCTACTGGGT GGATGGGCGC CAGAACATCA AGCGAGCCAA GGACGACGGG
ACCCAGCCCT TTGTTTTGAC CTCTCTGAGC CAAGGCCAAA ACCCAGACAG GCAGCCCCAC
GACCTCAGCA TCGACATCTA CAGCCGGACA CTGTTCTGGA CGTGCGAGGC CACCAATACC
ATCAACGTCC ACAGGCTGAG CGGGGAAGCC ATGGGGGTGG TGCTGCGTGG GGACCGCGAC
AAGCCCAGGG CCATCGTCGT CAACGCGGAG CGAGGGTACC TGTACTTCAC CAACATGCAG
GACCGGGCAG CCAAGATCGA ACGCGCAGCC CTGGACGGCA CCGAGCGCGA GGTCCTCTTC
ACCACCGGCC TCATCCGCCC TGTGGCCCTG GTGGTAGACA ACACACTGGG CAAGCTGTTC
TGGGTGGACG CGGACCTGAA GCGCATTGAG AGCTGTGACC TGTCAGGGGC CAACCGCCTG
ACCCTGGAGG ACGCCAACAT CGTGCAGCCT CTGGGCCTGA CCATCCTTGG CAAGCATCTC
TACTGGATCG ACCGCCAGCA GCAGATGATC GAGCGTGTGG AGAAGACCAC CGGGGACAAG
CGGACTCGCA TCCAGGGCCG TGTCGCCCAC CTCACTGGCA TCCATGCAGT GGAGGAAGTC
AGCCTGGAGG AGTTCTCAGC CCACCCATGT GCCCGTGACA ATGGTGGCTG CTCCCACATC
TGTATTGCCA AGGGTGATGG GACACCACGG TGCTCATGCC CAGTCCACCT CGTGCTCCTG
CAGAACCTGC TGACCTGTGG AGAGCCGCCC ACCTGCTCCC CGGACCAGTT TGCATGTGCC
ACAGGGGAGA TCGACTGTAT CCCCGGGGCC TGGCGCTGTG ACGGCTTTCC CGAGTGCGAT
GACCAGAGCG ACGAGGAGGG CTGCCCCGTG TGCTCCGCCGC CCAGTTCCCC TGCGCGCGG
GGTCAGTGTGT GGACCTGCGC CTGCGCTGCG ACGGCGAGGC AGACTGTCAG GACCGCTCA
GACGAGGCGGA CTGTGACGCC ATCTGCCTGC CCAACCAGTT CCGGTGTGCG AGCGGCCAG
TGTGTCCTCAT CAAACAGCAG TGCGACTCCT TCCCCGACTG TATCGACGGC TCCGACGAG
CTCATGTGTGA AATCACCAAG CCGCCCTCAG ACGACAGCCC GGCCCACAGC AGTGCCATC

FIG. 25D

ZMAXGI_3AS

GGGCCCGTCA TTGGCATCAT CCTCTCTCTC TTCGTCATGG GTGGTGTCTA TTTTGTGTGC
CAGCGCGTGG TGTGCCAGCG CTATGCGGGG GCCAACGGGC CCTTCCCGCA CGAGTATGTC
AGCGGGACCC CGCACGTGCC CCTCAATTTC ATAGCCCCGG GCGGTTCCCA GCATGGCCCC
TTCACAGGCA TCGCATGCGG AAAGTCCATG ATGAGCTCCG TGAGCCTGAT GGGGGGCCGG
GGCGGGGTGC CCCTCTACGA CCGGAACCAC GTCACAGGGG CCTCGTCCAG CAGCTCGTCC
AGCACGAAGG CCACGCTGTA CCCGCCGATC CTGAACCCGC CGCCCTCCCC GGCCACGGAC
CCCTCCCTGT ACAACATGGA CATGTTCTAC TCTTCAAACA TTCCGGCCAC TGCGAGACCG
TACAGGCCCT ACATCATTCG AGGAATGGCG CCCCCGACGA CGCCCTGCAG CACCGACGTG
TGTGACAGCG ACTACAGCGC CAGCCGCTGG AAGGCCAGCA AGTACTACCT GGATTTGAAC
TCGGACTCAG ACCCCTATCC ACCCCCACCC ACGCCCCACA GCCAGTACCT GTCGGCGGAG
GACAGCTGCC CGCCCTCGCC CGCCACCGAG AGGAGCTACT TCCATCTCTT CCCGCCCCCT
CCGTCCCCCT GCACGGACTC ATCCTGACCT CGGCCGGGCC ACTCTGGCTT CTCTGTGCCC
CTGTAAATAG TTTTAAATAT GAACAAAGAA AAAAATATAT TTTATGATTT AAAAAATAAA
TATAATTGGG ATTTTAAAAA CATGAGAAAT GTGAACTGTG ATGGGGTGGG CAGGGCTGGG
AGAACTTTGT ACAGTGGAGA AATATTTATA AACTTAATTT TGTAAAACAG AAAAAAAAAA
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA GCGGCCGC

FIG. 26A

Alignment of Human and Mouse Zmax1 (LRP5) Amino-Acid Sequences.

Top= Human, Bottom = Mouse

```
  1 MEA....APPGPPWPLLLLLLLLLALCGCPAPAAASPLLLFANRRDVRLV 46
    ||.    |||.|| |||||:     |:.: .|||||||||||||||||||
  1 METAPTRAPPPPPPPLLLLV.....LYCSLVPAAASPLLLFANRRDVRLV 45

 47 DAGGVKLESTIVVSGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTG 96
    ||||||||||||.|||||||||||||||||||||||||||||||||||||
 46 DAGGVKLESTIVASGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTG 95

 97 AAVQNVVISGLVSPDGLACDWVGKKLYWTDSETNRIEVANLNGTSRKVLF 146
    ||.||:||||||||||||||||||||||||||||||||||||||||||||
 96 AAAQNIVISGLVSPDGLACDWVGKKLYWTDSETNRIEVANLNGTSRKVLF 145

147 WQDLDQPRAIALDPAHGYMYWTDWGETPRIERAGMDGSTRKIIVDSDIYW 196
    ||||||||||||||||||||||.|||||||||||||||||||||||||||
146 WQDLDQPRAIALDPAHGYMYWTDWGEAPRIERAGMDGSTRKIIVDSDIYW 195

197 PNGLTIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLS 246
    ||||||||||||||||||||||||||||||||||||||||||||||||||
196 PNGLTIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLS 245

247 GDTLYWTDWQTRSIHACNKRTGGKRKEILSALYSPMDIQVLSQERQPFFH 296
    ||||||||||||||||||:||:.|||||||||||||||||||||||| ||
246 GDTLYWTDWQTRSIHACNKWTGEQRKEILSALYSPMDIQVLSQERQPPFH 295

297 TRCEEDNGGCSHLCLLSPSEPFYTCACPTGVQLQDNGRTCKAGAEEVLLL 346
    |.|||||||||||||||.||||.|||||||||||||:|||.|||||||||
296 TPCEEDNGGCSHLCLLSPREPFYSCACPTGVQLQDNGKTCKTGAEEVLLL 345

347 ARRTDLRRISLDTPDFTDIVLQVDDIRHAIAIDYDPLEGYVYWTDDEVRA 396
    ||||||||||||||||||||||:||||||||||||||||||||||||||
346 ARRTDLRRISLDTPDFTDIVLQVGDIRHAIAIDYDPLEGYVYWTDDEVRA 395

397 IRRAYLDGSGAQTLVNTEINDPDGIAVDWVARNLYWTDTGTDRIEVTRLN 446
    ||||||||||||||||||||||||||||||||||||||||||||||||||
396 IRRAYLDGSGAQTLVNTEINDPDGIAVDWVARNLYWTDTGTDRIEVTRLN 445

447 GTSRKILVSEDLDEPRAIALHPVMGLMYWTDWGENPKIECANLDGQERRV 496
    ||||||||||||||||||.|||||||||||||||||||||||.:|:|
446 GTSRKILVSEDLDEPRAIVLHPVMGLMYWTDWGENPKIECANLDGRDRHV 495

497 LVNASLGWPNGLALDLQEGKLYWGDAKTDKIEVINVDGTKRRTLLEDKLP 546
    |||.||||||||||||||||||||||||||||||:|||||:||||||||
496 LVNTSLGWPNGLALDLQEGKLYWGDAKTDKIEVINIDGTKRKTLLEDKLP 545

547 HIFGFTLLGDFIYWTDWQRRSIERVHKVKASRDVIIDQLPDLMGLKAVNV 596
    ||||||||||||||||||||||||||||||||||||||||||||||||||
546 HIFGFTLLGDFIYWTDWQRRSIERVHKVKASRDVIIDQLPDLMGLKAVNV 595

597 AKVVGTNPCADRNGGCSHLCFFTPHATRCGCPIGLELLSDMKTCIVPEAF 646
    |||||||||||  |||||||||||:||:||||||||||||||||:||||
596 AKVVGTNPCADGNGGCSHLCFFTPRATKCGCPIGLELLSDMKTCIIPEAF 645

647 LVFTSRAAIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIYWTDVSL 696
    |||||||.||||||||||||||||||||||||||||||||||||||||||
646 LVFTSRATIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIYWTDVSL 695
```

FIG. 26B

Alignment of Human and Mouse Zmax1 (LRP5) Amino-Acid Sequences.

```
697 KTISRAFMNGSSVEHVVEFGLDYPEGMAVDWMGKNLYWADTGTNRIEVAR 746
    ||||||||||||||||||:||||||||||||||||||||||||||||||||
696 KTISRAFMNGSSVEHVIEFGLDYPEGMAVDWMGKNLYWADTGTNRIEVAR 745

747 LDGQFRQVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNC 796
    ||||||||||||||||||||||||||||||||||||||||||||||||||
746 LDGQFRQVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNC 795

797 MTLVDKVGRANDLTIDYADQRLYWTDLDTNMIESSNMLGQERVVIADDLP 846
    ||||||||||||||||||||||||||||||||||||||||||:|||||||
796 MTLVDKVGRANDLTIDYADQRLYWTDLDTNMIESSNMLGQERMVIADDLP 845

847 HPFGLTQYSDYIYWTDWNLHSIERADKTSGRNRTLIQGHLDFVMDILVFH 896
    .|||||||||||||||||||||||||||||||||||||||||||||||||
846 YPFGLTQYSDYIYWTDWNLHSIERADKTSGRNRTLIQGHLDFVMDILVFH 895

897 SSRQDGLNDCMHNNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNCSPPTT 946
    ||||||||||:|.|||||||||||||||||||||||||||||||||.|
896 SSRQDGLNDCVHSNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNCSPPST 945

947 FLLFSQKSAISRMIPDDQHSPDLILPLHGLRNVKAIDYDPLDKFIYWVDG 996
    ||||||| |||||||||| |||||:||||||||||||:||||||||||||
946 FLLFSQKFAISRMIPDDQLSPDLVLPLHGLRNVKAINYDPLDKFIYWVDG 995

997 RQNIKRAKDDGTQPFVLTSLSQGQNPDRQPHDLSIDIYSRTLFWTCEATN 1046
    |||||||||||||| :||| ||: .|||||||||||||||||||||||||
 996 RQNIKRAKDDGTQPSMLTSPSQSLSPDRQPHDLSIDIYSRTLFWTCEATN 1045

1047 TINVHRLSGEAMGVVLRGDRDKPRAIVVNAERGYLYFTNMQDRAAKIERA 1096
     |||||||.|:|||||||||||||||.|||||||:|||||||:|||||||
1046 TINVHRLDGDAMGVVLRGDRDKPRAIAVNAERGYMYFTNMQDHAAKIERA 1095

1097 ALDGTEREVLFTTGLIRPVALVVDNTLGKLFWVDADLKRIESCDLSGANR 1146
     .|||||||||||||||||||||||.|||||||||||||||||||||||||
1096 SLDGTEREVLFTTGLIRPVALVVDNALGKLFWVDADLKRIESCDLSGANR 1145

1147 LTLEDANIVQPLGLTILGKHLYWIDRQQQMIERVEKTTGDKRTRIQGRVA 1196
     ||||||||||:|||:||:|||||||||||||||||||||:||||.
1146 LTLEDANIVQPVGLTVLGRHLYWIDRQQQMIERVEKTTGDKRTRVQGRVT 1195

1197 HLTGIHAVEEVSLEEFSAHPCARDNGGCSHICIAKGDGTPRCSCPVHLVL 1246
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1196 HLTGIHAVEEVSLEEFSAHPCARDNGGCSHICIAKGDGTPRCSCPVHLVL 1245

1247 LQNLLTCGEPPTCSPDQFACATGEIDCIPGAWRCDGFPECDDQSDEEGCP 1296
     ||||||||||||||||||.||||||||||||||||.||||||||||
1246 LQNLLTCGEPPTCSPDQFACTTGEIDCIPGAWRCDGFPECADQSDEEGCP 1295

1297 VCSAAQFPCARGQCVDLRLRCDGEADCQDRSDEADCDAICLPNQFRCASG 1346
     ||||.||||||||||||||||||||||||||||:|||:|||||||||.||
1296 VCSASQFPCARGQCVDLRLRCDGEADCQDRSDEANCDAVCLPNQFRCTSG 1345

1347 QCVLIKQQCDSFPDCIDGSDELMCEITKPPSDDSPAHSSAIGPVIGIILS 1396
     |||||||||||||| |||||||||.|||||| ||||||||||||||||
1346 QCVLIKQQCDSFPDCADGSDELMCEINKPPSDDIPAHSSAIGPVIGIILS 1395

1397 LFVMGGVYFVCQRVVCQRYAGANGPFPHEYVSGTPHVPLNFIAPGGSQHG 1446
     ||||||||||||||:||||.||.|||||||||:|.||||||||||||||
1396 LFVMGGVYFVCQRVMCQRYTGASGPFPHEYVGGAPHVPLNFIAPGGSQHG 1445
```

FIG. 26C

Alignment of Human and Mouse Zmax1 (LRP5) Amino-Acid Sequences.

```
1447 PFTGIACGKSMMSSVSLMGGRGGVPLYDRNHVTGASSSSSSSTKATLYPP 1496
     ||.||:|:||:|||:||:||||:||||||||||||||||||||||||||||
1446 PFPGIPCSKSVMSSMSLVGGRGSVPLYDRNHVTGASSSSSSSTKATLYPP 1495

1497 ILNPPPSPATDPSLYNMDMFYSSNIPATARPYRPYIIRGMAPPTTPCSTD 1546
     ||||||||||||||||:|:||||.||||||||||:|||||||||||||||
1496 ILNPPPSPATDPSLYNVDVFYSSGIPATARPYRPYVIRGMAPPTTPCSTD 1545

1547 VCDSDYSASRWKASKYYLDLNSDSDPYPPPPTPHSQYLSAEDSCPPSPAT 1596
     ||||||| ||||.||||||||||||||||||||||||||||||||||:|
1546 VCDSDYSISRWKSSKYYLDLNSDSDPYPPPPTPHSQYLSAEDSCPPSPGT 1595

1597 ERSYFHLFPPPPSPCTDSS 1615
     |||| |||||||||||||||
1596 ERSYCHLFPPPPSPCTDSS 1614
```

FIG. 27A

Alignment of Human LRP5 and LRP6

Top = LRP5, Bottom = LRP6

Underlined = region marked for antibody production

Percent Similarity: 83.542 Percent Identity: 70.901

```
  1 MEAAPPGPPWPLLLLLLLLALCGCPAPAAASPLLLFANRRDVRLVDAGG 50
                :  :|  ||| |: ..   |.||||:||||||:||||||..
  1 ...........MGAVLRSLLA.CSFCVLLRAAPLLLYANRRDLRLVDATN 38

 51 VKLESTIVVSGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTGAAVQ 100
    .| :.||||:|||||||||||| ||.| :||.|||||||||.| :|.|: .||
 39 GKENATIVVGGLEDAAAVDFVFSHGLIYWSDVSEEAIKRTEFNKTE.SVQ 87

101 NVVISGLVSPDGLACDWVGKKLYWTDSETNRIEVANLNGTSRKVLFWQDL 150
    |||:|||:||||||||||:|.|||||||||||||||.||:|. |||||||:|
 88 NVVVSGLLSPDGLACDWLGEKLYWTDSETNRIEVSNLDGSLRKVLFWQEL 137

151 DQPRAIALDPAHGYMYWTDWGETPRIERAGMDGSTRKIIVDSDIYWPNGL 200
    |||||||||||. |:|||||||.|:||||||||||.| ||::|:|||||||
138 DQPRAIALDPSSGFMYWTDWGEVPKIERAGMDGSSRFIIINSEIYWPNGL 187

201 TIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTL 250
    |:|.|||||||||||||.|||:.||||: || ||.|||.||||||| :|.|
188 TLDYEEQKLYWADAKLNFIHKSNLDGTNRQAVVKGSLPHPFALTLFEDIL 237

251 YWTDWQTRSIHACNKRTGGKRKEILSALYSPMDIQVLSQERQPFFHTRCE 300
    ||||| |:|| |||| ||:  :|| |.::||||||:.:||:|||    ..|:
238 YWTDWSTHSILACNKYTGEGLREIHSDIFSPMDIHAFSQQRQPNATNPCG 287

301 EDNGGCSHLCLLSPSEPFYTCACPTGVQLQDNGRTCKAGAEEVLLLARRT 350
     |||||||||||:|| .||| |||||||.| :||:|||.||.|:||||||||
288 IDNGGCSHLCLMSPVKPFYQCACPTGVKLLENGKTCKDGATELLLLARRT 337

351 DLRRISLDTPDFTDIVLQVDDIRHAIAIDYDPLEGYVYWTDDEVRAIRRA 400
    ||||||||||||||||||::|||||||||||||:|||:||||||||||||.
338 DLRRISLDTPDFTDIVLQLEDIRHAIAIDYDPVEGYIYWTDDEVRAIRRS 387

401 YLDGSGAQTLVNTEINDPDGIAVDWVARNLYWTDTGTDRIEVTRLNGTSR 450
    ::|||||.|  :|..:|..|||||||||||||||||||||||||||||| |
388 FIDGSGSQFVVTAQIAHPDGIAVDWVARNLYWTDTGTDRIEVTRLNGTMR 437

451 KILVSEDLDEPRAIALHPVMGLMYWTDWGENPKIECANLDGQERRVLVNA 500
    |||:||||:|||||.|.|::|.|||||||| |||| |.||| :| ||||.
438 KILISEDLEEPRAIVLDPMVGYMYWTDWGEIPKIERAALDGSDRVVLVNT 487

501 SLGWPNGLALDLQEGKLYWGDAKTDKIEVINVDGTKRRTLLEDKLPHIFG 550
    ||||||||||.:|||:||||||||||||:|.||| ||.|:|||:||||||
488 SLGWPNGLALDYDEGKIYWGDAKTDKIEVMNTDGTGRRVLVEDKIPHIFG 537

551 FTLLGDFIYWTDWQRRSIERVHKVKASRDVIIDQLPDLMGLKAVNVAKVV 600
    ||||||::|||||||||||||||| .|.|:|||||||||||||.|| :|:
538 FTLLGDYVYWTDWQRRSIERVHKRSAEREVIIDQLPDLMGLKATNVHRVI 587

601 GTNPCADRNGGCSHLCFFTPHATRCGCPIGLELLSDMKTCIVPEAFLVFT 650
    |.||||: |||||||:: |:: ||:|||||:||:||||||||||||:|.
588 GSNPCAEENGGCSHLCLYRPQGLRCACPIGFELISDMKTCIVPEAFLLFS 637
```

FIG. 27B

Alignment of Human LRP5 and LRP6

```
651 SRAAIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIYWTDVSLKTIS 700
    .||.|:||||||||:||||||||||||||||.:|:||||:||||||
638 RRADIRRISLETNNNNVAIPLTGVKEASALDFDVTDNRIYWTDISLKTIS 687

701 RAFMNGSSVEHVVEFGLDYPEGMAVDWMGKNLYWADTGTNRIEVARLDGQ 750
    |||||||.:|||||||||||||||||:||||||||||||||.:||||
688 RAFMNGSALEHVVEFGLDYPEGMAVDWLGKNLYWADTGTNRIEVSKLDGQ 737

751 FRQVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNCMTLV 800
     ||||||:|||.||.|||||..|::||||||||:| || |||.:  |||
738 HRQVLVWKDLDSPRALALDPAEGFMYWTEWGGKPKIDRAAMDGSERTTLV 787

801 DKVGRANDLTIDYADQRLYWTDLDTNMIESSNMLGQERVVIADDLPHPFG 850
    ..||||||||||..||||||||||:|||||||  :|  ||||||||||
788 PNVGRANGLTIDYAKRRLYWTDLDTNLIESSNMLGLNREVIADDLPHPFG 837

851 LTQYSDYIYWTDWNLHSIERADKTSGRNRTLIQGHLDFVMDILVFHSSRQ 900
    ||||  |||||||| .  :||||||||.|||:|||||||:||||||||||
838 LTQYQDYIYWTDWSRRSIERANKTSGQNRTIIQGHLDYVMDILVFHSSRQ 887

901 DGLNDCMHNNGQCGQLCLAIP.GGHRCGCASHYTLDPSSRNCSPPTTFLL 949
    .|:|:|  .||:|::||||:|  ||  |||:.||.|::..|.||:||||||
888 SGWNECASSNGHCSHLCLAVPVGGFVCGCPAHYSLNADNRTCSAPTTFLL 937

950 FSQKSAISRMIPDDQHSPDLILPLHGLRNVKAIDYDPLDKFIYWVDGRQN 999
    |||||||.||: |:|:|||:|||:|:||||:||||||||| :||:|:|||
938 FSQKSAINRMVIDEQQSPDIILPIHSLRNVRAIDYDPLDKQLYWIDSRQN 987

1000 .IKRAKDDGTQPF.VLTSLSQGQNPDRQPHDLSIDIYSRTLFWTCEATNT 1047
     |::|.:||.|.| |:.|  .:|| : ||.|||||||| ::||||||.
988 MIRKAQEDGSQGFTVVVSSVPSQNLEIQPYDLSIDIYSRYIYWTCEATNV 1037

1048 INVHRLSGEAMGVVLRGDRDKPRAIVVNAERGYLYFTNMQDRAAKIERAA 1097
     ||| ||.| .:||||:|:.|:||||||||:|:||:||||:|:|.:|||||
1038 INVTRLDGRSVGVVLKGEQDRPRAIVVNPEKGYMYFTNLQERSPKIERAA 1087

1098 LDGTEREVLFTTGLIRPVALVVDNTLGKLFWVDADLKRIESCDLSGANRL 1147
     |||||||||| .|| :|:||.:|. ||||||.|.||:||||:||||||:
1088 LDGTEREVLFFSGLSKPIALALDSRLGKLFWADSDLRRIESSDLSGANRI 1137

1148 TLEDANIVQPLGLTILGKHLYWIDRQQQMIERVEKTTGDKRTRIQGRVAH 1197
     .|||.||:||:|||:::. |||||:||||||:::.|. : ||::|:|:|:
1138 VLEDSNILQPVGLTVFENWLYWIDKQQQMIEKIDMTGREGRTKVQARIAQ 1187

1198 LTGIHAVEEVSLEEFSAHPCARDNGGCSHICIAKGDGTPRCSCPVHLVLL 1247
     |.:||||.|:.|:|:..||||.||||||||:.|||||.|||||:|||||
1188 LSDIHAVKELNLQEYRQHPCAQDNGGCSHICLVKGDGTTRCSCPMHLVLL 1237

1248 QNLLTCGEPPTCSPDQFACATGEIDCIPGAWRCDGFPECDDQSDEEGCPV 1297
     |: |.|||||||||:||.| ||||||||.|||||||.||:|:||| .|||
1238 QDELSCGEPPTCSPQQFTCFTGEIDCIPVAWRCDGFTECEDHSDELNCPV 1287

1298 CSAAQFPCARGQCVDLRLRCDGEADCQDRSDEADCDAICLPNQFRCASGQ 1347
     ||..||.||.|||:|  |||:|:|:|||:||| :|:.:|| :|||||.||
1288 CSESQFQCASGQCIDGALRCNGDANCQDKSDEKNCEVLCLIDQFRCANGQ 1337

1348 CVLIKQQCDSFPDCIDGSDELMCEITKPPSDDSPAHSSAIGPVIGIILSL 1397
     |:  ...||  .|| | |||| |  |..|  .|. ...:|.|||:|:.:
1338 CIGKHKKCDHNVDCSDKSDELDCYPTEEP...APQATNTVGSVIGVIVTI 1384
```

FIG. 27C

Alignment of Human LRP5 and LRP6

```
1398 FVMGGVYFVCQRVVCQRYAGANGPFPHEYVS.GTPHVPLNFIAPGGSQHG 1446
     ||  |.||||:|||::|.|   |.:.:.:::||   |.: |||.:::..:|  |
1385 FVSGTVYFICQRMLCPRMKGDGETMTNDYVVHGPASVPLGYVPHPSSLSG 1434

1447 PFTGIACGKSMMSSVSLMGGRGGVPLYDRNHVTGASSSSSSSTKATLYPP 1496
     .:.|:. ||||:||:|:|||.:|.| |||.||||||||||||||:|.:|:
1435 SLPGMSRGKSMISSLSIMGGSSGPP.YDRAHVTGASSSSSSSTKGTYFPA 1483

1497 ILNPPPSPATDPSLYNMDMFYSSNIPATAR..PYRPYIIRGMAPPTTPCS 1544
     ||||||||||:.| |.|:: |||| |.| |  .|||| .| :||||||||
1484 ILNPPPSPATERSHYTMEFGYSSNSPSTHRSYSYRPYSYRHFAPPTTPCS 1533

1545 TDVCDSDYSASRWKAS.....KYYLDLNSDSDPYPPPPTPHSQYLSAE.. 1587
     ||||||||.:||:..|       |  ||| ||:| |||||||:|||||||
1534 TDVCDSDYAPSRRMTSVATAKGYTSDLNYDSEPVPPPPTPRSQYLSAEEN 1583

1588 .DSCPPSPATERSY.FHLFPPPPSPCTDSS 1615
      :|||||| |||||  ||:|||||||||||
1584 YESCPPSPYTERSYSHHLYPPPPSPCTDSS 1613
```

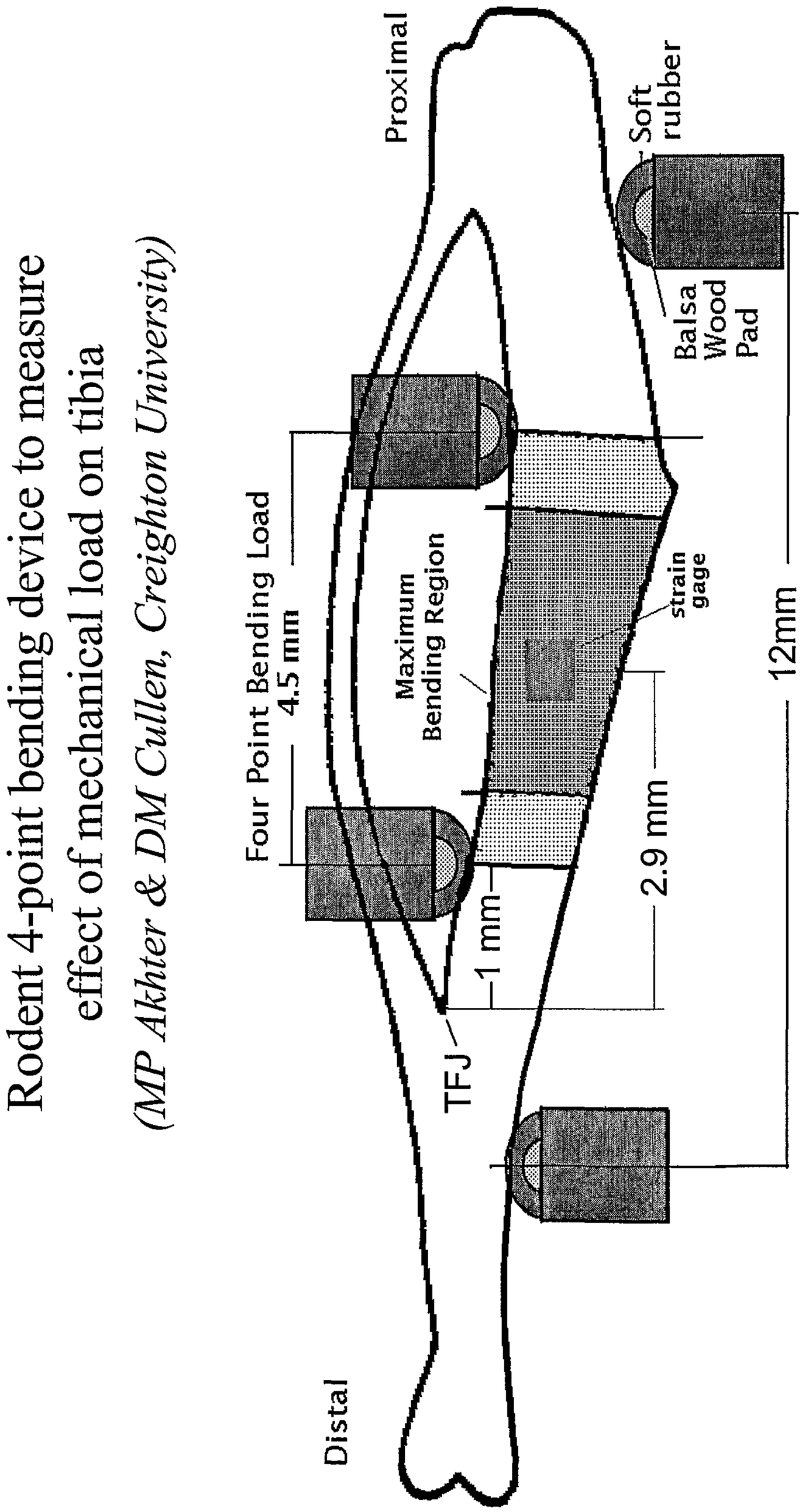

Rodent 4-point bending device to measure effect of mechanical load on tibia
(MP Akhter & DM Cullen, Creighton University)
Proximal
Soft rubber
Balsa Wood Pad
Four Point Bending Load
4.5 mm
Maximum Bending Region
strain gage
12mm
2.9 mm
1 mm
TFJ
Distal

FIG. 28

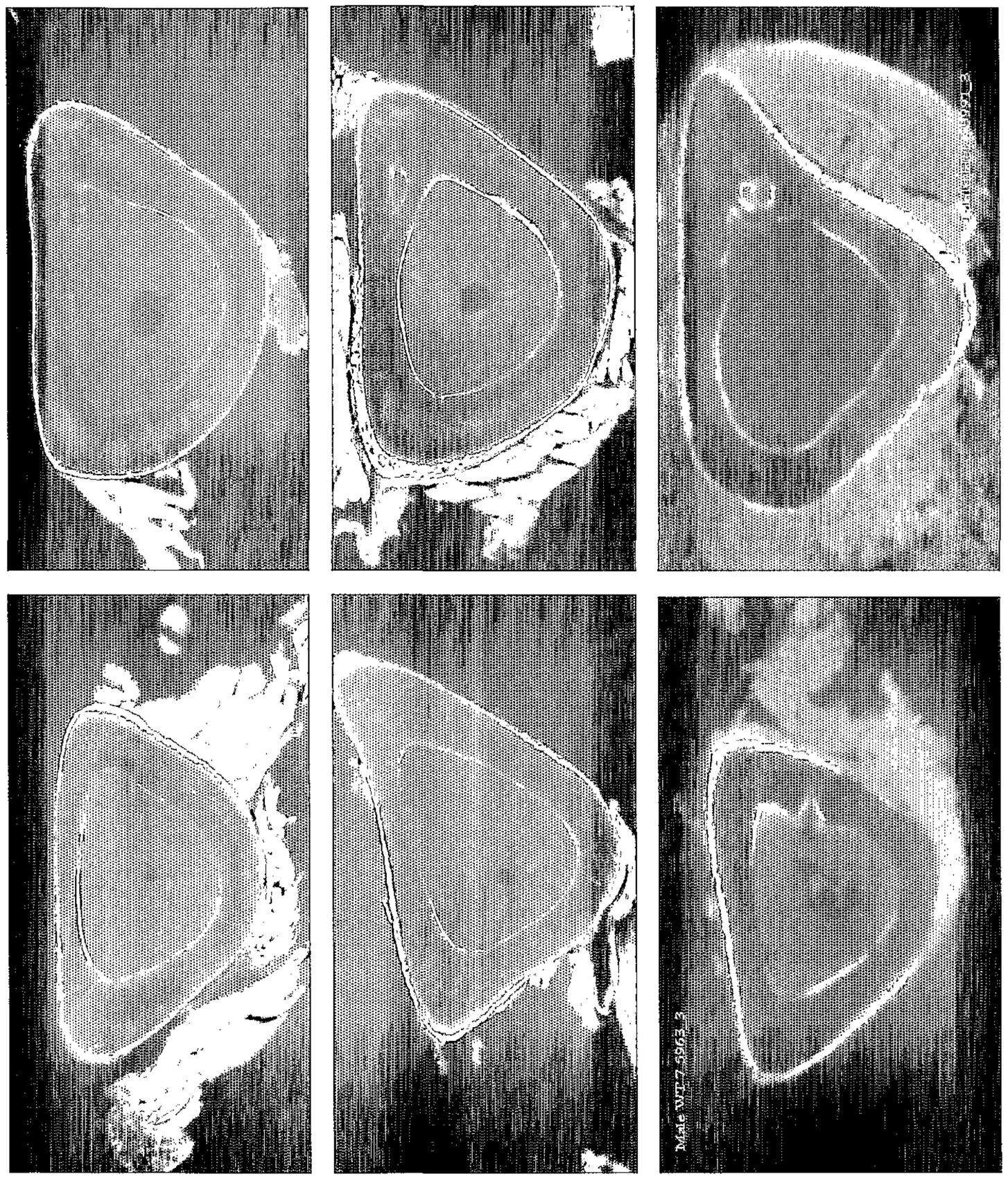
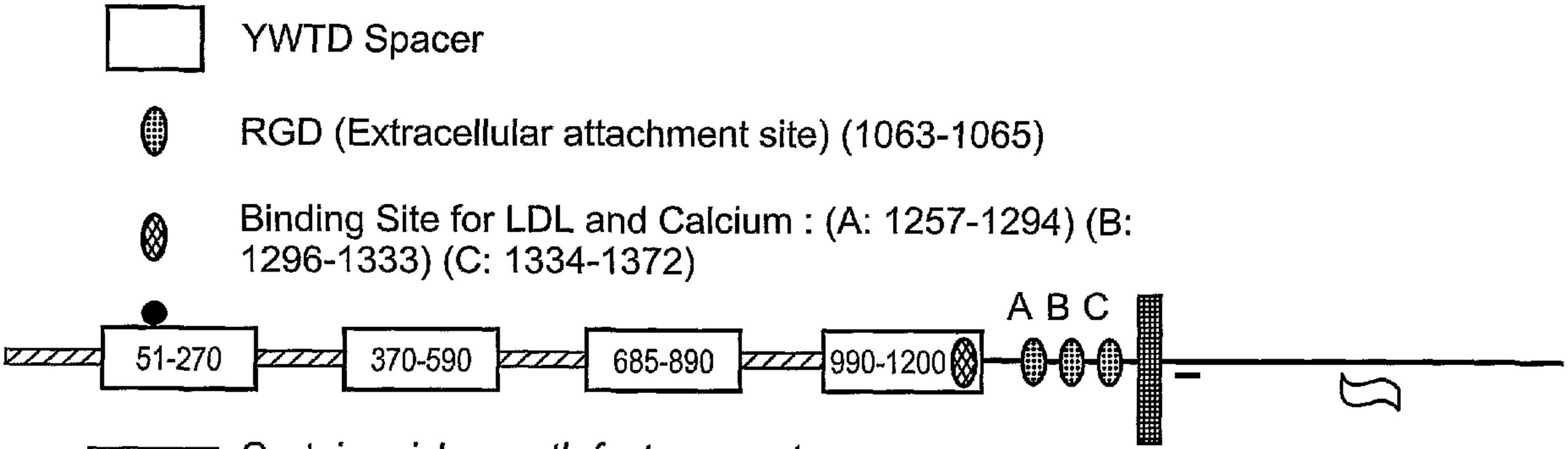

Enhanced bone formation in HBM1
transgenic mice: *Response to mechanical loading*
HET
NTG
Female 5N unloaded
Female 5N loaded
Male 7N loaded

FIG. 29

Model for a LDL Receptor-Related protein, Zmax1



YWTD Spacer

RGD (Extracellular attachment site) (1063-1065)

Binding Site for LDL and Calcium : (A: 1257-1294) (B: 1296-1333) (C: 1334-1372)

Cysteine-rich growth factor repeats

Transmembrane Region (1387-1408)

Ideal PEST region (With CK-II phosphorylation site)

Internalization Domain (1419-1422)

Site of Glycine to Valine change in HBM allele