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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 January 2009 (15.01.2009)

PCT

(10) International Publication Number
WO 2009/009431 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2008/069201

(22) International Filing Date: 3 July 2008 (03.07.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/958,629 6 July 2007 (06.07.2007) US

(71) Applicant *(for all designated States except US)*: **THE REGENTS OF THE UNIVERSITY OF MICHIGAN** [US/US]; 3003 S. State St., Ann Arbor, MI 48109 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHINNAIYAN, Arul, M.** [US/US]; 51300 Plymouth Valley Drive, Plymouth, MI 48170 (US). **TOMLINS, Scott** [US/US]; 1077 Barton Dr. #112, Ann Arbor, MI 48105 (US). **DHANASEKARAN, Saravana, Mohan** [IN/US]; 921 Church Street #110, Ann Arbor, MI 48104 (US).

(74) Agent: **ARENSON, Tanya, A.**; Casimir Jones, S.C., 440 Science Drive, Suite 203, Madison, WI 53711 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (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*

WO 2009/009431 A2

(54) Title: MIPOL1-ETV1 GENE REARRANGEMENTS

(57) Abstract: Compositions and methods associated with recurrent MIPOL1-ETV1 genetic rearrangements that are useful for cancer diagnosis and therapy are disclosed.

MIPOL1-ETV1 GENE REARRANGEMENTS

This application claims priority to U.S. Provisional Application Serial No.: 60/958,629, filed July 6, 2007, which is herein incorporated by reference in its entirety.

This invention was made with government support under CA069568 and CA111275 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE INVENTION

This invention relates to compositions and methods for cancer diagnosis, research and therapy, including but not limited to, cancer markers. In particular, this invention relates to recurrent MIPOL1-ETV1 genetic rearrangements that are useful as diagnostic markers and clinical targets for prostate cancer.

BACKGROUND OF THE INVENTION

Cancer research may identify altered genes that are causally implicated in oncogenesis. Several types of somatic mutations that result in altered activity of an oncogene or tumor suppressor gene have been identified, including base substitutions, insertions, deletions, translocations, and chromosomal gains and losses. Compelling evidence exists for a causal role for some chromosomal rearrangements in cancer (Rowley, *Nat Rev Cancer* 1: 245 (2001)). Recurrent chromosomal aberrations have been primarily characteristic of leukemias, lymphomas, and sarcomas. Less than 1% of the known, disease-specific chromosomal rearrangements are associated with epithelial tumors (carcinomas), although those cancers are much more common and contribute to a relatively large fraction of the morbidity and mortality associated with human cancer(Mitelman, *Mutat Res* 462: 247 (2000)). While hematological malignancies are often characterized by disease-specific chromosomal rearrangements, most solid tumors have a plethora of non-specific chromosomal aberrations. Karyotypic complexity of solid tumors is thought to result from secondary alterations acquired through cancer evolution or progression.

Cancer-related chromosomal rearrangements may result from two primary mechanisms. In one, promoter/enhancer elements of one gene are rearranged adjacent to a proto-oncogene, thus causing altered expression of an oncogenic protein. This

type of translocation is exemplified by the apposition of immunoglobulin (IG) and T-cell receptor (TCR) genes to the MYC oncogene, leading to oncogene activation in B- and T-cell malignancies, respectively (Rabbitts, *Nature* 372: 143 (1994)). In the other mechanism, rearrangement results in the fusion of two genes, which produces a fusion protein that may have a new function or altered activity. This type of translocation is exemplified by the BCR-ABL gene fusion in chronic myelogenous leukemia (CML) (Rowley, *Nature* 243: 290 (1973); de Klein et al., *Nature* 300: 765 (1982)), which led to the rational development of imatinib mesylate that successfully targets the BCR-ABL kinase (Deininger et al., *Blood* 105: 2640 (2005)).

Recurrent MIPOL1-ETV1 genetic rearrangements are described herein, which are useful for diagnosis and therapeutic applications related to human epithelial tumors.

SUMMARY OF THE INVENTION

A method is disclosed for diagnosing prostate cancer comprising detecting the presence or absence in a biological sample of a MIPOL1-ETV1 genetic rearrangement, wherein the presence in the sample of the genetic rearrangement is indicative of prostate cancer in the individual from whom the sample was obtained. In some embodiments, the sample is tissue, blood, plasma, serum, urine, semen, prostatic secretions or prostate cells. In some embodiments, the detecting step comprises detecting chromosomal rearrangements of genomic DNA encoding MIPOL1 and ETV1. The detecting step may use a nucleic acid sequencing technique or a nucleic acid hybridization technique, such as *in situ* hybridization (ISH), hybridization to one or more moieties in a microarray, or Southern blot analysis. In some embodiments, the detecting step further includes nucleic acid amplification, which may use known methods that include, but are not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA). In some embodiments, the detecting step detects mRNA associated with MIPOL1-ETV1 genetic rearrangements or protein expression resulting from MIPOL1-ETV1 genetic rearrangements. Such detection may include analysis of RNA and/or protein expression levels, or determination of sequence characteristics.

Compositions are disclosed for diagnosing prostate cancer comprising a reagent that directly or indirectly detects a junction between ETV1 genetic material and MIPOL1 genetic material associated with a MIPOL1-ETV1 genetic rearrangement. Embodiments of such reagents include: a probe comprising a sequence that hybridizes to the junction between ETV1 genetic material and MIPOL1 genetic material associated with a MIPOL1-ETV1 genetic rearrangement, which may be the junction at which a ETV1 gene is inserted into a MIPOL1 gene; a combination of first and second probes, in which a first probe comprises a sequence that hybridizes to the ETV1 gene and a second probe comprises a sequence that hybridizes to the MIPOL1 gene; and at least one first amplification oligonucleotide that comprises a sequence that hybridizes specifically to a ETV1 gene and at least one second amplification oligonucleotide that comprises a sequence that hybridizes specifically to an MIPOL1 gene.

In some embodiments, the composition of amplification oligonucleotides may also include a probe that hybridizes specifically to a sequence located between the sequences hybridized by the first amplificaiton oligonucleotide and the second oligonucleotide, which probe may hybridize specifically to a sequence in the ETV1 gene or in the MIPOL1 gene. All probes may be linked directly or indirectly to a label that provides a detectable signal.

DESCRIPTION OF THE FIGURES

FIG. 1 illustrates the entire ETV1 locus rearranged to 14q13.3-14q21.1 in LNCaP and MDA-PCa 2B prostate cancer cells. FIG. 1, a is a schematic illustration of the ETV1 locus on chromosome 7 and FIG. 1, b is a schematic illustration of 14q13.3-14q21.1 on chromosome 14 and BACs used as probes for fluorescence in situ hybridization (FISH) to detect genetic rearrangements. FIG. 1, c and –d illustrate FISH performed by using BACs indicated with the corresponding fluorescent label on metaphase spreads from LNCaP (tetraploid) cells (FIG. 1, c) and d) MDA-PCa 2B (diploid) cells (FIG. 1, d) to detect rearrangements at the ETV1 locus (left panel) and 14q13.3-14q21.1 (right panel). FIG. 1, e-g, illustrate structure of ETV1 and 14q13.3-14q21.1 in: normal cells (FIG. 1, e), LNCaP cells (FIG. 1, f), and MDA-PCa 2B cells (FIG. 1, g), as determined by FISH for all.

FIG. 2 shows that the ETV1 locus is rearranged to chromosome 14 in LNCaP cells. FIG. 2, a-b, are schematic illustrations of BACs used from chromosomes 7p

and 14q32, respectively (previously FISH mapped to chromosome 14). FIG. 2, c. illustrates FISH using BACs labeled with the indicated fluorescent label showed two copies of ETV1 on chromosome 7 and two copies on chromosome 14, as identified by RP11-483K13.

FIG. 3 shows the identification of the genomic breakpoint at the ETV1 locus in LNCaP cells. FIG. 3, a illustrates FISH used to narrow the breakpoint region between BACs 12 and 2, as shown in green in FIG. 3, b, as BACs 1 and 12 co-localized to chromosome 7 and chromosome 14. FIG. 3, b illustrates a series of 22 probes (A-Q) for Southern blotting that were designed to span the implicated region. FIG. 3, c illustrates a restriction map of the Probe A region, with restriction sites for PstI and EcoRI indicated, which was the only probe to show rearranged bands on Southern blotting. FIG. 3, d illustrates Southern blotting with Probe A which showed additional bands in LNCaP genomic DNA digested with EcoRI and PstI, but not in VCaP, normal male ("Nor male") genomic DNA or human placental DNA.

FIG. 4 shows inverse PCR that identifies the insertion of ETV1 into the MIPOL1 locus on chromosome 14. FIG. 4, a illustrates, a series of divergent PCR primers designed for inverse PCR (A3-B3) which were based on the detection of rearrangements in LNCaP genomic DNA digested with EcoRI and PstI using Probe A.. FIG. 4, b illustrates results obtained from PCR amplification using nested PCRs subsequently with A1 and B1 (1), A2 and B2 (2), and A3 and B3 (3) on PstI digested low (L) and high (H) passage LNCaP genomic DNA . FIG. 4, c illustrates the sequence of the nested PCR product of A3 and B3 (shown in FIG. 4, b), in which the DNA breakpoint is shown in red, and the partner sequence was intronic DNA from the MIPOL1 locus at 14q13.3-14q21.1. FIG. 4, d illustrates the breakpoint location shown by an asterisk. FIG. 4, e illustrates primers designed (as shown in FIG. 4, c) to confirm the fusion by PCR on undigested DNA, in which PCR confirmed the fusion in both low and high passage LNCaP cells, but not VcaP, normal male ("Nor male") genomic DNA or human placental DNA.

FIG. 5 shows that the 7p telomere is retained on all copies of chromosome 7 in LNCaP, by using FISH performed using chromosome 7 centromeric and 7p telomere probes on LNCaP metaphases. All four copies of chromosome 7 identified by the centromeric probes retained their 7p telomeric sequence, demonstrating that the ETV1 rearrangement is not a translocation nor involves a telomeric deletion at 7p.

FIG. 6 shows that ETV1 and the contiguous genes at 14q13.3-14q21.1 are coordinately regulated in prostate cancer, LNCaP and the androgen insensitive LNCaP derivative C4-2B. FIG. 6, a illustrates the tissue specificity of the 4 contiguous genes in the 14q13.3-14q21.1 region which was interrogated in the expo dataset (using Oncomine) and expression (in normalized expression units) is shown for 28 distinct tumor types and prostate cancer. FIG. 6, b and –c illustrate androgen regulation of the contiguous 14q13.3-14q21.1 genes (FIG. 6, b) and ETV1 (FIG. 6, c), as determined by qPCR in LNCaP cells with (+) or without (-) stimulation by the synthetic androgen R1881. Ratios of target gene to GAPDH (mean (n = 4) + S.E.) are shown. FIG. 6, d illustrates that LNCaP and its androgen insensitive derivative, C4-2B, were profiled using a microarray (Agilent Whole Genome Microarrays), in which the top ten most differentially expressed features (LNCaP/C4-2B) are shown. FIG. 6, e-g illustrate qPCR results that show decreased expression of androgen regulated genes in C4-2B compared to LNCaP, for markers ETV1 (FIG. 6, e) and PSA (FIG. 6, f), and the 4 contiguous transcripts at 14q13.3-14q21.1 (FIG. 6, g).

FIG. 7 shows that MDA-PCa 2B has outlier expression of ETV1, in which expression of ERG and ETV1 was determined by qPCR for the prostate cancer cell lines MDA-PCa 2B, LNCaP, PC3, NCI-H660, 22RV1, VCaP, DU145, LAPC4 and the immortalized benign prostate epithelial cell line RWPE, and the amounts of ERG and ETV1 for each sample were normalized to the average amount of GAPDH and HMBS for each sample.

FIG. 8 shows that the LNCaP androgen insensitive derivative C4-2B harbors the same ETV1 rearrangement as the parental LNCaP cell line. FIG. 8, a is a schematic of the ETV1 locus on chromosome 7 and BACs used as probes for fluorescence in situ hybridization (FISH) to detect rearrangements at the ETV1 locus, in which the breakpoint locations identified in LNCaP are indicated by an asterisk. FIG. 8, b illustrates results obtained by FISH performed by using BACs indicated with the corresponding fluorescent label on C4-2B (tetraploid). FIG. 8, c illustrates results obtained by PCR using primers from chromosomes 7 and 14 which confirmed the rearrangement of the ETV1 locus in C4-2B, similarly to low and high passage LNCaP, and in which no products were obtained from other prostate cancer cell lines, human placental DNA, or normal male human DNA.

FIG. 9 shows that the LNCaP androgen insensitive derivative C4-2B harbors the same ETV1 rearrangement as the parental LNCaP cell line. FIG. 9, a ia a

schematic illustration of the ETV1 locus (purple) on chromosome 7 (blue) and BACs used as probes for fluorescence in situ hybridization (FISH) to detect rearrangements at the ETV1 locus. FIG. 9, billustrates results obtained by FISH performed by using BACs indicated with the corresponding fluorescent label on C4-2B (tetraploid). FIG. 9, c illustrates results obtained by PCR using primers from chromosomes 7 and 14 that confirmed the rearrangement of the ETV1 locus in C4-2B, similarly to low and high passage LNCaP, and in which no products were obtained from other prostate cancer cell lines, human placental DNA, or normal male human DNA.

DEFINITIONS

To facilitate an understanding of this disclosure, terms are defined below:

As used herein, "gene rearrangement" or "genetic rearrangement" refers to any altered arrangement of genomic DNA resulting from the chromosomal rearrangement of two distinct genetic regions, two different genes, or portions of two different genes. The product of a genetic rearrangement includes a fusion of genetic material that does not exist in the absence of the genetic rearrangement. That is, the genetic rearrangement results in a junction that fuses one genetic sequence to another genetic sequence to produce a junction does not exist in a wild type non-rearranged genome. Examples of genetic rearrangements include, but are not limited to, insertions of all or part of a first gene or genetic locus into a second gene or genetic locus. In some cases, an insertion may result in a new configuration of the first and second genes or genetic loci that changes regulation of expression, e.g., resulting in increased or decreased expression of a wild type protein. In other cases, an insertion may result in a gene fusion or chimeric DNA that is transcribed to make a chimeric RNA, which may be translated to make altered protein(s) compared to wild type protein(s), i.e., resulting in a new product that is different than made in a wild type configuration.

As used herein, "gene fusion" may refer to a chimeric genomic DNA, a chimeric mRNA that is transcribed from a chimeric DNA (which may be referred to as "gene fusion RNA"), or a protein translated from chimeric mRNA which is transcribed from a chimeric DNA. Such a "gene fusion protein" (or "gene fusion polypeptide") may be a truncated protein compared to wild type protein, or a chimeric protein resulting from expression of at least a portion of a first gene fused to at least a

portion of a second gene. A gene fusion (or RNA transcribed from the gene fusion DNA) need not include entire genes or exons of genes.

As used herein, the term "transcriptional regulatory region" refers to the non-coding upstream regulatory sequence of a gene, also called the 5' untranslated region (5'UTR).

As used herein, the terms "detect", "detecting", or "detection" may describe either the general act of discovering or discerning or the specific observation of a detectably labeled composition.

As used herein, the term "inhibits at least one biological activity of a gene fusion" refers to any agent that decreases any activity of a gene fusion disclosed herein, its transcript(s) and/or translation products (*e.g.*, including, but not limited to, the activities described herein), via directly contacting a gene fusion protein, contacting gene fusion mRNA or genomic DNA, causing conformational changes of gene fusion polypeptides, decreasing gene fusion protein levels, or interfering with gene fusion interactions with signaling partners, and affecting the expression of target gene fusions (i.e., transcription or translation of the resulting transcripts). Inhibitors also include molecules that indirectly regulate gene fusion biological activity by intercepting upstream signaling molecules.

As used herein, the term "siRNAs" refers to small interfering RNAs. In some embodiments, siRNAs comprise a duplex, or double-stranded region, of about 18-25 nucleotides long; often siRNAs contain from about two to four unpaired nucleotides at the 3' end of each strand. At least one strand of the duplex or double-stranded region of a siRNA is substantially homologous to, or substantially complementary to, a target RNA molecule. The strand complementary to a target RNA molecule is the "antisense strand;" the strand homologous to the target RNA molecule is the "sense strand," and is also complementary to the siRNA antisense strand. siRNAs may also contain additional sequences; non-limiting examples of such sequences include linking sequences, or loops, as well as stem and other folded structures. siRNAs appear to function as key intermediaries in triggering RNA interference in invertebrates and in vertebrates, and in triggering sequence-specific RNA degradation during posttranscriptional gene silencing in plants.

The term "RNA interference" or "RNAi" refers to the silencing or decreasing of gene expression by siRNAs. It is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by siRNA that is

homologous in its duplex region to the sequence of the silenced gene. The gene may be endogenous or exogenous to the organism, present integrated into a chromosome or present in a transfection vector that is not integrated into the genome. The expression of the gene is either completely or partially inhibited. RNAi may also be considered to inhibit the function of a target RNA; the function of the target RNA may be complete or partial.

As used herein, the term "stage of cancer" refers to a qualitative or quantitative assessment of the level of advancement of a cancer. Criteria used to determine the stage of a cancer include, but are not limited to, the size of the tumor and the extent of metastases (*e.g.*, localized or distant).

As used herein, the term "gene transfer system" refers to any means of delivering a composition comprising a nucleic acid sequence to a cell or tissue. For example, gene transfer systems include, but are not limited to, vectors (*e.g.*, retroviral, adenoviral, adeno-associated viral, and other nucleic acid-based delivery systems), microinjection of naked nucleic acid, polymer-based delivery systems (*e.g.*, liposome-based and metallic particle-based systems), biolistic injection, and the like. As used herein, the term "viral gene transfer system" refers to gene transfer systems comprising viral elements (*e.g.*, intact viruses, modified viruses and viral components such as nucleic acids or proteins) to facilitate delivery of the sample to a desired cell or tissue. As used herein, the term "adenovirus gene transfer system" refers to gene transfer systems comprising intact or altered viruses belonging to the family Adenoviridae.

As used herein, the term "site-specific recombination target sequences" refers to nucleic acid sequences that provide recognition sequences for recombination factors and the location where recombination takes place.

As used herein, the term "nucleic acid molecule" refers to any nucleic acid containing molecule, including but not limited to, DNA or RNA. The term encompasses sequences that include any of the known base analogs of DNA and RNA including, but not limited to, 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxylmethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, dihydrouracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine,

5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine.

The term "gene" refers to a nucleic acid (*e.g.*, DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (*e.g.*, rRNA, tRNA). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (*e.g.*, enzymatic activity, ligand binding, signal transduction, immunogenicity, etc.) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and the sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length mRNA. Sequences located 5' of the coding region and present on the mRNA are referred to as 5' non-translated sequences. Sequences located 3' or downstream of the coding region and present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

As used herein, the term "heterologous gene" refers to a gene that is not in its natural environment. For example, a heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (*e.g.*, mutated, added in multiple copies, linked to non-native regulatory sequences, etc). Heterologous genes are distinguished from endogenous genes in that the heterologous gene sequences are

typically joined to DNA sequences that are not found naturally associated with the gene sequences in the chromosome or are associated with portions of the chromosome not found in nature (*e.g.*, genes expressed in loci where the gene is not normally expressed).

As used herein, the term "oligonucleotide," refers to a short length of single-stranded polynucleotide chain. Oligonucleotides are typically less than 200 residues long (*e.g.*, between 15 and 100), however, as used herein, the term is also intended to encompass longer polynucleotide chains. Oligonucleotides are often referred to by their length. For example a 24 residue oligonucleotide is referred to as a "24-mer". Oligonucleotides can form secondary and tertiary structures by self-hybridizing or by hybridizing to other polynucleotides. Such structures can include, but are not limited to, duplexes, hairpins, cruciforms, bends, and triplexes.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, the sequence "5'-A-G-T-3'," is complementary to the sequence "3'-T-C-A-5'." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (*i.e.*, identity). A partially complementary sequence is a nucleic acid molecule that at least partially inhibits a completely complementary nucleic acid molecule from hybridizing to a target nucleic acid is "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous nucleic acid molecule to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to

one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target that is substantially non-complementary (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

A gene may produce multiple RNA species that are generated by differential splicing of the primary RNA transcript. cDNAs that are splice variants of the same gene will contain regions of sequence identity or complete homology (representing the presence of the same exon or portion of the same exon on both cDNAs) and regions of complete non-identity (for example, representing the presence of exon "A" on cDNA 1 wherein cDNA 2 contains exon "B" instead). Because the two cDNAs contain regions of sequence identity they will both hybridize to a probe derived from the entire gene or portions of the gene containing sequences found on both cDNAs; the two splice variants are therefore substantially homologous to such a probe and to each other.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Under "low stringency conditions" a nucleic acid sequence of interest will hybridize to its exact

complement, sequences with single base mismatches, closely related sequences (*e.g.*, sequences with 90% or greater homology), and sequences having only partial homology (*e.g.*, sequences with 50-90% homology). Under 'medium stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, sequences with single base mismatches, and closely relation sequences (*e.g.*, 90% or greater homology). Under "high stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, and (depending on conditions such a temperature) sequences with single base mismatches. In other words, under conditions of high stringency the temperature can be raised so as to exclude hybridization to sequences with single base mismatches.

"High stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH_2PO_4 H_2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH_2PO_4 H_2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Low stringency conditions" comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH_2PO_4 H_2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharamcia), 5 g BSA (Fraction V; Sigma)] and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA,

RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (*e.g.*, increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.) (see definition above for "stringency").

As used herein, the term “amplification oligonucleotide” refers to an oligonucleotide that hybridizes to a target nucleic acid, or its complement, and participates in a nucleic acid amplification reaction. An example of an amplification oligonucleotide is a “primer” that hybridizes to a template nucleic acid and contains a 3' OH end that is extended by a polymerase in an amplification process. Another example of an amplification oligonucleotide is an oligonucleotide that is not extended by a polymerase (*e.g.*, because it has a 3' blocked end) but participates in or facilitates amplification. Amplification oligonucleotides may optionally include modified nucleotides or analogs, or additional nucleotides that participate in an amplification reaction but are not complementary to or contained in the target nucleic acid. Amplification oligonucleotides may contain a sequence that is not complementary to the target or template sequence. For example, the 5' region of a primer may include a promoter sequence that is non-complementary to the target nucleic acid (referred to as a “promoter-primer”). Those skilled in the art will understand that an amplification oligonucleotide that functions as a primer may be modified to include a 5' promoter sequence, and thus function as a promoter-primer. Similarly, a promoter-primer may be modified by removal of, or synthesis without, a promoter sequence and still function as a primer. A 3' blocked amplification oligonucleotide may provide a promoter sequence and serve as a template for polymerization (referred to as a “promoter-provider”).

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as

DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

As used herein, the term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to at least a portion of another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the disclosed methods may be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. The compositions and methods that use such compositions as disclosed herein are not limited to any particular detection system or label.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" or "isolated polynucleotide" refers to a nucleic acid sequence that is identified and separated from at least one component or contaminant with which it is ordinarily associated in its natural source. Isolated nucleic acid is such present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids as nucleic acids such as DNA and RNA found in the state they exist in nature. For example, a given DNA sequence (*e.g.*, a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. However, isolated nucleic acid encoding a given protein includes, by way of example, such nucleic acid in cells ordinarily expressing the given protein where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid, oligonucleotide, or polynucleotide may be present in single-stranded or double-

stranded form. When an isolated nucleic acid, oligonucleotide or polynucleotide is to be used to express a protein, the oligonucleotide or polynucleotide will contain at a minimum the sense or coding strand (*i.e.*, the oligonucleotide or polynucleotide may be single-stranded), but may contain both the sense and anti-sense strands (*i.e.*, the oligonucleotide or polynucleotide may be double-stranded).

As used herein, the term "purified" or "to purify" refers to the removal of components (*e.g.*, contaminants) from a sample. For example, antibodies are purified by removal of contaminating non-immunoglobulin proteins; they are also purified by the removal of immunoglobulin that does not bind to the target molecule. The removal of non-immunoglobulin proteins and/or the removal of immunoglobulins that do not bind to the target molecule results in an increase in the percent of target-reactive immunoglobulins in the sample. In another example, recombinant polypeptides are expressed in bacterial host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

DETAILED DESCRIPTION OF THE INVENTION

The disclosed compositions and methods are based on the discovery of recurrent gene rearrangements in prostate cancer. These compositions and methods are useful for diagnostic applications, therapeutic methods, and evaluation of therapeutic treatments that either directly or indirectly detect or target MIPOL1-ETV1 gene rearrangements.

I. Gene Rearrangements

Recurrent gene rearrangements disclosed herein are indicative of prostate cancer. These genetic rearrangements result from chromosomal rearrangements that insert MIPOL1 genetic material into the ETV1 genetic locus. These recurrent gene rearrangements are useful diagnostic markers and clinical targets for prostate cancer.

In some embodiments, all or a portion of the ETV1 is inserted into the MIPOL1 locus (e.g., into an intron). The gene rearrangement is detectable, for example, as a chromosomal rearrangement of genomic DNA having at least a portion of an ETV1 gene inserted into a MIPOL1 locus.

II. Antibodies

The gene rearrangements disclosed herein result in proteins, which include fragments, derivatives and analogs thereof, which may be used as immunogens to produce antibodies useful for diagnostic and therapeutic applications. Such antibodies may be polyclonal or monoclonal, chimeric, humanized, single chain or Fab fragments, which may be labeled or unlabeled, all of which may be produced by using well known procedures and standard laboratory practices. *See*, *e.g.*, Burns, ed., *Immunochemical Protocols*, 3rd ed., Humana Press (2005); Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory (1988); Kozbor et al., *Immunology Today* 4: 72 (1983); Köhler and Milstein, *Nature* 256: 495 (1975).

III. Diagnostic Applications

The disclosed MIPOL1-ETV1 genetic rearrangements provide DNA, RNA and protein based diagnostic methods that detect, either directly or indirectly, the gene rearrangements or a product made specifically as a result of the genetic rearrangement. The disclosed MIPOL1-ETV1 genetic rearrangements also provide compositions useful for diagnostic purposes, such as oligonucleotide probes that specifically detect all or part of the genetic rearrangement. Such compositions may be in the form of a kit.

The disclosed diagnostic methods may be qualitative or quantitative. Quantitative methods may, e.g., discriminate between indolent and aggressive cancers via a cutoff or threshold level where expression above that level provides information on the aggressiveness of the cancer which provides useful diagnostic and/or prognostic information to a treating physician or patient. Qualitative or quantitative diagnostic methods may include amplification of a target, signal or intermediary, such as by using a universal primer that amplifies a sequence that serves as an indicator for the presence or level of the specific target associated with the MIPOL1-ETV1 genetic rearrangement.

An initial assay may confirm the presence of a gene rearrangement but not identify the specific rearrangement. A secondary assay is then performed to determine the identity of the particular rearrangement, if desired. The second assay may use a different detection technology than the initial assay.

The disclosed MIPOL1-ETV1 genetic rearrangement may be detected along with other markers in a multiplex or panel format. Markers are selected for their predictive value alone or in combination with the gene rearrangements. Exemplary

prostate cancer markers include, but are not limited to: AMACR/P504S (U.S. Pat. No. 6,262,245); PCA3 (U.S. Pat. No. 7,008,765); PCGEM1 (U.S. Pat. No. 6,828,429); prostein/P501S, P503S, P504S, P509S, P510S, prostase/P703P, P710P (U.S. Publication No. 20030185830); and, those disclosed in U.S. Pat. Nos. 5,854,206 and 6,034,218, and U.S. Publication No. 20030175736, each of which is herein incorporated by reference in its entirety. Markers for other cancers, diseases, infections, and metabolic conditions are also contemplated for inclusion in a multiplex of panel format.

The diagnostic methods as disclosed herein may be modified with reference to data correlating particular gene rearrangements with the stage, aggressiveness or progression of the disease or the presence or risk of metastasis. The information provided by these diagnostic methods provide useful information to a physician who, based on that information, may choose an appropriate therapeutic treatment for a particular patient.

A. Sample

Any biological sample suspected of containing the MIPOL1-ETV1 gene rearrangements may be tested according to the disclosed methods. Such a sample may be tissue (*e.g.*, prostate biopsy sample or tissue obtained by prostatectomy), blood, urine, semen, prostatic secretions or a fraction thereof (*e.g.*, plasma, serum, urine supernatant, urine cell pellet or prostate cells), which may be obtained from a patient or other source of biological material, e.g., autopsy sample or forensic material. In preferred embodiments, a urine sample is collected immediately following an attentive digital rectal examination (DRE), which causes prostate cells from the prostate gland to shed into the urinary tract.

The sample may be processed to isolate or enrich the sample for the gene rearrangements or cells that contain the gene rearrangements. A variety of well techniques that use standard laboratory practices may be used for this purpose, such as, e.g., centrifugation, immunocapture, cell lysis, and nucleic acid target capture (*See, e.g.*, EP Pat. No. 1 409 727, herein incorporated by reference in its entirety).

B. DNA and RNA Detection

The disclosed MIPOL1-ETV1 genetic rearrangements may be detected as chromosomal rearrangements of genomic DNA or as chimeric mRNA produced from

a chromosomal rearrangement by using a variety of well known nucleic acid techniques that rely on standard laboratory methods, such as, e.g., nucleic acid sequencing, nucleic acid hybridization, and, nucleic acid amplification.

1. Sequencing

Illustrative non-limiting examples of nucleic acid sequencing techniques include, but are not limited to, chain terminator (Sanger) sequencing and dye terminator sequencing. Those of ordinary skill in the art will recognize that because RNA is less stable in the cell and more prone to nuclease attack experimentally RNA is usually reverse transcribed to DNA before sequencing.

Chain terminator sequencing uses sequence-specific termination of a DNA synthesis reaction using modified nucleotide substrates. Extension is initiated at a specific site on the template DNA by using a short radioactive, or other labeled, oligonucleotide primer complementary to the template at that region. The oligonucleotide primer is extended using a DNA polymerase, standard four deoxynucleotide bases, and a low concentration of one chain terminating nucleotide, most commonly a di-deoxynucleotide. This reaction is repeated in four separate tubes with each of the bases taking turns as the di-deoxynucleotide. Limited incorporation of the chain terminating nucleotide by the DNA polymerase results in a series of related DNA fragments that are terminated only at positions where that particular di-deoxynucleotide is used. For each reaction tube, the fragments are size-separated by electrophoresis in a slab polyacrylamide gel or a capillary tube filled with a viscous polymer. The sequence is determined by reading which lane produces a visualized mark from the labeled primer as you scan from the top of the gel to the bottom.

Dye terminator sequencing alternatively labels the terminators. Complete sequencing can be performed in a single reaction by labeling each of the di-deoxynucleotide chain-terminators with a separate fluorescent dye, which fluoresces at a different wavelength.

2. Hybridization

Illustrative non-limiting examples of nucleic acid hybridization techniques include, but are not limited to, *in situ* hybridization (ISH), microarray, and Southern or Northern blot.

In situ hybridization (ISH) is a type of hybridization that uses a labeled complementary DNA or RNA strand as a probe to localize a specific DNA or RNA

sequence in a portion or section of tissue (*in situ*), or, if the tissue is small enough, the entire tissue (whole mount ISH). DNA ISH can be used to determine the structure of chromosomes. RNA ISH is used to measure and localize mRNAs and other transcripts within tissue sections or whole mounts. Sample cells and tissues are usually treated to fix the target transcripts in place and to increase access of the probe. The probe hybridizes to the target sequence at elevated temperature, and then the excess probe is washed away. The probe that was labeled with either radio-, fluorescent- or antigen-labeled bases is localized and quantitated in the tissue using either autoradiography, fluorescence microscopy or immunohistochemistry, respectively. ISH can also use two or more probes, labeled with radioactivity or the other non-radioactive labels, to simultaneously detect two or more transcripts.

2.1 FISH

In some embodiments, gene rearrangements are detected using fluorescence *in situ* hybridization (FISH). The preferred FISH assays use bacterial artificial chromosomes (BACs), which have been used extensively in the human genome sequencing project (see *Nature* 409: 953-958 (2001)) and clones containing specific BACs are widely available or can be made by using standard laboratory practices. Each BAC clone from the human genome has been given a reference name that unambiguously identifies it. These names can be used to find a corresponding GenBank sequence and to order copies of the clone from a distributor.

Specific BAC clones that can be used in FISH protocols to detect MIPOL1-ETV1 genetic rearrangements are probes specific for RP11-124L22 and RP11-703A4. These probes, or other suitable probes, usually are labeled with appropriate fluorescent or other markers and then used in hybridizations. The Examples section provided herein sets forth one particular protocol that is effective for measuring rearrangements but one of skill in the art will recognize that many variations of this assay can be used equally well. Specific protocols are well known in the art and can be readily adapted for detecting MIPOL1-ETV1 rearrangements. Guidance regarding such methodology is provided in many references including: *In situ* Hybridization: Medical Applications (eds. G. R. Coulton and J. de Belleroche), Kluwer Academic Publishers, Boston (1992); *In situ* Hybridization: In Neurobiology; Advances in Methodology (eds. J. H. Eberwine, K. L. Valentino, and J. D. Barchas), Oxford University Press Inc., England (1994); *In situ* Hybridization: A Practical Approach

(ed. D. G. Wilkinson), Oxford University Press Inc., England (1992)); Kuo, *et al.*, *Am. J. Hum. Genet. 49*:112-119 (1991); Klinger, *et al.*, *Am. J. Hum. Genet. 51*:55-65 (1992); and Ward, *et al.*, *Am. J. Hum. Genet. 52*:854-865 (1993)). Patents providing guidance on such methodology include U.S. 5,225,326; 5,545,524; 6,121,489 and 6,573,043, and commercially available kits also provide protocols for performing FISH (*e.g.*, from Oncor, Inc., Gaithersburg, MD). All of these references are hereby incorporated by reference in their entirety and may be used along with similar references in the art and with the information provided in the Examples section herein to establish procedural steps convenient for a particular laboratory.

2.2 Microarrays

Different kinds of biological assays are called microarrays including, but not limited to: DNA microarrays (*e.g.*, cDNA microarrays and oligonucleotide microarrays); protein microarrays; tissue microarrays; transfection or cell microarrays; chemical compound microarrays; and, antibody microarrays. A DNA microarray, commonly known as gene chip, DNA chip, or biochip, is a collection of microscopic DNA spots attached to a solid surface (*e.g.*, glass, plastic or silicon chip) forming an array for the purpose of expression profiling or monitoring expression levels for thousands of genes simultaneously. The affixed DNA segments are known as probes, thousands of which can be used in a single DNA microarray. Microarrays can be used to identify disease genes by comparing gene expression in disease and normal cells. Microarrays can be fabricated using a variety of technologies, including but not limiting: printing with fine-pointed pins onto glass slides; photolithography using pre-made masks; photolithography using dynamic micromirror devices; ink-jet printing; or, electrochemistry on microelectrode arrays.

Southern and Northern blotting is used to detect specific DNA or RNA sequences, respectively. DNA or RNA extracted from a sample is fragmented, electrophoretically separated on a matrix gel, and transferred to a membrane filter. The filter bound DNA or RNA is subject to hybridization with a labeled probe complementary to the sequence of interest. Hybridized probe bound to the filter is detected. A variant of the procedure is the reverse Northern blot, in which the substrate nucleic acid that is affixed to the membrane is a collection of isolated DNA fragments and the probe is RNA extracted from a tissue and labeled.

3. Amplification

Chromosomal rearrangements of genomic DNA and chimeric mRNA may be amplified prior to or simultaneous with detection. Illustrative non-limiting examples of nucleic acid amplification techniques include, but are not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA). Those of ordinary skill in the art will recognize that certain amplification techniques (*e.g.*, PCR) require that RNA be reversed transcribed to DNA prior to amplification (*e.g.*, RT-PCR), whereas other amplification techniques directly amplify RNA (*e.g.*, TMA and NASBA).

The polymerase chain reaction (U.S. Pat. Nos. 4,683,195, 4,683,202, 4,800,159 and 4,965,188, each of which is herein incorporated by reference in its entirety), commonly referred to as PCR, uses multiple cycles of denaturation, annealing of primer pairs to opposite strands, and primer extension to exponentially increase copy numbers of a target nucleic acid sequence. In a variation called RT-PCR, reverse transcriptase (RT) is used to make a complementary DNA (cDNA) from mRNA, and the cDNA is then amplified by PCR to produce multiple copies of DNA. For other various permutations of PCR *see*, *e.g.*, U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159; Mullis et al., *Meth. Enzymol.* 155: 335 (1987); and, Murakawa et al., *DNA* 7: 287 (1988), each of which is herein incorporated by reference in its entirety.

Transcription mediated amplification (U.S. Pat. Nos. 5,480,784 and 5,399,491, each of which is herein incorporated by reference in its entirety), commonly referred to as TMA, synthesizes multiple copies of a target nucleic acid sequence autocatalytically under conditions of substantially constant temperature, ionic strength, and pH in which multiple RNA copies of the target sequence autocatalytically generate additional copies. *See*, *e.g.*, U.S. Pat. Nos. 5,399,491 and 5,824,518, each of which is herein incorporated by reference in its entirety. In a variation described in U.S. Publ. No. 20060046265 (herein incorporated by reference in its entirety), TMA optionally incorporates the use of blocking moieties, terminating moieties, and other modifying moieties to improve TMA process sensitivity and accuracy.

The ligase chain reaction (Weiss, R., *Science* 254: 1292 (1991), herein incorporated by reference in its entirety), commonly referred to as LCR, uses two sets

of complementary DNA oligonucleotides that hybridize to adjacent regions of the target nucleic acid. The DNA oligonucleotides are covalently linked by a DNA ligase in repeated cycles of thermal denaturation, hybridization and ligation to produce a detectable double-stranded ligated oligonucleotide product.

Strand displacement amplification (Walker, G. et al., *Proc. Natl. Acad. Sci. USA* 89: 392-396 (1992); U.S. Pat. Nos. 5,270,184 and 5,455,166, each of which is herein incorporated by reference in its entirety), commonly referred to as SDA, uses cycles of annealing pairs of primer sequences to opposite strands of a target sequence, primer extension in the presence of a dNTPαS to produce a duplex hemiphosphorothioated primer extension product, endonuclease-mediated nicking of a hemimodified restriction endonuclease recognition site, and polymerase-mediated primer extension from the 3' end of the nick to displace an existing strand and produce a strand for the next round of primer annealing, nicking and strand displacement, resulting in geometric amplification of product. Thermophilic SDA (tSDA) uses thermophilic endonucleases and polymerases at higher temperatures in essentially the same method (EP Pat. No. 0 684 315).

Other amplification methods include, for example: nucleic acid sequence based amplification (U.S. Pat. No. 5,130,238, herein incorporated by reference in its entirety), commonly referred to as NASBA; one that uses an RNA replicase to amplify the probe molecule itself (Lizardi et al., *BioTechnol.* 6: 1197 (1988), herein incorporated by reference in its entirety), commonly referred to as Qβ replicase; a transcription based amplification method (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86:1173 (1989)); and, self-sustained sequence replication (Guatelli et al., *Proc. Natl. Acad. Sci. USA* 87: 1874 (1990), each of which is herein incorporated by reference in its entirety). For further discussion of known amplification methods *see* Persing, David H., "In Vitro Nucleic Acid Amplification Techniques" in *Diagnostic Medical Microbiology: Principles and Applications* (Persing et al., Eds.), pp. 51-87 (American Society for Microbiology, Washington, DC (1993)).

4. Detection Methods

Non-amplified or amplified gene rearrangement nucleic acids can be detected by any conventional means. For example, the gene rearrangements can be detected by hybridization with a detectably labeled probe and measurement of the resulting hybrids. Illustrative non-limiting examples of detection methods are described below.

One illustrative detection method, the Hybridization Protection Assay (HPA) involves hybridizing a chemiluminescent oligonucleotide probe (*e.g.*, an acridinium ester-labeled (AE) probe) to the target sequence, selectively hydrolyzing the chemiluminescent label present on unhybridized probe, and measuring the chemiluminescence produced from the remaining probe in a luminometer. *See*, *e.g.*, U.S. Pat. No. 5,283,174 and Norman C. Nelson et al., Nonisotopic Probing, Blotting, and Sequencing, ch. 17 (Larry J. Kricka ed., 2d ed. 1995, each of which is herein incorporated by reference in its entirety).

Another illustrative detection method provides for quantitative evaluation of the amplification process in real-time. Evaluation of an amplification process in "real-time" involves determining the amount of amplicon in the reaction mixture either continuously or periodically during the amplification reaction, and using the determined values to calculate the amount of target sequence initially present in the sample. A variety of methods for determining the amount of initial target sequence present in a sample based on real-time amplification are well known in the art. These include methods disclosed in U.S. Pat. Nos. 6,303,305 and 6,541,205, each of which is herein incorporated by reference in its entirety. Another method for determining the quantity of target sequence initially present in a sample, but which is not based on a real-time amplification, is disclosed in U.S. Pat. No. 5,710,029, herein incorporated by reference in its entirety.

Amplification products may be detected in real-time through the use of various self-hybridizing probes, most of which have a stem-loop structure. Such self-hybridizing probes are labeled so that they emit differently detectable signals, depending on whether the probes are in a self-hybridized state or an altered state through hybridization to a target sequence. By way of non-limiting example, "molecular torches" are a type of self-hybridizing probe that includes distinct regions of self-complementarity (referred to as "the target binding domain" and "the target closing domain") which are connected by a joining region (*e.g.*, non-nucleotide linker) and which hybridize to each other under predetermined hybridization assay conditions. In a preferred embodiment, molecular torches contain single-stranded base regions in the target binding domain that are from 1 to about 20 bases in length and are accessible for hybridization to a target sequence present in an amplification reaction under strand displacement conditions. Under strand displacement conditions, hybridization of the two complementary regions, which may be fully or partially

complementary, of the molecular torch is favored, except in the presence of the target sequence, which will bind to the single-stranded region present in the target binding domain and displace all or a portion of the target closing domain. The target binding domain and the target closing domain of a molecular torch include a detectable label or a pair of interacting labels (*e.g.*, luminescent/quencher) positioned so that a different signal is produced when the molecular torch is self-hybridized than when the molecular torch is hybridized to the target sequence, thereby permitting detection of probe:target duplexes in a test sample in the presence of unhybridized molecular torches. Molecular torches and many types of interacting label pairs are known (e.g., U.S. Pat. No. 6,534,274, herein incorporated by reference in its entirety).

Another example of a detection probe having self-complementarity is a "molecular beacon" (see U.S. Pat. Nos. 5,925,517 and 6,150,097, herein incorporated by reference in entirety). Molecular beacons include nucleic acid molecules having a target complementary sequence, an affinity pair (or nucleic acid arms) holding the probe in a closed conformation in the absence of a target sequence present in an amplification reaction, and a label pair that interacts when the probe is in a closed conformation. Hybridization of the target sequence and the target complementary sequence separates the members of the affinity pair, thereby shifting the probe to an open conformation. The shift to the open conformation is detectable due to reduced interaction of the label pair, which may be, for example, a fluorophore and a quencher (*e.g.*, DABCYL and EDANS).

Other self-hybridizing probes are well known to those of ordinary skill in the art. By way of non-limiting example, probe binding pairs having interacting labels (e.g., *see* U.S. Pat. No. 5,928,862, herein incorporated by reference in its entirety) may be adapted for use in the compositions and methods disclosed herein. Probe systems used to detect single nucleotide polymorphisms (SNPs) might also be used. Additional detection systems include "molecular switches," (e.g., *see* U.S. Publ. No. 20050042638, herein incorporated by reference in its entirety). Other probes, such as those comprising intercalating dyes and/or fluorochromes, are also useful for detection of amplification products in the methods disclosed herein (e.g., *see* U.S. Pat. No. 5,814,447, herein incorporated by reference in its entirety).

C. Data Analysis

In some embodiments, a computer-based analysis program is used to translate the raw data generated by the detection assay (*e.g.*, the presence, absence, or amount of a given marker or markers) into data of predictive value for a clinician. The clinician can access the predictive data using any suitable means. Thus, in some preferred embodiments, the present invention provides the further benefit that the clinician, who is not likely to be trained in genetics or molecular biology, need not understand the raw data. The data is presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information in order to optimize the care of the subject.

Any method may be used that is capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information provides, medical personal, and subjects. For example, in some embodiments of the present invention, a sample (*e.g.*, a biopsy or a serum or urine sample) is obtained from a subject and submitted to a profiling service (*e.g.*, clinical lab at a medical facility, genomic profiling business, etc.), located in any part of the world (*e.g.*, in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center, or subjects may collect the sample themselves (*e.g.*, a urine sample) and directly send it to a profiling center. Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (*e.g.*, an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center using an electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced (*i.e.*, expression data), specific for the diagnostic or prognostic information desired for the subject.

The profile data is then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw expression data, the prepared format may represent a diagnosis or risk assessment (*e.g.*, likelihood of cancer being present) for the subject, along with recommendations for particular treatment options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can

be printed for the clinician (*e.g.*, at the point of care) or displayed to the clinician on a computer monitor.

In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data is then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data is stored in a central facility with uniform security protocols), speed, and uniformity of data analysis. The central processing facility can then control the fate of the data following treatment of the subject. For example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers.

In some embodiments, the subject is able to directly access the data using the electronic communication system. The subject may chose further intervention or counseling based on the results. In some embodiments, the data is used for research use. For example, the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition or stage of disease.

D. *In vivo* Imaging

The gene rearrangements disclosed herein may also be detected using *in vivo* imaging techniques, including but not limited to: radionuclide imaging; positron emission tomography (PET); computerized axial tomography, X-ray or magnetic resonance imaging method, fluorescence detection, and chemiluminescent detection. In some embodiments, *in vivo* imaging techniques are used to visualize the presence of or expression of cancer markers in an animal (*e.g.*, a human or non-human mammal). For example, in some embodiments, cancer marker mRNA or protein is labeled using a labeled antibody specific for the cancer marker. A specifically bound and labeled antibody can be detected in an individual using an *in vivo* imaging method, including, but not limited to, radionuclide imaging, positron emission tomography, computerized axial tomography, X-ray or magnetic resonance imaging method, fluorescence detection, and chemiluminescent detection. Methods for generating antibodies to the disclosed cancer markers are described above.

The *in vivo* imaging methods that use the compositions disclosed herein that detect MIPOL1-ETV1 rearrangements or products derived from them are useful in the

diagnosis of cancers, particularly prostate cancer, that express the cancer markers disclosed herein. *In vivo* imaging visualizes the presence of a marker indicative of the cancer, allowing diagnosis and/or prognosis without the use of an unpleasant biopsy. For example, the presence of a marker indicative of cancers likely to metastasize can be detected. The *in vivo* imaging methods can further be used to detect metastatic cancers in other parts of the body.

In some embodiments, reagents (*e.g.*, antibodies) specific for the cancer markers associated with MIPOL1-ETV1 genetic rearrangements are fluorescently labeled. The labeled antibodies are introduced into a subject (*e.g.*, orally or parenterally). Fluorescently labeled antibodies are detected using any suitable method or system (*e.g.*, *see* U.S. Pat. No. 6,198,107, herein incorporated by reference).

In other embodiments, antibodies are radioactively labeled. The use of antibodies for *in vivo* diagnosis is well known in the art, e.g., by using an antibody-based labeling system to image tumors (*see* Sumerdon *et al.*, Nucl. Med. Biol 17:247-254 [1990], Griffin *et al.*, J. Clin. Onc. 9:631-640 [1991], and Lauffer, Magnetic Resonance in Medicine 22:339-342 [1991]). The label used with an antibody-based system will depend on the imaging modality chosen, for example, radioactive labels such as Indium-111, Technetium-99m, or Iodine-131 for use with planar scans or single photon emission computed tomography (SPECT), positron emitting labels such as Fluorine-19 for use with positron emission tomography (PET), and paramagnetic ions such as Gadolinium (III) or Manganese (II) for use with MRI.

Radioactive metals with half-lives ranging from 1 hour to 3.5 days are available for conjugation to antibodies, such as scandium-47 (3.5 days) gallium-67 (2.8 days), gallium-68 (68 minutes), technetiium-99m (6 hours), and indium-111 (3.2 days), of which gallium-67, technetium-99m, and indium-111 are preferable for gamma camera imaging, gallium-68 is preferable for positron emission tomography.

A useful method of labeling antibodies with such radiometals is by means of a bifunctional chelating agent, such as diethylenetriaminepentaacetic acid (DTPA), as described, for example, by Khaw *et al.* (Science 209:295 [1980]) for In-111 and Tc-99m, and by Scheinberg *et al.* (Science 215:1511 [1982]). Other chelating agents may also be used, but the 1-(p-carboxymethoxybenzyl)EDTA and the carboxycarbonic anhydride of DTPA are advantageous because their use permits conjugation without affecting the antibody's immunoreactivity substantially.

Another method for coupling DPTA to proteins is by use of the cyclic anhydride of DTPA, as described by Hnatowich *et al.* (Int. J. Appl. Radiat. Isot. 33:327 [1982]) for labeling of albumin with In-111, but which can be adapted for labeling of antibodies. A suitable method of labeling antibodies with Tc-99m is known (e.g., *see* Crockford *et al.*, U.S. Pat. No. 4,323,546, herein incorporated by reference).

A preferred method of labeling immunoglobulins with Tc-99m is that described by Wong *et al.* (Int. J. Appl. Radiat. Isot., 29:251 [1978]) for plasma protein, and recently applied successfully by Wong *et al.* (J. Nucl. Med., 23:229 [1981]) for labeling antibodies.

In the case of the radiometals conjugated to the specific antibody, it is likewise desirable to introduce as high a proportion of the radiolabel as possible into the antibody molecule without destroying its immunospecificity. A further improvement may be achieved by effecting radiolabeling in the presence of the specific cancer marker associated with a MIPOL1-ETV1 genetic rearrangement, to insure that the antigen binding site on the antibody is protected. The antigen is separated after labeling.

In still further embodiments, *in vivo* biophotonic imaging (Xenogen, Almeda, CA) is used for *in vivo* imaging. This real-time *in vivo* imaging utilizes luciferase, an enzyme that catalyzes a light-emitting reaction. The luciferase gene is incorporated into cells, microorganisms, and animals (*e.g.*, to produce a fusion protein with a cancer marker associated with a MIPOL1-ETV1 genetic rearrangement), so that when the cancer marker is active a light emission occurs which is captured as an image and analyzed by using a CCD camera and appropriate software.

E. Compositions & Kits

Compositions for use in the disclosed diagnostic methods include, but are not limited to, probes, amplification oligonucleotides, and antibodies. The compositions detect a product only when a MIPOL1-ETV1 rearrangement is present, preferably ETV1 genetic material inserted into MIPOL1 genetic material. These compositions include, but are not limited to: a single labeled probe comprising a sequence that hybridizes to the junction at which ETV1 is inserted into the MIPOL1 locus (*i.e.*, spans the gene rearrangement junction); a pair of amplification oligonucleotides wherein the first amplification oligonucleotide comprises a sequence that hybridizes

to MIPOL1 and a second amplification oligonucleotide comprising a sequence that hybridizes to ETV1.

Other useful compositions, however, include: a pair of labeled probes wherein the first labeled probe comprises a sequence that hybridizes to MIPOL1 and the second labeled probe comprises a sequence that hybridizes to ETV1.

Any of these compositions, alone or in combination with other compositions disclosed herein or well known in the art, may be provided in the form of a kit. For example, the single labeled probe and pair of amplification oligonucleotides may be provided in a kit for the amplification and detection of MIPOL1-ETV1 genetic rearrangements. Kits may further comprise appropriate controls and/or detection reagents. Any one or more reagents that find use in any of the methods described herein may be provided in the kit.

The probe and antibody compositions may also be provided in the form of an array.

IV. Drug Screening Applications

In some embodiments, the disclosed compositions and methods are used in drug screening assays (*e.g.*, to screen for anticancer drugs). These screening methodsuse cancer markers that include those associated with MIPOL1-ETV1 gene rearrangements, but are not limited only to those genetic rearrangements. For example, an embodiment may screen for compounds that alter (*e.g.*, decrease) the expression of cancer marker genes, including those associated with MIPOL1-ETV1 gene rearrangements. Compounds or agents to be screened for may interfere with transcription (e.g., by interacting with a promoter region), may interfere with mRNA produced from the rearrangement (*e.g.*, by RNA interference, antisense technologies, etc.), or may interfere with pathways that are upstream or downstream of the biological activity of the gene rearrangement. In some embodiments, candidate compounds are antisense or interfering RNA agents (*e.g.*, oligonucleotides) directed against cancer markers. In other embodiments, candidate compounds are antibodies or small molecules that specifically bind to a cancer marker regulator or expression product associated with MIPOL1-ETV1 gene rearrangements and inhibit its biological function.

In some embodiments, candidate compounds are evaluated for their ability to alter cancer marker expression by contacting a compound with a cell expressing a

cancer marker and then assaying for the effect of the candidate compounds on expression. In some embodiments, the effect of candidate compounds on expression of a cancer marker gene is assayed for by detecting the level of cancer marker mRNA expressed by the cell. mRNA expression can be detected by any suitable method. In other embodiments, the effect of candidate compounds on expression of cancer marker genes is assayed by measuring the level of polypeptide encoded by the cancer markers. The level of polypeptide expressed can be measured using any suitable method, including but not limited to, those disclosed herein.

The test compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone, which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, *e.g.*, Zuckennann *et al.,* J. Med. Chem. 37: 2678-85 [1994]); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are preferred for use with peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) Anticancer Drug Des. 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.,* Proc. Natl. Acad. Sci. U.S.A. 90:6909 [1993]; Erb *et al.,* Proc. Nad. Acad. Sci. USA 91:11422 [1994]; Zuckermann *et al.,* J. Med. Chem. 37:2678 [1994]; Cho *et al.,* Science 261:1303 [1993]; Carrell *et al.,* Angew. Chem. Int. Ed. Engl. 33.2059 [1994]; Carell *et al.,* Angew. Chem. Int. Ed. Engl. 33:2061 [1994]; and Gallop *et al.,* J. Med. Chem. 37:1233 [1994].

Libraries of compounds may be presented in solution (*e.g.*, Houghten, Biotechniques 13:412-421 [1992]), or on beads (Lam, Nature 354:82-84 [1991]), chips (Fodor, Nature 364:555-556 [1993]), bacteria or spores (U.S. Pat. No. 5,223,409; herein incorporated by reference), plasmids (Cull *et al.,* Proc. Nad. Acad. Sci. USA 89:18651869 [1992]) or on phage (Scott and Smith, Science 249:386-390 [1990]; Devlin Science 249:404-406 [1990]; Cwirla *et al.,* Proc. NatI. Acad. Sci. 87:6378-6382 [1990]; Felici, J. Mol. Biol. 222:301 [1991]).

VI. Therapeutic Applications

Some embodiments provide therapies for cancer (*e.g.*, prostate cancer). Preferred therapy embodiments target directly or indirectly cancer markers, including those associated with MIPOL1-ETV1 genetic rearrangements.

A. RNA Interference and Antisense Therapies

Some embodimentstarget the expression of cancer markers associated with MIPOL1-ETV1 genetic rearrangements. Some embodiments employ compositions comprising oligomeric antisense or RNAi compounds, particularly oligonucleotides (*e.g.*, those identified in the drug screening methods described above), for use in modulating the function of nucleic acid molecules encoding cancer markers associated with MIPOL1-ETV1 genetic rearrangements, ultimately modulating the amount of cancer marker expressed.

1. RNA Interference (RNAi)

In some embodiments, RNAi is used to inhibit expression of MIPOL1-ETV1 gene rearrangements. RNAi represents an evolutionary conserved cellular defense for controlling the expression of foreign genes in most eukaryotes, including humans. RNAi is typically triggered by double-stranded RNA (dsRNA) and causes sequence-specific mRNA degradation of single-stranded target RNAs homologous in response to dsRNA. The mediators of mRNA degradation are small interfering RNA duplexes (siRNAs), which are normally produced from long dsRNA by enzymatic cleavage in the cell. siRNAs are generally approximately twenty-one nucleotides in length (*e.g.* 21-23 nucleotides in length), and have a base-paired structure characterized by two nucleotide 3'-overhangs. Following the introduction of a small RNA, or RNAi, into the cell, it is believed the sequence is delivered to an enzyme complex called RISC (RNA-induced silencing complex). RISC recognizes the target and cleaves it with an endonuclease. It is noted that if larger RNA sequences are delivered to a cell, RNase III enzyme (Dicer) converts longer dsRNA into 21-23 nt ds siRNA fragments. In some embodiments, RNAi oligonucleotides are designed to target the junction region of gene rearrangements.

Chemically synthesized siRNAs have become powerful reagents for genome-wide analysis of mammalian gene function in cultured somatic cells. Beyond their value for validation of gene function, siRNAs also hold great potential as gene-specific

therapeutic agents (Tuschl and Borkhardt, Molecular Intervent. 2002; 2(3):158-67, herein incorporated by reference).

The transfection of siRNAs into animal cells results in the potent, long-lasting post-transcriptional silencing of specific genes (Caplen et al, Proc Natl Acad Sci U.S.A. 2001; 98: 9742–7; Elbashir et al., Nature. 2001; 411:494–8; Elbashir et al., Genes Dev. 2001;15: 188–200; and Elbashir et al., EMBO J. 2001; 20: 6877–88, all of which are herein incorporated by reference). Methods and compositions for performing RNAi with siRNAs are known (e.g., *see* U.S. Pat. 6,506,559, herein incorporated by reference).

siRNAs are extraordinarily effective at lowering the amounts of targeted RNA, and by extension proteins, frequently to undetectable levels. The silencing effect can last several months, and is extraordinarily specific, because one nucleotide mismatch between the target RNA and the central region of the siRNA is frequently sufficient to prevent silencing (Brummelkamp et al, Science 2002; 296:550–3; and Holen et al, Nucleic Acids Res. 2002; 30:1757–66, both of which are herein incorporated by reference).

An important factor in the design of siRNAs is the presence of accessible sites for siRNA binding. Bahoia et al., (J. Biol. Chem., 2003; 278: 15991-15997; herein incorporated by reference) describe the use of a type of DNA array called a scanning array to find accessible sites in mRNAs for designing effective siRNAs. These arrays comprise oligonucleotides ranging in size from monomers to a certain maximum, usually Comers, synthesized using a physical barrier (mask) by stepwise addition of each base in the sequence. Thus the arrays represent a full oligonucleotide complement of a region of the target gene. Hybridization of the target mRNA to these arrays provides an exhaustive accessibility profile of this region of the target mRNA. Such data are useful in the design of antisense oligonucleotides (ranging from 7mers to 25mers), where it is important to achieve a compromise between oligonucleotide length and binding affinity, to retain efficacy and target specificity (Sohail et al, Nucleic Acids Res., 2001; 29(10): 2041- 2045). Additional methods and concerns for selecting siRNAs are described for example, in WO 05054270, WO05038054A1, WO03070966A2, J Mol Biol. 2005 May 13;348(4):883-93, J Mol Biol. 2005 May 13;348(4):871-81, and Nucleic Acids Res. 2003 Aug 1;31(15):4417-24, each of which is herein incorporated by reference in its entirety. In addition, software (*e.g.*, the

MWG online siMAX siRNA design tool) is commercially or publicly available for use in the selection of siRNAs.

2. Antisense

In other embodiments, expression of MIPOL1-ETV1 gene rearrangements is modulated using antisense compounds that specifically hybridize with one or more nucleic acids encoding cancer markers associated with MIPOL1-ETV1 genetic rearrangements. The specific hybridization of an oligomeric compound with its target nucleic acid interferes with the normal function of the nucleic acid. This modulation of function of a target nucleic acid by compounds that specifically hybridize to it is generally referred to as "antisense." The functions of DNA to be interfered with include replication and transcription. The functions of RNA to be interfered with include all vital functions such as, for example, translocation of the RNA to the site of protein translation, translation of protein from the RNA, splicing of the RNA to yield one or more mRNA species, and catalytic activity that may be engaged in or facilitated by the RNA. The overall effect of such interference with target nucleic acid function is modulation of the expression of cancer markers associated with MIPOL1-ETV1 genetic rearrangements. Herein, "modulation" means either an increase (stimulation) or a decrease (inhibition) in the expression of a gene. For example, expression may be inhibited to potentially prevent tumor proliferation.

Antisense methods preferably target specific nucleic acids. "Targeting" an antisense compound to a particular nucleic acid usually refers to a multistep process that begins with identification of a nucleic acid sequence whose function is to be modulated. This may be, e.g., a cellular gene (or mRNA transcribed from the gene) whose expression is associated with a particular disorder or disease state, or a nucleic acid molecule from an infectious agent. Herein, the target is a nucleic acid molecule encoding a cancer marker associated with MIPOL1-ETV1 genetic rearrangements. The targeting process also includes determination of a site or sites within this gene for the antisense interaction to occur such that the desired effect, *e.g.*, detection or modulation of expression of the protein, will result. Herein, a preferred intragenic site is the region encompassing the translation initiation or termination codon of the open reading frame (ORF) of the gene. Since the translation initiation codon is typically 5'-AUG (in transcribed mRNA molecules; 5'-ATG in the corresponding DNA molecule), the translation initiation codon is also referred to as the "AUG codon," the

"start codon" or the "AUG start codon". A minority of genes have a translation initiation codon having the RNA sequence 5'-GUG, 5'-UUG or 5'-CUG, and 5'-AUA, 5'-ACG and 5'-CUG have been shown to function *in vivo*. Thus, the terms "translation initiation codon" and "start codon" can encompass many codon sequences, even though the initiator amino acid in each instance is typically methionine (in eukaryotes) or formylmethionine (in prokaryotes). Eukaryotic and prokaryotic genes may have two or more alternative start codons, any one of which may be preferentially used for translation initiation in a particular cell type or tissue, or under a particular set of conditions. Herein, "start codon" and "translation initiation codon" refer to the codon or codons that are used *in vivo* to initiate translation of an mRNA molecule transcribed from a gene encoding a tumor antigen associated with MIPOL1-ETV1 genetic rearrangements, regardless of the sequence(s) of such codons.

Translation termination codon (or "stop codon") of a gene may have one of three sequences (*i.e.*, 5'-UAA, 5'-UAG and 5'-UGA; the corresponding DNA sequences are 5'-TAA, 5'-TAG and 5'-TGA, respectively). The terms "start codon region" and "translation initiation codon region" refer to a portion of such an mRNA or gene that encompasses from about 25 to about 50 contiguous nucleotides in either direction (*i.e.*, 5' or 3') from a translation initiation codon. Similarly, the terms "stop codon region" and "translation termination codon region" refer to a portion of such an mRNA or gene that encompasses from about 25 to about 50 contiguous nucleotides in either direction (*i.e.*, 5' or 3') from a translation termination codon.

The open reading frame (ORF) or "coding region," which refers to the region between the translation initiation codon and the translation termination codon, is also a region that may be targeted effectively. Other target regions include the 5' untranslated region (5' UTR), referring to the portion of an mRNA in the 5' direction from the translation initiation codon, and thus including nucleotides between the 5' cap site and the translation initiation codon of an mRNA or corresponding nucleotides on the gene, and the 3' untranslated region (3' UTR), referring to the portion of an mRNA in the 3' direction from the translation termination codon, and thus including nucleotides between the translation termination codon and 3' end of an mRNA or corresponding nucleotides on the gene. The 5' cap of an mRNA comprises an N7-methylated guanosine residue joined to the 5'-most residue of the mRNA via a 5'-5' triphosphate linkage. The 5' cap region of an mRNA is considered to include the

5' cap structure itself as well as the first 50 nucleotides adjacent to the cap. The cap region may also be a preferred target region.

Although some eukaryotic mRNA transcripts are directly translated, many contain one or more regions, known as "introns," that are excised from a transcript before it is translated. The remaining (and therefore translated) regions are known as "exons" and are spliced together to form a continuous mRNA sequence. mRNA splice sites (*i.e.*, intron-exon junctions) may also be preferred target regions, and are particularly useful *in situ*ations where aberrant splicing is implicated in disease, or where an overproduction of a particular mRNA splice product is implicated in disease. Aberrant fusion junctions due to rearrangements or deletions are also preferred targets. It has also been found that introns can also be effective, and therefore preferred, target regions for antisense compounds targeted, for example, to DNA or pre-mRNA.

In some embodiments, target sites for antisense inhibition are identified using commercially available software programs (*e.g.*, Biognostik, Gottingen, Germany; SysArris Software, Bangalore, India; Antisense Research Group, University of Liverpool, Liverpool, England; GeneTrove, Carlsbad, CA). In other embodiments, target sites for antisense inhibition are identified using the accessible site method described in PCT Publ. No. WO0198537A2, herein incorporated by reference.

Once one or more target sites have been identified, oligonucleotides are chosen that are sufficiently complementary to the target (*i.e.*, hybridize sufficiently well and with sufficient specificity) to give the desired effect. For example, in preferred embodiments, antisense oligonucleotides are targeted to or near the start codon associated with MIPOL1-ETV1 genetic rearrangements.

In the context of this invention, "hybridization," with respect to antisense compositions and methods, means hydrogen bonding, which may be Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary nucleoside or nucleotide bases. For example, adenine and thymine are complementary nucleobases that pair through the formation of hydrogen bonds. It is understood that the sequence of an antisense compound need not be 100% complementary to that of its target nucleic acid to be specifically hybridizable. An antisense compound is specifically hybridizable when binding of the compound to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA to cause a loss of utility, and there is a sufficient degree of complementarity

to avoid non-specific binding of the antisense compound to non-target sequences under conditions in which specific binding is desired (*i.e.*, under physiological conditions in the case of *in vivo* assays or therapeutic treatment, and in the case of *in vitro* assays, under conditions in which the assays are performed).

Antisense compounds are commonly used as research reagents and diagnostics. For example, antisense oligonucleotides, which are able to inhibit gene expression with specificity, can be used to elucidate the function of particular genes. Antisense compounds are also used, for example, to distinguish between functions of various members of a biological pathway.

The specificity and sensitivity of antisense is also applied for therapeutic uses. For example, antisense oligonucleotides have been employed as therapeutic moieties in the treatment of disease states in animals and man. Antisense oligonucleotides have been safely and effectively administered to humans and numerous clinical trials are presently underway. It is thus established that oligonucleotides are useful therapeutic modalities that can be configured to be useful in treatment regimes for treatment of cells, tissues, and animals, especially humans.

While antisense oligonucleotides are preferred, other oligomeric antisense compounds, including but not limited to oligonucleotide mimetics may be used, such as are described below. Preferred antisense compounds comprise from about 8 to about 30 nucleobases (*i.e.*, from about 8 to about 30 linked bases), although both longer and shorter sequences may be used. Particularly preferred antisense compounds are antisense oligonucleotides, even more preferably those comprising from about 12 to about 25 nucleobases.

Specific examples of preferred antisense compounds include oligonucleotides containing modified backbones or non-natural internucleoside linkages. As defined herein, oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone. For the purposes of this specification, modified oligonucleotides that do not have a phosphorus atom in their internucleoside backbone can also be considered to be oligonucleosides.

Preferred modified oligonucleotide backbones include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates

including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms are also included.

Preferred modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH2 component parts.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage (*i.e.*, the backbone) of the nucleotide units are replaced with novel groups. The base units are maintained for hybridization with an appropriate nucleic acid target compound. One such oligomeric compound, an oligonucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA compounds, the sugar-backbone of an oligonucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleobases are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. Methods for preparation of PNA compounds are well known (e.g., *see* U.S. Pat. Nos.: 5,539,082; 5,714,331; and 5,719,262, and Nielsen *et al.*, Science 254:1497 (1991), each of which is herein incorporated by reference). Most preferred embodiments of the invention are oligonucleotides with phosphorothioate backbones and oligonucleosides with heteroatom backbones, and in particular --CH2, --NH--O--CH2--, --CH2--N(CH3)--O--CH2-- [known as a methylene (methylimino) or MMI backbone], --CH2--O--N(CH3)--CH2--, --CH2--N(CH3)--N(CH3)--CH2--, and --O--N(CH3)--CH2--CH2-- [wherein the native phosphodiester backbone is

represented as --O--P--O--CH2--], amid backbone, and morpholino backbone structures, all of which are well known (e.g., *see* U.S. Pat. Nos. 5,489,677, 5,602,240, and 5,034,506).

Modified oligonucleotides may also contain one or more substituted sugar moieties. Preferred oligonucleotides comprise one of the following at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C1 to C10 alkyl or C2 to C10 alkenyl and alkynyl. Particularly preferred are O[(CH2)nO]mCH3, O(CH2)nOCH3, O(CH2)nNH2, O(CH2)nCH3, O(CH2)nONH2, and O(CH2)nON[(CH2)nCH3)]2, where n and m are from 1 to about 10. Other preferred oligonucleotides comprise one of the following at the 2' position: C1 to C10 lower alkyl, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH3, OCN, Cl, Br, CN, CF3, OCF3, SOCH3, SO2CH3, ONO2, NO2, N3, NH2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. A preferred modification includes 2'-methoxyethoxy (2'-O--CH2CH2OCH3, also known as 2'-O-(2-methoxyethyl) or 2'-MOE) (Martin *et al.*, Helv. Chim. Acta 78:486 [1995]) *i.e.*, an alkoxyalkoxy group. A further preferred modification includes 2'-dimethylaminooxyethoxy (*i.e.*, a O(CH2)2ON(CH3)2 group), also known as 2'-DMAOE, and 2'-dimethylaminoethoxyethoxy (also known in the art as 2'-O-dimethylaminoethoxyethyl or 2'-DMAEOE), *i.e.*, 2'-O--CH2--O--CH2--N(CH2)2.

Other preferred modifications include 2'-methoxy(2'-O--CH3), 2'-aminopropoxy(2'-OCH2CH2CH2NH2) and 2'-fluoro (2'-F). Similar modifications may also be made at other positions on the oligonucleotide, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Oligonucleotides may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar.

Oligonucleotides may also include nucleobase (often referred to in the art simply as "base") modifications or substitutions. As used herein, "unmodified" or "natural" nucleobases include the purine bases adenine (A) and guanine (G), and the

pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified nucleobases are well known (e.g., see U.S. Pat. No. 3,687,808) and include other synthetic and natural nucleobases (for which the A, G, T, C and U abbreviations for the bases are used in the following examples), such as 5-methylcytosine (5-me-C), 5-hydroxymethyl C, xanthine, hypoxanthine, 2-amino-A, 6-methyl or 2-propyl and other alkyl derivatives of A and G, 2-thio-U, 2-thio-T and 2-thio-C, 5-halo-U and -C, 5-propynyl U and C, 6-azo U, C and T, 5-uracil (pseudouracil), 4-thio-U, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted A and G, 5-halo substituted U and C, particularly 5-bromo, 5-trifluoromethyl and other 5-substituted U and C, 7-methyl-G and 7-methyl-A, 8-aza-G and 8-aza-A, 7-deaza-G and 7-deaza-A and 3-deaza-G and 3-deaza-A. Certain of these nucleobases are particularly useful for increasing the binding affinity of the oligomeric compounds of the invention. These include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyl-U and 5-propynyl-C. 5-methyl-C substitutions are known to increase nucleic acid duplex stability and are preferred base substitutions in some embodiments, even more particularly when combined with 2'-O-methoxyethyl sugar modifications.

Another modification of the oligonucleotides involves chemically linking to the oligonucleotide one or more moieties or conjugates that enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include but are not limited to lipid moieties such as a cholesterol moiety, cholic acid, a thioether, (*e.g.*, hexyl-S-tritylthiol), a thiocholesterol, an aliphatic chain, (*e.g.*, dodecandiol or undecyl residues), a phospholipid, (*e.g.*, di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate), a polyamine or a polyethylene glycol chain or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety.

One skilled in the relevant art knows well how to generate oligonucleotides containing the above-described modifications. The present invention is not limited to the antisense oligonucleotides described above. Any suitable modification or substitution may be used.

It is not necessary for all positions in a given compound to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single compound or even at a single nucleoside within an oligonucleotide. Antisense compounds may be chimeric compounds. "Chimeric"

antisense compounds or "chimeras," as used herein, are antisense compounds, particularly oligonucleotides, which contain two or more chemically distinct regions, each made up of at least one monomer unit, *i.e.*, a nucleotide in the case of an oligonucleotide compound. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the oligonucleotide increased resistance to nuclease degradation, increased cellular uptake, and/or increased binding affinity for the target nucleic acid. An additional region of the oligonucleotide may serve as a substrate for enzymes capable of cleaving RNA:DNA or RNA:RNA hybrids. By way of example, RNaseH is a cellular endonuclease that cleaves the RNA strand of an RNA:DNA duplex. Activation of RNase H, therefore, results in cleavage of the RNA target, thereby greatly enhancing the efficiency of oligonucleotide inhibition of gene expression. Consequently, comparable results can often be obtained with shorter oligonucleotides when chimeric oligonucleotides are used, compared to phosphorothioate deoxyoligonucleotides hybridizing to the same target region. Cleavage of the RNA target can be routinely detected by gel electrophoresis and, if necessary, associated nucleic acid hybridization techniques known in the art.

Chimeric antisense compounds may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, oligonucleosides and/or oligonucleotide mimetics as described above.

Other embodiments include pharmaceutical compositions and formulations that include the antisense compounds as described herein.

B. Gene Therapy

Embodiments may use any genetic manipulation to modulate the expression of cancer markers associated with MIPOL1-ETV1 genetic rearrangements described herein. Examples of genetic manipulation include, but are not limited to, gene knockout (such as by removing the genetic rearrangement from the chromosome using, e.g., by recombination), expression of antisense constructs with or without inducible promoters, and the like. Delivery of nucleic acid construct to cells *in vitro* or *in vivo* may be conducted using any suitable method. A suitable method is one that introduces the nucleic acid construct into the cell such that the desired event occurs (*e.g.*, expression of an antisense construct). Genetic therapy may also be used to

deliver siRNA or other interfering molecules that are expressed *in vivo* (*e.g.*, upon stimulation by an inducible promoter (*e.g.*, an androgen-responsive promoter)).

Introduction of molecules carrying genetic information into cells is achieved by any of various methods including, but not limited to, directed injection of naked DNA constructs, bombardment with gold particles loaded with said constructs, and macromolecule mediated gene transfer using, for example, liposomes, biopolymers, and the like. Preferred methods use gene delivery vehicles derived from viruses, including, but not limited to, adenoviruses, retroviruses, vaccinia viruses, and adeno-associated viruses. Because of the higher efficiency as compared to retroviruses, vectors derived from adenoviruses are the preferred gene delivery vehicles for transferring nucleic acid molecules into host cells *in vivo*. Adenoviral vectors and their use in gene transfer are well known (e.g., see PCT publications WO 00/12738 and WO 00/09675 and U.S. Pat. Nos. 6,033,908, 6,019,978, 6,001,557, 5,994,132, 5,994,128, 5,994,106, 5,981,225, 5,885,808, 5,872,154, 5,830,730, and 5,824,544, each of which is herein incorporated by reference in its entirety). Such vectors and methods have been shown to provide very efficient *in vivo* gene transfer into a variety of solid tumors in animal models and into human solid tumor xenografts in immune-deficient mice.

Vectors may be administered to subject in a variety of well known ways, e.g., administered into tumors or tissue associated with tumors by using direct injection or administration via the blood or lymphatic circulation (*See e.g.*, PCT publication 99/02685 herein incorporated by reference in its entirety). Exemplary dose levels of adenoviral vector are preferably 10^8 to 10^{11} vector particles added to the perfusate.

C. Antibody or Small Molecule Therapies

Some embodiments are or use antibodies and/or small molecules that target prostate tumors that express a cancer marker associated with MIPOL1-ETV1 genetic rearrangements. In some embodiments, the therapeutic regimen is selected based on a diagnostic result and uses a suitable antibody (*e.g.*, monoclonal, polyclonal, or synthetic) in the therapeutic methods. In preferred embodiments, the antibodies used for cancer therapy are humanized antibodies. Methods for humanizing antibodies are well known (*See e.g.*, U.S. Pat. Nos. 6,180,370, 5,585,089, 6,054,297, and 5,565,332; each of which is herein incorporated by reference).

In some embodiments, the therapeutic antibodies comprise an antibody generated against a cancer marker associated with MIPOL1-ETV1 genetic rearrangements, wherein the antibody is conjugated to a cytotoxic agent. In such embodiments, a tumor specific therapeutic agent is generated that does not target normal cells, thus reducing many of the detrimental side effects of traditional chemotherapy. For certain applications, it is envisioned that the therapeutic agents will be pharmacologic agents that will serve as useful agents for attachment to antibodies, particularly cytotoxic or otherwise anticellular agents having the ability to kill or suppress the growth or cell division of endothelial cells. Embodiments may use any pharmacologic agent that can be conjugated to an antibody, and delivered in active form. Exemplary anticellular agents include chemotherapeutic agents, radioisotopes, and cytotoxins. Such therapeutic antibodies may include a variety of cytotoxic moieties, including but not limited to, radioactive isotopes (*e.g.*, iodine-131, iodine-123, technicium-99m, indium-111, rhenium-188, rhenium-186, gallium-67, copper-67, yttrium-90, iodine-125 or astatine-211), hormones such as a steroid, antimetabolites such as cytosines (*e.g.*, arabinoside, fluorouracil, methotrexate or aminopterin; an anthracycline; mitomycin C), vinca alkaloids (*e.g.*, demecolcine; etoposide; mithramycin), and antitumor alkylating agent such as chlorambucil or melphalan. Other embodiments may include agents such as a coagulant, a cytokine, growth factor, bacterial endotoxin or the lipid A moiety of bacterial endotoxin. For example, in some embodiments, therapeutic agents will include plant-, fungus- or bacteria-derived toxin, such as an A chain toxins, a ribosome inactivating protein, α-sarcin, aspergillin, restrictocin, a ribonuclease, diphtheria toxin or pseudomonas exotoxin, to mention just a few examples. In some preferred embodiments, deglycosylated ricin A chain is used.

In any event, it is proposed that agents such as these may, if desired, be successfully conjugated to an antibody, in a manner that will allow their targeting, internalization, release or presentation to blood components at the site of the targeted tumor cells as required using known conjugation technology (*See, e.g.*, Ghose *et al.*, Methods Enzymol., 93:280 [1983]).

For example, some embodiments provide immunotoxins targeted against cancer marker associated with MIPOL1-ETV1 genetic rearrangements. Immunotoxins are conjugates of a specific targeting agent typically a tumor-directed antibody or fragment, with a cytotoxic agent, such as a toxin moiety. The targeting

agent directs the toxin to, and thereby selectively kills, cells carrying the targeted antigen. In some embodiments, therapeutic antibodies employ crosslinkers that provide high *in vivo* stability (Thorpe *et al.*, Cancer Res., 48:6396 [1988]).

In other embodiments, particularly those involving treatment of solid tumors, antibodies are designed to have a cytotoxic or otherwise anticellular effect against the tumor vasculature, by suppressing the growth or cell division of the vascular endothelial cells. This attack is intended to lead to a tumor-localized vascular collapse, depriving the tumor cells, particularly those tumor cells distal of the vasculature, of oxygen and nutrients, ultimately leading to cell death and tumor necrosis.

In preferred embodiments, antibody based therapeutics are formulated as pharmaceutical compositions as described herein. In preferred embodiments, administration of an antibody composition that targets a moiety associated with MIPOL1-ETV1 genetic rearrangements results in a measurable decrease in cancer (*e.g.*, decrease or elimination of tumor).

VII. Transgenic Animals

Embodiments include generation of transgenic animals comprising an exogenous cancer marker gene that is identical to or representative of a MIPOL1-ETV1 genetic rearrangement described herein, which includes mutants and variants thereof (*e.g.*, truncations or single nucleotide polymorphisms). In preferred embodiments, the transgenic animal displays an altered phenotype (*e.g.*, increased or decreased presence of markers associated with MIPOL1-ETV1 genetic rearrangements) as compared to wild-type animals. Methods for analyzing the presence or absence of such phenotypes include but are not limited to, those disclosed herein. In some preferred embodiments, the transgenic animals further display an increased or decreased growth of tumors or evidence of cancer.

Such transgenic animals are useful in drug (*e.g.*, cancer therapy) screens. In some embodiments, test compounds (*e.g.*, a drug that is suspected of being useful to treat cancer) and control compounds (*e.g.*, a placebo) are administered to the transgenic animals and the control animals and the effects evaluated.

The transgenic animals can be generated via a variety of methods. In some embodiments, embryonal cells at various developmental stages are used to introduce transgenes for the production of transgenic animals. Different methods are used

depending on the stage of development of the embryonal cell. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter that allows reproducible injection of 1-2 picoliters (pl) of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host genome before the first cleavage (Brinster *et al.,* Proc. Natl. Acad. Sci. USA 82:4438-4442 [1985]). As a consequence, all cells of the transgenic non-human animal carry the incorporated transgene. This is reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells harbor the transgene based on standard Mendelian genetics. Methods for making transgenics are well known (e.g., see U.S. Pat. No. 4,873,191, which is herein incorporated by reference in its entirety).

In other embodiments, retroviral infection is used to introduce transgenes into a non-human animal. In some embodiments, the retroviral vector is used to transfect oocytes by injecting the retroviral vector into the perivitelline space of the oocyte (U.S. Pat. No. 6,080,912, incorporated herein by reference). In other embodiments, the developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Janenich, Proc. Natl. Acad. Sci. USA 73:1260 [1976]). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan *et al.*, in *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [1986]). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner *et al.,* Proc. Natl. Acad Sci. USA 82:6927 [1985]). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Stewart, *et al.,* EMBO J., 6:383 [1987]). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner *et al.,* Nature 298:623 [1982]). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of cells that form the transgenic animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome that generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germline, albeit with low efficiency, by intrauterine retroviral infection of the midgestation embryo (Jahner *et al., supra* [1982]). Additional means of using retroviruses or retroviral vectors to create

transgenic animals known to the art involve the micro-injection of retroviral particles or mitomycin C-treated cells producing retrovirus into the perivitelline space of fertilized eggs or early embryos (PCT International Application WO 90/08832 [1990], and Haskell and Bowen, Mol. Reprod. Dev., 40:386 [1995]).

In other embodiments, the transgene is introduced into embryonic stem cells and the transfected stem cells are used to form an embryo. ES cells are obtained by culturing pre-implantation embryos *in vitro* under appropriate conditions (Evans *et al.,* Nature 292:154 [1981]; Bradley *et al.,* Nature 309:255 [1984]; Gossler *et al.*, Proc. Acad. Sci. USA 83:9065 [1986]; and Robertson *et al.*, Nature 322:445 [1986]). Transgenes can be efficiently introduced into the ES cells by DNA transfection by a variety of methods known to the art including calcium phosphate co-precipitation, protoplast or spheroplast fusion, lipofection and DEAE-dextran-mediated transfection. Transgenes may also be introduced into ES cells by retrovirus-mediated transduction or by micro-injection. Such transfected ES cells can thereafter colonize an embryo following their introduction into the blastocoel of a blastocyst-stage embryo and contribute to the germ line of the resulting chimeric animal (for review, *See*, Jaenisch, Science 240:1468 [1988]). Prior to the introduction of transfected ES cells into the blastocoel, the transfected ES cells may be subjected to various selection protocols to enrich for ES cells which have integrated the transgene assuming that the transgene provides a means for such selection. Alternatively, the polymerase chain reaction may be used to screen for ES cells that have integrated the transgene. This technique obviates the need for growth of the transfected ES cells under appropriate selective conditions prior to transfer into the blastocoel.

In still other embodiments, homologous recombination is used to knock-out gene function or create deletion mutants (*e.g.*, truncation mutants), using well known methods (*see* U.S. Pat. No. 5,614,396, incorporated herein by reference).

EXPERIMENTAL

The following examples are provided to demonstrate and illustrate certain preferred embodiments and aspects of the compositions and methods disclosed herein, but are not to be construed as limiting the scope of the claimed invention.

Example 1: ETV1 GENERIC REARRANGEMENTS IN CELL LINES

This example shows that LNCaP and MDA-PCa 2B cell lines harbor rearrangements that localize the ETV1 locus to 14q13.3-14q21.1.

Experiments were designed to identify cell line models of aberrant ETV1 activation. The prostate cancer cell line LNCaP has markedly over-expressed ETV1, but RLM-RACE revealed expression of only the wild type transcript (Tomlins et al., Science 310, 644-8 (2005)). These experiments were designed to identify in LNCaP cells a novel rearrangement affecting the expression of ETV1. Thus, a split-probe FISH strategy, consisting of one probe on and one probe 5' to the ETV1 locus, was used to look for gross rearrangements involving ETV1 (FIG. 1).

Interphase FISH on formalin-fixed paraffin-embedded (FFPE) tissue sections was performed using known methods (substantially as previously described in Tomlins et al., Cancer Res 66, 3396-400 (2006)). A minimum of 50 nuclei per assay were evaluated. Metaphase spreads of LNCaP and MDA-PCa 2B were prepared using standard cytogenetic techniques. Slides were pre-treated in 2x SSC for 2 min, 70% ethanol for 2 min and 100% ethanol for 2 min, and air dried. Slide samples and probes were co-denatured at 75°C for 2 min, and hybridized overnight at 37°C. Post-hybridization was in 0.5x SSC at 42°C for 5 min, followed by 3 washes in PBST. Fluorescent detection was performed using anti-digoxigenin conjugated to fluorescein (Roche Applied Science, Indianapolis, IN) and streptavidin conjugated to a fluorophore (Alexa Fluor 594, Invitrogen). Slides were counterstained and mounted using standard reagents and methods (in ProLong Gold Antifade Reagent with DAPI, Invitrogen). Slides were examined using a fluorescence microscope (Axio Imager Z1, Zeiss, Thornwood, NY) and imaged with a CCD camera using a standard algorithm to analyze results (ISIS software, Metasystems, Altlussheim, Germany). BACs (listed in Table 2, obtained from the BACPAC Resource Center (Oakland, CA)) were used to prepare probes using previously described methods (Tomline et al., 2006, supra). Pre-labeled chromosome 7 centromere and 7p telomeric probes were also used (from Vysis Corp., Des Plaines, IL). The integrity and correct localization of all probes were verified by hybridization to metaphase spreads of normal peripheral lymphocytes.

On LNCaP metaphases, this assay revealed two pairs of co-localizing signals at the ETV1 locus on 7p, and two split signals where the 5' signals remained on 7p, while two copies of the ETV1 locus were inserted into another chromosome (FIG. 1c). This rearrangement in 2 of 4 copies of chromosome 7 is consistent with other

rearrangements observed by G-banding or spectral karyotyping (SKY) in tetraploid LNCap cells (Beheshti et al., Mol Diagn 5, 23-32 (2000); Beheshti et al., Neoplasia 3, 62-9 (2001); Gibas et al., Cancer Genet Cytogenet 11, 399-404 (1984); van Bokhoven et al., Prostate 57, 226-44 (2003)). Cytogenetic analysis indicated that ETV1 was inserted into chromosome 14, and this was confirmed by using a FISH mapped BAC previously localized to 14q (FIG. 2). Subsequent FISH assays were used to determine that the break on chromosome 7 was localized completely 5' to the ETV1 locus, consistent with RACE revealing over-expression of full length ETV1 (FIG. 3).

To localize the breakpoint, Southern blotting was performed using 22 probes across the implicated region of 7p. Genomic DNA (10 μg) was digested with EcoRI or PstI (New England Biologicals, Ipswich, MA) overnight. Fragments were resolved on a 0.8% agarose gel at 40 V overnight, transferred to a nylon membrane (Hybond NX), prehybridized, hybridized with probe and washed according to standard laboratory procedures. A series of 22 probes spanning the region of chr 7 implicated by FISH (between RP11-313C20 and RP11-703A4) were generated by PCR amplification using pooled normal human male genomic DNA as template, a high fidelity polymerase enzyme (Platinum Taq High Fidelity), and primers listed in Table 1. Twenty-five ng of each probe was labeled with dCTP-P32 and used for hybridization.

A single probe (Probe A, see Table 1, "Southern Probe") showed evidence of a rearrangement, with both EcoRI and PstI digested LNCaP DNA showing additional bands (FIG. 3).

PstI digested DNA was then used for inverse PCR to identify the genomic breakpoint sequence on chromosome 7 as well as the partner sequence (FIG. 4). Primers A1, A2, A3, which are reverse complemented from the wildtype sequence and are divergent to primers B1, B2, B3, were used for inverse PCR on PstI digested DNA and religated (intramolecular ligation) using the LNCaP genomic DNA template. Nested PCRs were performed in the following order of primer combinations: A1 and B1, then A2 and B2, and finally A3 and B3. The fusion product was amplified using PCR amplification and standard laboratory procedures (Expand 20 kbplus PCR System, Roche Diagnostics GmbH, Mannheim, Germany, used according to the manufacturor's instructions). The enriched 3Kb band observed in nested PCRs was cloned into a vector (pCR8/GW/TOPO, Invitrogen), and DNA isolated from various clones were screened for inserts and positive clones were

sequenced using standard laboratory procedures (by the University of Michigan DNA Sequencing Core Laboratory, Ann Arbor, MI). Clones of the fusions sequences were confirmed by PCR amplification (Platinum Taq High Fidelity system) using fusion specific primers (genomic fusion f and r; Table 1).

Sequencing of the amplified product confirmed the breakpoint on chromosome 7p located in a region identified by using Southern blotting, and the insertion point was an intronic sequence from the MIPOL1 locus at chromosome 14q13.3-14q21.1 (FIGS. 1b and 4), consistent with the results obtained by FISH analysis described above. This rearrangement was confirmed by PCR amplification of a product made by using undigested LNCaP genomic DNA as template, isolated from two different passages, while no product was amplified from VCaP, normal male, or normal placental genomic DNA (FIG. 4). Consistent with cytogenetic and SKY data revealing no gross rearrangements in chromosomes 7 or 14 in LNCaP cells29-32, FISH analysis using split probes around the 14q13.3-14q21.1 locus showed co-localized signals (FIG. 1c). These results are consistent with the interpretation that a limited insertion occurred around ETV1, or the entire 7p arm telomeric to ETV1 (approximately 14 MB) was deleted. The latter possibility was ruled out because high density aCGH did not reveal the presence of any deletions telomeric to ETV1, and FISH demonstrated intact 7p telomeric sequence on all copies of chromosome 7 (FIG. 5). Therefore, these results confirm a cryptic insertion of a minimal region around ETV1 into 14q13.3-14q21.1 in LNCaP.

Additional prostate cancer cell lines were tested to screen for ERG and ETV1 expression by quantitative PCR (qPCR), to identify additional genetic rearrangements. The following cells were tested: an immortalized benign prostate epithelial cell line (RWPE) and 8 prostate cancer cell lines (MDA-PCa 2B, LNCaP, VCaP, LAPC4, 22Rv1, NCI-H660, PC3 and DU145). Quantitative PCR (QPCR) was performed by using the oligonucleotide primers shown in Table 1 ("Fusion QPCR" and "Androgen QPCR" sections) with standard laboratory procedures and reagents (Power SYBR Green Mastermix and 7300 Real Time PCR system, Applied Biosystems, Foster City, CA).. All oligonucleotide primers were synthesized by standard methods (performed by Integrated DNA Technologies, Coralville, IA). HMBS and GAPDH5, and PSA6 primers were as described previously (Vandesompele et al., Genome Biol. 3:RESEARCH0034 (2002); Specht et al., Am J.

Pathol. 158: 419 (2001)). Androgen stimulation reactions were performed in quadruplicate, and all other reactions were performed in duplicate.

Marked over-expression of ETV1 was detected in LNCaP3 cells, and over-expression of ERG was detected in VCaP and NCI- H660 cells, which are both TMPRSS2:ERG positive (Tomlins et al., Science 310, 644-8 (2005)). Results showed that MDA-PCa 2B cells expressed higher levels of ETV1 than LNCaP cells (FIG. 7).

MDA-PCa 2B cells were also analyzed for genetic rearrangement involving the ETV1 locus. Analysis by RLM-RACE revealed only full-length ETV1 made in MDA-PCa 2B cells, indicating that rearrangement involves the entire ETV1 locus. A previous SKY and G-banding analysis of DNA from MDA-PCa 2B cells demonstrated the presence of a balanced t(7;14)(p21;q21)32, the locations of the ETV1 and MIPOL1 loci. The same FISH analysis was performed for MDA-PCa 2B cells as described above for LNCaP cells, using split probes around the ETV1 locus and the 14q13.3-14q21.1 regions. The results demonstrated that MDA-PCa 2B cells also harbor a rearrangement involving ETV1, in which the whole ETV1 locus is translocated to the d14 chromosome (FIG. 1d). FISH revealed that the 1.5 MB 14q13.3-14q21.1 region is the partner of this balanced translocation, as the telomeric 14q13.3-14q21.1 probe is translocated to the d7 chromosome (FIG. 1d).

Example 2: ABERRANT EXPRESSION OF ETV1 ASSOCIATED WITH GENETIC REARRANGEMENTS

This example shows that region 14q13.3-14q21.1 is coordinately regulated in prostate cancer and LNCaP cells.

The existence of mechanistically distinct rearrangements resulting in the localization of ETV1 to 14q13.3-14q21.1 (FIG. 1, e-g) in prostate cancer cell lines with outlier expression of ETV1 indicates that elements at this region mediate aberrant ETV1 transcription. The structure of the genetic rearrangement in LNCaP cells indicates that ETV1 does not acquire a new proximal promoter after insertion into the 14q13.3-14q21.1 locus because approximately 200kb of the 5' upstream sequence is inserted along with the ETV1 locus (see FIG. 1a). Furthermore, while expression of MIPOL1 in the prostate has not been described, FOXA1, immediately adjacent to MIPOL1, is strongly expressed throughout human and murine prostate development and acts as a transcriptional potentiator for androgen-regulated gene expression (Mirosevich et al., Prostate 66, 1013-28 (2006); Mirosevich et al., Prostate

62, 339-52 (2005)). Over-expression of ETV1 in these lines may be driven by enhancer elements that confer coordinated prostate specificity or androgen regulation to this region. To test this hypothesis, the expression of four contiguous transcripts in the 1.5 MB 14q13.3-14q21.1 region (SLC25A21, MIPOL1, FOXA1 and TTC6) was assessed using data in the expO dataset (International Genomics Consortium (IGC) expression project for Oncology (expO)). The expO dataset is a multi-tumor gene expression dataset generated by a consortium of labs including TGEN and has been included in Oncomine (Rhodes et al., Neoplasia 6, 1-6 (2004)). All four transcripts showed significant over-expression in prostate cancer compared to all other cancers (FIG. 6a). Furthermore, when all measured genes on chromosome 14 were ranked by significant over-expression in prostate cancer compared to all other cancers in this data set, MIPOL1 ranked second, FOXA1 ranked fifth, and TTC6 ranked twenty-third.

Although all four transcripts were over-expressed in prostate cancer, stimulation of LNCaP with R1881 minimally increased the expression of FOXA1 (1.3 fold, $p = 0.05$), while SLC25A21 (0.75 fold, $p = 0.008$), MIPOL1 (1.2 fold, $p = 0.08$) and TTC6 (0.70 fold, $p = 0.10$) showed no significant change or reduced expression following R1881 stimulation (FIG. 6b). However, stimulation of LNCaP with R1881 resulted in significantly increased expression of ETV1 (1.8 fold, $p = 0.0004$, FIG. 6c), indicating that rearrangement to this region confers androgen-responsiveness to ETV1, in addition to aberrant expression. Together, these results show coordinated over-expression in prostate cancer for the genes in this region (and ETV1 when inserted therein), with additional regulation mediated by androgen.

To model progression to hormone refractory metastatic disease, LNCaP cells have previously been cultured in the absence of androgen and clones were selected that are insensitive to androgen. These model systems thus are related to expression levels of genetic components and androgen regulated expression of genetic components in the cells, which would include ETV1 and genes at the 14q13.3-14q21.1 region contained in the LNCaP cell line and the derived clones made from this cell line. Our investigation of all of these independently derived cell lines for which publicly available expression profiling data is available, ETV1 shows marked down-regulation in androgen insensitive derivatives of LNCaP36-39 cells. To confirm these findings, gene expression in LNCaP and its androgen insensitive derivative C4-2B40 were profiled by using a hybridization based assay to examine

gene expression (Agilent Whole Genome Microarrays). As shown in FIG. 6d, ETV1 was the fourth most over-expressed feature in LNCaP compared to C4-2B (307-fold difference). When measured by using qPCR, the marked down-regulation of ETV1 in C4-2B (about 22,500 fold), compared to the parental LNCaP cell line was confirmed (FIG. 6e). Furthermore, qPCR also demonstrated that C4-2B cells express less PSA (3-fold less) and the four genes at 14q13.3-14q21.1, when compared to LNCaP cells (FIG. 6, f-g). These results demonstrate coordinated regulation of androgen induced genes and genes at 14q13.3-14q21.1. These results indicate that ETV1 is markedly down-regulated in androgen insensitive LNCaP derivatives either due to deletion or transcriptional mechanisms, or clones that do not harbor the ins(7;14) are selected by using the androgen deprivation method used to derive these clones from the LNCaP cell line. By using FISH and PCR analysis on genomic DNA, the data confirmed that C4-2B cells harbor the same ETV1 rearrangement as LNCaP cells (FIGS. 8 and 9), leading to the conclusion that the down-regulation of ETV1 is due to mutation or transcriptional changes in the derivative clones.

Table 1

Assay	Gene/Region	Sequence (Accession No.)	Bases Within Sequence	Primer Name	Oligonucleotide Sequence (5' to 3')	SEQ ID NO
Androgen / Expression QPCR	ETV1	NM_004956.3	624-645	ETV1_exon 6-7_f	CTACCCCATGGACCACAGATTT	1
Androgen / Expression QPCR	ETV1	NM_004956.3	771-750	ETV1_exon 6-7_r	CTTAAAGCCTTGTGCTGGGAAG	2
Expression QPCR	ERG	NM_004449.3	574-597	ERG_exon 5-6_f	CGCAGAGTTATCGTGCCAGCAGAT	3
Expression QPCR	ERG	NM_004449.3	659-636	ERG_exon 5-6_r	CCATATTCTTTCACCGCCCACTCC	4
RLM-RACE	NA		NA	Generacer 5'_f	CGACTGGAGCACGAGGACACTGA	5
RLM-RACE	ETV1	NM_004449.3	374-351	ETV1_exon 4-5_r	CATGGACTGTGGGGTTCTTTCTTG	6
RLM-RACE	ETV1	NM_004449.3	735-710	ETV1_exon 7-r	AGACATCTGGCGTTGGTACATAGGAC	7
Fusion QPCR	HERV-K_22q11.23	BC020811.1	303-327	HERV-K:ETV1-f	GAGTCCCAAGTACGTCCACGGTCAG	8
Fusion QPCR	ETV1	NM_004956.3	371-345	HERV-K:ETV1-r	GGACTGTGGGGTTCTTTCTTGATTTC	9
Fusion QPCR	HNRPA2B1	NM_002137.2	136-155	HNRPA2B1:ETV1-f	TGCGGGAAATCGGGCTGAAG	10
Fusion QPCR	ETV1	NM_004956.3	181-154	HNRPA2B1:ETV1-r	TTTTCCTGACATTTGTTGGTTTCTCGTT	11
Fusion QPCR	SLC45A3	NM_033102.2	74-92	SLC45A3:ETV1-f	CGCTGGCTCCGGGTGACAG	12
Fusion QPCR	ETV1	NM_004956.3	366-340	SLC45A3:ETV1-r	GTGGGGTTCTTTCTTGATTTTCAGTGG	13
Fusion QPCR	C15ORF21	NM_001005266.1	313-336	C15ORF21:ETV1-f	CAACTAACACTGCGGCTTCCTGAG	14
Fusion QPCR	ETV1	NM_004956.3	483-461	C15ORF21:ETV1-r	CATTCCCACTTGTGGCTTCTGAT	15
Androgen QPCR	TMPRSS2	NM_005656.2	1539-1563	TMPRSS2-f	CAGGAGTGTACGGGAATGTGATGGT	16
Androgen QPCR	TMPRSS2	NM_005656.2	1608-1585	TMPRSS2-r	GATTAGCCGTCTGCCCTCATTTGT	17
Androgen QPCR	TTC6	NM_001007795.1	1080-1108	TTC6-f	TGCCATGAAGATCAGTAC	18

					TACAGCAGAAT	
Androgen QPCR	TTC6	NM_001007795.1	1150-1125	TTC6-r	GTGGCCCATAAACTCATG AATCACC	19
Androgen QPCR	SLC25A21	NM_030631.1	356-377	SLC25A21-f	CAGATCGTGGCCGGTGGT TCT	20
Androgen QPCR	SLC25A21	NM_030631.1	408-483	SLC25A21-r	GGGTGCATCAGGCAAATT TCTACAAG	21
Androgen QPCR	MIPOL1	NM_138731.2	1607-1633	MIPOL1-f	CAACAACAAAATGAGGA ACTGGCTACT	22
Androgen QPCR	MIPOL1	NM_138731.2	1673-1649	MIPOL1-r	ATTCCATATTTGCTCGCTC TGTCAG	23
Androgen QPCR	FOXA1	NM_004496.2	327-350	FOXA1-f	GAAGATGGAAGGGCATG AAACCAG	24
Androgen QPCR	FOXA1	NM_004496.2	408-389	FOXA1-r	GCTGACCGGGACGGAGG AGT	25
Androgen QPCR	SLC45A3	NM_033102.2	1223-1242	SLC45A3-f	TCGTGGGCGAGGGGCTGT A	26
Androgen QPCR	SLC45A3	NM_033102.2	1308-1284	SLC45A3-r	CATCCGAACGCCTTCATC ATAGTGT	27
Androgen QPCR	HERV-K_22q11.23	BC020811.1	168-194	HERV-K_22q11.23-f	CTTTTCTCTAGGGTGAAG GGACTCTCG	28
Androgen QPCR	HERV-K_22q11.23	BC020811.1	263-238	HERV-K_22q11.23-r	CTTCACCCACAAGGCTCA CTGTAGAC	29
Androgen QPCR	HNRPA2B1	NM_002137.2	594-620	HNRPA2B1-f	GCTTTGGCTTTGTTACTTT TGATGACC	30
Androgen QPCR	HNRPA2B1	NM_002137.2	693-665	HNRPA2B1-r	GCCTTTCTTACTTCTGCAT TATGACCATT	31
Androgen QPCR	C15ORF21	NM_001005266.1	219-243	C15ORF21-f	AAGGACGTGCAAGGATG TTTTTATT	32
Androgen QPCR	C15ORF21	NM_001005266.1	293-274	C15ORF21-r	ATGGGAAGATGGGGGCT GTT	33
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,685,335-13,685,364	LNCAP_A-f	GTCAATGGCTAAAAGATG GATAAAAGTGGA	34
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,686,833-13,686,804	LNCAP_A-r	CAGATAGAAGAGGGGTT AGCAAAATGTGTT	35
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,690,754-13,690,779	LNCAP_B1-f	CAGAAGGCAAATGTGAG AGGATAGTC	36
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,691,679-13,691,657	LNCAP_B1-r	CTGGATCTGTAACACCCG TGAGC	37
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,693,600-13,693,625	LNCAP_B2-f	AAAAAGCAAAGACAAGA CCGTGGATT	38
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,695,054-13,695,028	LNCAP_B2-r	GAACTACCTGCGTGCTGA CTTGGAGAT	39
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,699,772-13,699,801	LNCAP_C-f	AAAAGGCAAAGAGGGGT TAAAACATACATA	40
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,700,641-13,700,618	LNCAP_C-r	AACCCCTCCTTCCACTTCT CCACT	41
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,705,005-13,705,034	LNCAP_D-f	TGGAGGCATAGAAAAGC TGAGAAATAAG	42
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,705,856-13,705,830	LNCAP_D-r	TTGGTGCTAGAAGAACTG GCAGAAAC	43
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,711,263-13,711,291	LNCAP_F-f	GAAAGTCAGGGGCACAT ATAGATTAGAG	44
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,712,112-13,712,089	LNCAP_F-r	GCCTTCCCCATACAGTTT CTCCTT	45
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,708,247-13,708,275	LNCAP_E-f	AAGTTCCTTAAGCCCAGG ATCGTAGGTA	46
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,709,539-13,709,514	LNCAP_E-r	ATATGAAGCCAGCAGCCA GGTAGCA	47
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,719,560-13,719,587	LNCAP_G-f	TTAGATAAACTGAAAGCC GAACCTGAAC	48
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,720,465-13,720,441	LNCAP_G-r	CAAACTCGCAAGCAATGT GAACTGT	49
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,723,583-13,723,615	LNCAP_H-f	TCACCGACAAAACCCATA GACAAAGAGT	50
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,724,850-13,724,823	LNCAP_H-r	TTAAATGGTGAGGCAATG AGGAAAGTG	51
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,727,707-13,727,735	LNCAP_I-f	TTGCTCATTCTCTTTCTCC CCTACACTAA	52
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,729,407-13,729,387	LNCAP_I-r	TCCCCACCACCAACCATC CTC	53
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,730,208-13,730,233	LNCAP_J1-f	CTGGGGGAAAAGCAAGT AGGAAAGTA	54
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,731,073-13,731,047	LNCAP_J1-r	ACAAGAGTTAGTCACGGC AAAGGAGTT	55

Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,733,199-13,733,220	LNCAP_J2-f	GCCCTTTGCCCATGAGAACTAA	56
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,734,028-13,734,003	LNCAP_J2-r	TCCCAGAAGAGATGATATGAGGTGTC	57
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,737,858-13,737,881	LNCAP_K-f	TCAGTCCCATCTCCCCTAAACCA	58
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,739,203-13,739,180	LNCAP_K-r	CACCATTCTCACCCGACCACATTG	59
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,743,031-13,743,060	LNCAP_L-f	TGTAAACTGCAATGAAAAGAAAAGAAAAAG	60
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,744,016-13,743,987	LNCAP_L-r	CAAGAGATGGGAGAGGAAGAATGAATAATA	61
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,746,316-13,746,343	LNCAP_M-f	CTATCTAGTCCCTTACGCTTTCCCTGTG	62
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,747,138-13,747,116	LNCAP_M-r	CATTAGCATTTGGCCTTTGGTCA	63
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,754,560-13,754,584	LNCAP_N1-f	TGCCTCCCCATAAGTCACCAATCTC	64
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,755,734-13,755,705	LNCAP_N1-r	CCTGTATTCTAACCCTGGACTTCTCATCAA	65
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,756,219-13,756,246	LNCAP_N2-f	CTTGTTTATTGGCCTAGTCCTTTGTGCT	66
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,757,204-13,757,179	LNCAP_N2-r	GCTTTGTGGGTAGTCCTGTCTGAGTG	67
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,760,623-13,760,642	LNCAP_O1-f	GGCCCATCCCGGTTTGCTAA	68
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,761,824-13,761,798	LNCAP_O1-r	GTTTCCCCACCACTTCCTTTCTATGTC	69
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,764,545-13,764,569	LNCAP_O2-f	GCACAAGACATACACGCAGATACAC	70
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,765,703-13,765,678	LNCAP_O2-r	AACGCTCGACTATGGAACTTTACCTG	71
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,767,646-13,767,674	LNCAP_P1-f	TCCTCTCATTCATTTGCATTCGTGTTAG	72
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,768,845-13,768,819	LNCAP_P1-r	GGCTTTGAGGGATTACTGGGTTGTTCT	73
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,770,885-13,770,905	LNCAP_P2-f	GCAGGGCAAAGAAGCAGTAGG	74
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,772,453-13,772,428	LNCAP_P2-r	GGATCCCAATTTAGTTTCAAGTTACG	75
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,774,561-13,774,586	LNCAP_Q-f	ATGTGCTGGCTAGATTGGACTGAAAA	76
Southern probe	Chr 7 (5' to ETV1)	NT_007819.16	13,775,272-13,775,244	LNCAP_Q-r	CAATAAAGCTGGAGGGGTGATAAATAAAT	77
Inverse PCR	Chr 7 (Probe A)	NT_007819.16	13,685,817-13,685,793	Inverse A1	TTAGAAGGAGACAATCTTATTCCAG	78
Inverse PCR	Chr 7 (Probe A)	NT_007819.16	13685,657-13,685,634	Inverse A2	CTCTTAAAGAGATGAAGCAGGGAG	79
Inverse PCR	Chr 7 (Probe A)	NT_007819.16	13,685,626-13,685,603	Inverse A3	TTGGCTAGATACAGGGTGAATATT	80
Inverse PCR	Chr 7 (Probe A)	NT_007819.16	13,685,833-13,685,856	Inverse B1	TGAATTCATGTGTGTAGCTGAGCC	81
Inverse PCR	Chr 7 (Probe A)	NT_007819.16	13,685,860-13,685,883	Inverse B2	TGACAGCGGGAATAAAGTACATGC	82
Inverse PCR	Chr 7 (Probe A)	NT_007819.16	13,686,095-13,686,118	Inverse B3	GTTGGGAGGTTTACTTGCCAATTA	83
Genomic fusion	Chr 7 (5' to ETV1)	NT_007819.16	13,686,807-13,686,832	Genomic fusion-f	ACATTTTGCTAACCCCTCTTCTATCT	84
Genomic fusion	Chr 14 (MIPOL1)	NT_026437.11	18,985,248-18,985,272	Genomic fusion-r	TCAACCTCAAAAATAAATGGCATCT	85
RWPE-*ETV1*	SERPINE1	NM_000602.1	1181-1200	SERPINE1-f	GCATGGCCCCCGAGGAGAT	86
RWPE-*ETV1*	SERPINE1	NM_000602.1	1270-1248	SERPINE1-r	CTTGGCCCATGAAAAGGACTGTT	87
RWPE-*ETV1*	TGFBI	NM_000358.1	1506-1528	TGFBI-f	AGGTACGGGACCCTGTTCACGAT	88
RWPE-*ETV1*	TGFBI	NM_000358.1	1605-1580	TGFBI-r	CTACCAGCATGCTAAAGCGATTGTCT	89
RWPE-*ETV1*	IGFBP3	NM_000598.4	738-762	IGFBP3-f	CGAGTCCAAGCGGGAGACAGAATA	90
RWPE-*ETV1*	IGFBP3	NM_000598.4	837-814	IGFBP3-r	TACACCCCTGGGACTCAGCACATT	91
RWPE-*ETV1*	MMP3	NM_002422.3	1055-1080	MMP3-f	TTCATTTTGGCCATCTCTTCCTTCAG	92
RWPE-*ETV1*	MMP3	NM_002422.3	1181-1155	MMP3-r	TATCCAGCTCGTACCTCA	93

					TTTCCTCT	
RWPE-*ETV1*	SPOCK1	NM_004598.3	829-848	SPOCK1-f	GCCCACCAGCTCCAACAC AG	94
RWPE-*ETV1*	SPOCK1	NM_004598.3	951-928	SPOCK1-r	GAAGGGTCAAGCAGGAG GTCATAG	95
RWPE-*ETV1*	BCL2	NM_000633.2	1014-1039	BCL2-f	CCCTGTGGATGACTGAGT ACCTGAAC	96
RWPE-*ETV1*	BCL2	NM_000633.2	1084-1064	BCL2-r	GGCATCCCAGCCTCCGTT ATC	97
RWPE-*ETV1*	MMP14	NM_004995.2	1036-1059	MMP14-f	AATTTTGTGCTGCCCGAT GATGAC	98
RWPE-*ETV1*	MMP14	NM_004995.2	1151-1129	MMP14-r	GGAACAGAAGGCCGGGA GGTAGT	99
RWPE-*ETV1*	MMP2	NM_004530.2	953-974	MMP2-f	GAAGGCCAAGTGGTCCGT GTGA	100
RWPE-*ETV1*	MMP2	NM_004530.2	1044-1019	MMP2-r	CAGCTGTTGTACTCCTTG CCATTGAA	101
RWPE-*ETV1*	ADAM19	NM_023038.3	2146-2165	ADAM19-f	GCCTATGCCCCCTGAGAG TG	102
RWPE-*ETV1*	ADAM19	NM_023038.3	2271-2245	ADAM19-r	GCTTGAGTTGGCCTAGTT TGTTGTTC	103
RWPE-*ETV1*	MMP9	NM_004994.2	1181-1201	MMP9-f	TGCCCGGACCAAGGATAC AGT	104
RWPE-*ETV1*	MMP9	NM_004994.2	1239-1221	MMP9-r	AGCGCGTGGCCGAACTCA T	105
RWPE-*ETV1*	PLAU	NM_002658.2	1169-1194	PLAU-f	TACGGCTCTGAAGTCACC ACCAAAAT	106
RWPE-*ETV1*	PLAU	NM_002658.2	1308-1286	PLAU-r	CCCCAGCTCACAATTCCA GTCAA	107

Table 2

Probe #	Gene/Region	Localization	Probe
1	ETV1	3'	RP11-124L22
2	ETV1	5'	RP11-703A4
3	Chr 14q13.3-14q21.1	C	RP11-945C4
4	Chr 14q13.3-14q21.1	T	RP11-107E23
5	HNRPA2B1 3'	3'	RP11-11F13
6	HNRPA2B1 5'	5'	RP11-379M24
7	HERV-K_22q11.23 5'	5'	RP11-61N10
8	HERV-K_22q11.23 3'	3'	RP11-71G19
9	SLC45A3	3'	RP11-249h15
10	SLC45A3	5'	RP11-131E5
11	C15ORF21	5'	RP11-474E1
12	C15ORF21	3'	RP11-626F7
13	14q32	3'	RP11-483K13
14	ETV1	5'	RP11-313C20
15	Chr 7 centromere	C	CEP 7
16	Chr 7p telomere	T	TelVysion 7p
17	TMPRSS2	5'	RP11-35C4

All publications, patents, patent applications and sequences identified by accession numbers mentioned in the above specification are herein incorporated by reference in their entirety. Although the invention has been described in connection with specific embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Modifications and

variations of the described compositions and methods of the invention that do not significantly change the functional features of the compositions and methods described herein are intended to be within the scope of the following claims.

5

CLAIMS

We claim:

1. A method for detecting prostate cancer comprising detecting the presence of a MIPOL1-ETV1 genetic rearrangement in a biological sample, wherein the presence in the sample of the genetic rearrangement is indicative of prostate cancer in an individual from whom the sample was derived.

2. The method of claim 1, wherein detecting the MIPOL1-ETV1 genetic rearrangement comprises detecting a chromosomal rearrangement of genomic DNA that comprises MIPOL1 genetic material and ETV1 genetic material in the same genomic region.

3. The method of claim 2, wherein detecting the chromosomal rearrangement of genomic DNA uses a nucleic acid sequencing technique.

4. The method of claim 2, wherein detecting the chromosomal rearrangement of genomic DNA uses a nucleic acid hybridization technique.

5. The method of claim 4, wherein detecting the chromosomal rearrangement of genomic DNA uses a nucleic acid hybridization technique selected from the group consisting of: *in situ* hybridization (ISH), analysis of hybridization to a microarray of probes, and Southern blot analysis.

6. The method of claim 2, wherein detecting the chromosomal rearrangement of genomic DNA uses a nucleic acid amplification method.

7. The method of claim 6, wherein the nucleic acid amplification method used is selected from the group consisting of: polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA).

8. The method of claim 1, wherein the sample is selected from the group consisting of tissue, blood, plasma, serum, urine, urine supernatant, urine cell pellet, semen, prostatic secretions and prostate cells.

9. A composition for detecting a MIPOL1-ETV1 genetic rearrangement associated with prostate cancer comprising at least one of the following:

(a) a probe comprising a sequence that hybridizes specifically to a junction at which an ETV1 gene is inserted into a MIPOL1 gene;

(b) a first probe comprising a sequence that hybridizes specifically to a MIPOL1 gene and a second probe comprising a sequence that hybridizes specifically to an ETV1 gene; or

(c) a first amplification oligonucleotide comprising a sequence that hybridizes specifically to a MIPOL1 gene and a second amplification oligonucleotide comprising a sequence that hybridizes specifically to an ETV1 gene and a means for detecting an amplified product.

10. The composition of claim 10, wherein the means for detecting an amplified product in (c) is selected from the group consisting of probes, intercalating dyes and labeled amplification oligonucleotides.

Figure 1

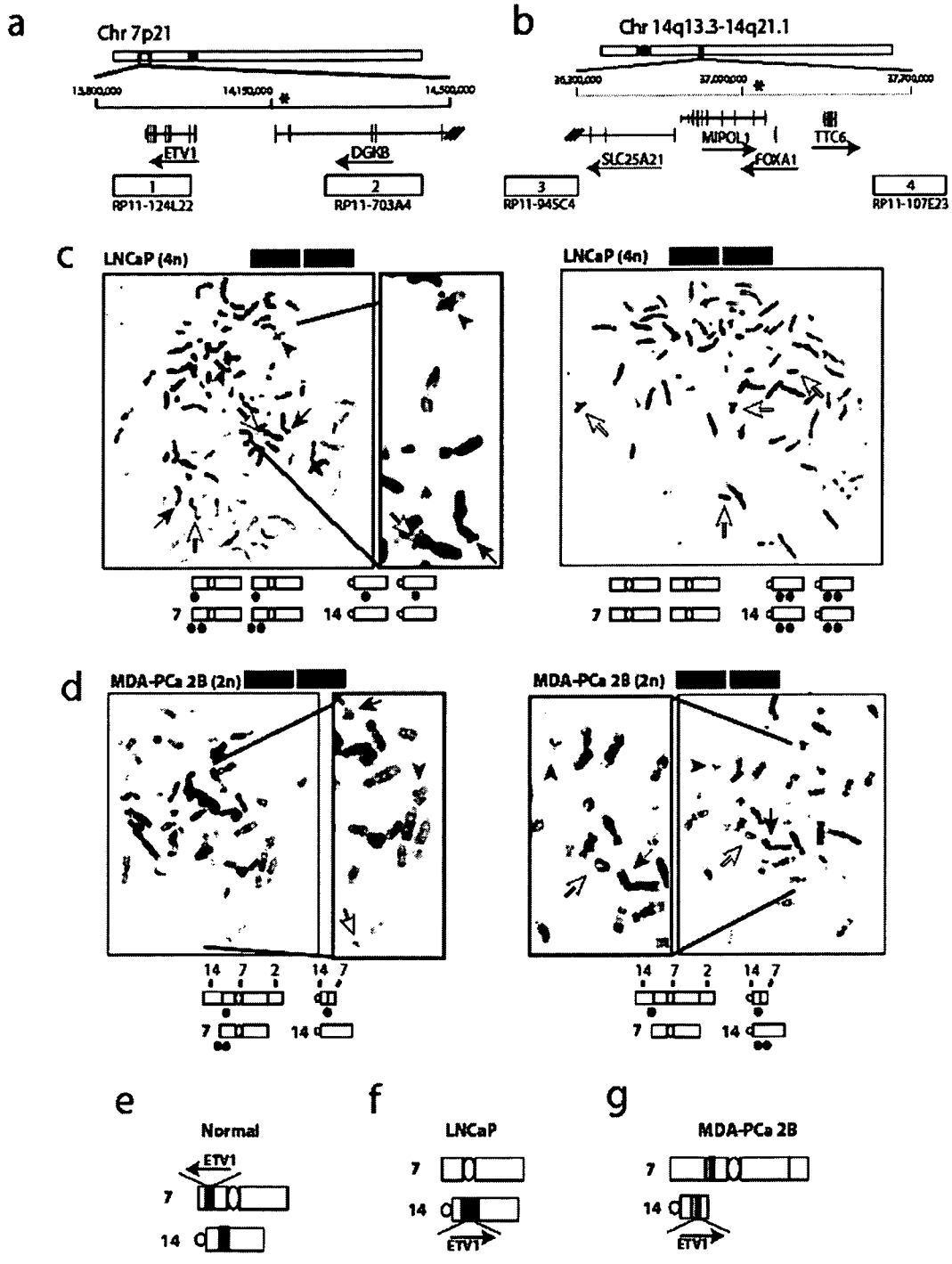

a
Chr 7p21
13,800,000
14,150,000
14,500,000
ETV1
DGKB
1
RP11-124L22
2
RP11-703A4
b
Chr 14q13.3-14q21.1
36,300,000
37,000,000
37,700,000
MIPOL1
TTC6
SLC25A21
FOXA1
3
RP11-94SC4
4
RP11-107E23
c
LNCaP (4n)
LNCaP (4n)
7
14
d
MDA-PCa 2B (2n)
MDA-PCa 2B (2n)
14 7 2 14 7
7
14
e
Normal
ETV1
7
14
f
LNCaP
7
14
ETV1
g
MDA-PCa 2B
7
14
ETV1

Figure 2

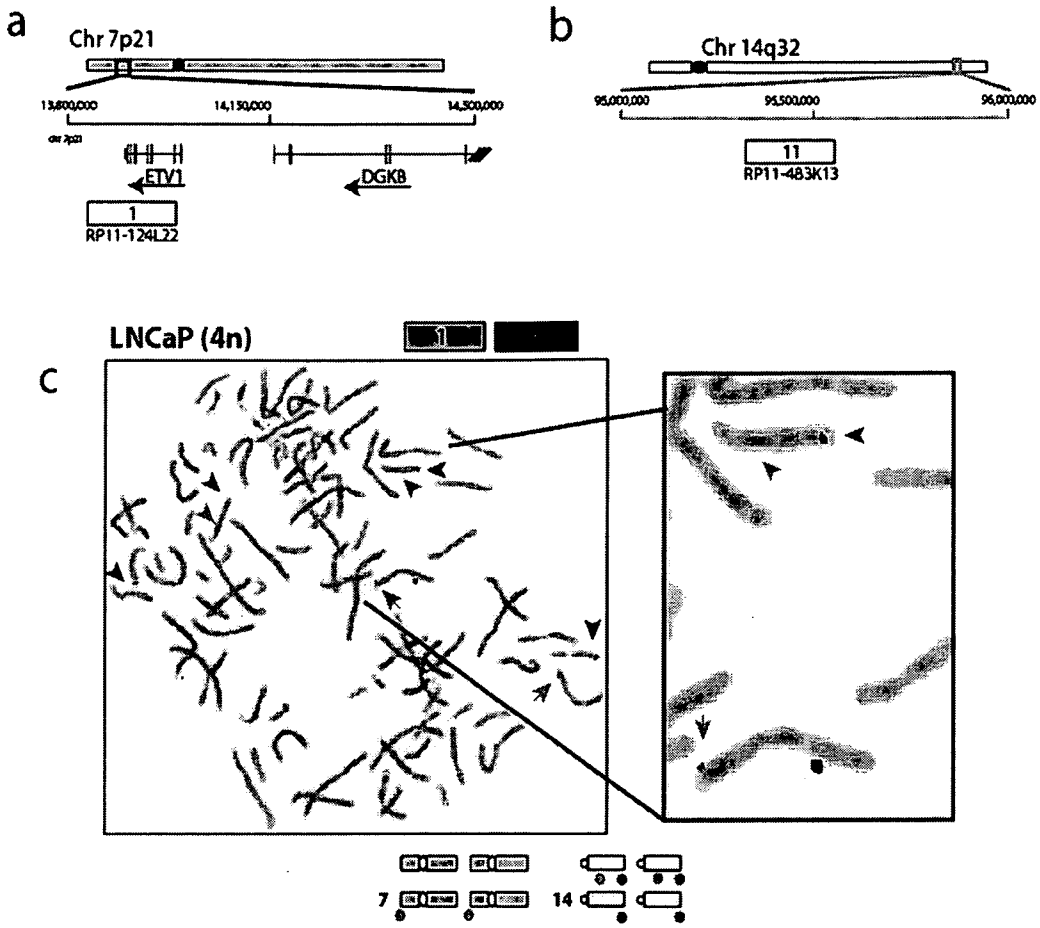

a
Chr 7p21
13,600,000
14,150,000
14,500,000
ETV1
DGKB
1
RP11-124L22
b
Chr 14q32
95,000,000
95,500,000
96,000,000
11
RP11-483K13
c
LNCaP (4n)
7
14

Figure 3

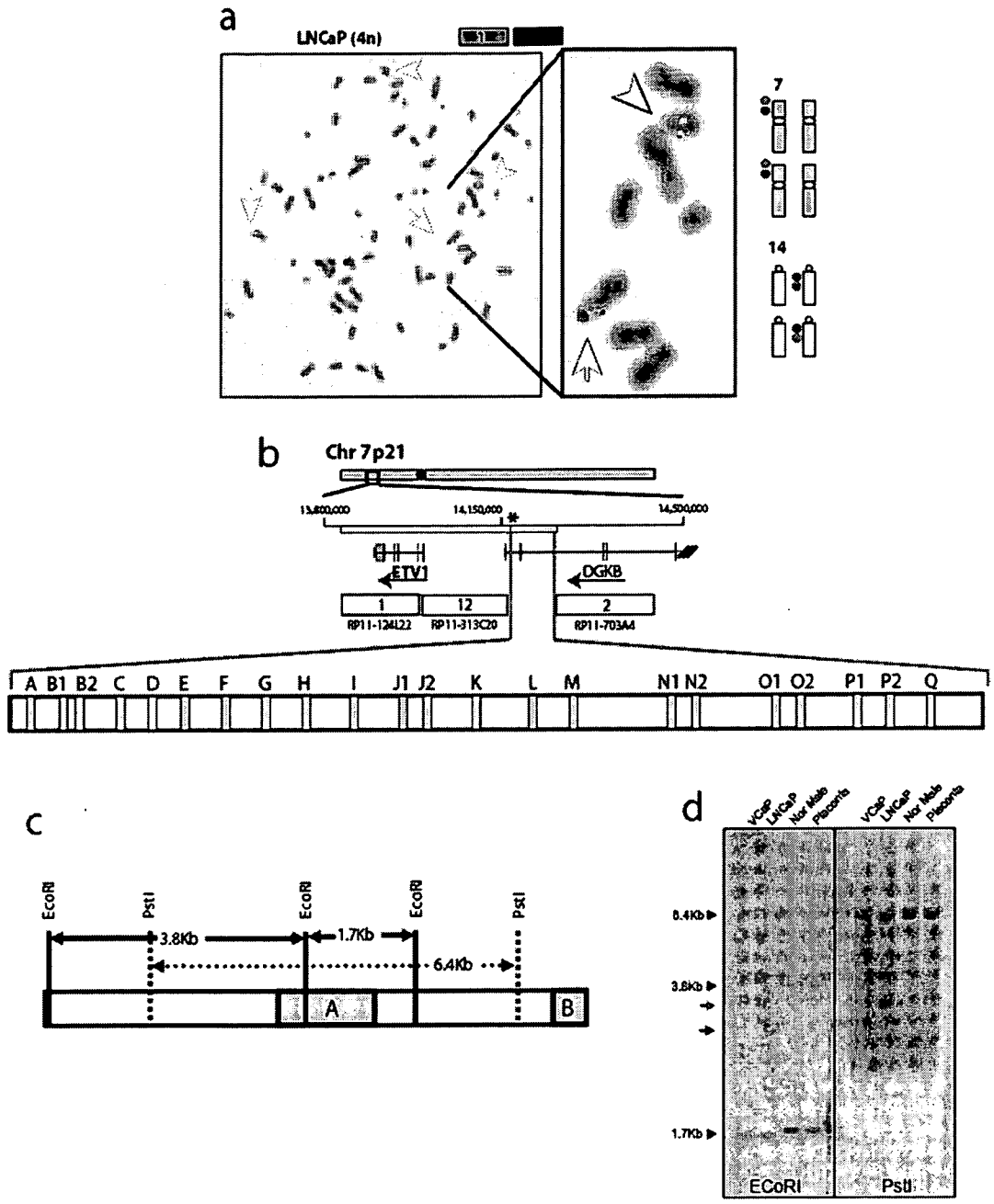

a
LNCaP (4n)
7
14
b
Chr 7p21
13,800,000
14,150,000
14,500,000
ETV1
DGKB
1
12
2
RP11-124L22
RP11-313C20
RP11-703A4
A B1 B2 C D E F G H I J1 J2 K L M N1 N2 O1 O2 P1 P2 Q
c
EcoRI
PstI
EcoRI
EcoRI
PstI
3.8Kb
1.7Kb
6.4Kb
A
B
d
VCaP
LNCaP
Nor Male
Placenta
VCaP
LNCaP
Nor Male
Placenta
8.4Kb
3.8Kb
1.7Kb
EcoRI
PstI

Figure 4

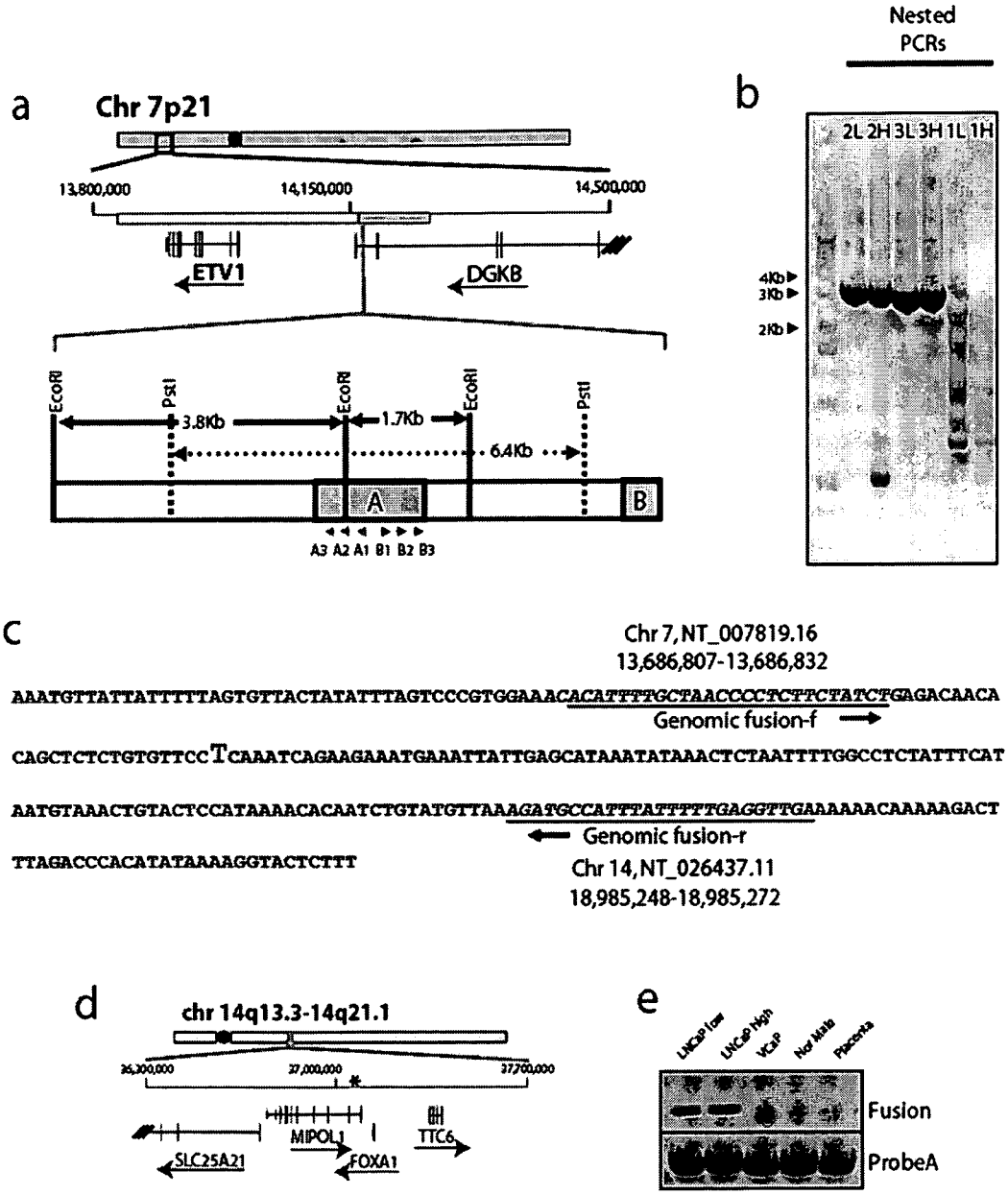

a
Chr 7p21
13,800,000
14,150,000
14,500,000
ETV1
DGKB
EcoRI
PstI
EcoRI
EcoRI
PstI
3.8Kb
1.7Kb
6.4Kb
A
B
A3 A2 A1 B1 B2 B3
b
Nested PCRs
2L 2H 3L 3H 1L 1H
4Kb
3Kb
2Kb
c
Chr 7, NT_007819.16
13,686,807-13,686,832
AAATGTTATTATTTTTAGTGTTACTATATTTAGTCCCGTGGAAACACATTTTGCTAACCCCTCTTCTATCTGAGACAACA
Genomic fusion-f
CAGCTCTCTGTGTTCCTCAAATCAGAAGAAATGAAATTATTGAGCATAAATATAAACTCTAATTTTGGCCTCTATTTCAT
AATGTAAACTGTACTCCATAAAACACAATCTGTATGTTAAAGATGCCATTTATTTTTGAGGTTGAAAAAACAAAAAGACT
Genomic fusion-r
TTAGACCCACATATAAAAGGTACTCTTT
Chr 14, NT_026437.11
18,985,248-18,985,272
d
chr 14q13.3-14q21.1
36,300,000
37,000,000
37700,000
MIPOL1
TTC6
SLC25A21
FOXA1
e
LNCaP low
LNCaP High
VCaP
Nor Male
Placenta
Fusion
ProbeA

Figure 5

Figure 6

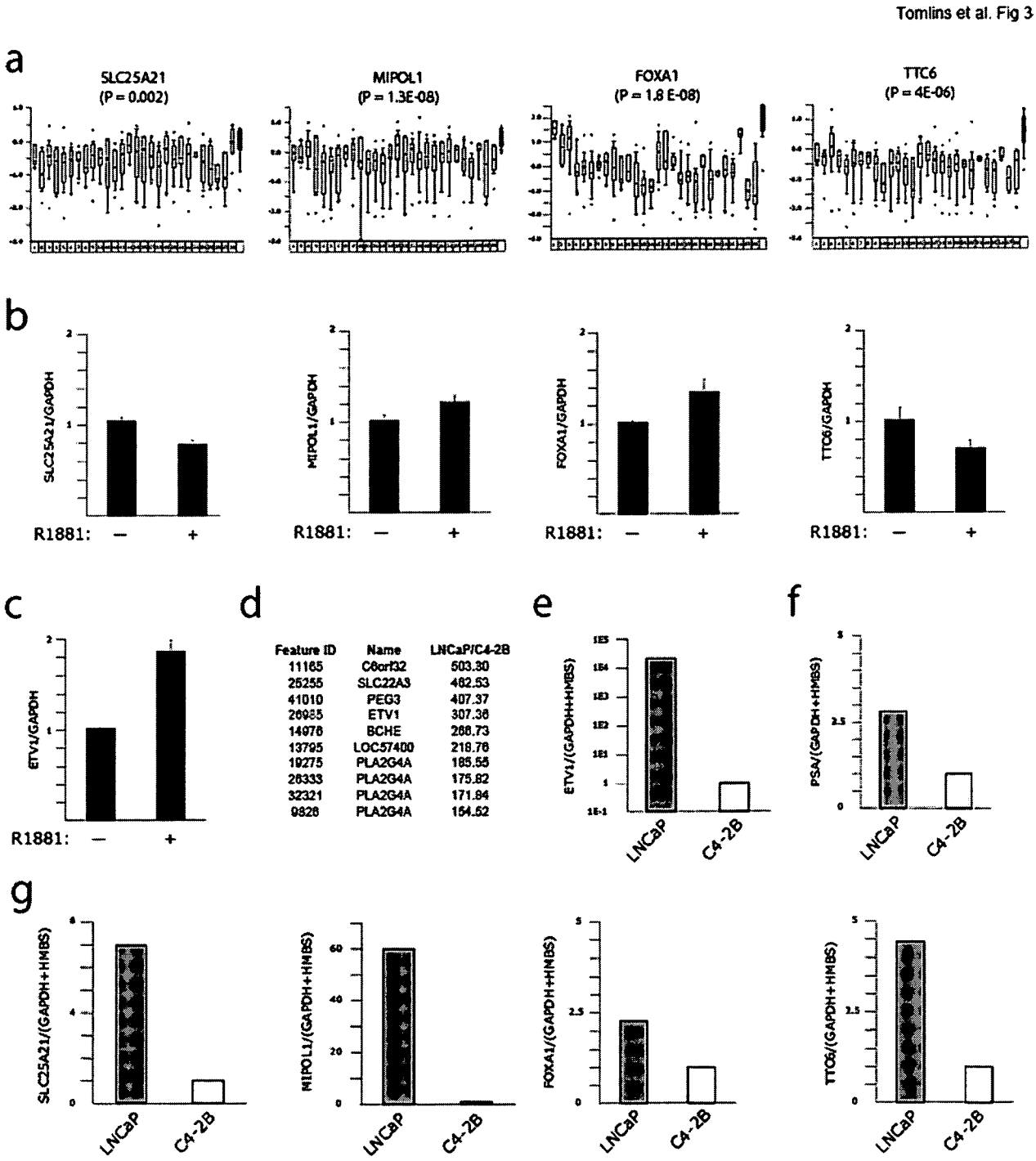

Tomlins et al. Fig 3
a
SLC25A21 (P = 0.002)
MIPOL1 (P = 1.3E-08)
FOXA1 (P = 1.8 E-08)
TTC6 (P = 4E-06)
b
SLC25A21/GAPDH
MIPOL1/GAPDH
FOXA1/GAPDH
TTC6/GAPDH
R1881: – +
c
ETV1/GAPDH
R1881: – +
d
Feature ID Name LNCaP/C4-2B
11165 C6orf32 503.30
25255 SLC22A3 462.53
41010 PEG3 407.37
26985 ETV1 307.36
14976 BCHE 268.73
13795 LOC57400 218.76
19275 PLA2G4A 185.55
28333 PLA2G4A 175.82
32321 PLA2G4A 171.84
9828 PLA2G4A 154.52
e
ETV1/(GAPDH+HMBS)
LNCaP C4-2B
f
PSA/(GAPDH+HMBS)
LNCaP C4-2B
g
SLC25A21/(GAPDH+HMBS)
MIPOL1/(GAPDH+HMBS)
FOXA1/(GAPDH+HMBS)
TTC6/(GAPDH+HMBS)
LNCaP C4-2B

Figure 7

Figure 8

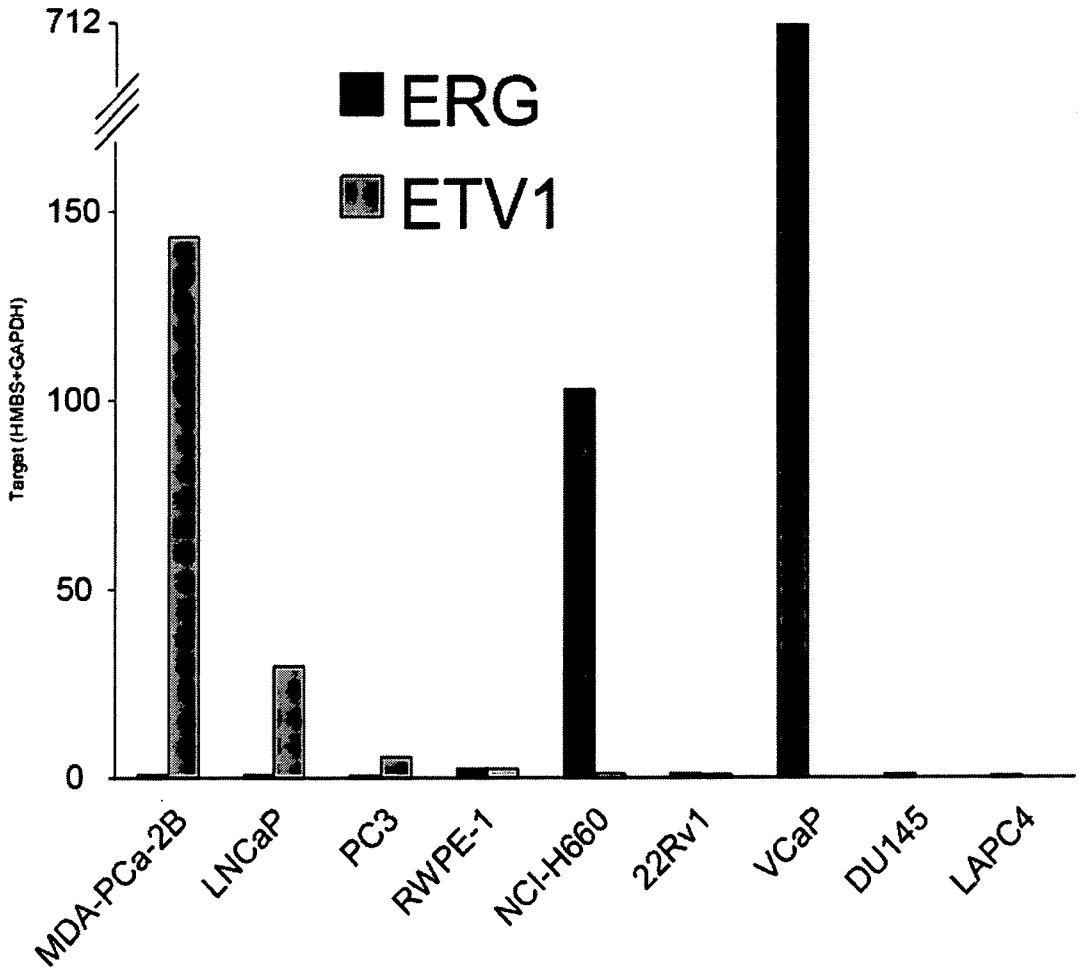
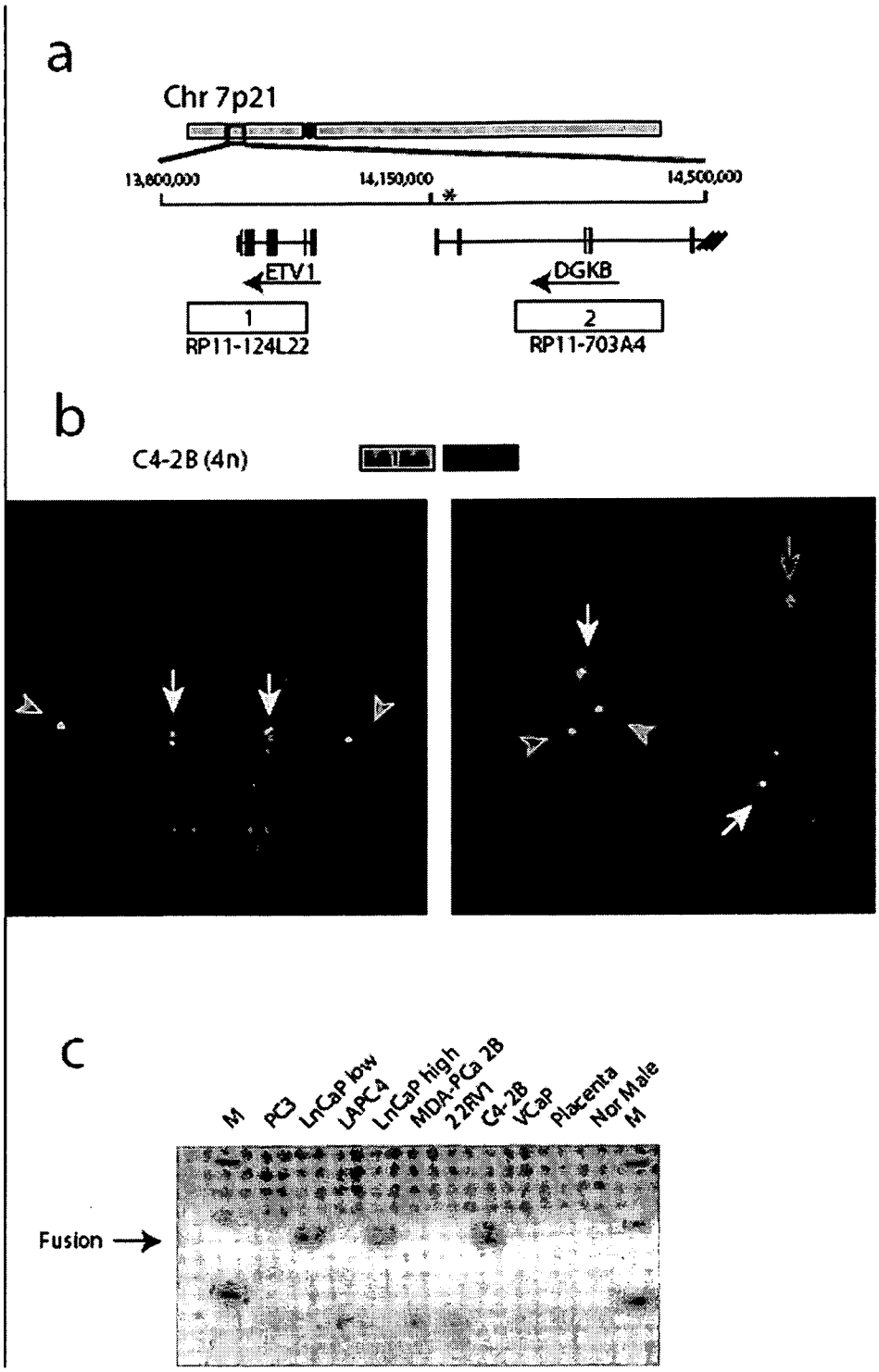

712
150
100
50
0
Target (HMBS+GAPDH)
ERG
ETV1
MDA-PCa-2B
LNCaP
PC3
RWPE-1
NCI-H660
22Rv1
VCaP
DU145
LAPC4

Figure 9


a
Chr 7p21
13,600,000
14,150,000
14,500,000
ETV1
DGKB
1
RP11-124L22
2
RP11-703A4
b
C4-2B (4n)
c
M
PC3
LnCaP low
LAPC4
LnCaP high
MDA-PCa 2B
22RV1
C4-2B
VCaP
Placenta
Nor Male
M
Fusion