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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 February 2008 (28.02.2008)

PCT

(10) International Publication Number
WO 2008/024114 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *A61K 49/00* (2006.01)

(21) International Application Number:
PCT/US2006/033147

(22) International Filing Date: 24 August 2006 (24.08.2006)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **GENIZON BIOSCIENCES INC.** [CA/CA]; 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA).

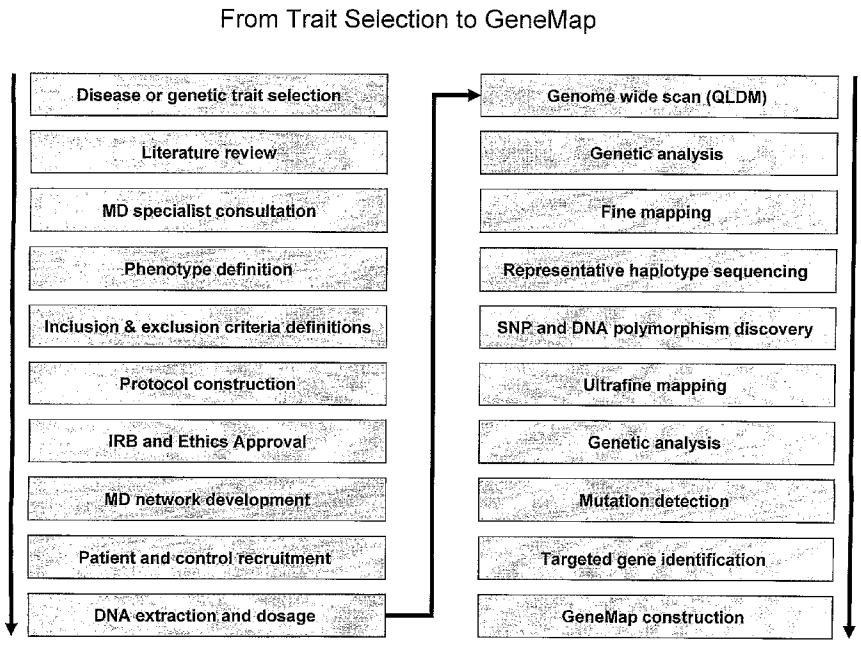
(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BELOUCHI, Abdelmajid** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **RAELSON, John Verner** [US/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **BRADLEY, Walter Edward** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **PAQUIN, Bruno** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **PEREPETCHAI, Valeri** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **NGUYEN-HUU, Quynh** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **CROTEAU, Pascal** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **ALLARD, Rene** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **PAQUIN, Nouzha** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **VAN EERDEWEGH, Paul** [US/US]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **LITTLE, Randall David** [US/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **KEITH, Tim** [US/US]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **SEGAL, Jonathan** [US/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA). **BRUAT, Vanessa** [CA/CA]; c/o GENIZON BIOSCIENCES INC., 880 McCaffrey, Ville St-laurent, Québec H4T 2C7 (CA).

(74) Agents: **TUSCAN, Michael S.** et al.; Cooley Godward Kronish LLP, Attn: Patent Group, 11951 Freedom Drive, Reston, VA 20190 (US).

[Continued on next page]

(54) Title: GENEMAP OF THE HUMAN GENES ASSOCIATED WITH SCHIZOPHRENIA



(57) Abstract: The present invention relates to the selection of a set of SNP, allele and haplotype markers for use in genome wide association studies based on linkage disequilibrium mapping. In particular, the invention relates to the fields of pharmacogenomics, diagnostics, patient therapy and the use of genetic haplotype information to predict an individual's susceptibility to schizophrenia and/or their response to a particular drug or drugs.

WO 2008/024114 A1

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, 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, 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, HU, IE, IS, IT, LT, LU, LV, MC, NL, 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:

— *with international search report*

GeneMap of the human genes associated with schizophrenia

INVENTORS: Abdelmajid Belouchi, John Verner Raelson, Walter Edward Bradley, Valeri Perepetchai, Vanessa Bruat, Bruno Paquin, Randall D. Little, Quynh Nguyen-Huu, Pascal Croteau, Rene Allard, Nouzha Paquin, Paul Van Eerdewegh, Jonathan Segal and Tim Keith.

FIELD OF THE INVENTION

The invention relates to the field of genomics and genetics, including genome analysis and the study of DNA variations. In particular, the invention relates to the fields of pharmacogenomics, diagnostics, patient therapy and the use of genetic haplotype information to predict an individual's susceptibility to schizophrenia and/or their response to a particular drug or drugs, so that drugs tailored to genetic differences of population groups may be developed and/or administered to the appropriate population.

The invention also relates to a GeneMap for schizophrenia disease, which links variations in DNA (including both genic and non-genic regions) to an individual's susceptibility to schizophrenia and/or response to a particular drug or drugs. The invention further relates to the genes disclosed in the GeneMap (see Tables 5-7), which is related to methods and reagents for detection of an individual's increased or decreased risk for schizophrenia by identifying at least one polymorphism in one or a combination of the genes from the GeneMap. Also related are the candidate regions identified in Table 1, which are associated with schizophrenia. In addition, the invention further relates to nucleotide sequences of those genes including genomic DNA sequences, cDNA sequences, single nucleotide polymorphisms (SNPs), alleles and haplotypes (see Sequence Listing and Tables 2-7).

The invention further relates to isolated nucleic acids comprising these nucleotide sequences and isolated polypeptides or peptides encoded thereby. Also related, are expression vectors and host cells comprising the disclosed nucleic acids or fragments thereof, as well as antibodies that bind to the encoded polypeptides or peptides.

The present invention further relates to ligands that modulate the activity of the disclosed genes or gene products. In addition, the invention relates to diagnostics and therapeutics for schizophrenia, utilizing the disclosed nucleic acids, SNPs, chromosomal regions, gene maps, polypeptides or peptides, antibodies and/or ligands and small molecules that activate or repress relevant signaling events.

BACKGROUND OF THE INVENTION

Schizophrenia is a severe psychiatric condition that affects approximately one percent of the population worldwide (Lewis *et al.*, 2000). People with schizophrenia often experience both "positive" symptoms (*e.g.*, delusions, hallucinations, paranoia, psychosis, disorganized thinking, and agitation) and "negative" symptoms (*e.g.*, lack of drive or initiative, social withdrawal, apathy, impaired attention, cognitive impairments and emotional unresponsiveness).

There are an estimated 45 million people with schizophrenia in the world, with more than 33 million of them in developing countries. This disease places a heavy burden on the patient's family and relatives, both in terms of the direct and indirect costs involved, and the social stigma associated with the illness, sometimes over generations. Moreover, schizophrenia accounts for one fourth of all mental health costs and takes up one in three psychiatric hospital beds. Most schizophrenia patients are never able to work. The cost of schizophrenia to society is enormous. The most common cause of death among schizophrenic patients is suicide (in 10% of patients) which represents a 20 times higher risk than for the general population. Deaths from heart disease and from diseases of the respiratory and digestive system are also increased among schizophrenic patients.

Studies of the inheritance of schizophrenia have revealed that it is a multi- factorial disease characterized by multiple genetic susceptibility elements; each likely contributing a modest increase in risk (Karayiorgou *et al.*, 1997). The DNA sequences between two human genomes are 99.9% identical. The variations in DNA sequence between individuals can be, for example, deletions of small or large stretches of DNA, insertions of stretches of DNA, variations in the number of repetitive DNA elements in non-coding regions, or changes in single base positions

in the genome called "single nucleotide polymorphisms" (SNPs). Human DNA sequence variations account for a large fraction of observed differences between individuals, including susceptibility to disorders or a genetic link to disorders, such as schizophrenia.

Complex disorders such as schizophrenia are believed to involve several genes rather than single genes, as observed in rare disorders. This makes detection of any particular gene substantially more difficult than in a rare disorder, where a single gene mutation segregating according to a Mendelian inheritance pattern is the causative mutation. Any one of the multiple interacting gene mutations involved in the etiology of a complex and common disorder will impart a lower relative risk for the disorder than will the single gene mutation involved in a simple genetic disorder. Low relative risk alleles are more difficult to detect and, as a result, the success of positional cloning using linkage mapping that was achieved for simple genetic disorder genes has not been repeated for complex disorders.

Several approaches have been proposed to discover and characterize multiple genes in complex genetic disorders. These gene discovery methods can be subdivided into hypothesis-free disorder association studies and hypothesis-driven candidate gene or region studies. The candidate gene approach relies on the analysis of a gene in patients who have a disorder or a genetic disorder in which the gene is thought to play a role. This approach is limited in utility because it only provides for the investigation of genes with known functions. Although variant sequences of candidate genes may be identified using this approach, it is inherently limited by the fact that variant sequences in other genes that contribute to the phenotype will be necessarily missed when the technique is employed. A genome-wide scan (GWS) has been shown to be efficient in identifying schizophrenia susceptibility markers, such as the NRG1 gene on chromosome 8. In contrast to the candidate gene approach, a GWS searches throughout the genome without any a priori hypothesis and consequently can identify genes that are not obvious candidates for the complex genetic disorder as well as genes that are relevant candidates for the disorder. Furthermore, it can identify structurally important chromosomal regions a "that can influence the expression of specific, disorder-related genes.

Family-based linkage mapping methods were initially used for disorder locus identification. This technique locates genes based on the relatively limited number of genetic recombination events within the families used in the study, and results in large chromosomal regions containing hundreds of genes, any one of which could be the disorder-causing gene. Population-based, or linkage disequilibrium (LD) mapping is based on the premise that regions adjacent to a gene of interest are co-transmitted through the generations along with the gene. As a result, LD extends over shorter genetic regions than does linkage (Hewett *et al.*, 2002), and can facilitate detection of genes with lower relative risk than family linkage mapping approaches. It also defines much smaller candidate regions which may contain only a few genes, making the identification of the actual disorder gene much easier.

It has been estimated that a GWS that uses a general population and case/control association (LD) analysis would require approximately 700,000 SNP markers (Carlson *et al.*, 2003). The cost of a GWS at this marker density for a sufficient sample size for statistical power is economically prohibitive. The use of a founder population (genetic isolates), such as the French Canadian population of Quebec, is one solution to the problem with LD analysis. The French Canadian population in Quebec (Quebec Founder Population - QFP) provides one of the best resources in the world for gene discovery based on its high levels of genetic sharing and genetic homogeneity. By combining DNA collected from the QFP, high throughput genotyping capabilities and proprietary algorithms for genetic analysis, a comprehensive genome-wide association study is facilitated. The present invention relates specifically to a set of schizophrenia-related genes (GeneMap) and targets which present attractive points of therapeutic intervention for schizophrenia.

In view of the foregoing, identifying susceptibility genes associated with schizophrenia and elucidating their respective biochemical pathways will facilitate the development of effective treatments for schizophrenia. This will also lead to the identification of diagnostic markers, which will predict the propensity for any such disease and allow therapeutic intervention before such disease occurs. The identification of genetic markers associated with schizophrenia will lead to the development of effective therapeutic interventions for a much greater proportion of the individuals affected by schizophrenia. Knowledge of schizophrenia-associated

polymorphisms provides the benefit of predicting an individual susceptibility to schizophrenia. The present invention satisfies this need and provides related advantages as well.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1a. Method employed by the inventors to permit the identification of genes predisposing to a particular genetic disorder, such as schizophrenia. The method can be applied for any given disorder and the end result is the construction of a GeneMap for a particular disorder. Briefly, a genetically heritable disorder selected followed by the preparation of an in-depth literature review on the prevalence, phenotypes, and available treatments (if relevant) of that disorder. The literature review includes a list and description of candidate genes and regions associated with the disorder. A clinical specialist in the field of the genetic disorder is consulted for the definition of phenotype. Inclusion and exclusion criteria are then set and a study protocol is prepared. IRB and ethical approval are sought prior to patient recruitment. A network of physicians is required to recruit the necessary cases and controls for the study from the Quebec Founder Population. Individuals (cases and controls) are then recruited and DNA extraction and dosage is performed from the blood samples obtained. Samples are pooled into several cases and control pools. A GWS is performed on the pooled case and control samples using, as a minimum, the marker density determined from a study of linkage disequilibrium in the Quebec Founder Population; a study that lead to the formulation of the Quebec linkage disequilibrium map (QLDM). The results from the GWS genotyping are analyzed and candidate regions are selected for confirmatory mapping followed by fine mapping at a higher marker density in individual case and control samples in order to validate and/or refine the signal. The gene content of the candidate regions is analyzed and characterized. The representative haplotypes are then selected and sequenced. Once polymorphisms are identified by sequencing efforts the frequencies of genotypes and haplotypes in individual cases and controls are analyzed in a similar manner as for the GWS and fine mapping data. Ultrafine mapping is performed on all the samples to identify the polymorphisms that are most associated with the disorder phenotype as part of the search for the actual DNA polymorphisms that confer susceptibility to the disorder. The genes found associated with the disorder, are then corroborated in a different population. The corroborated genes are used for the construction of a GeneMap.

DESCRIPTION OF THE FILES CONTAINED ON THE CD-R

The sequence listing submitted concurrently herewith on compact disc under 37 C.F.R. §§1.821(c) and 1.821(e) is herein incorporated by reference in its entirety. A compact disc copy of the Sequence Listing (COPY 1) (filename: GENI 019 00WO SeqList.ST25.txt, date recorded: August 24, 2006, file size 1,681 kilobytes); a duplicate compact disc copy of the Sequence Listing (COPY 2) (filename: GENI 019 00WO SeqList.ST25.txt, date recorded: August 24, 2006, file size 1,681 kilobytes); a duplicate compact disc copy of the Sequence Listing (COPY 3) (filename: GENI 019 00WO SeqList.ST25.txt, date recorded: August 24, 2006, file size 1,681 kilobytes); a computer readable format copy of the Sequence Listing (CRF COPY) (filename: GENI 019 00WO SeqList.ST25.txt, date recorded: August 24, 2006, file size 1,681 kilobytes).

DEFINITIONS

Throughout the description of the present invention, several terms are used that are specific to the science of this field. For the sake of clarity and to avoid any misunderstanding, these definitions are provided to aid in the understanding of the specification and claims:

Allele: One of a pair, or series, of forms of a gene or non-genic region that occur at a given locus in a chromosome. Alleles are symbolized with the same basic symbol (*e.g.*, B for dominant and b for recessive; B1, B2, Bn for n additive alleles at a locus). In a normal diploid cell there are two alleles of any one gene (one from each parent), which occupy the same relative position (locus) on homologous chromosomes. Within a population there may be more than two alleles of a gene. See multiple alleles. SNPs also have alleles, *i.e.*, the two (or more) nucleotides that characterize the SNP

Amplification of nucleic acids: refers to methods such as polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known in the art and are described, for example, in U.S. Patent Nos. 4,683,195 and 4,683,202.

Reagents and hardware for conducting PCR are commercially available. Primers useful for amplifying sequences from the disorder region, are preferably complementary to, and preferably hybridize specifically to, sequences in the disorder region or in regions that flank a target region therein. Genes from Tables 5-7 generated by amplification may be sequenced directly. Alternatively, the amplified sequence(s) may be cloned prior to sequence analysis.

Antigenic component: is a moiety that binds to its specific antibody with sufficiently high affinity to form a detectable antigen-antibody complex.

Antibodies: refer to polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, that can bind to proteins and fragments thereof or to nucleic acid sequences from the disorder region, particularly from the disorder gene products or a portion thereof. The term antibody is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Proteins may be prepared synthetically in a protein synthesizer and coupled to a carrier molecule and injected over several months into rabbits. Rabbit sera are tested for immunoreactivity to the protein or fragment. Monoclonal antibodies may be made by injecting mice with the proteins, or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with protein or fragments thereof (Harlow *et al.* 1988, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). These antibodies will be useful in assays as well as therapeutics.

Associated allele: refers to an allele at a polymorphic locus that is associated with a particular phenotype of interest, *e.g.*, a predisposition to a disorder (*e.g.*, schizophrenia) or a particular drug response.

cDNA: refers to complementary or copy DNA produced from an RNA template by the action of RNA-dependent DNA polymerase (reverse transcriptase). Thus, a cDNA clone is a duplex DNA sequence complementary to an RNA molecule of interest, included in a cloning vector or amplified by PCR. This term includes genes from which the intervening sequences have been removed.

cDNA library: refers to a collection of recombinant DNA molecules containing cDNA inserts that together comprise essentially all of the expressed genes of an organism

or tissue. A cDNA library can be prepared by methods known to one skilled in the art (see, *e.g.*, Cowell and Austin, 1997, "DNA Library Protocols," Methods in Molecular Biology). Generally, RNA is first isolated from the cells of the desired organism, and the RNA is used to prepare cDNA molecules.

Cloning: refers to the use of recombinant DNA techniques to insert a particular gene or other DNA sequence into a vector molecule. In order to successfully clone a desired gene, it is necessary to use methods for generating DNA fragments, for joining the fragments to vector molecules, for introducing the composite DNA molecule into a host cell in which it can replicate, and for selecting the clone having the target gene from amongst the recipient host cells.

Cloning vector: refers to a plasmid or phage DNA or other DNA molecule that is able to replicate in a host cell. The cloning vector is typically characterized by one or more endonuclease recognition sites at which such DNA sequences may be cleaved in a determinable fashion without loss of an essential biological function of the DNA, and which may contain a selectable marker suitable for use in the identification of cells containing the vector.

Coding sequence or a protein-coding sequence: is a polynucleotide sequence capable of being transcribed into mRNA and/or capable of being translated into a polypeptide or peptide. The boundaries of the coding sequence are typically determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus.

Complement of a nucleic acid sequence: refers to the antisense sequence that participates in Watson-Crick base-pairing with the original sequence.

Disorder region: refers to the portions of the human chromosomes displayed in Table 1 bounded by the markers from Tables 2-7.

Disorder-associated nucleic acid or polypeptide sequence: refers to a nucleic acid sequence that maps to a region of Table 1 or the polypeptides encoded therein (Tables 2-7 SNPs, nucleic acids, and polypeptides). For nucleic acids, this encompasses sequences that are identical or complementary to the gene sequences from Tables 5-7, as well as sequence-conservative, function-conservative, and non-

conservative variants thereof. For polypeptides, this encompasses sequences that are identical to the polypeptide, as well as function-conservative and non-conservative variants thereof. Included are the alleles of naturally-occurring polymorphisms causative of schizophrenia such as, but not limited to, alleles that cause altered expression of genes of Tables 5-7 and alleles that cause altered protein levels or stability (*e.g.*, decreased levels, increased levels, expression in an inappropriate tissue type, increased stability, and decreased stability).

Expression vector: refers to a vehicle or plasmid that is capable of expressing a gene that has been cloned into it, after transformation or integration in a host cell. The cloned gene is usually placed under the control of (*i.e.*, operably linked to) a regulatory sequence.

Function-conservative variants: are those in which a change in one or more nucleotides in a given codon position results in a polypeptide sequence in which a given amino acid residue in the polypeptide has been replaced by a conservative amino acid substitution. Function-conservative variants also include analogs of a given polypeptide and any polypeptides that have the ability to elicit antibodies specific to a designated polypeptide.

Founder population: also called a population isolate, this is a large number of people who have mostly descended, in genetic isolation from other populations, from a much smaller number of people who lived many generations ago.

Gene: refers to a DNA sequence that encodes through its template or messenger RNA a sequence of amino acids characteristic of a specific peptide, polypeptide, or protein. The term "gene" also refers to a DNA sequence that encodes an RNA product. The term gene as used herein with reference to genomic DNA includes intervening, non-coding regions, as well as regulatory regions, and can include 5' and 3' ends. A gene sequence is wild-type if such sequence is usually found in individuals unaffected by the disorder or condition of interest, *e.g.*, schizophrenia. However, environmental factors and other genes can also play an important role in the ultimate determination of the genetic disorder. In the context of complex disorders involving multiple genes (oligogenic disorders), the wild type, or normal

sequence can also be associated with a measurable risk or susceptibility, receiving its reference status based on its frequency in the general population.

GeneMaps are defined as groups of gene(s) that are directly or indirectly involved in at least one phenotype of a disorder, *e.g.*, schizophrenia. As such, GeneMaps enable the development of synergistic diagnostic products, creating "theranostics".

Genotype: set of alleles at a specified locus or loci.

Haplotype: The allelic pattern of a group of (usually contiguous) DNA markers or other polymorphic loci along an individual chromosome or double helical DNA segment. Haplotypes identify individual chromosomes or chromosome segments. The presence of shared haplotype patterns among a group of individuals implies that the locus defined by the haplotype has been inherited, identical by descent (IBD), from a common ancestor. Detection of identical by descent haplotypes is the basis of linkage disequilibrium (LD) mapping. Haplotypes are broken down through the generations by recombination and mutation. In some instances, a specific allele or haplotype may be associated with susceptibility to a disorder or condition of interest, *e.g.*, schizophrenia. In other instances, an allele or haplotype may be associated with a decrease in susceptibility to a disorder or condition of interest, *i.e.*, a protective sequence (see table 4 for haplotypes associated with Schizophrenia).

Host: includes prokaryotes and eukaryotes. The term includes an organism or cell that is the recipient of an expression vector (*e.g.*, autonomously replicating or integrating vector).

Hybridizable: nucleic acids are hybridizable to each other when at least one strand of the nucleic acid can anneal to another nucleic acid strand under defined stringency conditions. In some embodiments, hybridization requires that the two nucleic acids contain at least 10 substantially complementary nucleotides; depending on the stringency of hybridization, however, mismatches may be tolerated. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementarity, and can be determined in accordance with the methods described herein.

Identity by descent (IBD): Identity among DNA sequences for different individuals that is due to the fact that they have all been inherited from a common ancestor. LD mapping identifies IBD haplotypes as the likely location of disorder genes shared by a group of cases.

Identity: as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, identity also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. Identity and similarity can be readily calculated by known methods, including but not limited to those described in A.M. Lesk (ed), 1988, Computational Molecular Biology, Oxford University Press, NY; D.W. Smith (ed), 1993, Biocomputing. Informatics and Genome Projects, Academic Press, NY; A.M. Griffin and H.G. Griffin, H. G (eds), 1994, ComputerAnalysis of Sequence Data, Part 1, Humana Press, NJ; G. von Heinje, 1987, Sequence Analysis in Molecular Biology, Academic Press; and M. Gribskov and J. Devereux (eds), 1991, Sequence Analysis Primer, M Stockton Press, NY; H. Carillo and D. Lipman, 1988, SIAM J. Applied Math., 48:1073.

Immunogenic component: is a moiety that is capable of eliciting a humoral and/or cellular immune response in a host animal.

Isolated nucleic acids: are nucleic acids separated away from other components (*e.g.*, DNA, RNA, and protein) with which they are associated (*e.g.*, as obtained from cells, chemical synthesis systems, or phage or nucleic acid libraries). Isolated nucleic acids are at least 60% free, preferably 75% free, and most preferably 90% free from other associated components. In accordance with the present invention, isolated nucleic acids can be obtained by methods described herein, or other established methods, including isolation from natural sources (*e.g.*, cells, tissues, or organs), chemical synthesis, recombinant methods, combinations of recombinant and chemical methods, and library screening methods.

Isolated polypeptides or peptides: are those that are separated from other components (*e.g.*, DNA, RNA, and other polypeptides or peptides) with which they are associated (*e.g.*, as obtained from cells, translation systems, or chemical

synthesis systems). In a preferred embodiment, isolated polypeptides or peptides are at least 10% pure; more preferably, 80% or 90% pure. Isolated polypeptides and peptides include those obtained by methods described herein or other established methods, including isolation from natural sources (*e.g.*, cells, tissues, or organs), chemical synthesis, recombinant methods, or combinations of recombinant and chemical methods. Proteins or polypeptides referred to herein as recombinant are proteins or polypeptides produced by the expression of recombinant nucleic acids. A portion as used herein with regard to a protein or polypeptide refers to fragments of that protein or polypeptide. The fragments can range in size from 5 amino acid residues to all but one residue of the entire protein sequence. Thus, a portion or fragment can be at least 5, 5-50, 50-100, l00-200, 200-400, 400-800, or more consecutive amino acid residues of a protein or polypeptide.

Linkage disequilibrium (LD): a statistical association between particular alleles at separate but linked loci, normally the result of a particular ancestral haplotype being common in the population studied. LD can also be defined as the situation in which the alleles for two or more loci do not occur together in individuals sampled from a population at frequencies predicted by the product of their individual allele frequencies. In other words, markers that are in LD do not follow Mendel's second law of independent random segregation. LD can be caused by any of several demographic or population artifacts as well as by the presence of genetic linkage between markers. However, when these artifacts are controlled and eliminated as sources of LD, then LD results directly from the fact that the loci involved are located close to each other on the same chromosome so that specific combinations of alleles for different markers (haplotypes) are inherited together. Markers that are in high LD can be assumed to be located near each other and a marker or haplotype that is in high LD with a genetic disorder can be assumed to be located near the gene that affects that disorder. The physical proximity of markers can be measured in family studies where it is called linkage or in population studies where it is called linkage disequilibrium.

LD mapping: population based gene mapping, which locates disorder genes by identifying regions of the genome where haplotypes or marker variation patterns are shared statistically more frequently among cases compared to healthy controls. This

method is based upon the assumption that many of the cases will have inherited an allele associated with the disorder from a common ancestor (IBD), and that this allele will be in LD with the disorder gene.

Locus: a specific position along a chromosome or DNA sequence. Depending upon context, a locus could be a gene, a marker, a chromosomal band or a specific sequence of one or more nucleotides.

Minor allele frequency (MAF): the population frequency of one of the alleles for a given polymorphism, which is equal or less than 50%. The sum of the MAF and the Major allele frequency equals one.

Markers: an identifiable DNA sequence that is variable (polymorphic) for different individuals within a population. These sequences facilitate the study of inheritance of a disorder or a gene. Such markers are used in mapping the order of genes along chromosomes and in following the inheritance of particular genes; genes closely linked to the marker or in LD with the marker will generally be inherited with it. Two types of markers are commonly used in genetic analysis, microsatellites and SNPs.

Microsatellite: DNA of eukaryotic cells comprising a repetitive, short sequence of DNA that is present as tandem repeats and in highly variable copy number, flanked by sequences unique to that locus.

Mutant sequence: a sequence that differs from one or more wild-type sequences. For example, a nucleic acid from a gene listed in Tables 5-7 containing a particular allele of a single nucleotide polymorphism may be a mutant sequence. In some cases, the individual carrying this allele has increased susceptibility toward the disorder, or condition of interest. In other cases, the mutant sequence might also refer to an allele that decreases the susceptibility toward a disorder or condition of interest and thus acts in a protective manner. The term mutation may also be used to describe a specific allele at a polymorphic locus.

Non-conservative variants: are those in which a change in one or more nucleotides in a given codon position results in a polypeptide sequence in which a given amino acid residue in a polypeptide has been replaced by a non-conservative amino acid

substitution. Non-conservative variants also include polypeptides comprising non-conservative amino acid substitutions.

Nucleic acid or polynucleotide: purine- and pyrimidine-containing polymers of any length, either polyribonucleotides or polydeoxyribonucleotide or mixed polyribo polydeoxyribonucleotides. This includes single-and double-stranded molecules, *i.e.*, DNA-DNA, DNA-RNA and RNA-RNA hybrids, as well as protein nucleic acids (PNA) formed by conjugating bases to an amino acid backbone. This also includes nucleic acids containing modified bases.

Nucleotide: consist of a ribose or deoxyribose sugar joined to a purine or pyrimidine base and to a phosphate group and that are the basic structural units of RNA and DNAF or its incorporation in DNA, nucleotides need to possess three phosphate esters but they are converted into monoesters in the process of incorporation.

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

Ortholog: denotes a gene or polypeptide obtained from one species that has homology to an analogous gene or polypeptide from a different species.

Paralog: denotes a gene or polypeptide obtained from a given species that has homology to a distinct gene or polypeptide from that same species.

Phenotype: any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to, a disorder.

Polymorphism: occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals at a single locus. A polymorphic site thus refers specifically to the locus at which the variation occurs. In some cases, an individual carrying a particular allele of a polymorphism has an increased or decreased susceptibility toward a disorder or condition of interest.

Portion and fragment: are synonymous. A portion as used with regard to a nucleic acid or polynucleotide refers to fragments of that nucleic acid or polynucleotide. The fragments can range in size from 8 nucleotides to all but one nucleotide of the entire gene sequence. Preferably, the fragments are at least about 8 to about 10 nucleotides in length; at least about 12 nucleotides in length; at least about 15 to about 20 nucleotides in length; at least about 25 nucleotides in length; or at least about 35 to about 55 nucleotides in length.

Probe or primer: refers to a nucleic acid or oligonucleotide that forms a hybrid structure with a sequence in a target region of a nucleic acid due to complementarity of the probe or primer sequence to at least one portion of the target region sequence.

Protein and polypeptide: are synonymous. Peptides are defined as fragments or portions of polypeptides, preferably fragments or portions having at least one functional activity (*e.g.*, proteolysis, adhesion, fusion, antigenic, or intracellular activity) as the complete polypeptide sequence.

Recombinant nucleic acids: nuclei acids which have been produced by recombinant DNA methodology, including those nucleic acids that are generated by procedures which rely upon a method of artificial replication, such as the polymerase chain reaction (PCR) and/or cloning into a vector using restriction enzymes. Portions of recombinant nucleic acids which code for polypeptides can be identified and isolated by, for example, the method of M. Jasin *et al.*, U.S. Patent No. 4,952,501.

Regulatory sequence: refers to a nucleic acid sequence that controls or regulates expression of structural genes when operably linked to those genes. These include, for example, the *lac* systems, the *trp* system, major operator and promoter regions of the phage lambda, the control region of fd coat protein and other sequences known to control the expression of genes in prokaryotic or eukaryotic cells. Regulatory sequences will vary depending on whether the vector is designed to express the operably linked gene in a prokaryotic or eukaryotic host, and may contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements and/or translational initiation and termination sites.

Sample: as used herein refers to a biological sample, such as, for example, tissue or fluid isolated from an individual or animal (including, without limitation, plasma, serum, cerebrospinal fluid, lymph, stools, tears, nails, hair, saliva, milk, pus, and tissue exudates and secretions) or from *in vitro* cell culture-constituents, as well as samples obtained from, for example, a laboratory procedure.

Schizophrenia: the term includes any type of schizophrenia such as paranoid schizophrenia, disorganized schizophrenia, catatonic schizophrenia, and a type of schizophrenia incapable to be classified.

Single nucleotide polymorphism (SNP): variation of a single nucleotide. This includes the replacement of one nucleotide by another and deletion or insertion of a single nucleotide. Typically, SNPs are biallelic markers although tri- and tetra-allelic markers also exist. For example, SNP A\C may comprise allele C or allele A (Tables 2 and 3). Thus, a nucleic acid molecule comprising SNP A\C may include a C or A at the polymorphic position. For a combination of SNPs, the term "haplotype" is used, *e.g.* the genotype of the SNPs in a single DNA strand that are linked to one another. In certain embodiments, the term "haplotype" is used to describe a combination of SNP alleles, *e.g.*, the alleles of the SNPs found together on a single DNA molecule. In specific embodiments, the SNPs in a haplotype are in linkage disequilibrium with one another.

Sequence-conservative: variants are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position (*i.e.*, silent mutation).

Substantially homologous: a nucleic acid or fragment thereof is substantially homologous to another if, when optimally aligned (with appropriate nucleotide insertions and/or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least 60% of the nucleotide bases, usually at least 70%, more usually at least 80%, preferably at least 90%, and more preferably at least 95-98% of the nucleotide bases. Alternatively, substantial homology exists when a nucleic acid or fragment thereof will hybridize, under selective hybridization conditions, to another nucleic acid (or a complementary strand thereof). Selectivity of hybridization exists when hybridization which is

substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% sequence identity over a stretch of at least about nine or more nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% (M. Kanehisa, 1984, NucL Acids Res. 11:203-213). The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least 14 nucleotides, usually at least 20 nucleotides, more usually at least 24 nucleotides, typically at least 28 nucleotides, more typically at least 32 nucleotides, and preferably at least 36 or more nucleotides.

Wild-type gene from Tables 5-7: refers to the reference sequence. The wild-type gene sequences from Tables 5-7 used to identify the variants (single nucleotide polymorphisms, alleles, and haplotypes) described in detail herein.

Technical and scientific terms used herein have the meanings commonly understood by one of ordinary skill in the art to which the present invention pertains, unless otherwise defined. Reference is made herein to various methodologies known to those of skill in the art. Publications and other materials setting forth such known methodologies to which reference is made are incorporated herein by reference in their entireties as though set forth in full. Standard reference works setting forth the general principles of recombinant DNA technology include J. Sambrook *et al.*, 1989, Molecular Cloning: A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; P.B. Kaufman *et al.*, (eds), 1995, Handbook of Molecular and Cellular Methods in Biology and Medicine, CRC Press, Boca Raton; M.J. McPherson (ed), 1991, Directed Mutagenesis: A Practical Approach, IRL Press, Oxford; J. Jones, 1992, Amino Acid and Peptide Synthesis, Oxford Science Publications, Oxford; B.M. Austen and O.M.R. Westwood, 1991, Protein Targeting and Secretion, IRL Press, Oxford; D.N Glover (ed), 1985, DNA Cloning, Volumes I and 11; M.J. Gait (ed), 1984, Oligonucleotide Synthesis; B.D. Hames and S.J. Higgins (eds), 1984, Nucleic Acid Hybridization; Quirke and Taylor (eds), 1991, PCR-A Practical Approach; Harries and Higgins (eds), 1984, Transcription and Translation; R.I. Freshney (ed), 1986, Animal Cell Culture; Immobilized Cells and Enzymes, 1986, IRL Press; Perbal, 1984, A Practical Guide to Molecular Cloning, J. H. Miller and M. P. Calos (eds), 1987, Gene Transfer Vectors for Mammalian Cells,

Cold Spring Harbor Laboratory Press; M.J. Bishop (ed), 1998, Guide to Human Genome Computing, 2d Ed., Academic Press, San Diego, CA; L.F. Peruski and A.H. Peruski, 1997, The Internet and the New Biology. Tools for Genomic and Molecular Research, American Society for Microbiology, Washington, D.C. Standard reference works setting forth the general principles of immunology include S. Sell, 1996, Immunology, Immunopathology & Immunity, 5th Ed., Appleton & Lange, Publ., Stamford, CT; D. Male *et al.*, 1996, Advanced Immunology, 3d Ed., Times Mirror Int'l Publishers Ltd., Publ., London; D.P. Stites and A.L Terr, 1991, Basic and Clinical Immunology, 7th Ed., Appleton & Lange, Publ., Norwalk, CT; and A.K. Abbas *et al.*, 1991, Cellular and Molecular Immunology, W. B. Saunders Co., Publ., Philadelphia, PA. Any suitable materials and/or methods known to those of skill can be utilized in carrying out the present invention; however, preferred materials and/or methods are described. Materials, reagents, and the like to which reference are made in the following description and examples are generally obtainable from commercial sources, and specific vendors are cited herein.

DETAILED DESCRIPTION OF THE INVENTION

Schizophrenia is one of the most severe and debilitating of the major psychiatric diseases. It usually starts in late adolescence or early adult life and often becomes chronic and disabling. Men and women are at equal risk of developing this illness; however, most males become ill between 16 and 25 years old, while females develop symptoms between 25 and 30. Schizophrenia is a devastating form of psychopathology, with prevalence worldwide of 0.5%-1%. Twin and adoption studies suggest that both genetic and environmental factors influence susceptibility (see, *e.g.*, Tsuang *et al.*, 1998 and 2005; Takahash *et al.*, 2005). Among first- degree relatives, the risk has been reported to vary from 6% in parents, to 10% in siblings, and to 13% in children of schizophrenic individuals; if one of the parents is also schizophrenic, the risk to siblings increases to 17%, and children of two schizophrenics have a risk of 46% of developing the illness (McGue *et al.*, 1991).

Reports of suggestive linkage to several loci have been published, including loci on chromosomes 3, 5, 6, 8, 10, 13, 20, 22 and the X chromosome (see, *e.g.* for

chromosomes 3p and 8p, Pulver *et al.*, 1995; for chromosomes 5q, 6p and 8p, Kendler *et al.*, 1999; for chromosomes 5q, 6p, 8p, 20p and 22q, Hovatta *et al.*, 1998; for chromosome 10, Skaub, *et al.*, 1998; for chromosome 13, Lin *et al.*, 1995; for chromosome 22, Gill, *et al.*, 1996; and for the X chromosome, Milunsky *et al.*, 1999).

The aggregation of schizophrenia disorder in families, the evidence from twin and adoption studies, and the lack of variation in incidence worldwide, indicate that schizophrenia is primarily a genetic condition, although environmental risk factors are also involved at some level as necessary, sufficient, or interactive causes.

As there are currently no cures for schizophrenia, the objective of treatment is to reduce the severity of the symptoms, if possible to the point of remission. Due to the similarities in symptoms, schizophrenia and bipolar disorder are often treated with some of the same medicaments. Both diseases are often treated with antipsychotics and neuroleptics.

For schizophrenia, for example, antipsychotic medications are the most common and most valuable treatments. There are four main classes of antipsychotic drugs which are commonly prescribed for schizophrenia. The first, neuroleptics, exemplified by chlorpromazine (Thorazine), has revolutionized the treatment of schizophrenic patients by reducing positive (psychotic) symptoms and preventing their recurrence. Patients receiving chlorpromazine have been able to leave mental hospitals and live in community programs or their own homes. But these drugs are far from ideal. Some 20% to 30% of patients do not respond to them at all, and others eventually relapse. These drugs were named neuroleptics because they produce serious neurological side effects, including rigidity and tremors in the arms and legs, muscle spasms, abnormal body movements, and akathisia (restless pacing and fidgeting). These side effects are so troublesome that many patients simply refuse to take the drugs. Besides, neuroleptics do not improve the so-called negative symptoms of schizophrenia and the side effects may even exacerbate these symptoms. Thus, despite the clear beneficial effects of neuroleptics, even some patients who have a good short-term response will ultimately deteriorate in overall functioning.

The well known deficiencies in the standard neuroleptics have stimulated a search for new treatments and have led to a new class of drugs termed atypical

neuroleptics. The first atypical neuroleptic, Clozapine, is effective for about one third of patients who do not respond to standard neuroleptics. It seems to reduce negative as well as positive symptoms, or at least exacerbates negative symptoms less than standard neuroleptics do. Moreover, it has beneficial effects on overall functioning and may reduce the chance of suicide in schizophrenic patients. It does not produce the troubling neurological symptoms of the standard neuroleptics, or raise blood levels of the hormone prolactin, excess of which may cause menstrual irregularities and infertility in women, and impotence or breast enlargement in men. Many patients who cannot tolerate standard neuroleptics have been able to take clozapine. However, clozapine has serious limitations. It was originally withdrawn from the market because it can cause agranulocytosis, a potentially lethal inability to produce white blood cells. Agranulocytosis remains a threat that requires careful monitoring and periodic blood tests. Clozapine can also cause seizures and other disturbing side effects (*e.g.*, drowsiness, lowered blood pressure, drooling, bed-wetting, and weight gain). Thus it is usually taken only by patients who do not respond to other drugs.

Researchers have developed a third class of antipsychotic drugs that have the virtues of clozapine without its drawbacks. One of these drugs is risperidone (Risperdal). Early studies suggest that it is as effective as standard neuroleptic drugs for positive symptoms and may be somewhat more effective for negative symptoms. It produces more neurological side effects than clozapine but fewer than standard neuroleptics. However, it also raises prolactin levels. Risperidone is now prescribed for a broad range of psychotic patients, and many clinicians seem to use it before clozapine for patients who do not respond to standard drugs, because they regard it as safer. Another new drug is Olanzapine (Zyprexa) which is at least as effective as standard drugs for positive symptoms and more effective for negative symptoms. It has few neurological side effects at ordinary clinical doses, and it does not significantly raise prolactin levels. Although it does not produce most of clozapine's most troubling side effects, including agranulocytosis, some patients taking Olanzapine may become sedated or dizzy, develop dry mouth, or gain weight. In rare cases, liver function tests become transiently abnormal.

Outcome studies in schizophrenia are usually based on hospital treatment studies and may not be representative of the population of schizophrenia patients. At the extremes of outcome, 20% of patients seem to recover completely after one episode of psychosis, whereas 14-19% of patients develop a chronic unremitting psychosis and never fully recover. In general, clinical outcome at five years seems to follow the rule of thirds: with about 35% of patients in the poor outcome category; 36% in the good outcome category, and the remainder with intermediate outcome. Prognosis in schizophrenia does not seem to worsen after five years.

Whatever the reasons, there is increasing evidence that leaving schizophrenia untreated for long periods early in course of the illness may negatively affect the outcome. However, the use of drugs is often delayed for patients experiencing a first episode of the illness. The patients may not realize that they are ill, or they may be afraid to seek help; family members sometimes hope the problem will simply disappear or cannot persuade the patient to seek treatment; clinicians may hesitate to prescribe antipsychotic medications when the diagnosis is uncertain because of potential side effects. Indeed, at the first manifestation of the disease, schizophrenia is difficult to distinguish from bipolar manic-depressive disorders, severe depression, drug-related disorders, and stress-related disorders. Since the optimum treatments differ among these diseases, the long term prognosis of the disorder also differs the beginning of the treatment.

All the known molecules used for the treatment of schizophrenia have side effects and act only against the symptoms of the disease. There is a strong need for new molecules without associated side effects and directed against targets which are involved in the causal mechanisms of schizophrenia. Therefore, tools facilitating the discovery and characterization of these targets are necessary and useful.

Thus, while there are a number of genetic studies and markers known, schizophrenia is not well understood. There is also a continuing need in the medical arts for genetic markers of schizophrenia and guidance for the use of such markers. The knowledge of these markers will allow researchers to understand the etiology of schizophrenia and could lead to drugs and medications which are directed against the cause of the disease, not just against their symptoms. Compositions and methods of the invention are useful for predicting the propensity for schizophrenia in

humans. Additionally, compositions and methods of the invention are useful for predicting the propensity for schizophrenia. Additionally, compositions and methods of the invention are useful for indicating possible early therapeutic intervention to prevent or to lessen the effects of schizophrenia. The present invention fulfills this need and provides further related advantages.

Genome wide association study to construct a GeneMap for schizophrenia

The present invention is based on the discovery of genes associated with schizophrenia. In the preferred embodiment, disorder-associated loci (candidate regions; Table 1) are therefore identified by the statistically significant differences in allele frequencies between the cases and the controls. For the purpose of the present invention, 27 candidate regions are identified.

The invention provides a method for the discovery of genes associated with schizophrenia and the construction of a GeneMap for schizophrenia in a human population, comprising the following steps (see Figure 1 and Example section herein):

Step 1: Recruit cases and controls

In the preferred embodiment, 500 cases along with 500 control individuals are recruited from the Quebec Founder Population (QFP).

In another embodiment, 470 cases (311 males and 159 females) along with 469 control individuals (311 males and 158 females) are recruited from the Quebec Founder Population (QFP).

In another embodiment, the present invention is performed as a whole or partially with DNA samples from individuals of another founder population than the Quebec population or from the general population.

Step 2: DNA extraction and dosage

Any sample comprising cells or nucleic acids from patients or controls may be used. Preferred samples are those easily obtained from the patient or control. Such

samples include, but are not limited to blood, peripheral lymphocytes, buccal swabs, epithelial cell swabs, nails, hair, bronchoalveolar lavage fluid, sputum, or other body fluid or tissue obtained from an individual.

In the preferred embodiment, DNA is extracted from such samples in the quantity and quality necessary to perform the invention using conventional DNA extraction and dosage techniques. The present invention is not linked to any DNA extraction or dosage platform in particular.

The extracted DNA from case and control samples from recruited individuals is pooled together in various pools. Pools are designed to segregate cases from controls, and males from females.

In the preferred embodiment, proband pools, consist of DNA extracted from recruited cases and control pools consist of DNA extracted from control individuals. The probands are then segregated according to gender.

In the preferred embodiment, the samples were collected as 470 cases (311 males and 159 females) and 469 controls (311 males and 158 females). The DNA extracted from cases and control samples was pooled into various case and control pools. Pools were designed to isolate cases from controls, and males from females. Pools of around 50 randomized individuals were created, resulting in 6 pools of affected males and 3 pools of affected females. In the same manner, 6 male control pools and 3 female control pools were created.

Step 3: Genotype the proband and control pools

In the preferred embodiment, assay specific and/or locus-specific and/or allele-specific oligonucleotides for every SNP marker of the present invention (Tables 2, 3 and 4) are organized onto one or more arrays. The genotype at each SNP locus is revealed by hybridizing short PCR fragments comprising each SNP locus onto these arrays. The arrays permit a high-throughput genome wide association study using DNA samples from individuals of the Quebec founder population. Such assay-specific and/or locus-specific and/or allele-specific oligonucleotides necessary for scoring each SNP of the present invention are preferably organized onto a solid

support. Such supports can be arrayed on wafers, glass slides, beads or any other type of solid support.

In another embodiment, the assay-specific and/or locus-specific and/or allele-specific oligonucleotides are not organized onto a solid support but are still used as a whole, in panels or one by one. The present invention is therefore not linked to any genotyping platform in particular.

In another embodiment, one or more portions of the SNPs maps (publicly available maps, proprietary maps from Perlegen Sciences, Inc. (Mountain View, CA, USA), and our own proprietary QLDM map) are used to screen the whole genome, a subset of chromosomes, a chromosome, a subset of genomic regions or a single genomic region.

Step 4: Exclude the markers that did not pass the quality control of the assay.

Preferably, the quality controls consist of, but are not limited to, the following criteria: eliminate SNPs that are non-polymorphic in the Quebec founder population or have ≤ 10% minor allele frequency (MAF).

Step 5: Perform the genetic analysis on the results obtained

In the preferred embodiment, genetic analysis is performed on all the genotypes from step 3.

In another embodiment, genetic analysis is performed on a total of 248,535 SNPs.

In one embodiment, the data analysis compares the relative fluorescence intensities of features corresponding to the reference allele of a given SNP with those corresponding to the alternate allele, to calculate a p-hat value. The latter is proportional to the fluorescence signal from perfect match features for the reference allele divided by the sum of fluorescence signals from perfect match features for the reference plus the alternate alleles. P-hat assumes values close to 1 (typically 0.9) for pure reference samples and close to 0 (typically 0.1) for pure alternate samples, and can be used as a measured estimate of the reference allele frequency of a SNP in a DNA pool. The difference between case and control pools, delta p-hat, is

calculated using the weighted average of case and control p-hats. Delta p-hat is a reliable estimate of the allele frequency difference between the cases and controls.

In yet another embodiment, the data is analyzed according to the p-hat value obtained from the previous embodiment on each pool. Single marker P values are calculated for all markers within the genome wide scan map as described in example 3 herein, using the p-hat value.

In another embodiment, the combined P values across multi-marker sliding windows are calculated after the method of Fisher (described in Example 3 herein).

In yet another embodiment, the exact P values for the observed case and control combination are calculated as the number of combinations with a value greater than or equal to the actual observed value, divided by 48620 combinations (see Example 3 herein, *Permutation Test for Exact P values).*

Step 6: Fine Mapping

During this step, the candidate regions that were identified by step 5 are further mapped and confirmed for the purpose of refinement and validation.

In the preferred embodiment, the cases and controls are individually genotyped to confirm the candidate regions. The confirmed candidate regions are processed by fine mapping to refine the candidate regions.

In the preferred embodiment, this fine mapping is performed with a density of genetic markers higher than in the genome wide scan (step 3) using any genotyping platform available in the art. Such fine mapping can also be performed with fewer genetic markers than in the GWS. Such fine mapping can be, but is not limited to, typing the allele via an allele-specific elongation assay that is then ligated to a locus-specific oligonucleotide. Such assays can be performed directly on the genomic DNA at a highly multiplex level and the products can be amplified using universal oligonucleotides. For each candidate region, the density of genetic markers can be, but is not limited to, a set of SNP markers with an average inter-marker distance of 1-4 Kb distributed over about 400 Kb to 1 Mb, roughly centered at the highest point

of the GWS association. The preferred samples are those obtained from schizophrenia samples including the ones used for the GWS.

In the preferred embodiment, the genetic analysis of the results obtained using haplotype information (available after confirmatory mapping of individual samples, see Examples section herein) as well as single-marker association (as performed as in step 5, described herein) are performed as described herein (step 5 and Example section). The candidate regions that are validated and confirmed after this analysis proceed to a gene mining step described in Example 5, herein, to characterize their marker and genetic content.

Step 7: SNP and DNA polymorphism discovery

In the preferred embodiment, all the candidate genes and regions identified in step 6 are sequenced for polymorphism identification.

In another embodiment, the entire region, including all introns, is sequenced to identify all polymorphisms.

In yet another embodiment, the candidate genes are prioritized for sequencing, and only functional gene elements (promoters, exons and splice sites) are sequenced.

In yet another embodiment, previously identified polymorphisms in the candidate regions can also be used. For example, SNPs from dbSNP, Perlegen Sciences, Inc., or others can also be used rather than resequencing the candidate regions to identify polymorphisms.

The discovery of SNPs and DNA polymorphisms generally comprises a step consisting of determining the major haplotypes in the region to be sequenced. The preferred samples are selected according to which haplotypes contribute to the association signal observed in the region to be sequenced. The purpose is to select a set of samples that covers all the major haplotypes in the given region. Each major haplotype is preferably analyzed in at least a few individuals.

Any analytical procedure may be used to detect the presence or absence of variant nucleotides at one or more polymorphic positions of the invention. In general, the detection of allelic variation requires a mutation discrimination technique, optionally

an amplification reaction and optionally a signal generation system. Any means of mutation detection or discrimination may be used. For instance, DNA sequencing, scanning methods, hybridization, extension based methods, incorporation based methods, restriction enzyme-based methods and ligation-based methods may be used in the methods of the invention.

Sequencing methods include, but are not limited to, direct sequencing, and sequencing by hybridization. Scanning methods include, but are not limited to, protein truncation test (PTT), single-strand conformation polymorphism analysis (SSCP), denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE), cleavage, heteroduplex analysis, chemical mismatch cleavage (CMC), and enzymatic mismatch cleavage. Hybridization-based methods of detection include, but are not limited to, solid phase hybridization such as dot blots, multiple allele specific diagnostic assay (MASDA), reverse dot blots, and oligonucleotide arrays (DNA Chips). Solution phase hybridization amplification methods may also be used, such as Taqman. Extension based methods include, but are not limited to, amplification refraction mutation systems (ARMS), amplification refractory mutation systems (ALEX), and competitive oligonucleotide priming systems (COPS). Incorporation based methods include, but are not limited to, mini-sequencing and arrayed primer extension (APEX). Restriction enzyme-based detection systems include, but are not limited to, restriction site generating PCR. Lastly, ligation based detection methods include, but are not limited to, oligonucleotide ligation assays (OLA). Signal generation or detection systems that may be used in the methods of the invention include, but are not limited to, fluorescence methods such as fluorescence resonance energy transfer (FRET), fluorescence quenching, fluorescence polarization as well as other chemiluminescence, electrochemiluminescence, Raman, radioactivity, colometric methods, hybridization protection assays and mass spectrometry methods. Further amplification methods include, but are not limited to self sustained replication (SSR), nucleic acid sequence based amplification (NASBA), ligase chain reaction (LCR), strand displacement amplification (SDA) and branched DNA (B-DNA).

Step 8: Ultrafine Mapping

This step further maps the candidate regions and genes confirmed in the previous step to identify and validate the responsible polymorphisms associated with schizophrenia in the human population.

In a preferred embodiment, the discovered SNPs and polymorphisms of step 7 are ultrafine mapped at a higher density of markers than the fine mapping described herein using the same technology described in step 6.

Step 9: GeneMap construction

The confirmed variations in DNA (including both genic and non-genic regions) are used to build a GeneMap for schizophrenia. The gene content of this GeneMap is described in more detail below. Such GeneMap can be used for other methods of the invention comprising the diagnostic methods described herein, the susceptibility to schizophrenia, the response to a particular drug, the efficacy of a particular drug, the screening methods described herein and the treatment methods described herein.

As is evident to one of ordinary skill in the art, all of the above steps or the steps of Figure 1 do not need to be performed, or performed in a given order to practice or use the SNPs, genomic regions, genes, proteins, *etc.* in the methods of the invention.

Genes from the GeneMap

In the preferred embodiment the GeneMap consists of genes and targets, in a variety of combinations, identified from the candidate regions listed in Table 1. In the preferred embodiment, all genes from Tables 5-7 are present in the GeneMap. In a more preferred embodiment, all genes from Tables 5-7, as well as proteins involved in their respective signaling pathways, are present in the GeneMap.

Nucleic acid sequences

The nucleic acid sequences of the present invention may be derived from a variety of sources including DNA, cDNA, synthetic DNA, synthetic RNA, derivatives, mimetics or combinations thereof. Such sequences may comprise genomic DNA, which may or may not include naturally occurring introns, genic regions, nongenic regions, and regulatory regions. Moreover, such genomic DNA may be obtained in association with promoter regions or poly (A) sequences. The sequences, genomic DNA, or cDNA may be obtained in any of several ways. Genomic DNA can be extracted and purified from suitable cells by means well known in the art. Alternatively, mRNA can be isolated from a cell and used to produce cDNA by reverse transcription or other means. The nucleic acids described herein are used certain embodiments of in the methods of the present invention for production of RNA, proteins or polypeptides, through incorporation into cells, tissues, or organisms. In one embodiment, DNA containing all or part of the coding sequence for the genes described in Tables 5-7, or the SNP markers described in Tables 2 and 3, is incorporated into a vector for expression of the encoded polypeptide in suitable host cells. The invention also comprises the use of the nucleotide sequence of the nucleic acids of this invention to identify DNA probes for the genes described in Tables 5-7 or the SNP markers described in Tables 2 and 3, PCR primers to amplify the genes described in Tables 5-7 or the SNP markers described in Tables 2 and 3, nucleotide polymorphisms in the genes described in Tables 5-7, and regulatory elements of the genes described in Tables 5-7. The nucleic acids of the present invention find use as primers and templates for the recombinant production of schizophrenia-associated peptides or polypeptides, for chromosome and gene mapping, to provide antisense sequences, for tissue distribution studies, to locate and obtain full length genes, to identify and obtain homologous sequences (wild-type and mutants), and in diagnostic applications.

Antisense oligonucleotides

In a particular embodiment of the invention, an antisense nucleic acid or oligonucleotide is wholly or partially complementary to, and can hybridize with, a target nucleic acid (either DNA or RNA) having the sequence of SEQ ID NO:1, NO:3 or any SEQ ID from Tables 2-7. For example, an antisense nucleic acid or oligonucleotide comprising 16 nucleotides can be sufficient to inhibit expression of at

least one gene from Tables 5-7. Alternatively, an antisense nucleic acid or oligonucleotide can be complementary to 5' or 3' untranslated regions, or can overlap the translation initiation codon (5' untranslated and translated regions) of at least one gene from Tables 5-7, or its functional equivalent. In another embodiment, the antisense nucleic acid is wholly or partially complementary to, and can hybridize with, a target nucleic acid that encodes a polypeptide from a gene described in Tables 5-7.

In addition, oligonucleotides can be constructed which will bind to duplex nucleic acid (*i.e.*, DNA:DNA or DNA:RNA), to form a stable triple helix containing or triplex nucleic acid. Such triplex oligonucleotides can inhibit transcription and/or expression of a gene from Tables 5-7, or its functional equivalent (M.D. Frank-Kamenetskii *et al.*, 1995). Triplex oligonucleotides are constructed using the basepairing rules of triple helix formation and the nucleotide sequence of the genes described in Tables 5-7.

The present invention encompasses methods of using oligonucleotides in antisense inhibition of the function of the genes from Tables 5-7. In the context of this invention, the term "oligonucleotide" refers to naturally-occurring species or synthetic species formed from naturally-occurring subunits or their close homologs. The term may also refer to moieties that function similarly to oligonucleotides, but have non-naturally-occurring portions. Thus, oligonucleotides may have altered sugar moieties or inter-sugar linkages. Exemplary among these are phosphorothioate and other sulfur containing species which are known in the art. In preferred embodiments, at least one of the phosphodiester bonds of the oligonucleotide has been substituted with a structure that functions to enhance the ability of the compositions to penetrate into the region of cells where the RNA whose activity is to be modulated is located. It is preferred that such substitutions comprise phosphorothioate bonds, methyl phosphonate bonds, or short chain alkyl or cycloalkyl structures. In accordance with other preferred embodiments, the phosphodiester bonds are substituted with structures which are, at once, substantially non-ionic and non-chiral, or with structures which are chiral and enantiomerically specific. Persons of ordinary skill in the art will be able to select other linkages for use in the practice of the invention. Oligonucleotides may also include species that include at least some modified base forms. Thus, purines and pyrimidines other than those normally found in nature may

be so employed. Similarly, modifications on the furanosyl portions of the nucleotide subunits may also be effected, as long as the essential tenets of this invention are adhered to. Examples of such modifications are 2'-O-alkyl- and 2'-halogen-substituted nucleotides. Some non-limiting examples of modifications at the 2' position of sugar moieties which are useful in the present invention include OH, SH, SCH3, F, OCH3, OCN, O(CH2), NH2 and O(CH2)n CH3, where n is from 1 to about 10. Such oligonucleotides are functionally interchangeable with natural oligonucleotides or synthesized oligonucleotides, which have one or more differences from the natural structure. All such analogs are comprehended by this invention so long as they function effectively to hybridize with at least one gene from Tables 3 or 4's DNA or RNA to inhibit the function thereof.

The oligonucleotides in accordance with this invention preferably comprise from about 3 to about 50 subunits. It is more preferred that such oligonucleotides and analogs comprise from about 8 to about 25 subunits and still more preferred to have from about 12 to about 20 subunits. As defined herein, a "subunit" is a base and sugar combination suitably bound to adjacent subunits through phosphodiester or other bonds. Antisense nucleic acids or oligonulcleotides can be produced by standard techniques (see, *e.g.*, Shewmaker *et al.*, U.S. Patent No. 6,107,065). The oligonucleotides used in accordance with this invention may be conveniently and routinely made through the well-known technique of solid phase synthesis. Any other means for such synthesis may also be employed; however, the actual synthesis of the oligonucleotides is well within the abilities of the practitioner. It is also well known to prepare other oligonucleotide such as phosphorothioates and alkylated derivatives.

The oligonucleotides of this invention are designed to be hybridizable with RNA (*e.g.*, mRNA) or DNA from genes described in Tables 5-7. For example, an oligonucleotide (*e.g.*, DNA oligonucleotide) that hybridizes to mRNA from a gene described in Tables 5-7 can be used to target the mRNA for RnaseH digestion. Alternatively, an oligonucleotide that can hybridize to the translation initiation site of the mRNA of a gene described in Tables 5-7 can be used to prevent translation of the mRNA. In another approach, oligonucleotides that bind to the double-stranded DNA of a gene from Tables 5-7 can be administered. Such oligonucleotides can form a triplex

construct and inhibit the transcription of the DNA encoding polypeptides of the genes described in Tables 5-7. Triple helix pairing prevents the double helix from opening sufficiently to allow the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described (see, *e.g.*, J.E. Gee *et al.*, 1994, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY).

As non-limiting examples, antisense oligonucleotides may be targeted to hybridize to the following regions: mRNA cap region; translation initiation site; translational termination site; transcription initiation site; transcription termination site; polyadenylation signal; 3' untranslated region; 5' untranslated region; 5' coding region; mid coding region; and 3' coding region. Preferably, the complementary oligonucleotide is designed to hybridize to the most unique 5' sequence of a gene described in Tables 5-7, including any of about 15-35 nucleotides spanning the 5' coding sequence. In accordance with the present invention, the antisense oligonucleotide can be synthesized, formulated as a pharmaceutical composition, and administered to a subject. The synthesis and utilization of antisense and triplex oligonucleotides have been previously described (*e.g.*, Simon *et al.*, 1999; Barre *et al.*, 2000; Elez *et al.*, 2000; Sauter *et al.*, 2000).

Alternatively, expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express nucleic acid sequence that is complementary to the nucleic acid sequence encoding a polypeptide from the genes described in Tables 5-7. These techniques are described both in Sambrook *et al.*, 1989 and in Ausubel *et al.*, 1992. For example, expression of at least one gene from Tables 5-7 can be inhibited by transforming a cell or tissue with an expression vector that expresses high levels of untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a nonreplicating vector, and even longer if appropriate replication elements are included in the vector system. Various assays may be used

to test the ability of gene-specific antisense oligonucleotides to inhibit the expression of at least one gene from Tables 5-7. For example, mRNA levels of the genes described in Tables 5-7 can be assessed by Northern blot analysis (Sambrook *et al.*, 1989; Ausubel *et al.*, 1992; J.C. Alwine *et al.* 1977; I.M. Bird, 1998), quantitative or semi-quantitative RT-PCR analysis (see, *e.g.*, W.M. Freeman *et al.*, 1999; Ren *et al.*, 1998; J.M. Cale *et al.*, 1998), or in situ hybridization (reviewed by A.K. Raap, 1998). Alternatively, antisense oligonucleotides may be assessed by measuring levels of the polypeptide from the genes described in Tables 5-7, *e.g.*, by western blot analysis, indirect immunofluorescence and immunoprecipitation techniques (see, *e.g.*, J.M. Walker, 1998, Protein Protocols on CD-ROM, Humana Press, Totowa, NJ). Any other means for such detection may also be employed, and is well within the abilities of the practitioner.

Mapping Technologies

The present invention includes various methods which employ mapping technologies to map SNPs and polymorphisms. For purpose of clarity, this section comprises, but is not limited to, the description of mapping technologies that can be utilized to achieve the embodiments described herein. Mapping technologies may be based on amplification methods, restriction enzyme cleavage methods, hybridization methods, sequencing methods, and cleavage methods using agents.

Amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990), transcriptional amplification system (Kwoh *et al.*, 1989), Q-Beta Replicase (Lizardi *et al.*, 1988), isothermal amplification (*e.g.* Dean *et al.*, 2002; and Hafner *et al.*, 2001), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of ordinary skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low number.

Restriction enzyme cleavage methods include: isolating sample and control DNA, amplification (optional), digestion with one or more restriction endonucleases, determination of fragment length sizes by gel electrophoresis and comparing samples and controls. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, sequence specific

ribozymes (see, *e.g.*, U.S. Pat. No. 5,498,531) or DNAzyme (*e.g.* U.S. Pat. No. 5,807,718) can be used to score for the presence of specific mutations by development or loss of a ribozyme or DNAzyme cleavage site.

SNPs and SNP maps of the invention can be identified or generated by hybridizing sample nucleic acids, *e.g.*, DNA or RNA, to high density arrays or bead arrays containing oligonucleotide probes corresponding to the SNPs of Table 1 (see the Affymetrix arrays and Illumina bead sets at www.affymetrix.com and www.illumina.com and see Cronin *et al.*, 1996; or Kozal *et al.*, 1996).

A variety of sequencing reactions known in the art can be used to directly sequence nucleic acids for the presence or the absence of one or more SNPs of Table 1. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert (1977) or Sanger (1977). It is also contemplated that any of a variety of automated sequencing procedures can be utilized, including sequencing by mass spectrometry (see, *e.g.* PCT International Publication No. WO 94/16101; Cohen *et al.*, 1996; and Griffin *et al.*, 1993), real-time pyrophosphate sequencing method (Ronaghi *et al.*, 1998; and Permutt *et al.*, 2001) and sequencing by hybridization (see *e.g.* Drmanac *et al.*, 2002).

Other methods of detecting SNPs include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA, DNA/DNA or RNA/DNA heteroduplexes (Myers *et al.*, 1985). In general, the technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing a wild-type sequence with potentially mutant RNA or DNA obtained from a sample. The double-stranded duplexes are treated with an agent who cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of a mutation or SNP. (see, for

example, Cotton *et al.*, 1988; and Saleeba *et al.*, 1992). In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping SNPs. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches (Hsu *et al.*, 1994). Other examples include, but are not limited to, the MutHLS enzyme complex of E. coli (Smith and Modrich Proc. 1996) and Cel 1 from the celery (Kulinski *et al.*, 2000) both cleave the DNA at various mismatches. According to an exemplary embodiment, a probe based on a polymorphic site corresponding to a SNP of Tables 2 and 3 is hybridized to a cDNA or other DNA product from a test cell or cells. The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039. Alternatively, the screen can be performed *in vivo* following the insertion of the heteroduplexes in an appropriate vector. The whole procedure is known to those ordinary skilled in the art and is referred to as mismatch repair detection (see *e.g.* Fakhrai-Rad *et al.*, 2004).

In other embodiments, alterations in electrophoretic mobility can be used to identify SNPs in a sample. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.*, 1989; Cotton *et al.*, 1993; and Hayashi 1992). Single-stranded DNA fragments of case and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence. The resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Kee *et al.*, 1991).

In yet another embodiment, the movement of mutant or wild-type fragments in a polyacrylamide gel containing a gradient of denaturant is assayed using denaturing

gradient gel electrophoresis (DGGE) (Myers *et al.*, 1985). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum *et al.*, 1987). In another embodiment, the mutant fragment is detected using denaturing HPLC (see *e.g.* Hoogendoorn *et al.*, 2000).

Examples of other techniques for detecting SNPs include, but are not limited to, selective oligonucleotide hybridization, selective amplification, selective primer extension, selective ligation, single-base extension, selective termination of extension or invasive cleavage assay. For example, oligonucleotide primers may be prepared in which the SNP is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.*, 1986; Saiki *et al.*, 1989). Such oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA. Alternatively, the amplification, the allele-specific hybridization and the detection can be done in a single assay following the principle of the 5' nuclease assay (*e.g.* see Livak *et al.*, 1995). For example, the associated allele, a particular allele of a polymorphic locus, or the like is amplified by PCR in the presence of both allele-specific oligonucleotides, each specific for one or the other allele. Each probe has a different fluorescent dye at the 5' end and a quencher at the 3' end. During PCR, if one or the other or both allele-specific oligonucleotides are hybridized to the template, the Taq polymerase via its 5' exonuclease activity will release the corresponding dyes. The latter will thus reveal the genotype of the amplified product.

Hybridization assays may also be carried out with a temperature gradient following the principle of dynamic allele-specific hybridization or like *e.g.* Jobs *et al.*, (2003); and Bourgeois and Labuda, (2004). For example, the hybridization is done using one of the two allele-specific oligonucleotides labeled with a fluorescent dye, an intercalating quencher under a gradually increasing temperature. At low temperature, the probe is hybridized to both the mismatched and full-matched template. The probe melts at a lower temperature when hybridized to the template with a

mismatch. The release of the probe is captured by an emission of the fluorescent dye, away from the quencher. The probe melts at a higher temperature when hybridized to the template with no mismatch. The temperature-dependent fluorescence signals therefore indicate the absence or presence of an associated allele, a particular allele of a polymorphic locus, or the like (*e.g.* Jobs *et al.*, 2003). Alternatively, the hybridization is done under a gradually decreasing temperature. In this case, both allele-specific oligonucleotides are hybridized to the template competitively. At high temperature none of the two probes are hybridized. Once the optimal temperature of the full-matched probe is reached, it hybridizes and leaves no target for the mismatched probe (*e.g.* Bourgeois and Labuda, 2004). In the latter case, if the allele-specific probes are differently labeled, then they are hybridized to a single PCR-amplified target. If the probes are labeled with the same dye, then the probe cocktail is hybridized to twice to identical templates with only one labeled probes, different in the two cocktails, in the presence of the unlabeled competitive probe.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the present invention. Oligonucleotides used as primers for specific amplification may carry the associated allele, a particular allele of a polymorphic locus, or the like, also referred to as "mutation" of interest in the center of the molecule, so that amplification depends on differential hybridization (Gibbs *et al.*, 1989) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner, 1993). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.*, 1992). It is anticipated that in certain embodiments, amplification may also be performed using Taq ligase for amplification (Barany, 1991). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known associated allele, a particular allele of a polymorphic locus, or the like at a specific site by looking for the presence or absence of amplification. The products of such an oligonucleotide ligation assay can also be detected by means of gel electrophoresis. Furthermore, the oligonucleotides may contain universal tags used in PCR amplification and zip code tags that are different

for each allele. The zip code tags are used to isolate a specific, labeled oligonucleotide that may contain a mobility modifier (*e.g.* Grossman *et al.*, 1994).

In yet another alternative, allele-specific elongation followed by ligation will form a template for PCR amplification. In such cases, elongation will occur only if there is a perfect match at the 3' end of the allele-specific oligonucleotide using a DNA polymerase. This reaction is performed directly on the genomic DNA and the extension/ligation products are amplified by PCR. To this end, the oligonucleotides contain universal tags allowing amplification at a high multiplex level and a zip code for SNP identification. The PCR tags are designed in such a way that the two alleles of a SNP are amplified by different forward primers, each having a different dye. The zip code tags are the same for both alleles of a given SNPs and they are used for hybridization of the PCR-amplified products to oligonucleotides bound to a solid support, chip, bead array or like. For an example of the procedure, see Fan *et al.* (Cold Spring Harbor Symposia on Quantitative Biology, Vol. LXVIII, pp. 69-78 2003).

Another alternative includes the single-base extension/ligation assay using a molecular inversion probe, consisting of a single, long oligonucleotide (see *e.g.* Hardenbol *et al.*, 2003). In such an embodiment, the oligonucleotide hybridizes on both side of the SNP locus directly on the genomic DNA, leaving a one-base gap at the SNP locus. The gap-filling, one-base extension/ligation is performed in four tubes, each having a different dNTP. Following this reaction, the oligonucleotide is circularized whereas unreactive, linear oligonucleotides are degraded using an exonuclease such as exonuclease I of E. coli. The circular oligonucleotides are then linearized and the products are amplified and labeled using universal tags on the oligonucleotides. The original oligonucleotide also contains a SNP-specific zip code allowing hybridization to oligonucleotides bound to a solid support, chip, and bead array or like. This reaction can be performed at a high multiplexed level.

In another alternative, the associated allele, a particular allele of a polymorphic locus, or the like is scored by single-base extension (see *e.g.* U.S. Pat. No. 5,888,819). The template is first amplified by PCR. The extension oligonucleotide is then hybridized next to the SNP locus and the extension reaction is performed using a thermostable polymerase such as ThermoSequenase (GE Healthcare) in the presence of labeled ddNTPs. This reaction can therefore be cycled several times.

The identity of the labeled ddNTP incorporated will reveal the genotype at the SNP locus. The labeled products can be detected by means of gel electrophoresis, fluorescence polarization (*e.g.* Chen *et al.*, 1999) or by hybridization to oligonucleotides bound to a solid support, chip, and bead array or like. In the latter case, the extension oligonucleotide will contain a SNP-specific zip code tag.

In yet another alternative, a SNP is scored by selective termination of extension. The template is first amplified by PCR and the extension oligonucleotide hybridizes in vicinity to the SNP locus, close to but not necessarily adjacent to it. The extension reaction is carried out using a thermostable polymerase such as Thermo Sequenase (GE Healthcare) in the presence of a mix of dNTPs and at least one ddNTP. The latter has to terminate the extension at one of the allele of the interrogated SNP, but not both such that the two alleles will generate extension products of different sizes. The extension product can then be detected by means of gel electrophoresis, in which case the extension products need to be labeled, or by mass spectrometry (see *e.g.* Storm *et al.*, 2003).

In another alternative, SNPs are detected using an invasive cleavage assay (see U.S. Pat. No. 6,090,543). There are five oligonucleotides per SNP to interrogate but these are used in a two step-reaction. During the primary reaction, three of the designed oligonucleotides are first hybridized directly to the genomic DNA. One of them is locus-specific and hybridizes up to the SNP locus (the pairing of the 3' base at the SNP locus is not necessary). There are two allele-specific oligonucleotides that hybridize in tandem to the locus-specific probe but also contain a 5' flap that is specific for each allele of the SNP. Depending upon hybridization of the allele-specific oligonucleotides at the base of the SNP locus, this creates a structure that is recognized by a cleavase enzyme (U.S. Pat. No. 6,090,606) and the allele-specific flap is released. During the secondary reaction, the flap fragments hybridize to a specific cassette to recreate the same structure as above except that the cleavage will release a small DNA fragment labeled with a fluorescent dye that can be detected using regular fluorescence detector. In the cassette, the emission of the dye is inhibited by a quencher.

Methods to identify agents that modulate the expression of a nucleic acid encoding a gene involved in schizophrenia.

The present invention provides methods for identifying agents that modulate the expression of a nucleic acid encoding a gene from Tables 5-7. Such methods may utilize any available means of monitoring for changes in the expression level of the nucleic acids of the invention. As used herein, an agent is said to modulate the expression of a nucleic acid of the invention if it is capable of up- or down- regulating expression of the nucleic acid in a cell. Such cells can be obtained from any parts of the body such as the scalp, blood, brain, colon, dermis, epidermis and other skin cells, cutaneous surfaces, intertrigious areas, genitalia, vessels and endothelium. Some non-limiting examples of cells that can be used are red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

In one assay format, the expression of a nucleic acid encoding a gene of the invention (see Tables 5-7) in a cell or tissue sample is monitored directly by hybridization to the nucleic acids of the invention. Cell lines or tissues are exposed to the agent to be tested under appropriate conditions and time and total RNA or mRNA is isolated by standard procedures such as those disclosed in Sambrook *et al.*, (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press).

Probes to detect differences in RNA expression levels between cells exposed to the agent and control cells may be prepared as described above. Hybridization conditions are modified using known methods, such as those described by Sambrook *et al.*, and Ausubel *et al.*, as required for each probe. Hybridization of total cellular RNA or RNA enriched for polyA RNA can be accomplished in any available format. For instance, total cellular RNA or RNA enriched for polyA RNA can be affixed to a solid support and the solid support exposed to at least one probe comprising at least one, or part of one of the sequences of the invention under conditions in which the probe will specifically hybridize. Alternatively, nucleic acid fragments comprising at least one, or part of one of the sequences of the invention

can be affixed to a solid support, such as a silicon chip or a porous glass wafer. The chip or wafer can then be exposed to total cellular RNA or polyA RNA from a sample under conditions in which the affixed sequences will specifically hybridize to the RNA. By examining for the ability of a given probe to specifically hybridize to an RNA sample from an untreated cell population and from a cell population exposed to the agent, agents which up or down regulate expression are identified.

Methods to identify agents that modulate the activity of a protein encoded by a gene involved in schizophrenia.

The present invention provides methods for identifying agents that modulate at least one activity of the proteins described in Tables 5-7. Such methods may utilize any means of monitoring or detecting the desired activity. As used herein, an agent is said to modulate the expression of a protein of the invention if it is capable of up- or down- regulating expression of the protein in a cell. Such cells can be obtained from any parts of the body such as the scalp, blood, brain, colon, dermis, epidermis and other skin cells, cutaneous surfaces, intertrigious areas, genitalia, vessels and endothelium. Some non-limiting examples of cells that can be used are red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

In one format, the specific activity of a protein of the invention, normalized to a standard unit, may be assayed in a cell population that has been exposed to the agent to be tested and compared to an unexposed control cell population may be assayed. Cell lines or populations are exposed to the agent to be tested under appropriate conditions and time. Cellular lysates may be prepared from the exposed cell line or population and a control, unexposed cell line or population. The cellular lysates are then analyzed with the probe.

Antibody probes can be prepared by immunizing suitable mammalian hosts utilizing appropriate immunization protocols using the proteins of the invention or antigen-containing fragments thereof. To enhance immunogenicity, these proteins or

fragments can be conjugated to suitable carriers. Methods for preparing immunogenic conjugates with carriers such as BSA, KLH or other carrier proteins are well known in the art. In some circumstances, direct conjugation using, for example, carbodiimide reagents may be effective; in other instances linking reagents such as those supplied by Pierce Chemical Co. (Rockford, IL) may be desirable to provide accessibility to the hapten. The hapten peptides can be extended at either the amino or carboxy terminus with a cysteine residue or interspersed with cysteine residues, for example, to facilitate linking to a carrier. Administration of the immunogens is conducted generally by injection over a suitable time period and with use of suitable adjuvants, as is generally understood in the art. During the immunization schedule, titers of antibodies are taken to determine adequacy of antibody formation. While the polyclonal antisera produced in this way may be satisfactory for some applications, for pharmaceutical compositions, use of monoclonal preparations is preferred. Immortalized cell lines which secrete the desired monoclonal antibodies may be prepared using standard methods, see *e.g.*, Kohler & Milstein (1992) or modifications which affect immortalization of lymphocytes or spleen cells, as is generally known. The immortalized cell lines secreting the desired antibodies can be screened by immunoassay in which the antigen is the peptide hapten, polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production in ascites fluid. The desired monoclonal antibodies may be recovered from the culture supernatant or from the ascites supernatant. Fragments of the monoclonal antibodies or the polyclonal antisera which contain the immunologically significant portion(s) can be used as antagonists, as well as the intact antibodies. Use of immunologically reactive fragments, such as Fab or Fab' fragments, is often preferable, especially in a therapeutic context, as these fragments are generally less immunogenic than the whole immunoglobulin. The antibodies or fragments may also be produced, using current technology, by recombinant means. Antibody regions that bind specifically to the desired regions of the protein can also be produced in the context of chimeras derived from multiple species. Antibody regions that bind specifically to the desired regions of the protein can also be produced in the context of chimeras from multiple species, for instance, humanized antibodies. The antibody can therefore be a humanized antibody or a human antibody, as described in U.S. Patent 5,585,089 or Riechmann *et al.* (1988).

Agents that are assayed in the above method can be randomly selected or rationally selected or designed. As used herein, an agent is said to be randomly selected when the agent is chosen randomly without considering the specific sequences involved in the association of the a protein of the invention alone or with its associated substrates, binding partners, etc. An example of randomly selected agents is the use of a chemical library or a peptide combinatorial library, or a growth broth of an organism. As used herein, an agent is said to be rationally selected or designed when the agent is chosen on a non-random basis which takes into account the sequence of the target site or its conformation in connection with the agent's action. Agents can be rationally selected or rationally designed by utilizing the peptide sequences that make up these sites. For example, a rationally selected peptide agent can be a peptide whose amino acid sequence is identical to or a derivative of any functional consensus site. The agents of the present invention can be, as examples, oligonucleotides, antisense polynucleotides, interfering RNA, peptides, peptide mimetics, antibodies, antibody fragments, small molecules, vitamin derivatives, as well as carbohydrates. Peptide agents of the invention can be prepared using standard solid phase (or solution phase) peptide synthesis methods, as is known in the art. In addition, the DNA encoding these peptides may be synthesized using commercially available oligonucleotide synthesis instrumentation and produced recombinantly using standard recombinant production systems. The production using solid phase peptide synthesis is necessitated if non-gene-encoded amino acids are to be included.

Another class of agents of the present invention includes antibodies or fragments thereof that bind to a protein encoded by a gene in Tables 5-7. Antibody agents can be obtained by immunization of suitable mammalian subjects with peptides, containing as antigenic regions, those portions of the protein intended to be targeted by the antibodies (see section above of antibodies as probes for standard antibody preparation methodologies).

In yet another class of agents, the present invention includes peptide mimetics that mimic the three-dimensional structure of the protein encoded by a gene from Tables 5-7. Such peptide mimetics may have significant advantages over naturally occurring peptides, including, for example: more economical production, greater chemical

stability, enhanced pharmacological properties (half-life, absorption, potency, efficacy, etc.), altered specificity (*e.g.*, a broad-spectrum of biological activities), reduced antigenicity and others. In one form, mimetics are peptide-containing molecules that mimic elements of protein secondary structure. The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen. A peptide mimetic is expected to permit molecular interactions similar to the natural molecule. In another form, peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compounds are also referred to as peptide mimetics or peptidomimetics (Fauchere, 1986; Veber & Freidinger, 1985; Evans *et al.*, 1987) which are usually developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to therapeutically useful peptides may be used to produce an equivalent therapeutic or prophylactic effect. Generally, peptide mimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage using methods known in the art. Labeling of peptide mimetics usually involves covalent attachment of one or more labels, directly or through a spacer (*e.g.*, an amide group), to non-interfering position(s) on the peptide mimetic that are predicted by quantitative structure-activity data and molecular modeling. Such non-interfering positions generally are positions that do not form direct contacts with the macromolecule(s) to which the peptide mimetic binds to produce the therapeutic effect. Derivitization (*e.g.*, labeling) of peptide mimetics should not substantially interfere with the desired biological or pharmacological activity of the peptide mimetic. The use of peptide mimetics can be enhanced through the use of combinatorial chemistry to create drug libraries. The design of peptide mimetics can be aided by identifying amino acid mutations that increase or decrease binding of the protein to its binding partners. Approaches that can be used include the yeast two hybrid method (see Chien *et al.*, 1991) and the phage display method. The two hybrid method detects protein-protein interactions in yeast (Fields *et al.*, 1989). The phage display method detects the interaction between an immobilized protein and a protein that is expressed on the surface of phages such as lambda and M13 (Amberg *et al.*, 1993; Hogrefe *et al.*, 1993). These

methods allow positive and negative selection for protein-protein interactions and the identification of the sequences that determine these interactions.

Method to diagnose schizophrenia

The present invention also relates to methods for diagnosing schizophrenia disorder or a related disorder, a disposition to such disorder, predisposition to such a disorder and/or disorder progression. In some methods, the steps comprise contacting a target sample with (a) nucleic molecule(s) or fragments thereof and comparing the concentration of individual mRNA(s) with the concentration of the corresponding mRNA(s) from at least one healthy donor. An aberrant (increased or decreased) mRNA level of at least one gene from Tables 5-7, at least 5 or 10 genes from Tables 5-7, or at least 50 genes from Tables 5-7, determined in the sample in comparison to the control sample is an indication of schizophrenia or a related disorder or a disposition to such kinds of disorders. For diagnosis, samples are from any parts of the body such as the scalp, brain, colon, blood, dermis, epidermis and other skin cells, cutaneous surfaces, intertrigious areas, genitalia, vessels and endothelium. Some non-limiting examples of cells that can be used are red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

For analysis of gene expression, total RNA is obtained from cells according to standard procedures and, preferably, reverse-transcribed. Preferably, a DNAse treatment (in order to get rid of contaminating genomic DNA) is performed. Some non-limiting examples of cells that can be used are: red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

The nucleic acid molecule or fragment is typically a nucleic acid probe for hybridization or a primer for PCR. The person skilled in the art is in a position to design suitable nucleic acids probes based on the information provided in the Tables of the present invention. The target cellular component, *i.e.* mRNA, *e.g.*, in skin, may be detected directly in situ, *e.g.* by in situ hybridization or it may be isolated from other cell components by common methods known to those skilled in the art before contacting with a probe. Detection methods include Northern blot analysis, RNase protection, in situ methods, *e.g.* in situ hybridization, *in vitro* amplification methods (PCR, LCR, QRNA replicase or RNA-transcription/amplification (TAS, 3SR), reverse dot blot disclosed in EP-B10237362) and other detection assays that are known to those skilled in the art. Products obtained by *in vitro* amplification can be detected according to established methods, *e.g.* by separating the products on agarose or polyacrylamide gels and by subsequent staining with ethidium bromide. Alternatively, the amplified products can be detected by using labeled primers for amplification or labeled dNTPs. Preferably, detection is based on a microarray.

The probes (or primers) (or, alternatively, the reverse-transcribed sample mRNAs) can be detectably labeled, for example, with a radioisotope, a bioluminescent compound, a chemiluminescent compound, a fluorescent compound, a metal chelate, or an enzyme.

The present invention also relates to the use of the nucleic acid molecules or fragments described above for the preparation of a diagnostic composition for the diagnosis of schizophrenia or a disposition to such a disorder.

The present invention also relates to the use of the nucleic acid molecules of the present invention for the isolation or development of a compound which is useful for therapy of schizophrenia. For example, the nucleic acid molecules of the invention and the data obtained using said nucleic acid molecules for diagnosis of schizophrenia disorder might allow for the identification of further genes which are specifically dysregulated, and thus may be considered as potential targets for therapeutic interventions.

The invention further provides prognostic assays that can be used to identify subjects having or at risk of developing schizophrenia. In such method, a test sample

is obtained from a subject and the amount and/or concentration of the nucleic acid described in Tables 5-7 is determined; wherein the presence of an associated allele, a particular allele of a polymorphic locus, or the likes in the nucleic acids sequences of this invention (see SEQ ID from Tables 2-7) can be diagnostic for a subject having or at risk of developing schizophrenia. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid, a cell sample, or tissue. A biological fluid can be, but is not limited to saliva, serum, mucus, urine, stools, spermatozoids, vaginal secretions, lymph, amiotic liquid, pleural liquid and tears. Some non-limiting examples of cells that can be used are red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, polypeptide, nucleic acid such as antisense DNA or interfering RNA (RNAi), small molecule or other drug candidate) to treat schizophrenia. Specifically, these assays can be used to predict whether an individual will have an efficacious response or will experience adverse events in response to such an agent. For example, such methods can be used to determine whether a subject can be effectively treated with an agent that modulates the expression and/or activity of a gene from Tables 5-7, or the nucleic acids described herein. In another example, an association study may be performed to identify polymorphisms from Tables 2 and 3 that are associated with a given response to the agent *e.g.*, an efficacious response or the likelihood of one or more adverse events. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a gene from Tables 5-7 in which a test sample is obtained and nucleic acids or polypeptides from Tables 2 or 3 are detected (*e.g.*, wherein the presence of a particular level of expression of a gene from Tables 5-7 or a particular allelic variant of such gene, such as polymorphism from Tables 2 or 3, is diagnostic for a subject that can be

administered an agent to treat a disorder, such as schizophrenia). In one embodiment, the method includes obtaining a sample from a subject suspected of having schizophrenia or an affected individual and exposing such sample to an agent. The expression and/or activity of the nucleic acids and or genes of the invention are monitored before and after treatment with such agent to assess the effect of such agent. After analysis of the expression values, one skilled in the art can determine whether such agent can effectively treat such subject. In another embodiment, the method includes obtaining a sample from a subject having or susceptible to developing schizophrenia and determining the allelic constitution of polymorphism from Tables 2 or 3 that are associated with a particular response to an agent. After analysis of the allelic constitution of the individual at the associated polymorphisms, one skilled in the art can determine whether such agent can effectively treat such subject.

The methods of the invention can also be used to detect genetic alterations in a gene from Tables 5-7, thereby determining if a subject with the lesioned gene is at risk for schizophrenia. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one alteration linked to or affecting the integrity of a gene from Tables 5-7 encoding a polypeptide or the misexpression of such gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of: (1) a deletion of one or more nucleotides from a gene from Tables 5-7; (2) an addition of one or more nucleotides to a gene from Tables 5-7; (3) a substitution of one or more nucleotides of a gene from Tables 5-7; (4) a chromosomal rearrangement of a gene from Tables 5-7; (5) an alteration in the level of a messenger RNA transcript of a gene from Tables 5-7; (6) aberrant modification of a gene from Tables 5-7, such as of the methylation pattern of the genomic DNA, (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a gene from Tables 5-7; (8) inappropriate post-translational modification of a polypeptide encoded by a gene from Tables 5-7; and (9) alternative promoter use. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in a gene from Tables 5-7. A preferred biological sample is a peripheral blood sample obtained by conventional means from a subject. Another preferred biological sample is a buccal swab. Other biological

samples can be, but is not limited to, urine, stools, spermatozoids, vaginal secretions, lymph, amiotic liquid, pleural liquid and tears.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.*, 1988; and Nakazawa *et al.*, 1994), the latter of which can be particularly useful for detecting point mutations in a gene from Tables 5-7 (see Abavaya *et al.*, 1995). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic DNA, mRNA, or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene from Tables 5-7 under conditions such that hybridization and amplification of the nucleic acid from Tables 5-7 (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with some of the techniques used for detecting an associated allele, a particular allele of a polymorphic locus, or the like described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990), transcriptional amplification system (Kwoh *et al.*, 1989), Q-Beta Replicase (Lizardi *et al.*, 1988), isothermal amplification (*e.g.* Dean *et al.*, 2002); and Hafner *et al.*, 2001), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of ordinary skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low number.

In an alternative embodiment, alterations in a gene from Tables 5-7, from a sample cell can be identified by identifying changes in a restriction enzyme cleavage pattern. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates an associated allele, a particular

allele of a polymorphic locus, or the like, in the sample DNA. Moreover, sequence specific ribozymes (see, *e.g.*, U.S. Pat. No. 5,498,531 or DNAzyme *e.g.* U.S. Pat. No. 5,807,718) can be used to score for the presence of specific associated allele, a particular allele of a polymorphic locus, or the likes by development or loss of a ribozyme or DNAzyme cleavage site.

The present invention also relates to further methods for diagnosing the schizophrenia or a related disorder, a disposition to such disorder, and predisposition to such a disorder and/or disorder progression. In some methods, the steps comprise contacting a target sample with (a) nucleic molecule(s) or fragments thereof and determining the presence or absence of a particular allele of a polymorphism that confers a disorder-related phenotype (*e.g.*, predisposition to such a disorder and/or disorder progression). The presence of at least one allele from Tables 2 and 3 that is associated with schizophrenia disorder ("associated allele"), at least 5 or 10 associated alleles from Tables 2 and 3, at least 50 associated alleles from Tables 2 and 3, or at least 100 associated alleles from Tables 2 and 3, determined in the sample is an indication of the schizophrenia disorder or a related disorder, a disposition or predisposition to such kinds of disorders, or a prognosis for such disorder progression. Samples may be obtained from any parts of the body such as the scalp, blood, brain colon, dermis, epidermis and other skin cells, cutaneous surfaces, intertrigious areas, genitalia, vessels and endothelium. Some non-limiting examples of cells that can be used are red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, lymphocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and artherial cells.

In other embodiments, alterations in a gene from Tables 5-7 can be identified by hybridizing sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays or bead arrays containing tens to thousands of oligonucleotide probes (Cronin *et al.*, 1996; Kozal *et al.*, 1996). For example, alterations in a gene from Tables 5-7 can be identified in two dimensional arrays containing light-generated DNA probes

as described in Cronin *et al.*, (1996). Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations, associated alleles or alleles of a particular polymorphic locus, by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene or associated alleles or particular allele of a polymorphic locus.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence a gene from Tables 5-7 and detect an associated allele, a particular allele of a polymorphic locus, or the like by comparing the sequence of the sample gene from Tables 5-7 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert (1977) or Sanger (1977). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Bio/Techniques 19:448, 1995) including sequencing by mass spectrometry (see, *e.g.* PCT International Publication No. WO 94/16101; Cohen *et al.*, 1996; and Griffin *et al.* 1993), real-time pyrophosphate sequencing method (Ronaghi *et al.*, 1998; and Permutt *et al.*, 2001) and sequencing by hybridization (see *e.g.* Drmanac *et al.*, 2002).

Other methods of detecting an associated allele, a particular allele of a polymorphic locus, or the likes in a gene from Tables 5-7 include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA, DNA/DNA or RNA/DNA heteroduplexes (Myers *et al.*, 1985). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type gene from Tables 5-7 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and

DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of an associated allele, a particular allele of a polymorphic locus, or the like. (see, for example, Cotton *et al.*, 1988; Saleeba *et al.*, 1992). In a preferred embodiment, the control DNA or RNA can be labeled for detection, as described herein.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point an associated allele, a particular allele of a polymorphic locus, or the likes in a gene from Tables 5-7 cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches (Hsu *et al.*, 1994). Other examples include, but are not limited to, the MutHLS enzyme complex of E. coli (Smith and Modrich., 1996) and Cel 1 from the celery (Kulinski *et al.*, 2000) both cleave the DNA at various mismatches. According to an exemplary embodiment, a probe based on a gene sequence from Tables 5-7 is hybridized to a cDNA or other DNA product from a test cell or cells. The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected using electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039. Alternatively, the screen can be performed *in vivo* following the insertion of the heteroduplexes in an appropriate vector. The whole procedure is known to those ordinary skilled in the art and is referred to as mismatch repair detection (see *e.g.* Fakhrai-Rad *et al.*, 2004).

In other embodiments, alterations in electrophoretic mobility can be used to identify an associated allele, a particular allele of a polymorphic locus, or the likes in genes from Tables 5-7. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.*, 1993; see also Cotton, 1993; and Hayashi *et al.*, 1992). Single-stranded DNA fragments of sample and control nucleic acids from Tables 5-7 will be denatured and allowed to renature. The secondary structure of

single-stranded nucleic acids varies according to sequence; the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Kee *et al.*, 1991).

In yet another embodiment, the movement of mutant or wild-type fragments in a polyacrylamide gel containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.*, 1985). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum *et al.*, 1987). In another embodiment, the mutant fragment is detected using denaturing HPLC (see *e.g.* Hoogendoorn *et al.*, 2000).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, selective primer extension, selective ligation, single-base extension, selective termination of extension or invasive cleavage assay. For example, oligonucleotide primers may be prepared in which the known associated allele, particular allele of a polymorphic locus, or the like is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.*, 1986; Saiki *et al.*, 1989). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different associated allele, a particular allele of a polymorphic locus, or the likes where the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA. Alternatively, the amplification, the allele-specific hybridization and the detection can be done in a single assay following the principle of the 5' nuclease assay (*e.g.* see Livak *et al.*, 1995). For example, the associated allele, a particular allele of a polymorphic locus, or the like locus is amplified by PCR in the presence of both allele-specific

oligonucleotides, each specific for one or the other allele. Each probe has a different fluorescent dye at the 5' end and a quencher at the 3' end. During PCR, if one or the other or both allele-specific oligonucleotides are hybridized to the template, the Taq polymerase via its 5' exonuclease activity will release the corresponding dyes. The latter will thus reveal the genotype of the amplified product.

The hybridization may also be carried out with a temperature gradient following the principle of dynamic allele-specific hybridization or like (*e.g.* Jobs *et al.*, 2003; and Bourgeois and Labuda, 2004). For example, the hybridization is done using one of the two allele-specific oligonucleotides labeled with a fluorescent dye, an intercalating quencher under a gradually increasing temperature. At low temperature, the probe is hybridized to both the mismatched and full-matched template. The probe melts at a lower temperature when hybridized to the template with a mismatch. The release of the probe is captured by an emission of the fluorescent dye, away from the quencher. The probe melts at a higher temperature when hybridized to the template with no mismatch. The temperature-dependent fluorescence signals therefore indicate the absence or presence of the associated allele, particular allele of a polymorphic locus, or the like (*e.g.* Jobs *et al.* supra). Alternatively, the hybridization is done under a gradually decreasing temperature. In this case, both allele-specific oligonucleotides are hybridized to the template competitively. At high temperature none of the two probes is hybridized. Once the optimal temperature of the full-matched probe is reached, it hybridizes and leaves no target for the mismatched probe. In the latter case, if the allele-specific probes are differently labeled, then they are hybridized to a single PCR-amplified target. If the probes are labeled with the same dye, then the probe cocktail is hybridizes twice to identical templates with only one labeled probes, different in the two cocktails, in the presence of the unlabeled competitive probe.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the present invention. Oligonucleotides used as primers for specific amplification may carry the associated allele, particular allele of a polymorphic locus, or the like of interest in the center of the molecule, so that amplification depends on differential hybridization (Gibbs *et al.*, 1989) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can

prevent, or reduce polymerase extension (Prossner, 1993). In addition it may be desirable to introduce a novel restriction site in the region of the associated allele, particular allele of a polymorphic locus, or the like to create cleavage-based detection (Gasparini *et al.*, 1992). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany, 1991). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known associated allele, a particular allele of a polymorphic locus, or the like at a specific site by looking for the presence or absence of amplification. The products of such an oligonucleotide ligation assay can also be detected by means of gel electrophoresis. Furthermore, the oligonucleotides may contain universal tags used in PCR amplification and zip code tags that are different for each allele. The zip code tags are used to isolate a specific, labeled oligonucleotide that may contain a mobility modifier (*e.g.* Grossman *et al.*, 1994).

In yet another alternative, allele-specific elongation followed by ligation will form a template for PCR amplification. In such cases, elongation will occur only if there is a perfect match at the 3' end of the allele-specific oligonucleotide using a DNA polymerase. This reaction is performed directly on the genomic DNA and the extension/ligation products are amplified by PCR. To this end, the oligonucleotides contain universal tags allowing amplification at a high multiplex level and a zip code for SNP identification. The PCR tags are designed in such a way that the two alleles of a SNP are amplified by different forward primers, each having a different dye. The zip code tags are the same for both alleles of a given SNP and they are used for hybridization of the PCR-amplified products to oligonucleotides bound to a solid support, chip, bead array or like. For an example of the procedure, see Fan *et al.* (Cold Spring Harbor Symposia on Quantitative Biology, Vol. LXVIII, pp. 69-78, 2003).

Another alternative includes the single-base extension/ligation assay using a molecular inversion probe, consisting of a single, long oligonucleotide (see *e.g.* Hardenbol *et al.*, 2003). In such an embodiment, the oligonucleotide hybridizes on both side of the SNP locus directly on the genomic DNA, leaving a one-base gap at the SNP locus. The gap-filling, one-base extension/ligation is performed in four tubes, each having a different dNTP. Following this reaction, the oligonucleotide is

circularized whereas unreactive, linear oligonucleotides are degraded using an exonuclease such as exonuclease I of E. coli. The circular oligonucleotides are then linearized and the products are amplified and labeled using universal tags on the oligonucleotides. The original oligonucleotide also contains a SNP-specific zip code allowing hybridization to oligonucleotides bound to a solid support, chip, bead array or like. This reaction can be performed at a highly multiplexed level.

In another alternative, the associated allele, particular allele of a polymorphic locus, or the like is scored by single-base extension (see *e.g.* U.S. Pat. No. 5,888,819). The template is first amplified by PCR. The extension oligonucleotide is then hybridized next to the SNP locus and the extension reaction is performed using a thermostable polymerase such as ThermoSequenase (GE Healthcare) in the presence of labeled ddNTPs. This reaction can therefore be cycled several times. The identity of the labeled ddNTP incorporated will reveal the genotype at the SNP locus. The labeled products can be detected by means of gel electrophoresis, fluorescence polarization (*e.g.* Chen *et al.*, 1999) or by hybridization to oligonucleotides bound to a solid support, chip, bead array or like. In the latter case, the extension oligonucleotide will contain a SNP-specific zip code tag.

In yet another alternative, the variant is scored by selective termination of extension. The template is first amplified by PCR and the extension oligonucleotide hybridizes in vicinity to the SNP locus, close to but not necessarily adjacent to it. The extension reaction is carried out using a thermostable polymerase such as Thermo Sequenase (GE Healthcare) in the presence of a mix of dNTPs and at least one ddNTP. The latter has to terminate the extension at one of the alleles of the interrogated SNP, but not both such that the two alleles will generate extension products of different sizes. The extension product can then be detected by means of gel electrophoresis, in which case the extension products need to be labeled, or by mass spectrometry (see *e.g.* Storm *et al.*, 2003).

In another alternative, the associated allele, particular allele of a polymorphic locus, or the like is detected using an invasive cleavage assay (see U.S. Pat. No. 6,090,543). There are five oligonucleotides per SNP to interrogate but these are used in a two step-reaction. During the primary reaction, three of the designed oligonucleotides are first hybridized directly to the genomic DNA. One of them is

locus-specific and hybridizes up to the SNP locus (the pairing of the 3' base at the SNP locus is not necessary). There are two allele-specific oligonucleotides that hybridize in tandem to the locus-specific probe but also contain a 5' flap that is specific for each allele of the SNP. Depending upon hybridization of the allele-specific oligonucleotides at the base of the SNP locus, this creates a structure that is recognized by a cleavase enzyme (U.S. Pat. No. 6,090,606) and the allele-specific flap is released. During the secondary reaction, the flap fragments hybridize to a specific cassette to recreate the same structure as above except that the cleavage will release a small DNA fragment labeled with a fluorescent dye that can be detected using regular fluorescence detector. In the cassette, the emission of the dye is inhibited by a quencher.

Other types of markers can also be used for diagnostic purposes. For example, microsatellites can also be useful to detect the genetic predisposition of an individual to a given disorder. Microsatellites consist of short sequence motifs of one or a few nucleotides repeated in tandem. The most common motifs are polynucleotide runs, dinucleotide repeats (particularly the CA repeats) and trinucleotide repeats. However, other types of repeats can also be used. The microsatellites are very useful for genetic mapping because they are highly polymorphic in their length. Microsatellite markers can be typed by various means, including but not limited to DNA PCR fragment sizing, oligonucleotide ligation assay and mass spectrometry. For example, the locus of the microsatellite is amplified by PCR and the size of the PCR fragment will be directly correlated to the length of the microsatellite repeat. The size of the PCR fragment can be detected by regular means of gel electrophoresis. The fragment can be labeled internally during PCR or by using end-labeled oligonucleotides in the PCR reaction (*e.g.* Mansfield *et al.*, 1996). Alternatively, the size of the PCR fragment is determined by mass spectrometry. In such a case, however, the flanking sequences need to be eliminated. This can be achieved by ribozyme cleavage of an RNA transcript of the microsatellite repeat (Krebs *et al.*, 2001). For example, the microsatellite locus is amplified using oligonucleotides that include a T7 promoter on one end and a ribozyme motif on the other end. Transcription of the amplified fragments will yield an RNA substrate for the ribozyme, releasing small RNA fragments that contain the repeated region. The size of the latter is determined by mass spectrometry. Alternatively, the flanking

sequences are specifically degraded. This is achieved by replacing the dTTP in the PCR reaction by dUTP. The dUTP nucleosides are then removed by uracyl DNA glycosylases and the resulting abasic sites are cleaved by either abasic endonucleases such as human AP endonuclease or chemical agents such as piperidine. Bases can also be modified post-PCR by chemical agents such as dimethyl sulfate and then cleaved by other chemical agents such as piperidine (see *e.g.* Maxam and Gilbert, 1977; U.S. Pat. No. 5,869,242; and U.S. Patent pending serial No. 60/335,068).

In another alternative, an oligonucleotide ligation assay can be performed. The microsatellite locus is first amplified by PCR. Then, different oligonucleotides can be submitted to ligation at the center of the repeat with a set of oligonucleotides covering all the possible lengths of the marker at a given locus (Zirvi *et al.*, 1999). Another example of design of an oligonucleotide assay comprises the ligation of three oligonucleotides; a 5' oligonucleotide hybridizing to the 5' flanking sequence, a repeat oligonucleotide of the length of the shortest allele of the marker hybridizing to the repeated region and a set of 3' oligonucleotides covering all the existing alleles hybridizing to the 3' flanking sequence and a portion of the repeated region for all the alleles longer than the shortest one. For the shortest allele, the 3' oligonucleotide exclusively hybridizes to the 3' flanking sequence (U.S. Pat. No. 6,479,244).

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid selected from the SEQ ID of Tables 2-7, or antibody reagent described herein, which may be conveniently used, for example, in a clinical setting to diagnose patient exhibiting symptoms or a family history of a disorder, or disorder involving abnormal activity of genes from Tables 5-7.

Method to treat an animal suspected of having schizophrenia

The present invention provides methods of treating a disorder associated with schizophrenia, by expressing *in vivo* the nucleic acids of at least one gene from Tables 5-7. These nucleic acids can be inserted into any of a number of well-known vectors for the transfection of target cells and organisms as described below. The nucleic acids are transfected into cells, *ex vivo* or *in vivo*, through the interaction of

the vector and the target cell. The nucleic acids encoding a gene from Tables 5-7, under the control of a promoter, then expresses the encoded protein, thereby mitigating the effects of absent, partial inactivation, or abnormal expression of a gene from Tables 5-7.

Such gene therapy procedures have been used to correct acquired and inherited genetic defects, cancer, and viral infection in a number of contexts. The ability to express artificial genes in humans facilitates the prevention and/or cure of many important human disorders, including many disorders which are not amenable to treatment by other therapies (for a review of gene therapy procedures, see Anderson, 1992; Nabel & Felgner, 1993; Mitani & Caskey, 1993; Mulligan, 1993; Dillon, 1993; Miller, 1992; Van Brunt, 1998; Vigne, 1995; Kremer & Perricaudet 1995; Doerfler & Bohm 1995; and Yu *et al.*, 1994).

Delivery of the gene or genetic material into the cell is the first critical step in gene therapy treatment of genetic disorders. A large number of delivery methods are well known to those of skill in the art. Preferably, the nucleic acids are administered for *in vivo* or *ex vivo* gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see the references included in the above section.

The use of RNA or DNA viral based systems for the delivery of nucleic acids take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in vitro* and the modified cells are administered to patients (*ex vivo*). Conventional viral based systems for the delivery of nucleic acids could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high

transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vector that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with pack capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and pack of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, *e.g.*, Buchscher *et al.*, 1992; Johann *et al.*, 1992; Sommerfelt *et al.*, 1990; Wilson *et al.*, 1989; Miller *et al.*, 1999; and PCT/US94/05700).

In applications where transient expression of the nucleic acid is preferred, adenoviral based systems are typically used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, *e.g.*, in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (see, *e.g.*, West *et al.*, 1987; U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, 1994; Muzyczka, 1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al.*, 1985; Tratschin, *et al.*, 1984; Hermonat & Muzyczka, 1984; and Samulski *et al.*, 1989.

In particular, numerous viral vector approaches are currently available for gene transfer in clinical trials, with retroviral vectors by far the most frequently used system. All of these viral vectors utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing

agent. pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar *et al.*, 1995; Kohn *et al.*, 1995; Malech *et al.*, 1997). PA317/pLASN was the first therapeutic vector used in a gene therapy trial (Blaese *et al.*, 1995). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors (Ellem *et al.*, 1997; and Dranoff *et al.*, 1997).

Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner *et al.*, 1998, Kearns *et al.*, 1996).

Replication-deficient recombinant adenoviral vectors (Ad) are predominantly used in transient expression gene therapy; because they can be produced at high titer and they readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiple types of tissues *in vivo,* including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman *et al.*, 1998). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker *et al.*, 1996; Sterman *et al.*, 1998; Welsh *et al.*, 1995; Alvarez *et al.*, 1997; Topf *et al.*, 1998.

Pack cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for pack and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are

supplied in trans by the pack cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for pack and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment to which adenovirus is more sensitive than AAV.

In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han *et al.*, 1995, reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (*e.g.*, Fab or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (*e.g.*, intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (*e.g.*, lymphocytes, bone marrow aspirates, and tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector. *Ex vivo* cell

transfection for diagnostics, research, or for gene therapy (*e.g.*, via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with a nucleic acid (gene or cDNA), and re-infused back into the subject organism (*e.g.*, patient). Various cell types suitable for *ex vivo* transfection are well known to those of skill in the art (see, *e.g.*, Freshney *et al.*, 1994; and the references cited therein for a discussion of how to isolate and culture cells from patients). In one embodiment, stem cells are used in *ex vivo* procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types *in vitro*, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells *in vitro* into clinically important immune cell types using cytokines such a GM-CSF, IFN-γ and TNF-α are known (see Inaba *et al.*, 1992). Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and lad (differentiated antigen presenting cells). Vectors (*e.g.*, retroviruses, adenoviruses, liposomes, etc.) containing therapeutic nucleic acids can be also administered directly to the organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells, as described above. The nucleic acids from Tables 2-7 are administered in any suitable manner, preferably with the pharmaceutically acceptable carriers described above. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route (see Samulski *et al.*, 1989). The present invention is not limited to any method of administering such nucleic acids, but preferentially uses the methods described herein.

The present invention further provides other methods of treating disorders, such as schizophrenia, by for example administering to an individual having schizophrenia (or suspected of having schizophrenia) an effective amount of an agent that regulates the expression, activity or physical state of at least one gene from Tables

5-7. An "effective amount" of an agent is an amount that modulates a level of expression or activity of a gene from Tables 5-7, in a cell in the individual at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80% or more, compared to a level of the respective gene from Tables 5-7 in a cell in the individual in the absence of the compound. The preventive or therapeutic agents of the present invention may be administered, either orally or parenterally, systemically or locally. For example, intravenous injection such as drip infusion, intramuscular injection, intraperitoneal injection, subcutaneous injection, suppositories, intestinal lavage, oral enteric coated tablets, and the like can be selected, and the method of administration may be chosen, as appropriate, depending on the age and the conditions of the patient. The effective dosage is chosen from the range of 0.01 mg to 100 mg per kg of body weight per administration. Alternatively, the dosage in the range of 1 to 1000 mg, preferably 5 to 50 mg per patient may be chosen. The therapeutic efficacy of the treatment may be monitored by observing various parts of the body and mind, such as the response to a mental status examination or a smell test (capability of identifying smells, such as smoke). Others ways of monitoring efficacy can be, but are not limited to monitoring well-being, memory, mental state and psychological state of the patient, or using a perceptual tests such as the HOD test (Hoffer-Osmond Diagnostic).

The present invention further provides a method of treating an individual clinically diagnosed with a schizophrenia or related disorder. The methods generally comprises analyzing a biological sample that includes a cell, in some cases, a skin cell, from an individual clinically diagnosed with an schizophrenia for the presence of modified levels of expression of at least 2 gene, at least 10 genes, or at least 50 genes from Tables 5-7. A treatment plan that is most effective for individuals clinically diagnosed as having a condition associated with a disorder or schizophrenia is then selected on the basis of the detected expression of such genes in a cell. Treatment may include administering a composition that includes an agent that modulates the expression or activity of a protein from Tables 5-7 in the cell. Information obtained as described in the methods above can also be used to predict the response of the individual to a particular agent. Thus, the invention further provides a method for predicting a patient's likelihood to respond to a drug treatment

for a condition associated with the schizophrenia disorder, comprising determining whether modified levels of a gene from Tables 5-7 is present in a cell, wherein the presence of protein is predictive of the patient's likelihood to respond to a drug treatment for the condition. Examples of the prevention or improvement of symptoms accompanied by schizophrenia that can monitor for effectiveness include response to a mental status examination or a smell test (capability of identifying smells, such as smoke). Others ways of monitoring efficacy can be, but are not limited to monitoring well-being, memory, mental state and psychological state of the patient, or using a perceptual tests such as the HOD test (Hoffer-Osmond Diagnostic).

The invention also provides a method of predicting a response to therapy in a subject having schizophrenia by determining the presence or absence in the subject of one or more markers associated with the schizophrenia described in Tables 2-4 and/or 5-7, diagnosing the subject in which the one or more markers are present as having schizophrenia, and predicting a response to a therapy based on the diagnosis *e.g.*, response to therapy may include an efficacious response and/or one or more adverse events. The invention also provides a method of optimizing therapy in a subject having schizophrenia disorder by determining the presence or absence in the subject of one or more markers associated with a clinical subtype of schizophrenia, diagnosing the subject in which the one or more markers are present as having a particular clinical subtype of schizophrenia, and treating the subject having a particular clinical subtype of schizophrenia based on the diagnosis. Thus, while there are a number of treatments for schizophrenia currently available, they all are accompanied by various side effects, high costs, and long complicated treatment protocols, which are often not available and effective in a large number of individuals. Accordingly, there remains a need in the art for more effective and otherwise improved methods for treating and preventing schizophrenia. Thus, there is a continuing need in the medical arts for genetic markers of schizophrenia and guidance for the use of such markers. The present invention fulfills this need and provides further related advantages.

EXAMPLES

Example 1: *Identification of cases and controls*

All individuals were sampled from the Quebec founder population (QFP). Membership in the founder population was defined as having four grandparents with French Canadian family names who were born in the Province of Quebec, Canada or in adjacent areas of the Provinces of New Brunswick and Ontario or in New England or New York State. The Quebec founder population has two distinct advantages over general populations for LD mapping. Because it is relatively young, about 12 to 15 generations from mid-17th century to present, and because it has a limited but sufficient number of founders, approximately 2600 effective founders (Charbonneau *et al.* 1987), the Quebec population is characterized both by extended LD and by decreased genetic heterogeneity. The increased extent of LD allows the detection of disease associated genes using a reasonable marker density, while still allowing the increased meiotic resolution of population-based mapping. The number of founders is small enough to result in increased LD and reduced allelic heterogeneity, yet large enough to insure that all of the major disease associated genes involved in general populations are present in Quebec. Reduced allelic heterogeneity should act to increase relative risk imparted by the remaining alleles and so increase the power of case/control studies to detect genes and disorder associated alleles within the Quebec population. The specific combination of age in generations, optimal number of founders and large present population size makes the QFP optimal for LD-based gene mapping.

The family relationships among samples are routinely examined using proprietary algorithms and information from the genealogical data bases. When two subjects are found to be too closely related for LD analysis, one of them is removed from the sample.

Case inclusion criteria are based upon a DSM-IV diagnosis (reviewed in James Morrison 1995) for the study and include having at least 2 of the following 5 symptoms: hallucinations, delusions, disorganized and incoherent speech, grossly disorganized behavior or catatonic behavior, odd beliefs and unusual perceptual experiences. If the patient has hallucinations, and these hallucinations consist of

auditory hallucinations, the case is included in the study even if the case does not have another one of the symptoms listed above. Also, if the case has delusions and these delusions are prominent and bizarre, the case is admitted to the study even if the case does not have another one of the symptom listed above. In addition to the first set of symptoms, the case also has to have at least one of the following symptoms: being dysfunctional at work, in interpersonal relationships or in personal care. The exclusion criteria are: mood disorders with a psychotic character, a schizo-affective disorder, substance abuse, a severe developmental disorder such as autism. Control inclusion criteria for the study requires the absence of obvious mental disorders and/or a history of depression. The controls are matched to the cases with respect to gender.

All human sampling was subject to ethical review procedures.

All QFP subjects (cases and controls) provided a 30 ml blood sample (3 barcoded tubes of 10 ml). Samples were processed immediately upon arrival at Genizon's laboratory. All samples were scanned and logged into a LabVantage Laboratory Information Management System (LIMS), which served as a hub between the clinical data management system and the genetic analysis system. Following centrifugation, the buffy coat containing the white blood cells was isolated from each tube. Genomic DNA was extracted from the buffy coat from one of the tubes, and stored at 4°C until required for genotyping. DNA extraction was performed with a commercial kit using a guanidine hydrochloride based method (FlexiGene, Qiagen) according to the manufacturer's instructions. The extraction method yielded high molecular weight DNA, and the quality of every DNA sample was verified by agarose gel electrophoresis. Genomic DNA appeared on the gel as a large band of very high molecular weight. The remaining two buffy coats were stored at –80°C as backups.

The samples were collected as 470 cases (311 males and 159 females) and 469 controls (311 males and 158 females). The DNA extracted from cases and control samples was pooled into various case and control pools. Pools were designed to isolate cases from controls, and males from females. Pools of around 50 randomized individuals were created, resulting in 6 pools of affected males and 3 pools of affected females. In the same manner, 6 male control pools and 3 female control pools were created.

Example 2: *Genome Wide Association*

Genotyping was performed using Perlegen Life Sciences ultra-high-throughput platform. Loci of interest were amplified and hybridized to wafers containing arrays of oligonucleotides. Allele discrimination was performed through allele-specific hybridization. In total, 248,535 SNPs, spread over 3 microarrays, were genotyped. The markers were selected from various databases including the ~1.6 million SNP database of Perlegen Life Sciences (Patil, 2001), the hapmap consortium database and dbSNP at NCBI and our proprietary markers selected from the Quebec founder population. The SNPs were chosen to maximize uniformity of genetic coverage and as much as possible, with a minor allele frequency of 10% or higher.

The genotyping information was entered into a Unified Genotype Database (a proprietary database under development) from which it was accessed using custom-built programs for export to the genetic analysis pipeline. Analyses of these genotypes were performed with the statistical tools described in Example 3. The GWS permitted the identification of 27 candidate regions that are further analyzed by the Fine Mapping approach described below.

Example 3: *Genetic Analysis*

The raw data generated by the GWS approach (Example 2 herein) was analyzed by various means to identify candidate regions (see also Fine Mapping described in Example 5).

Raw data analysis by Perlegen

The data analysis process compares the relative fluorescence intensities of features corresponding to the reference allele of a given SNP with those corresponding to the alternate allele, to calculate a p-hat value. The latter is proportional to the fluorescence signal from perfect match features for the reference allele divided by the sum of fluorescence signals from perfect match features for the reference plus the alternate alleles. P-hat assumes values close to 1 (typically 0.9) for pure reference samples and close to 0 (typically 0.1) for pure alternate samples, and can

be used as a measured estimate of the reference allele frequency of a SNP in a DNA pool. The difference between case and control pools, delta p-hat, is calculated using the weighted average of case and control p-hats. Delta p-hat is a reliable estimate of the allele frequency difference between the cases and controls.

Data analysis by Genizon

Analysis of the data by Genizon was based upon Perlegen's p-hat values.

Data analysis of p-hat

The data consisted of 9 pools of cases and 9 pools of controls, with a p-hat value for each pool (provided by Perlegen's analysis methods described above). The weighted mean p-hat value for all case pools was calculated as $\overline{X} = \sum N_i Phat_i / \sum N_i$ where N_i is the number of alleles (2 X the number of people) in each case pool and the weighted p-hat value for all control pools were calculated as $\overline{Y} = \sum L_j Phat_j / \sum L_j$ where L_i is the number of alleles in each control pool. The estimate of delta-phat, the difference in allele frequency between cases and controls was $\overline{X} - \overline{Y}$ and the significance of delta-phat was determined using a Student's t statistic where $t = \frac{\overline{X} - \overline{Y}}{\sqrt{S_p\left(\left(\sum N_i\right)^{-1} + \left(\sum L_i\right)^{-1}\right)}}$ where S_p is the pooled variance under the assumption that both case and control variances are equal. The variance of $\overline{X}$ was calculated as $S_X = \left(1/8 - 1\right)\sum N_i\left(Phat_i - \overline{X}\right)^2$ and similarly the variance of $\overline{Y}$ was calculated as $S_Y = \left(1/8 - 1\right)\sum L_j\left(Phat_j - \overline{Y}\right)^2$. The Student's t test was performed with 8+8-2 = 14 degrees of freedom. Single Marker P values were calculated for all markers within the genome wide scan map.

Combined P values

In addition to single marker P values, combined P values across multi-marker sliding windows were calculated after the method of Fisher (*Statistical Methods for Research Workers,* 14th edition Hafner Press N.Y. 1970 pp 99-100) where $\chi^2 = \sum \ln P_i$ with 2k degrees of freedom where P_i is the P value for association of

each of k markers within the sliding window. The combined P values identify regions of multiple single marker associations.

Permutation Test for Exact P values

There were 9 case and 9 control pools, each giving an estimate of p-*hat.* Therefore there are 18! (18 factorial) / 9!X9! = 48620 ways to group the pools into 2 groups of 9. The POOLEX exact test calculates the mean difference in phat for single markers between the two groups for all of these possible arrangements of the pools into two groups as well as the combined P values for 1, 5, 9 and 15 marker windows for all combinations. These values are then arranged in order in each case and the relative rank among all combinations of the actually observed combination for real cases and controls is determined. The exact P value for the observed case and control combination is calculated as:

$$\frac{\text{No. of combinations with a value greater than or equal to the actual observed value}}{\text{48620 total combinations}}$$

For example, if the actually observed combined P value for a 5 marker window corresponding to the actual cases and controls combination, were the largest combined P value of all 48620 possible combinations of 2 groups of 9, then the exact probability (P value) for observing this event by chance given no difference between cases and controls would be 1/48620 = 0.0000205.

Example 4: *Fine Mapping*

The top regions identified as being associated with schizophrenia by the GWS were further analyzed by Fine Mapping (genotyping all cases and controls samples individually) using a denser set of markers, in order to validate and/or refine the signal. Fine mapping was carried out using the Illumina BeadStation 500GX SNP genotyping platform. Alleles are genotyped using an allele-specific elongation assay that involves ligation to a locus-specific oligonucleotide. The assay is performed directly on genomic DNA at a highly multiplex level and the products are amplified

using universal oligonucleotides. For each candidate region, a set of SNP markers is selected with an average inter-marker distance varying with the mean extent of LD throughout the region as determined by delta-M (Δ_M), where M is the number of markers present in the 300 kb window centered at each reference marker is defined as the square root of the average r^2 or Δ^2_{ij} measures of LD between all ({M(M-1)}/2) pairwise comparisons of all (M) markers within the 300 kb window (Dawson *et al.*, 2002). This produces an average multi-marker measure of LD analogous to Hill's Δ statistic for two marker LD. Regions with a signal harboring a high $-Log_{10}$ P value and with mean delta-M of 0.3 or below are mapped with a target density of one marker per 10 kb. Regions showing a signal with a high $-Log_{10}$ P value and with mean delta-M between 0.3 and 0.35 as well as selected regions with a signal with a lower $-Log_{10}$ P value and a delta-M value below 0.35 are mapped with a target density of one marker per 10-20 kb. The principle is that low-LD regions will be mapped at a higher SNP density. Selected regions with a delta-M value above 0.35 are mapped with a density of one marker per 20-30 kb, including the markers used in the GWS. The selected regions are delimited by the location where the LDSTATS - Log_{10} P values reach the background level. The cohort consists of 470 cases and 469 controls (as used for the GWS).

Example 5: *Gene identification and characterization*

Any gene or EST mapping to the interval based on public map data or proprietary map data was considered as a candidate schizophrenia gene. The approach used to identify all genes located in the critical regions is described below.

Public gene mining

Once regions were identified using the analyses described above, a series of public data mining efforts were undertaken, with the aim of identifying all genes located within the critical intervals as well as their respective structural elements (*i.e.*, promoters and other regulatory elements, UTRs, exons and splice sites). The initial analysis relied on annotation information stored in public databases (*e.g.* NCBI,

UCSC Genome Bioinformatics, Entrez Human Genome Browser, OMIM – see below for database URL information).

Database URLs	
Name	URL
Biocarta	http://www.biocarta.com/
BioCyc	http://www.biocyc.org/
Bimolecular Interaction Network Database (BIND)	http://bind.ca/
Database of Interacting Proteins	http://dip.doe-mbi.ucla.edu/
ECgene EST clustering	http://genome.ewha.ac.kr/ECgene/, also available at http://genome.ucsc.edu/index.html?org=Human
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo/
Human Genome Browser	http://www.ensembl.org/Homo_sapiens/
Interdom	http://interdom.lit.org.sg/help/term.php
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/
Molecular Interactions Database (MINT)	http://mint.bio.uniroma2.it/mint/
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/
Online Mendelian Inheritance in Man (OMIM)	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
OmniViz	http://www.omniviz.com/applications/omni_viz.htm
Pathway Enterprise	http://www.omniviz.com/applications/pathways.htm
Reactome	http://www.reactome.org/
Transpath	http://www.biobase.de/pages/products/transpath.html
UCSC Genome Bioinformatics	http://genome.ucsc.edu/index.html?org=Human
UniGene	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?d

	b=unigene

For some genes the available public annotation was extensive, whereas for others very little was known about a gene's function. A more detailed analysis was therefore performed to characterize genes that corresponded to this latter class. Importantly, the presence of rare splice variants and artifactual ESTs was carefully evaluated. The public data derived from the EST clustering software ECgene (Kim et al., 2005) were also examined for additional gene models. This tool combines genome-based EST clustering and transcript assembly to generate a group of alternatively spliced consensi. This analysis of novel ESTs provided an indication of additional gene content in some regions. The resulting clusters were graphically displayed against the genomic sequence, providing indications of separate clusters that may contribute to the same gene, thereby facilitating development of confirmatory experiments in the laboratory.

A unique consensus sequence was constructed for each splice variant and a trained reviewer assessed each alignment. This assessment included examination of all putative splice junctions for consensus splice donor/acceptor sequences, putative start codons, consensus Kozak sequences and upstream in-frame stops, and the location of polyadenylation signals. In addition, conserved noncoding sequences (CNSs) that could potentially be involved in regulatory functions were included as important information for each gene. The genomic reference and exon sequences were then archived for future reference. A master assembly that included all splice variants, exons and the genomic structure was used in subsequent analyses (*i.e.*, analysis of polymorphisms).

An important component of these efforts was the ability to visualize and store the results of the data mining efforts. A customized version of the highly versatile genome browser GBrowse (http://www.gmod.org/) was implemented in order to permit the visualization of several types of information against the corresponding genomic sequence. In addition, the results of the statistical analyses were plotted against the genomic interval, thereby greatly facilitating focused analysis of gene content.

2. *Computational Analysis of Genes and GeneMaps*

In order to assist in the prioritization of candidate genes for which minimal annotation existed, a series of computational analyses were performed that included basic BLAST searches and alignments to identify related genes. In some cases this provided an indication of potential function. In addition, protein domains and motifs were identified that further assisted in the understanding of potential function, as well as predicted cellular localization.

A comprehensive review of the public literature was also performed in order to facilitate identification of information regarding the potential role of candidate genes in the pathophysiology of schizophrenia. In addition to the standard review of the literature, public resources (Medline and other online databases) were also mined for information regarding the involvement of candidate genes in specific signaling pathways. A variety of pathway and yeast two hybrid databases were mined for information regarding protein-protein interactions. These included BIND, MINT, DIP, Interdom, and Reactome, among others. By identifying homologues of genes in the schizophrenia candidate regions and exploring whether interacting proteins had been identified already, knowledge regarding the GeneMaps for schizophrenia was advanced. The pathway information gained from the use of these resources was also integrated with the literature review efforts, as described above.

3. *Expression Studies*

In order to determine the expression patterns for genes, relevant information was first extracted from public databases. The UniGene database, for example, contains information regarding the tissue source for ESTs and cDNAs contributing to individual clusters. This information was extracted and summarized to provide an indication in which tissues the gene was expressed. Particular emphasis was placed on annotating the tissue source for bona fide ESTs, since many ESTs mapped to Unigene clusters are artifactual. In addition, SAGE and microarray data, also curated at NCBI (Gene Expression Omnibus), provided information on expression profiles for individual genes. Particular emphasis was placed on identifying genes that were expressed in tissues known to be involved in the pathophysiology of schizophrenia disorder and/or schizophrenias.

4. *Polymorphism analysis*

Polymorphisms identified in candidate genes, including those from the public domain as well as those identified by sequencing candidate genes and regions, are evaluated for potential function. Initially, polymorphisms are examined for potential impact upon encoded proteins. If the protein is a member of a gene family with reported 3-dimensional structural information, this information is used to predict the location of the polymorphism with respect to protein structure. This information provided insight into the potential role of polymorphisms in altering protein or ligand interactions, as well as suitability as a drug target. In a second phase of analysis we evaluate the potential role of polymorphisms in other biological phenomena, including regulation of transcription, splicing and mRNA stability, etc. There are many examples of the functional involvement of naturally occurring polymorphisms in these processes. As part of this analysis, polymorphisms located in promoter or other regulatory elements, canonical splice sites, exonic and intronic splice enhancers and repressors, conserved noncoding sequences and UTRs are localized.

Example 6: *SNP and Polymorphism Discovery (SNPD)*

Candidate genes and regions are selected for sequencing in order to identify all polymorphisms. In cases where the critical interval, identified by fine mapping, was relatively small (~50 kb), the entire region, including all introns, is sequenced to identify polymorphisms. In situations where the region is large (>50 kb), candidate genes are prioritized for sequencing, and/or only functional gene elements (promoters, exons and splice sites) are sequenced.

The samples to be sequenced are selected according to which haplotypes contribute to the association signal observed in the region. The purpose is to select a set of samples that covered all the major haplotypes in the given region. Each major haplotype must be present in a few copies. The first step therefore consisted of determining the major haplotypes in the region to be sequenced.

Once a region is defined with the two boundary markers, all the markers used in fine mapping that are located within the region are used to determine the major haplotypes. Long haplotypes covering the whole region are thus inferred using the middle marker as an anchor. The results included two series of haplotype themes that define the major haplotypes, comparing the cases and the controls. This exercise is repeated using an anchor in the peripheral regions to ensure that major haplotype subsets that are not anchored at the original middle marker are not missed.

Once the major haplotypes are determined as described above, appropriate genomic DNA samples are selected such that each major haplotype and haplotype subset are represented in at least two to four copies.

The protocol includes the following steps, once a region is delimited:

1. Primer design

The design of the primers is performed using a proprietary primer design tool. A primer quality control is included in the primer design process. Primers that successfully passed the control quality process were synthesized by Integrated DNA Technologies (IDT). The sense and anti-sense oligos are separated such that the sense oligos are placed on one plate in the same position as their anti-sense counterparts are on another plate. Two additional plates are created from each storage plate, one for use in PCR and the other for sequencing. For PCR, the sense and anti-sense oligos of the same pair are combined in the same well to achieve a final concentration of 1.5 μM for each oligonucleotide.

2. PCR optimization

PCR conditions are optimized by testing a variety of conditions that included varying salt concentrations and temperatures, as well as including various additives. PCR products are checked for robust amplification and minimal background by agarose gel electrophoresis.

3. PCR on selected samples

PCR products to be used for sequencing are amplified using the conditions chosen during optimization. The PCR products are purified free of salts, dNTPs and unincorporated primers by use of a MultiScreen PCR384 filter plate manufactured by Millipore. Following PCR, the amplicons are quantified by use of a lambda/Hind III standard curve. This is done to ensure that the quantity of PCR product required for sequencing had been generated. The raw data was measured against the standard curve data in Excel by use of a macro.

4. Sequencing

Sequencing of PCR products is performed by DNA Landmarks using ABI 3730 capillary sequencing instruments.

5. Sequence analysis

The ABI Prism SeqScape software (Applied Biosystems) is used for SNP identification. The chromatogram trace files were imported into a SeqScape sequencing project and the base calling is automatically performed. Sequences are then aligned and compared to each other using the SeqScape program. The base calling is checked manually, base by base; editing was performed if needed.

Example 7: *Ultra Fine Mapping (UFM)*

Once polymorphisms are identified by sequencing efforts as described in Example 6, additional genotyping of all newly found polymorphisms is performed on the samples used in the fine mapping studies. Various types of genotyping assays may need to be utilized based on the type of polymorphism identified (*i.e.*, SNP, indel, microsatellite). The assay type can be, but is not restricted to, Sentrix Assay Matrix on Illumina BeadStations, microsatellite on MegaBACE, SNP on ABI or Orchid. The frequencies of genotypes and haplotypes in cases and controls are analyzed in a similar manner as the GWS and fine mapping data. By examining all SNPs in a region, polymorphisms are identified that increase an individual's susceptibility to schizophrenia. The goal of ultra-fine mapping is to identify the polymorphism that is most associated with disorder phenotype as part of the search for the actual DNA

polymorphism that confers susceptibility to disorder. This statistical identification may need to be corroborated by functional studies.

Example 8: *Confirmation of Candidate regions and genes in a general population*

The confirmation of any putative associations described in Example 7 is performed in an independent general population patient sample. These DNA samples consist of at least 400 controls and 400 patients with schizophrenia.

All publications, patents and patent applications mentioned in the specification and reference list are herein incorporated by reference in their entirety for all purposes. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in molecular biology, genetics, or related fields are intended to be within the scope of the following claims.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and H (D. N. Glover ed., 4); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis *et al.* U.S. Patent No. 4.683,195; Nucleic Acid Hybridization (B.D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Haines & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. 1. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J.H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu *et al.* eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker,

eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

REFERENCES

Alwine, J. C., D. J. Kemp, *et al.* (1977). "Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes." Proc Natl Acad Sci U S A 74(12): 5350-4.

Alwine, J. C., Reed, S. I., & Stark, G. R. (1977). Characterization of the autoregulation of simian virus 40 gene A. J Virol, 24(1), 22-27.

Anderson, W. F. (1992). "Human gene therapy." Science 256(5058): 808-13.

Barany, F. (1991). "Genetic disease detection and DNA amplification using cloned thermostable ligase." Proc Natl Acad Sci U S A 88(1): 189-93.

Barre, F. X., S. Ait-Si-Ali, *et al.* (2000). "Unambiguous demonstration of triple-helix-directed gene modification." Proc Natl Acad Sci U S A 97(7): 3084-8.

Bourgeois, S. and D. Labuda (2004). "Dynamic allele-specific oligonucleotide hybridization on solid support." Anal Biochem 324(2): 309-11.

Cale, J. M., C. E. Shaw, *et al.* (1998). "Optimization of a reverse transcription-polymerase chain reaction (RT-PCR) mass assay for low-abundance mRNA." Methods Mol Biol 105: 351-71.

Chen, X., L. Levine, *et al.* (1999). "Fluorescence polarization in homogeneous nucleic acid analysis." Genome Res 9(5): 492-8.

Chien, C. T., P. L. Bartel, *et al.* (1991). "The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest." Proc Natl Acad Sci U S A 88(21): 9578-82.

Cohen, A. S., D. L. Smisek, *et al.* (1996). "Emerging technologies for sequencing antisense oligonucleotides: capillary electrophoresis and mass spectrometry." Adv Chromatogr 36: 127-62.

Cotton, R. G., N. R. Rodrigues, *et al.* (1988). "Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations." Proc Natl Acad Sci U S A 85(12): 4397-401.

Cronin, M. T., R. V. Fucini, *et al.* (1996). "Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays." Hum Mutat 7(3): 244-55.

Dean, F. B., S. Hosono, *et al.* (2002). "Comprehensive human genome amplification using multiple displacement amplification." Proc Natl Acad Sci U S A 99(8): 5261-6.

Dillon, N. (1993). "Regulating gene expression in gene therapy." Trends Biotechnol 11(5): 167-73.

Drmanac, R., S. Drmanac, *et al.* (2002). "Sequencing by hybridization (SBH): advantages, achievements, and opportunities." Adv Biochem Eng Biotechnol 77: 75-101.

Elez, R., A. Piiper, *et al.* (2000). "Polo-like kinase1, a new target for antisense tumor therapy." Biochem Biophys Res Commun 269(2): 352-6.

Evans, B. E., K. E. Rittle, *et al.* (1987). "Design of nonpeptidal ligands for a peptide receptor: cholecystokinin antagonists." J Med Chem 30(7): 1229-39.

Fackler, O. T. and B. M. Peterlin (2000). "Endocytic entry of HIV-1." Curr Biol 10(16): 1005-8.

Fakhrai-Rad, H., J. Zheng, *et al.* (2004). "SNP discovery in pooled samples with mismatch repair detection." Genome Res 14(7): 1404-12.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature 340(6230): 245-6.

Frank-Kamenetskii, M. D. and S. M. Mirkin (1995). "Triplex DNA structures." Annu Rev Biochem 64: 65-95.

Freeman, W. M., S. J. Walker, *et al.* (1999). "Quantitative RT-PCR: pitfalls and potential." Biotechniques 26(1): 112-22, 124-5.

Gasparini, P., A. Bonizzato, *et al.* (1992). "Restriction site generating-polymerase chain reaction (RG-PCR) for the probeless detection of hidden genetic variation: application to the study of some common cystic fibrosis mutations." Mol Cell Probes 6(1): 1-7.

Gibbs, R. A., P. N. Nguyen, *et al.* (1989). "Detection of single DNA base differences by competitive oligonucleotide priming." Nucleic Acids Res 17(7): 2437-48.

Glatt, S. J., Wang, R. S., Yeh, Y. C., Tsuang, M. T., & Faraone, S. V. (2005). Five NOTCH4 polymorphisms show weak evidence for association with schizophrenia: evidence from meta-analyses. Schizophr Res, 73(2-3), 281-290.

Griffin, H. G. and A. M. Griffin (1993). "DNA sequencing. Recent innovations and future trends." Appl Biochem Biotechnol 38(1-2): 147-59.

Grossman, P. D., W. Bloch, *et al.* (1994). "High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation." Nucleic Acids Res 22(21): 4527-34.

Guatelli, J. C., K. M. Whitfield, *et al.* (1990). "Isothermal, *in vitro* amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication." Proc Natl Acad Sci U S A 87(5): 1874-8.

Hafner, G. J., I. C. Yang, *et al.* (2001). "Isothermal amplification and multimerization of DNA by Bst DNA polymerase." Biotechniques 30(4): 852-6, 858, 860 passim.

Hardenbol, P., J. Baner, *et al.* (2003). "Multiplexed genotyping with sequence-tagged molecular inversion probes." Nat Biotechnol 21(6): 673-8. Epub 2003 May 5.

Hayashi, K. (1992). "PCR-SSCP: a method for detection of mutations." Genet Anal Tech Appl 9(3): 73-9.

Hermonat, P. L. and N. Muzyczka (1984). "Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells." Proc Natl Acad Sci U S A 81(20): 6466-70.

Hogrefe, H. H., R. L. Mullinax, *et al.* (1993). "A bacteriophage lambda vector for the cloning and expression of immunoglobulin Fab fragments on the surface of filamentous phage." Gene 128(1): 119-26.

Hovatta, I., Lichtermann, D., Juvonen, H., Suvisaari, J., Terwilliger, J. D., Arajarvi, R., Kokko-Sahin, M. L., Ekelund, J., Lonnqvist, J., & Peltonen, L. (1998). Linkage analysis of putative schizophrenia gene candidate regions on chromosomes 3p, 5q, 6p, 8p, 20p and 22q in a population-based sampled Finnish family set. Mol Psychiatry, 3(5), 452-457

Hoogendoorn, B., N. Norton, *et al.* (2000). "Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools." Hum Genet 107(5): 488-93.

Hsu, I. C., Q. Yang, *et al.* (1994). "Detection of DNA point mutations with DNA mismatch repair enzymes." Carcinogenesis 15(8): 1657-62.

Jobs, M., W. M. Howell, *et al.* (2003). "DASH-2: flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays." Genome Res 13(5): 916-24.

Kanazawa, N., M. Kurosaki, *et al.* (2004). "Regulation of hepatitis C virus replication by interferon regulatory factor 1." J Virol 78(18): 9713-20.

Kanehisa, M. (1984). "Use of statistical criteria for screening potential homologies in nucleic acid sequences." Nucleic Acids Res 12(1 Pt 1): 203-13.

Karayiorgou, M., & Gogos, J. A. (1997a). Dissecting the genetic complexity of schizophrenia. Mol Psychiatry, 2(3), 211-223.

Karayiorgou, M., & Gogos, J. A. (1997b). A turning point in schizophrenia genetics. Neuron, 19(5), 967-979.

Kendler, K. S., MacLean, C. J., Ma, Y., O'Neill, F. A., Walsh, D., & Straub, R. E. (1999). Marker-to-marker linkage disequilibrium on chromosomes 5q, 6p, and 8p in Irish high-density schizophrenia pedigrees. Am J Med Genet, 88(1), 29-33.

Kim, N, S. Shin, et al. (2005). "ECgene: Genome-based EST clustering and gene modeling for alternative splicing." Genome Research 15:566-576.

Kohler, G. and C. Milstein (1992). "Continuous cultures of fused cells secreting antibody of predefined specificity. 1975." Biotechnology 24: 524-6.

Kotin, R. M. (1994). "Prospects for the use of adeno-associated virus as a vector for human gene therapy." Hum Gene Ther 5(7): 793-801.

Kozal, M. J., N. Shah, *et al.* (1996). "Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays." Nat Med 2(7): 753-9.

Krebs, S., D. Seichter, *et al.* (2001). "Genotyping of dinucleotide tandem repeats by MALDI mass spectrometry of ribozyme-cleaved RNA transcripts." Nat Biotechnol 19(9): 877-80.

Kremer, E. J. and M. Perricaudet (1995). "Adenovirus and adeno-associated virus mediated gene transfer." Br Med Bull 51(1): 31-44.

Kulinski, J., D. Besack, *et al.* (2000). "CEL I enzymatic mutation detection assay." Biotechniques 29(1): 44-6, 48.

Lewis, D. A., & Lieberman, J. A. (2000). Catching up on schizophrenia: natural history and neurobiology. Neuron, 28(2), 325-334

Livak, K. J., J. Marmaro, *et al.* (1995). "Towards fully automated genome-wide polymorphism screening." Nat Genet 9(4): 341-2.

Lizardi, P. M., X. Huang, *et al.* (1998). "Mutation detection and single-molecule counting using isothermal rolling-circle amplification." Nat Genet 19(3): 225-32.

Mansfield, E. S., M. Vainer, *et al.* (1996). "Sensitivity, reproducibility, and accuracy in short tandem repeat genotyping using capillary array electrophoresis." Genome Res 6(9): 893-903.

Maxam, A. M. and W. Gilbert (1977). "A new method for sequencing DNA." Proc Natl Acad Sci U S A 74(2): 560-4.

McGue, M., & Gottesman, II. (1991). The genetic epidemiology of schizophrenia and the design of linkage studies. Eur Arch Psychiatry Clin Neurosci, 240(3), 174-181.

Miller, A. D. (1992). "Human gene therapy comes of age." Nature 357(6378): 455-60.

Miller, A. D., J. V. Garcia, *et al.* (1991). "Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus." J Virol 65(5): 2220-4.

Mitani, K. and C. T. Caskey (1993). "Delivering therapeutic genes--matching approach and application." Trends Biotechnol 11(5): 162-6.

Muzyczka, N. (1994). "Adeno-associated virus (AAV) vectors: will they work?" J Clin Invest 94(4): 1351.

Myers, R. M., Z. Larin, *et al.* (1985). "Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes." Science 230(4731): 1242-6.

Myers, R. M., N. Lumelsky, *et al.* (1985). "Detection of single base substitutions in total genomic DNA." Nature 313(6002): 495-8.

Nabel, G. J. and P. L. Felgner (1993). "Direct gene transfer for immunotherapy and immunization." Trends Biotechnol 11(5): 211-5.

Nagley, P. and Y. H. Wei (1998). "Ageing and mammalian mitochondrial genetics." Trends Genet 14(12): 513-7.

Nathan, C. and M. U. Shiloh (2000). "Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens." Proc Natl Acad Sci U S A 97(16): 8841-8.

Niu, T., Z. S. Qin, *et al.* (2002). "Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms." Am J Hum Genet 70(1): 157-69.

Orita, M., H. Iwahana, *et al.* (1989). "Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms." Proc Natl Acad Sci U S A 86(8): 2766-70.

Permutt, M. A., Wasson, J. C., Suarez, B. K., Lin, J., Thomas, J., Meyer, J., Lewitzky, S., Rennich, J. S., Parker, A., DuPrat, L., Maruti, S., Chayen, S., & Glaser, B. (2001). A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. Diabetes, 50(3), 681-685.

Pulver, A. E., Lasseter, V. K., Kasch, L., Wolyniec, P., Nestadt, G., Blouin, J. L., Kimberland, M., Babb, R., Vourlis, S., Chen, H., & *et al.* (1995). Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet, 60(3), 252-260

Raap, A. K. (1998). "Advances in fluorescence in situ hybridization." Mutat Res 400(1-2): 287-98.

Remick, D. G. (1995). "Applied molecular biology of sepsis." J Crit Care 10(4): 198-212.

Ronaghi, M., M. Uhlen, *et al.* (1998). "A sequencing method based on real-time pyrophosphate." Science 281(5375): 363, 365.

Rosenecker, J., K. H. Harms, *et al.* (1996). "Adenovirus infection in cystic fibrosis patients: implications for the use of adenoviral vectors for gene transfer." Infection 24(1): 5-8.

Saiki, R. K., T. L. Bugawan, *et al.* (1986). "Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes." Nature 324(6093): 163-6.

Saiki, R. K., P. S. Walsh, *et al.* (1989). "Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes." Proc Natl Acad Sci U S A 86(16): 6230-4.

Saleeba, J. A. and R. G. Cotton (1993). "Chemical cleavage of mismatch to detect mutations." Methods Enzymol 217: 286-95.

Samulski, R. J., L. S. Chang, *et al.* (1989). "Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression." J Virol 63(9): 3822-8.

Sanger, F., S. Nicklen, *et al.* (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A 74(12): 5463-7.

Sauter, E. R., M. Herlyn, *et al.* (2000). "Prolonged response to antisense cyclin D1 in a human squamous cancer xenograft model." Clin Cancer Res 6(2): 654-60.

Simon, H., I. Fortsch, *et al.* (1999). "Triple helix formation inhibits DNA gyrase activity." Antisense Nucleic Acid Drug Dev 9(6): 527-31.

Sommerfelt, M. A. and R. A. Weiss (1990). "Receptor interference groups of 20 retroviruses plating on human cells." Virology 176(1): 58-69.

Sterman, D. H., J. Treat, *et al.* (1998). "Adenovirus-mediated herpes simplex virus thymidine kinase/ganciclovir gene therapy in patients with localized malignancy: results of a phase I clinical trial in malignant mesothelioma." Hum Gene Ther 9(7): 1083-92.

Storm, N., B. Darnhofer-Patel, *et al.* (2003). "MALDI-TOF mass spectrometry-based SNP genotyping." Methods Mol Biol 212: 241-62.

Takahashi, S., Faraone, S. V., Lasky-Su, J., & Tsuang, M. T. (2005). Genome-wide scan of homogeneous subtypes of NIMH genetics initiative schizophrenia families. Psychiatry Res, 133(2-3), 111-122.

Tsuang, M. T. (1998a). Genetic epidemiology of schizophrenia: review and reassessment. Kaohsiung J Med Sci, 14(7), 405-412.

Tsuang, M. T. (1998b). Recent advances in genetic research on schizophrenia. J Biomed Sci, 5(1), 28-30.

Tsuang, M. T., Nossova, N., Yager, T., Tsuang, M. M., Guo, S. C., Shyu, K. G., Glatt, S. J., & Liew, C. C. (2005). Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. Am J Med Genet B Neuropsychiatr Genet, 133(1), 1-5

Tornillo, L., Carafa, V., Richter, J., Sauter, G., Moch, H., Minola, E., Gambacorta, M., Bianchi, L., Vecchione, R., & Terracciano, L. M. (2000). Marked genetic similarities

between hepatitis B virus-positive and hepatitis C virus-positive hepatocellular carcinomas. J Pathol, 192(3), 307-312.

Tratschin, J. D., M. H. West, *et al.* (1984). "A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase." Mol Cell Biol 4(10): 2072-81.

Walker, J. M. (1998). Pathomechanics and classification of cartilage lesions, facilitation of repair. J Orthop Sports Phys Ther, 28(4), 216-231.

West, A. B., Lockhart, P. J., O'Farell, C., & Farrer, M. J. (2003). Identification of a novel gene linked to parkin via a bi-directional promoter. J Mol Biol, 326(1), 11-19.

West, M. H., J. P. Trempe, *et al.* (1987). "Gene expression in adeno-associated virus vectors: the effects of chimeric mRNA structure, helper virus, and adenovirus VA1 RNA." Virology 160(1): 38-47.

Wilson, C., M. S. Reitz, *et al.* (1989). "Formation of infectious hybrid virions with gibbon ape leukemia virus and human T-cell leukemia virus retroviral envelope glycoproteins and the gag and pol proteins of Moloney murine leukemia virus." J Virol 63(5): 2374-8.

Yajnik, V., Paulding, C., Sordella, R., McClatchey, A. I., Saito, M., Wahrer, D. C., Reynolds, P., Bell, D. W., Lake, R., van den Heuvel, S., Settleman, J., & Haber, D. A. (2003). DOCK4, a GTPase activator, is disrupted during tumorigenesis. Cell, 112(5), 673-684.

Yang, B., Wen, X., Kodali, N. S., Oleykowski, C. A., Miller, C. G., Kulinski, J., Besack, D., Yeung, J. A., Kowalski, D., & Yeung, A. T. (2000). Purification, cloning, and characterization of the CEL I nuclease. Biochemistry, 39(13), 3533-3541

Yu, M., E. Poeschla, *et al.* (1994). "Progress towards gene therapy for HIV infection." Gene Ther 1(1): 13-26.

Books

Austen BM and. Westwood OMR. Protein Targeting and Secretion. Oxford: IRL Press; 1991. 85 p.

Bishop MJ, editor. Guide to Human Genome Computing, 2d ed. San Diego: Academic Press; 1998. 306 p.

Cowell IG, Austin CA, editors. DNA Library Protocols. Methods in Molecular Biology. Vol. 69 Totowa, N.J.:Humana Press; 1997. 321p.

Freshney RI, editor. Animal Cell Culture: A Practical Approach. Oxford: IRL Press; 1986.

Freshney RI. Culture Of Animal Cells: A Manual of Basic Technique. New York: AR Liss; 1987. 397 p.

Glover DM, editor. DNA Cloning: A Pratical Approach. Vols 1 & 2. Oxford; Washington: IRL Press; 1985.

Gribskov M, Devereux J, editors. Sequence Analysis Primer. Oxford University Press; 1994. 296 p.

Griffin AM, Griffin HG, editors. ComputerAnalysis of Sequence Data, Part 1. Totowa, N.J. Humana Press; 1994. 392 p.

Hames BD, Higgins SJ, editors. Nucleic Acid Hybridization: A Practical Approach. Oxford: IRL Press; 1985. 245 p.

Hames BD, Higgins SJ, editors. Transcription and Translation: A Practical Approach. Oxford: IRL Press; 1984. 328 p.

Harlow Ed, Lane D. Antibodies: A Laboratory Manual. New York: Cold Spring Harbor Laboratory; 1988. 726 p.

Heinje G. von. Sequence Analysis in Molecular Biology. San Diego: Academic Press; 1987. 188 p.

Hogan B, Costantini F, Lacy E, editors. Manipulating the Mouse Embryo: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press; 1986. 332 p.

Huber BE, Carr BI. Molecular and Immunologic Approaches. Mt. Kisco, NY: Futura Publishing Co; 1994.

Jones J. Amino Acid and Peptide Synthesis. Oxford; New York: Oxford Science Publications; 1992. 86 p.

Kaufman PB, William W, Donghern K, editors. Handbook of Molecular and Cellular Methods in Biology and Medicine. Boca Raton: CRC Press; 1995. 484 p.

Lesk AM, editor. Computational Molecular Biology: sources and methods for sequence analysis. New York: Oxford University Press; 1988. 254p.

Male D, Cooke A, Owen M, Trowsdale J, Champion B, editors. Advanced Immunology. 3rd ed. London; Baltimore: Mosby; 1996. 273 p.

McPherson MJ, editor. Directed Mutagenesis: A Practical Approach. New York: IRL Press; 1991. 257 p.

McPherson MJ, Quirke P, Taylor JR, editors. PCR: A Practical Approach. Oxford; New York: IRL Press; 1991. 253 p.

Miller JH, Calos MP, editors. Gene Transfer Vectors for Mammalian Cells. New York: Cold Spring Harbor Laboratory Press; 1987. 169 p.

Miller JH, Calos MP, editors. Gene Transfer Vectors For Mammalian Cells. New York: Cold Spring Harbor Laboratory; 1987. 169p.

Morrison James, editors Guilford Publications; 1995.

Pawlowitzki IH, Edwards JH, Thompson EA, editors. Genetic Mapping of Disease Genes. Academic Press London; 1997. 288 p.

Perbal BV. A Practical Guide to Molecular Cloning. 1st ed. New York: Wiley Interscience Publication; 1984. 554 p.

Perbal BV. A Practical Guide To Molecular Cloning. New York: Wiley; 1984. 554 p.

Peruski LF, Peruski AH. The Internet and the New Biology. Tools for Genomic and Molecular Research. Washington, D.C.: American Society for Microbiology Press; 1997.

Sambrook J. Molecular Cloning: A Laboratory Manual. 2nd ed. 3 vols. New York: Cold Spring Harbor Laboratory Press; 1989.

Sell S. Immunology, Immunopathology & Immunity. 5th ed. Stamford, CT: Appleton & Lange; 1996. 1014 p.

Smith DW, editor. Biocomputing. Informatics and Genome Projects, New York: Academic Press; 1993. 336p.

Stites DP, Terr AT, editors. Basic and Clinical Immunology. 7th ed. Norwalk, CT: Appleton & Lange; 1991. 870 p.

Walker JM. Protein Protocols on CD-ROM, Humana Press, Totowa, NJ.

Weir DM, Herzenberg LA, Blackwell C, editors. Handbook Of Experimental Immunology. 4 vols. Oxford: Blackwell; 1986.

Woodward J. Immobilized Cells And Enzymes: A Practical Approach. Oxford: IRL Press; 1986.

Wu R, Grossman L, editors. Methods in Enzymoloqy: Recombinant DNA Part E. Vol. 154. Amsterdam: Elsevier Science; 1987. 576 p.

Wu R, Grossman L, editors. Methods in Enzymoloqy: Recombinant DNA Part F. Vol. 155. Amsterdam: Elsevier Science; 1987. 628 p.

Patent

WO 30/42661A2

WO 1997US0005216

PCT/US94/05700

WO 0181628A1

WO 94/16101

WO 93/24641

U.S. Patent No. 60/335,068

U.S. Patent No. 5,869,242

U.S. Patent No. 6,090,606

U.S. Patent No. 5,888,819

U.S. Patent No. 5,459,039

U.S. Patent No. 5,498,531

U.S. Patent No. 5,807,718

U.S. Patent No. 4,683,195

U.S. Patent No. 4,683,202

U.S. Patent No. 4,797,368

U.S. Patent No. 6,090,543

U.S. Patent No. 6,107,065

U.S. Patent No. 4,952,501.

U.S. Patent No. 20040009479A1

U.S. Patent No. 5,315,000

U.S. Patent No. 5,585,089

U.S. Patent No. 6,090,606

U.S. Patent No. 5,869,242

U.S. Patent No. 6,479,244

U.S. Patent No. 4,797,368

U.S. Patent No. 5,173,414

Table 1. List of Schizophrenia candidate regions identified from the genome wide scan and Fine Mapping association analyses. The first column denotes the region identifier. The second and third columns correspond to the chromosome and cytogenetic band, respectively. The fourth and fifth columns correspond to the chromosomal start and end coordinates of the NCBI genome assembly derived from build 35 (B35).

Region	Chromosome	Cytogenetic Band	B35 Start	B35 End
1	1	1p32.2	55851773	56007133
2	1	1p31.1	75819327	75964032
3	1	1q32.1	201596441	202046899
4	2	2p16.1	60119625	60217515
5	2	2q14.1	115289729	115654508
6	2	2q32.1	183286723	183338600
7	2	2q37.2	236419935	236549343
8	3	3p26.1	6425027	6497659
9	3	3p12.3	76609002	76654002
10	4	4p15.33	14244317	14351953
11	5	5q13.3	75934367	75985978
12	5	5q21.3	106905473	106953865
13	6	6q23.2	133696033	133771621
14	7	7p15.2	26476334	26594560
15	7	7q21.3	97454046	97711910
16	7	7q35	144659113	144705856
17	10	10p12.31	20042742	20176365
18	10	10q21.1	53978045	54075328
19	11	11q21	92824911	92937042
20	11	11q24.3	128340968	128702726
21	13	13q21.2	60636792	60893563
22	13	13q31.1	81061202	81120127
23	14	14q21.3	46953692	47137969
24	16	16q22.3	72104472	72149472
25	20	20p11.21	25536407	25821023
26	X	Xp22.31	8227264	8331203
27	X	Xq21.32	92543632	92746053

Table 2. Schizophrenia genome wide association study results in the Quebec Founder Population (QFP). SNP markers found to be associated with Schizophrenia from the analysis of genome wide scan (GWS) data. Columns include: Region ID; Chromosome; Build 35 location in base pairs (bp); rs#, dbSNP data base (NCBI) reference number; Sequence ID, unique numerical identifier for this patent application; Sequence, 21 bp of sequence covering 10 base pair of unique sequence flanking either side of central polymorphic SNP; - log10 P values for GWS, - log10 of the P value for statistical significance from the GWS (see Example section) for the most highly associated multi-marker haplotypes centered at the reference marker and defined by the sliding windows of

						Poolex			
Region ID	Chr	B35 Position	RS#	Seq ID	Flanking Sequence	W01	W05	W09	W15
1	1	55851773	9436525	502	TTTAATTTGTRTTTGGCTCAC	1.032	1.689	1.996	2.062
1	1	55852379	9326036	503	TACTCATAACRAATGACTTAA	1.180	3.406	2.308	2.027
1	1	55864746	2767503	504	CATCCTTTTAKGGTTCTTCAC	2.909	3.143	2.566	1.813
1	1	55875479	9326037	505	TATTATTCACRTCCCTCTTCT	2.901	2.861	2.963	2.823
1	1	55882116	17112658	506	CCACTAAGAGKTCATTGGAAA	0.008	2.496	2.988	3.842
1	1	55890082	6672047	507	GCTGTGTTCTKAACAGGAATC	0.292	2.308	2.812	3.466
1	1	55899881	2671226	508	GATAAATATGWCACATTGACT	1.265	0.786	3.196	3.733
1	1	55913045	2767499	509	TCATGTTTCCYAGAGCCTTCA	0.954	1.014	3.842	4.210
1	1	55920285	1165512	510	TGCATTTCAARCTCTTTGTGG	0.000	2.700	3.456	4.210
1	1	55927431	1165504	511	CTGCCAGCTGRATCAGAAGAG	0.305	4.687	4.210	3.842
1	1	55936001	17112758	512	TTCCAATTTGRGTTGACTGAC	3.195	3.645	4.210	3.365
1	1	55946188	17112768	513	AACACAGTTTYGAAAAATTAG	3.842	3.289	4.210	3.365
1	1	55955692	-	514	GTATTTACCCYTGCTGTCCCC	0.008	3.511	4.210	3.842
1	1	55981978	17112821	515	ATATTAACCTWAACTCCAAAC	1.048	3.456	3.325	3.325
1	1	55993259	10888939	516	TCTTATAACAMCTCAGTTAAA	0.471	1.550	3.195	3.034
1	1	56000137	12077235	517	GCCCAATGAGRCCCATAGACA	2.597	1.812	2.444	2.800
1	1	56007133	12092701	518	CTACAGTTTCYATCTCGTCAT	0.119	1.357	0.856	2.800
2	1	75819327	814873	519	ATTTGGGTCAYGGTTCAAAAT	0.113	1.662	1.182	2.386
2	1	75828083	1739897	520	ACTGAAAACCWTTATTTTGGG	1.533	1.213	1.153	2.472
2	1	75847967	7552404	521	CTAAAAGAAARCCTTCCCAAT	1.262	0.970	2.737	2.881
2	1	75871299	699825	522	GTTTCTTATCSCAAGGCAGTA	0.082	2.916	2.737	3.541
2	1	75884005	1498317	523	ACTTGTTATTMGAGAACACAG	0.709	3.074	3.074	3.573
2	1	75888530	11164037	524	GTGCCAAAGTWTAATACGGTC	4.687	2.887	3.325	3.388
2	1	75901466	1251079	525	ACTTTGAATAYAGCGCTTAAT	1.508	3.255	3.842	3.408
2	1	75907563	1313071	526	TTTTCAAGCASTGTGAGGATA	1.068	3.842	3.733	3.541
2	1	75922498	11161511	527	GCCGGGTCTCYGTTTTATATT	1.435	2.401	3.733	3.408
2	1	75931822	1694419	528	TTTTTAATTGRTAGGTACAGT	1.320	2.467	3.573	3.325
2	1	75939299	12073631	529	TGGTACAGACRTGTGTATTAC	1.554	2.842	2.354	3.511
2	1	75947105	6664254	530	TTCTTTGAAGKGGGATCATTT	0.000	2.454	2.071	3.511
2	1	75954184	1146614	531	TAAAATAAATYTGTGGAAATT	0.000	1.569	2.237	3.409
2	1	75964032	1146634	532	CAGACAGGGTRGAGTAGGGTT	1.157	0.428	1.958	2.449
3	1	201596941	10900430	533	AGCTGAATGCMGAGAACTGAG	2.446	1.084	0.844	0.471
3	1	201602715	2802850	534	GGCACCTGTGSAGGTTCTGTT	0.215	1.124	0.822	0.790
3	1	201608623	17336419	535	CTTGGGCTCAKGATGCCTTGG	1.069	1.286	0.564	0.730
3	1	201616694	16854693	536	CTGTAAAGAAYCCCCAGACAG	0.117	0.306	0.843	0.795
3	1	201626873	2802851	537	CAAGATCTTARCCCTAGGTCT	0.224	0.204	1.034	0.689
3	1	201633065	2802854	538	TATAACAACCRTCTGAAAAAG	0.390	0.132	0.421	0.515
3	1	201642233	2802829	539	TGGCTTCCTARTTCTATCAAG	0.143	0.533	0.530	0.622
3	1	201651151	2066306	540	TCACTGGGTCYGTTCCCCAAC	0.348	0.482	0.265	0.683
3	1	201658030	17415030	541	TGAACAGATCRTGGTCCCTGG	1.415	0.590	0.262	0.214
3	1	201664462	6685118	542	TGTCCTCAATYCAGGGTAGGA	0.179	0.545	0.301	0.179
3	1	201675604	2246662	543	GCCTGGAGGTMAAAGGTAAAG	0.647	0.421	0.258	0.199
3	1	201686084	-	544	CAAGCTCTGCYACCATGCAGC	0.028	0.114	0.238	0.867
3	1	201692587	857450	545	ATGGGGGCCCRGCACCGGCTC	0.052	0.134	0.177	1.481
3	1	201706165	16854930	546	GTATTGCCCGYTTCTGAGTGG	0.483	0.029	0.138	1.612
3	1	201714370	2802846	547	TTCTATACCCRTGCTCCTGTC	0.281	0.029	1.152	2.045
3	1	201720684	-	548	AGAGAGGCACRGGCTCCCCTG	0.133	0.351	1.710	1.964
3	1	201730032	4951155	549	CAGTACTTTAYGCCTCTTGTA	0.054	1.717	1.943	3.064
3	1	201738263	-	550	TAAATGAATGYATGTAGGTAT	1.275	2.204	2.259	2.988
3	1	201754973	9787334	551	GGACTCTCCAYGGTCCTGGAG	2.874	2.432	2.107	2.954

3	1	201760531	871423	552	CTGGACAGTAMCTCTCTGGGT	1.878	3.043	3.784	3.182
3	1	201773122	6662930	553	CTTTGGATTTYCATGGGCTTT	0.000	2.723	3.608	3.043
3	1	201780273	16855053	554	TCCTGAATGTSTCAGCAAGTT	1.129	3.386	3.764	3.165
3	1	201786819	2197909	555	CTTTCCACACYGAAAGATCTA	0.284	2.388	3.608	3.609
3	1	201792849	10158848	556	CCCTCCAAGAYGACCACGTTA	2.265	2.747	2.923	3.210
3	1	201798691	2029139	557	CATCAAGTATYTCGGGGGTCA	0.279	3.024	2.838	3.272
3	1	201804282	17416419	558	AAGTCTCAGTRCTGTAAGTTA	1.573	2.547	2.829	3.272
3	1	201831616	3852948	559	AATAAACAACMACAGGAACAA	2.015	1.997	2.706	3.024
3	1	201839154	10751430	560	GTCACCTGGCRCCTGTCCTGA	0.388	2.553	2.894	2.413
3	1	201846756	11240369	561	CACTTCATGARGAACCCTCCA	2.446	2.105	2.041	2.179
3	1	201854164	-	562	TACATCTCTGMTCAAACCTTA	1.024	1.993	2.213	2.864
3	1	201859716	10751431	563	CTTTATCCTCYTACAGCAAAA	0.045	2.399	1.689	3.182
3	1	201867091	4950984	564	TGTGAATATCRTGATACCTAG	1.328	0.930	1.156	2.971
3	1	201878117	11240376	565	ACACAACTATRTATTAGTAAT	0.847	0.517	2.390	2.923
3	1	201888765	6593929	566	AGTGAGAAAARGCTCTTCACA	0.219	0.629	2.346	2.864
3	1	201896307	12129162	567	CTCATTTAAGRCAAACAGATG	0.075	1.626	2.723	2.784
3	1	201904758	11240385	568	GATTACCCCCRGAAAGGAAGA	0.348	2.480	2.954	3.483
3	1	201911654	17345153	569	CCTCTGCTCCRTCTCTAAACT	3.842	3.155	2.404	3.344
3	1	201922859	7531256	570	CAACTGGACTKTAAGGCAGGG	3.143	3.687	2.687	3.182
3	1	201938152	1172161	571	GGAATGTTGGYGACCTAGGCT	2.427	4.085	3.366	4.386
3	1	201944581	1172140	572	GACAGACCCARCTGACATGTA	0.058	2.566	3.842	3.909
3	1	201953518	1172150	573	AAATTTAGCTRCTTTTTAGCG	0.251	2.662	3.842	3.909
3	1	201964104	1151784	574	AGGATTTGCAYGCCCTCAGAT	1.598	2.525	3.909	3.541
3	1	201971436	1668869	575	TCACATAACASTCATCTTACG	3.645	2.931	4.386	3.687
3	1	201976925	9326600	576	CAACACCCCARCTCCACAAGT	1.198	3.909	3.107	4.386
3	1	201990099	1172125	577	CATCCTGTTGWCTAACCCAAA	1.314	3.608	3.024	3.240
3	1	201994779	17345837	578	AGGTCTTCAGRTGAGCCCGGC	3.784	2.773	2.988	3.240
3	1	202003987	16855350	579	GACTCTGTCGYGGGCATTCCC	0.271	2.188	3.131	3.034
3	1	202011269	17345997	580	GTTCTTATGTYTGTTGAATCT	0.138	1.942	2.140	3.265
3	1	202019092	11801560	581	GCTGGGAAAGRCAGCGGGAGC	0.022	0.336	2.348	3.687
3	1	202038247	-	582	CACCTCCCACRGCGAGGAGGC	0.610	0.437	2.511	3.408
3	1	202046899	11590464	583	TTCGAATTCTSTGTACTTACA	1.046	1.049	0.914	2.784
4	2	60119625	17050506	584	GCTCTTCGTGWTATATTCTGA	0.177	0.239	1.265	1.844
4	2	60125286	17050520	585	CTCTAGGTCTRTCTGCCTTCC	0.560	0.596	2.408	2.009
4	2	60136951	-	586	GAAGGCAGTCSTGGTGAGCTG	0.326	2.377	2.714	2.028
4	2	60145978	11125823	587	TCTTCTCATGYCCCCTCTCAT	1.286	3.909	2.714	1.799
4	2	60165881	13003893	588	TGAGTGGTCASTAGGCAAATG	3.386	4.386	3.165	2.074
4	2	60172290	2215897	589	GAATATCCAGMAGCAATGCCT	3.096	4.386	4.086	2.100
4	2	60182556	4672357	590	TTCCTCGTTTKTAAAGAGTTC	0.269	3.608	3.240	1.913
4	2	60191069	994972	591	AAGGAAACTTYCTGCCTGGCC	0.442	1.729	3.240	2.034
4	2	60206928	17030159	592	CATGATTCACRGATGGAGTGA	0.342	0.199	2.842	2.472
4	2	60217515	17011925	593	GGGGAGATACYCTATAAAATA	0.351	0.378	0.979	2.737
5	2	115289729	908931	594	AGACTAATAAYTTCATCTAGT	1.408	1.126	1.703	2.057
5	2	115299278	17043624	595	TATAATTTCARTCATGCATAC	0.357	1.087	2.566	2.563
5	2	115307588	7558870	596	AGCCTGGCACRTGATAAGAAT	0.666	2.494	2.742	2.681
5	2	115318871	17731643	597	CCAAGTATGTRTGATTCCAAC	0.608	2.380	3.307	3.386
5	2	115325639	11695954	598	CACGCATTGCMAGGAGACTTG	2.525	2.415	3.165	3.483
5	2	115333911	17732092	599	GATTGCTGAGRAACCTGGCTT	1.212	3.119	3.024	3.483
5	2	115343640	9308704	600	GTTTAAGGGAYGGCCTCCAGA	0.990	3.096	3.143	3.386
5	2	115353658	10192492	601	GAGCTTCCCTYAAGGCTGGAT	2.939	3.165	3.168	3.541
5	2	115359858	12053455	602	TCCATTATAARCTTATAGGTC	1.161	2.728	3.195	3.386
5	2	115369080	11867172	603	CAATCTATGAMGAATTACACA	1.981	2.823	3.289	3.511
5	2	115384367	1401755	604	TTTTTACTACWTTCTGATTTC	1.017	2.550	3.408	3.573
5	2	115395748	-	605	TTCTGAGCCAYCTACTCTACT	0.006	2.641	3.456	3.645
5	2	115407040	11691241	606	AGAATTATGAYAGATCCAGTA	1.416	2.778	3.365	3.733
5	2	115414003	10496481	607	ATGCACAATCKAATTTTTTCC	2.083	2.789	3.325	3.325
5	2	115421231	17043823	608	ATTTGAGAGTYAATTTTTTTA	3.396	3.307	3.325	3.573
5	2	115437126	2175176	609	GGACACAAATKTCAGTATCTG	1.946	3.131	3.408	3.733
5	2	115443201	6742611	610	TAAGGTCATTKTTTAAATTTT	1.462	3.645	3.256	3.609
5	2	115446541	10460456	611	AATCAAGAAAKTGTTCTATGA	0.000	2.456	3.224	4.085
5	2	115461547	17452458	612	AATATCTTGGRTCTTATTGAG	1.324	1.771	3.784	3.909

5	2	115467666	-	613	CTATTATAATYGAAGTCCAAA	0.394	2.186	3.156	4.085
5	2	115474886	7423711	614	TGTTTTACTCWACCTGTAACA	0.862	3.006	3.541	4.085
5	2	115483935	1396921	615	CCAATTAAGTRAAGCTAATCC	2.508	2.386	2.867	4.085
5	2	115510653	2175175	616	CATTAGAATCKAATTAAAATA	3.386	3.043	2.773	4.085
5	2	115521332	-	617	TGTCTTTTTCRTTATTTTAAA	0.025	3.240	3.608	3.902
5	2	115532844	17452532	618	AATCCATTCGYGTCTATCTTG	2.413	2.573	3.541	3.902
5	2	115539945	11892585	619	GGCAGAAGCTKTTTATCTTAG	0.000	3.687	4.366	3.902
5	2	115554230	17635169	620	GGCTCATGAGYACAGCTGGAA	2.089	3.784	3.608	4.085
5	2	115568001	1876861	621	ACTTTGATGTRCAAATTAAAA	4.386	3.608	3.541	4.085
5	2	115574437	12467030	622	GCTTAGGTATMCAAAGTGGCG	0.137	3.784	3.764	4.085
5	2	115581028	17040479	623	ATTTGTGTATKATAACATGAA	0.346	4.085	3.687	3.784
5	2	115592076	6747204	624	GTAGTCTGAAMCAATATTATC	0.913	2.491	4.085	3.483
5	2	115598082	7579207	625	AGACCATTCARTGACTTCCTA	3.645	2.459	4.366	3.541
5	2	115611026	17044133	626	TTCTGATAGGSAGTGATAGAA	0.479	3.168	2.113	3.432
5	2	115619855	7569587	627	GTCAGATACAYCATTATTATG	0.039	2.823	2.649	3.483
5	2	115630326	13021135	628	ATTAGTCAAGMTCTTACAGTA	1.212	0.758	2.550	3.609
5	2	115645471	10445863	629	ACAAAGCCTTYGGTGGTGTCA	1.039	0.843	2.044	1.614
5	2	115654508	6738642	630	ATGACATCCTSTCACTAGAAA	0.288	0.982	0.416	1.689
6	2	183286723	17265503	631	ATTTGAGCTAYTGATCCTGTT	0.570	0.654	1.946	1.362
6	2	183294740	13031786	632	CTTAAAATGASAAAAGTAATC	0.411	2.333	1.705	1.279
6	2	183303162	1547586	633	TTATTCGTTTYTGCAGTTTTT	0.941	2.687	1.804	1.262
6	2	183311989	13403435	634	GGGTTTCAAARTATTTAACTC	4.386	2.881	1.770	1.228
6	2	183324819	13023449	635	TTTTCTGAGAYGCTAAGGATT	0.025	2.823	1.856	1.255
6	2	183331487	13432321	636	ACTAGGCATAYGGGGGAAAAG	0.440	2.200	1.668	1.272
6	2	183338600	13399625	637	GCAAGAGCTTRAGTAGATGCA	0.457	0.226	1.480	1.245
7	2	236419935	7579748	638	GTGTTTGGGCYGTCTAGAACC	0.615	1.097	3.240	2.714
7	2	236457082	2696387	639	CTGCGCAAATKTTTTCCTTTT	0.822	1.519	3.240	2.773
7	2	236469334	6761949	640	ATGATAGAGGYCCTGCCAGTG	1.926	3.687	3.307	2.985
7	2	236473982	1878154	641	ATTTAGATGTRTGCTTAATGT	0.830	3.493	3.107	3.131
7	2	236491123	6760477	642	CAAAAATGCTRCCAGCCTGTC	4.388	3.687	3.541	3.240
7	2	236524181	908444	643	TCTCCTCATCRCCCCGCCCCC	0.000	3.573	3.182	3.386
7	2	236549343	4663621	644	TAATGTATCCRTCTACCATGA	0.402	3.365	3.483	3.432
8	3	6425027	11917330	645	TGAAAGATACRTAATTTGTTA	0.037	0.025	1.806	1.755
8	3	6434012	155420	646	CACAGTATTAYGGCCTTGCAG	0.048	2.286	2.252	1.903
8	3	6439294	155421	647	TGATCTTTTAYTACAGAGCAT	0.545	2.752	2.333	1.811
8	3	6450006	156488	648	TGGTGAACAARTGAATTGCAT	3.909	2.829	2.237	1.850
8	3	6460400	266409	649	ATAGACAGGCRTACAGATGTA	1.874	2.909	2.462	1.887
8	3	6473283	425914	650	CTTTATCTCAYGTGCAATAAC	1.507	2.705	2.566	1.834
8	3	6474205	432398	651	AATCAGAAAGYGTGTTAATGG	1.362	1.681	2.678	2.026
8	3	6488129	17824633	652	TAGGCTGCTARTAACAAAAAT	0.135	1.150	2.399	1.906
8	3	6497659	17288247	653	AATCTAAGGCRCACAGATTCC	0.877	0.564	1.254	1.936
9	3	76613862	-	654	AAGTCAGAATSCCCAACCAGA	0.218	1.681	1.952	0.969
9	3	76621782	6789109	655	ACATTAGGACRTGCCTTAAAT	0.262	2.254	1.872	1.274
9	3	76631502	17014282	656	GGTAACTATTYGACCAAACAC	4.687	2.077	1.552	1.244
9	3	76651471	3849502	657	TCACTTTTGTRTACCCGGGGA	1.013	1.936	1.440	1.472
10	4	14244317	16890676	658	ACATAATTCCKTGTAAATAAA	0.261	1.074	1.420	1.627
10	4	14251039	16890707	659	ACACATCTTGWCATGACTGGC	1.698	1.085	2.467	2.563
10	4	14257831	17354181	660	CAAATGTAAAYGGTAATATGA	0.130	1.388	2.316	3.064
10	4	14263213	13148803	661	TCTTCTAGTGWCTTCTGCCAA	1.198	2.686	1.938	3.169
10	4	14268676	959541	662	AATAAAAGTAYACCTAGGCAG	1.046	2.260	3.408	3.255
10	4	14274320	13105355	663	AGCCCACATCRTCTTCTCCTG	2.404	2.383	3.198	3.511
10	4	14281206	4383614	664	TTCTACTCCTYTTGAATTTCC	0.198	3.842	3.168	3.408
10	4	14296397	16890866	665	TATAATAAGTWCAGGTCAATA	0.390	3.265	3.096	3.325
10	4	14302700	16890883	666	GGTAGGAATAYGGTGTCCTTA	4.085	2.931	3.325	3.224
10	4	14306017	4358395	667	CTTCATAAGAMGACAAGAGAG	1.543	2.649	3.456	3.119
10	4	14322707	16890892	668	CATTCGCCTTSAGCAACAGCG	0.141	2.916	2.718	3.015
10	4	14329096	6449061	669	TGGATTAGTAYATTGAAAGAA	0.724	1.516	2.811	3.143
10	4	14332793	12649601	670	TCAGAACAATRGGAGATGGAA	1.760	0.832	2.486	2.946
10	4	14351953	16890931	671	CTGTCATCCARGTAACACTCA	0.449	0.912	0.777	1.680
11	5	75934367	16873519	672	TTGAGAAACAYGGCTCTGTCC	0.103	0.060	2.819	1.543
11	5	75949061	2069685	673	TGAAGCATATYTCTTAGGAGC	0.000	0.118	1.932	1.968

11	5	75955505	2069641	674	GTGAGGATGGYGTTTGTTCAG	0.007	2.742	1.897	2.214
11	5	75962568	3797400	675	TGATGACAACRAATTGTAATT	0.686	2.285	1.744	1.669
11	5	75968473	2455229	676	TTCATTCCATRTTAGTGGGTC	4.085	2.198	2.100	1.067
11	5	75974377	2455234	677	GAGGCCAGCTRTACTCCGGCT	0.371	2.285	1.946	0.981
11	5	75985978	3815774	678	TAATGGATCTKAGAACTTCAT	0.348	2.369	1.807	0.628
12	5	106905473	163893	679	TCTGTTCTACYATCTTTAAAC	0.019	1.839	1.493	1.464
12	5	106915238	17413725	680	TAGCAACTCCSTGCCGATCTC	1.401	2.109	1.374	1.378
12	5	106924516	2484103	681	AGTTCTGTAAYGTGAGCGCCA	4.687	2.635	1.359	1.475
12	5	106941488	597083	682	TTTTTGGAGCRCGGGGGTTCA	0.163	2.705	1.931	1.365
12	5	106953865	3844311	683	TATTCTAAACYCTACCTTCCA	0.885	1.736	1.852	1.316
13	6	133696033	6924629	684	GAAGGCAAGGRTCATGCTGAA	1.079	1.099	0.966	4.687
13	6	133705030	9375956	685	TGGGTTTCTTYGGTAGCAGGG	0.817	1.248	2.576	3.989
13	6	133722330	7743259	686	TGATTTCAGCWATTCCCTTAT	1.539	1.304	3.224	3.733
13	6	133730946	212826	687	TACCATTTCTSTGTTTCTAGG	0.429	3.285	4.210	2.946
13	6	133737200	9402508	688	TCCTGGAAGGYCCTCGCATTA	0.277	4.687	3.988	2.464
13	6	133750355	211602	689	GGTTGTACAGWCCACAGAATT	4.386	3.365	3.511	2.117
13	6	133756903	3777811	690	CAATAAGCCARTGGATGAATG	1.292	3.119	2.823	2.180
13	6	133764928	2327322	691	GATATCTAGAYCATTTCAGCA	1.469	2.901	2.223	2.568
13	6	133771621	211623	692	ATTCATGGTAKTCACCAGACA	0.133	1.125	1.762	2.415
14	7	26476334	10271515	693	AGACCTTGTAYGTACTTATGT	0.000	0.780	1.769	2.909
14	7	26483805	10231161	694	AAGGTATAAASGCAAAATATT	0.916	2.272	2.580	3.043
14	7	26487530	7786456	695	TAGTGGTGTAYGGAGATCAAT	0.109	1.909	2.530	3.784
14	7	26499560	212853	696	GATTAAATCARTTGGTAAATG	2.615	3.155	3.107	3.609
14	7	26510882	10277439	697	TATCTAGTGAYAAATAGGCTG	0.287	2.653	3.541	3.240
14	7	26520304	2299102	698	TAGCACAGAAYCTTGCCTATA	3.053	3.784	4.366	3.344
14	7	26527922	17310851	699	CCTTCTGGTAYAGTGTTAGGA	0.032	3.107	4.386	3.687
14	7	26536185	1024454	700	CTCAGTTACAYTGGGGAAGAA	2.700	3.986	4.085	3.733
14	7	26543634	12700757	701	CAGTTGGCTCYGATTTCCTAC	1.980	3.325	3.541	3.483
14	7	26550239	7810127	702	ATTACTACATYTGAACATACT	2.939	3.541	3.456	3.240
14	7	26556296	-	703	TAAGCATTGTRGCTACTTCTA	0.057	2.686	3.034	3.307
14	7	26561463	10486476	704	TAGCAAGTAAMACATTAAATT	0.039	3.085	2.778	2.535
14	7	26573185	7784257	705	AGAGAGTAAAYGAAATCTTTT	0.589	1.670	2.287	2.670
14	7	26587870	10261915	706	TACCAAAGTCYGTCACAGCCA	2.812	1.743	1.976	2.292
14	7	26594560	10250223	707	TTGATACCTCMCTTTGCTAAA	1.298	1.707	1.130	2.409
15	7	97454046	9641232	708	TAGAGAAAACRTAAAATGCCT	0.177	1.285	1.748	1.763
15	7	97459994	13221480	709	ATGTCAGGCAYTTAGCACCTG	2.210	1.510	1.653	1.670
15	7	97466861	3801294	710	AGTCTTCCCCRGAAGTGCAGG	1.138	2.321	1.716	3.107
15	7	97479927	2291749	711	CGAACTGCTTYCTGTAATTGC	0.949	2.249	1.763	3.131
15	7	97483791	17169462	712	CACTGGAAAAYCAAACCTGTT	2.024	1.647	2.123	3.064
15	7	97507796	2291750	713	CTGGTCTCTCRACTTCCGAGG	0.909	1.517	3.432	3.483
15	7	97530376	12536125	714	AACAAGGATCRGGAGTAGACA	0.020	1.698	3.541	3.386
15	7	97557847	4506131	715	CGAGGGATGGSCCAGATATCC	0.967	3.573	3.483	3.386
15	7	97583094	1006169	716	TCTCTACTTCRCTGGTGTTGA	1.285	3.998	3.483	3.307
15	7	97589637	6963210	717	ATCAGATAGCRTGCAGCCTGA	4.687	4.210	3.909	3.182
15	7	97596322	10808107	718	GACGTCCCATYGGGCACACCT	2.217	4.687	3.645	3.240
15	7	97602273	7807455	719	CTATAGAGAAMTCAAGGACAC	2.573	4.085	3.909	3.107
15	7	97608368	6465672	720	TTGAACTCATRCTTCTAAAAT	1.426	2.240	3.784	3.024
15	7	97613791	7812180	721	AAATCCACACYTATTTGTAAG	1.288	2.273	3.541	3.064
15	7	97622938	7795176	722	TCCTGGGTCAMAGGATCTTTC	1.068	2.299	2.456	3.307
15	7	97630066	6944262	723	CCCTTCGAACKGATTCTGAAG	2.630	2.148	2.427	3.307
15	7	97636746	3801256	724	GGTTCTCATTSTAGTGTTAAG	2.568	2.259	2.390	3.131
15	7	97650113	2107718	725	ACTTCTACACRGTATCCCTTT	0.215	2.371	2.321	2.511
15	7	97661068	6465582	726	GAGACAAAACMAAGTCCAAAT	1.894	2.194	2.305	2.511
15	7	97673117	3779193	727	TAGGAATTTCYCAAATTTCGA	1.627	1.651	2.159	2.314
15	7	97687178	7783611	728	GAATGCAACCRTTTCACGAAA	0.462	1.739	2.097	2.135
15	7	97711910	4729442	729	AAAGTCTAATYTCTGAATAGG	0.643	0.679	1.388	2.000
16	7	144659113	2372055	730	TCGTGTCAATSGCAAAAATGG	0.059	1.801	1.565	1.061
16	7	144667900	6958091	731	TAGATGATGGRGTTAACCACC	0.328	2.946	1.450	1.296
16	7	144674138	6952652	732	GGCTAATGATRTTTAAGTATG	4.687	2.429	1.716	1.263
16	7	144680315	17169752	733	ATTCTTCACARTATTTTTGTA	2.070	2.641	1.781	1.263
16	7	144685963	6946575	734	CGTTGCTCACWTATATGCTCA	0.213	2.686	1.849	1.368

16	7	144705856	6943077	735	GCACTATTGAYTTACAATTTC	0.440	1.054	1.884	1.605
17	10	20042742	2025789	736	AGTCTGGTTCYGGAGTTAGCA	0.277	0.211	0.737	1.689
17	10	20043985	2478746	737	AAATATGTACRTGCACTCTTT	0.846	0.456	2.025	1.699
17	10	20050326	16919309	738	CAACTCAGATYCACCATTAAT	0.298	0.740	2.071	1.863
17	10	20057858	7100283	739	ACACAACTTAMTCAAATTAAT	0.835	2.849	2.319	2.272
17	10	20069776	7073601	740	TGGTCTGGCTYTTTTAGAGGT	0.809	3.143	2.384	2.742
17	10	20106875	4748610	741	CTCAGTGCTTYGATATAGGAG	3.988	3.385	2.842	2.615
17	10	20119156	12763948	742	GGTCTTTCAAYTGAAACTAGA	0.443	3.784	3.608	3.064
17	10	20125405	12263244	743	CCTATCCCCARTTTGATACTA	0.833	3.784	3.453	2.884
17	10	20131169	16919362	744	CAGCAGTCAGYATTGAATGGT	0.076	0.948	3.541	3.008
17	10	20140825	11011581	745	CACTCTGCATYGTACAGCATA	0.886	1.235	4.086	3.307
17	10	20150341	11011636	746	GACGTGTTGTYTTAGAATTCT	1.210	1.089	1.916	3.344
17	10	20159308	12263862	747	TTCAGGTCAGYATCTACCATG	1.407	2.243	1.832	3.155
17	10	20170064	333693	748	TCCTTCTCTTYAAATCTTACA	0.442	2.082	1.773	3.344
17	10	20176365	7900249	749	GTCATTATCTSTGGTGGAAAA	2.670	1.783	2.256	1.984
18	10	53978045	2463963	750	ATTGGCTCATRTATACTAAAA	0.441	0.849	1.313	1.628
18	10	53987152	1950067	751	GGACCCAAGGMAAATGGAATT	1.332	1.829	2.881	1.634
18	10	53997189	7072039	752	CCGGAAATTTRGTACAGTGCT	0.152	2.009	3.608	1.821
18	10	54029767	12415510	753	AAAGTGGTCCRTGGGAAGCAA	2.981	3.909	2.580	1.731
18	10	54040012	2136892	754	TCTTTTGTTGWAGGAATTCAG	0.643	4.386	2.764	1.773
18	10	54053922	16931762	755	TGTCTGTCACYGAACATATTT	3.784	3.608	2.829	1.632
18	10	54061798	1919780	756	GGATGGAAGTKGTGGCTGTCT	0.241	2.146	1.998	1.953
18	10	54075328	12218308	757	CTAAAAGTCTRAATTCCTCTG	0.007	1.689	1.967	2.252
19	11	92824911	2658786	758	GTGAGATTAAMCCAGAAGCTG	1.254	1.159	1.936	3.195
19	11	92838631	2658791	759	TCTGTTTTGGYCACTGTGCAG	0.275	1.286	3.511	3.195
19	11	92849120	2608280	760	AATTCAGATARTGTCGGTTTC	0.118	1.252	3.256	3.168
19	11	92858899	897906	761	CACCCTACAAWTAAGAACAGC	1.260	2.590	2.901	3.365
19	11	92883170	2605618	762	TTTGGCCTTAKGGCTACAGTC	1.203	3.015	2.757	3.224
19	11	92898509	986134	763	CAGGATTGGAYCCGGAAATTA	4.687	3.325	2.251	3.063
19	11	92903685	2605629	764	GACTCTTCCASTTTTTCAGCA	0.701	3.224	2.619	2.773
19	11	92917120	7128507	765	TCTTCGAAATRTAAAATATTG	0.178	2.519	2.718	2.723
19	11	92923152	4753472	766	ACCACACTAARATCACAACTC	0.416	0.503	3.195	2.209
19	11	92930862	11020413	767	TTTCATGAATYATTCTATGCT	0.788	0.324	2.381	2.108
19	11	92937042	7127955	768	CTAAGGTTCAYCTGGACCTGC	0.432	1.172	0.868	1.923
20	11	128340968	546239	769	CCTACTCCTCYGTCCCCACTC	1.311	1.192	1.023	0.694
20	11	128375055	511296	770	GTGTAAAAGCRGTTTGCATTT	0.462	1.281	0.933	1.100
20	11	128393921	7942958	771	TTAAAGGAACYGAACAAATAC	0.600	1.101	1.001	1.143
20	11	128404165	7936496	772	AAGATTTTTTWAAATTAAAGT	0.685	0.545	1.001	1.146
20	11	128420261	619325	773	AGCTAAACTAKACAGTTTCCC	0.590	0.586	0.946	1.027
20	11	128443394	10893967	774	TGCATGTATTYGATTCATGAG	0.258	0.489	0.581	2.083
20	11	128469870	556655	775	CATCTTTGGAYGTCAATTTGC	0.500	0.492	0.928	1.988
20	11	128498038	626999	776	CACATTAATAYATATATAATG	0.000	0.426	0.716	1.640
20	11	128503422	-	777	AGGATAAATTYCTTGAGCATG	0.848	0.944	1.900	1.958
20	11	128519254	4937400	778	TAACAAAATARACTTAATGCT	0.416	0.777	1.716	2.590
20	11	128559769	11221596	779	CTAACACAAASGTTTGACTCT	1.330	2.733	1.606	2.511
20	11	128565670	1473299	780	GCTCATCCTTMCACAAATTTG	0.118	2.229	2.641	2.429
20	11	128580286	2511818	781	TAGAACCAGAYATCAATATTA	4.687	1.986	3.988	2.439
20	11	128587624	1031063	782	TTATTTAAGCYGTACCAATAT	0.248	2.556	2.747	2.682
20	11	128610465	6590369	783	ACATTTACTGKGCTCCTCCCT	0.124	3.842	2.649	2.916
20	11	128625622	6590371	784	AAAAAATTAAYCAGGTTGAAT	2.292	2.054	2.887	2.874
20	11	128665973	4644651	785	CCTTAAAATCRTAGACCTGAT	2.388	2.342	3.255	2.728
20	11	128682460	1493537	786	TAAGCTATGCYGCAATAACAA	0.747	2.823	1.914	2.657
20	11	128702726	4936074	787	TCTTCCCATTSCCCATTCCCC	0.000	1.950	2.009	2.200
21	13	60636792	969147	788	TTGGAATCATYCTTATGTTTT	0.326	1.887	1.905	2.098
21	13	60644803	3102213	789	TGTTGCTTGAYAGACTCTCCA	1.398	2.284	2.143	2.045
21	13	60652316	3102221	790	TTGTCTTACARTCGTTCTCTG	1.179	2.142	2.444	1.978
21	13	60667281	9539057	791	ATGAGATGGTRCCCGGTGATC	2.054	2.163	2.494	2.314
21	13	60676634	3127284	792	CATAAATTTCYGCGTGGTTTA	2.020	2.511	2.282	3.131
21	13	60687408	9570517	793	TCTACGTCTGKTTTTTATCAT	0.319	2.422	2.310	3.909
21	13	60714270	7332634	794	CAGCAAATGTYAGGTAGATAA	2.778	2.148	2.723	3.909
21	13	60727881	17059540	795	ACGGGTTTTAYAGATCTTTCA	0.833	1.592	4.386	4.388

21	13	60735640	-	796	ATCATACTACYTGTACCCCCT	0.428	2.733	4.386	3.541
21	13	60744842	17059565	797	AGGTAGGTCAYTTGTCTCAGA	0.460	3.784	4.386	4.085
21	13	60746124	2128881	798	TATATCTATAYGGTAACTCCA	2.218	3.909	4.386	4.085
21	13	60780785	1158312	799	CAACAATTATKTCTGATTTAT	3.024	3.606	4.085	3.687
21	13	60808070	7989120	800	CCTGTATAAAYGCATAACAAT	1.150	3.784	3.784	3.687
21	13	60819009	1028965	801	CACAAGTATTRTACATTTAAG	0.061	2.752	2.954	2.818
21	13	60826241	12430643	802	TTCTCTTACTYTATACAGCTT	0.916	0.655	2.586	1.910
21	13	60833710	17253874	803	CAAAACAAAGSGCTATAGAAA	0.288	0.218	1.838	1.647
21	13	60846575	9528326	804	ATAAAATGATYTCTTTTTGCT	0.368	0.218	0.515	1.647
21	13	60860097	10507666	805	TCCTCGTTCAWATGGAATGCC	0.148	0.282	0.228	1.691
21	13	60866066	1023113	806	TACATTGATARCAGAGGCCAT	0.052	0.314	0.467	1.178
21	13	60872154	17059730	807	CAATGTGCTARGACTAGGAAA	1.020	0.246	0.354	0.495
21	13	60879638	9570568	808	CATTAAAAGASTTGTAATATG	0.408	0.549	0.355	0.298
21	13	60886154	3812873	809	AATAGGGGTCSGAGAGAGAAA	0.105	0.685	0.341	0.309
21	13	60893563	4366613	810	AATCAAAATTKATGCTCAAGT	0.966	0.412	0.470	0.224
22	13	81061202	17074500	811	TGTGAGTGCAWTGATTTCATC	0.765	0.962	1.741	1.782
22	13	81068953	-	812	ATAAAATCTTRTTCCAAATCA	0.341	2.494	2.187	1.997
22	13	81075765	17074505	813	TCTGGAGTTASAGGTGAAATT	0.735	2.553	2.434	2.022
22	13	81084875	17074524	814	ACTGTATAGCMAAGACCTACT	4.085	2.705	2.413	2.325
22	13	81092175	9805522	815	CTGCCAGCTCYTTCTTCATTG	0.752	2.630	2.399	2.752
22	13	81099123	17074541	816	TATGCTTAGCRAGAATGAAAA	0.979	3.182	2.630	2.687
22	13	81106262	10507923	817	TCTTCAAACAYACCATACATC	1.062	1.844	2.733	2.489
22	13	81113141	9574936	818	TTGTAACAGCRAAGTGGTATC	1.568	1.740	3.131	2.365
22	13	81120127	9531231	819	TAAGCCTCCTRTAAGTCACAG	0.928	1.913	1.992	2.517
23	14	46953692	17743577	820	GAACATTTGAYCATTGTTGAG	1.012	1.848	1.920	1.969
23	14	46960712	1815636	821	AACAGGATAAYAGTAATGGAA	1.732	2.043	2.020	2.249
23	14	46967812	4627244	822	CCTTAAGAACMAAATTTTTGT	0.583	1.796	2.259	2.608
23	14	46974907	17118654	823	AAAATCATGGSCTTTTTTCTG	0.237	2.144	2.272	2.737
23	14	46981666	4424837	824	CACCTTCCTGYGGATACCACA	1.559	2.005	2.432	2.768
23	14	46992516	4900764	825	TTCTCACTCTWCTTGGTAATT	1.770	1.885	2.954	2.963
23	14	47002220	-	826	CTATCTACAAWTTCCTGCTAG	2.182	2.319	2.823	3.095
23	14	47007856	1950385	827	AAATGTCAAAYCTGCTCCATA	0.549	3.155	2.874	2.687
23	14	47015329	12892144	828	CACGTAAACAMCTTAGGTTTT	3.085	3.385	2.963	3.455
23	14	47021190	11624882	829	CCCAGGATGAYAGAAATTTAG	3.224	3.224	3.119	3.265
23	14	47027014	17118776	830	GGGGCTCTCTRGGTTATCTCA	1.311	3.456	2.709	3.265
23	14	47033221	8019096	831	TTGTGTAACAYATGAAATTGA	0.832	2.946	3.196	3.511
23	14	47046266	2017732	832	GAATTATGAAYATGGTTTGTG	1.421	1.180	3.645	3.842
23	14	47053268	-	833	TGACACTGACSAGCCATCTTT	0.509	1.674	3.842	3.365
23	14	47059554	10220462	834	AGTCAGAAAAYAGATACAGAA	0.041	2.036	3.196	3.456
23	14	47065968	6572434	835	ACAGTATCAGYATGCTCACTG	2.462	3.733	2.508	3.733
23	14	47072828	7156218	836	TATTATATTGKTATTTGATTA	0.000	4.386	2.742	3.224
23	14	47080731	4900780	837	TCCATTTTATMCTCTTCTTAA	2.939	4.386	3.764	2.747
23	14	47086453	10483565	838	TTGGTCTGAAYGGAATACAAA	1.810	2.742	3.156	2.547
23	14	47094500	10136935	839	ATATCCTCAGRAGGAAGACTC	0.810	3.307	3.541	3.085
23	14	47112076	1955798	840	GCGTAATCCCRTAACCAAATA	0.727	1.871	3.043	2.964
23	14	47124732	17686922	841	ATATTTTCCTMATATTCCTAT	1.233	0.981	2.818	2.773
23	14	47137969	17118957	842	AGTAGAAAGTRTCTGAGTTCA	0.598	1.385	1.850	2.789
24	16	72110200	2432545	843	ACAAAGTCTTYGAAAAGCATT	0.145	1.121	1.841	1.033
24	16	72120092	11643638	844	ACTCTTATTCYTGAAGTAGTG	0.103	1.535	1.264	1.491
24	16	72126972	6564305	845	CAGGGTAGTARGAAAAGAGTC	4.687	1.742	0.904	1.557
24	16	72134302	16972157	846	GGAACTTCACYCTGTTGTGTA	0.623	1.630	0.885	1.186
24	16	72144267	956065	847	CAGTTAACACRCAACCAGTAA	0.865	1.486	1.103	0.982
25	20	25536407	6115224	848	AGTTTGTAAAYGTACTCAAGA	0.653	0.506	1.502	1.990
25	20	25542783	3752273	849	GTAAACACGAWAGGGCATTTG	0.690	0.513	1.566	3.842
25	20	25548835	6138616	850	AAAATGGGAGWATTGATTGAA	0.181	2.531	1.734	3.307
25	20	25554369	4815438	851	TGCTTACCTCRTTGACGTATA	1.007	1.788	2.047	3.432
25	20	25574907	6050754	852	ATCACAGTGCRTCTTTCCATT	3.842	1.751	4.085	3.608
25	20	25588732	6115241	853	AATAAGCATTSAGGCATCTCA	0.128	2.408	3.687	3.687
25	20	25597055	6515674	854	ATTTTTATGGKTGTACAATTA	0.808	4.386	3.432	3.272
25	20	25614807	926486	855	GTTGCACTCASGCAGTTGGAG	1.504	3.210	3.541	3.344
25	20	25621409	6050793	856	TTAGTTACCARTAAATTAGAT	4.210	3.272	3.608	3.307

25	20	25653889	6050851	857	ATATAGTTGGKTTATGGTTTT	1.749	3.784	3.131	3.272
25	20	25660656	6050858	858	AAGTAACTTCYATGGATTTTT	2.535	3.432	3.107	3.272
25	20	25666766	8122516	859	TTTTCTTTTAWAATACTTTTG	2.422	2.494	3.432	3.344
25	20	25674034	16988014	860	GGCTCAGCCAYGGTGGAAGTA	1.515	2.617	3.156	2.988
25	20	25687466	4990081	861	CCCAGCCTTARTGATGGCAAG	1.040	2.227	2.432	3.024
25	20	25808999	-	862	GTGCTGAGAARACATGACGCA	0.692	1.269	2.427	3.064
25	20	25821023	6037266	863	ATCTTTAAAGSCTTCAAAATC	0.000	0.726	1.582	2.723
26	X	8227264	941108	864	AGTATCATTTKAGTGTGTCTT	1.274	0.714	0.480	1.220
26	X	8247729	6640134	865	ATGAATATTAYGGAGCAACGT	0.193	0.825	2.178	1.195
26	X	8255800	6638816	866	GCATGTGTGTYTTAATTCACA	0.346	0.680	2.494	1.187
26	X	8275717	2078361	867	CCACAATTGCWACAATAATGC	0.719	2.384	2.560	1.295
26	X	8282105	-	868	TTATTTCCCTYGTTCCCACAC	0.322	2.806	2.653	1.649
26	X	8286079	5978889	869	CAAAGCTCCAYAAAGGTTGTT	4.388	2.696	1.791	1.694
26	X	8294959	-	870	CAAATATTTTKCTGCAGTCCA	0.590	2.432	1.760	2.600
26	X	8302230	5934440	871	GGCCCTTAGTYATGATGCAAA	0.282	2.187	1.937	2.346
26	X	8314388	17306755	872	AACACATTCAYTCCTACATCT	0.591	0.161	1.871	2.021
26	X	8315933	808118	873	AGCCAAGGCCRCCATGTTTCT	0.098	0.259	2.854	2.502
26	X	8331203	6640185	874	TGCATGACAAWATAGCAGCTC	0.054	0.386	0.761	2.368
27	X	92543632	12854147	875	TTATAGGGAAMCAACATTATA	0.577	1.587	1.621	2.031
27	X	92553275	17248183	876	AATGGGTAAGSAAACAAACCA	1.594	1.538	2.214	2.342
27	X	92567045	4893028	877	TAGAATGTAAMCCTAGAAAAT	1.242	1.567	2.206	2.375
27	X	92579096	1014621	878	TGCTGCTTTCRTAGGAATATT	0.895	2.681	2.250	2.495
27	X	92584665	6522704	879	CTACAGAATCRCTGAGGTCAT	1.074	2.284	2.974	2.665
27	X	92601672	7877669	880	ATTTTTAAACSGTGTATGCTA	2.848	2.291	3.365	2.963
27	X	92610848	5940244	881	AATATTTTATSACTGTGTCTA	0.506	3.015	2.800	3.224
27	X	92616967	5940245	882	GGGCAGCGTTYATCTCCCACC	1.294	2.857	2.997	2.916
27	X	92621684	-	883	GTTTAACTGCYTCCACAGGGT	1.727	1.986	4.687	3.365
27	X	92639462	17003593	884	AATTCGCTTCYAGAGGCACTC	0.000	2.284	4.687	3.289
27	X	92657382	4892963	885	AGAAACAATAMAAATTATCAA	1.598	3.182	2.560	3.645
27	X	92671201	16982665	886	ATAATCCTGGRATAATCCTGA	0.487	2.733	2.954	3.408
27	X	92682855	1326176	887	GTGTAGCATGRATTCTATAGG	1.681	2.292	3.687	3.511
27	X	92687969	1326175	888	AAAATGTTTCRTTCAACAGGT	1.922	1.775	2.505	2.166
27	X	92703599	2312032	889	GACATTTCATSAAAGATATTA	0.088	2.446	2.107	2.176
27	X	92718933	4893037	890	AAACTTTCATRCATTACAAAG	0.408	1.338	1.721	2.065
27	X	92729569	5983240	891	ATATACCTTTYCCCAGAGCCT	1.329	0.407	1.549	1.383
27	X	92746053	-	892	TTCACTTGAAYAGCACTCATA	0.334	0.796	1.008	1.240

Table 3. SNP markers found to be associated with Schizophrenia in Fine Mapping studies in the QFP. Individual SNP markers genotyped in the genome wide scan are presented in each row of the table. The corresponding region ID and chromosome is presented as identified in Table 1 (columns 1 and 2). The coordinate of the SNP according to the NCBI genome assembly build 35 is indicated in column 3. The RS# column corresponds to the NCBI dbSNP identifier for the SNP (column 4). The Seq ID is the unique numerical identifier for this SNP in the sequence listing for this patent, and is indicated in column 5. Column 6, labeled Flanking Sequence, corresponds to 21 bp of nucleotide sequence centered at the SNP, which is coded using the standard degenerate naming system. The remainder of the table lists -log10 p values for association of the indicated haplotype centered at the corresponding SNP with the disease as described in the text, using LDSTATS V2, V4 and SingleType. Values for the association of single markers, as well as 3, 5, 7, 9, 11, 13, 15, 17 marker haplotype windows are shown for LDSTATS V4 and LDSTATS V2 analysis (see text for explanation of statistical calculations). The last two columns represent the -log10 P values for single marker genotype and allele association using Single Type analysis (see EXAMPLE section for description of analysis).

						LDSTATS v2								LDSTATS v4								SingleType	
Region ID	Chr	B35 Position	RS#	Seq ID	Flanking Sequence	Single Marker	W03	W05	W07	W09	W11	W13	W15	W17	Single Marker	W03	W05	W07	W09	W11	W13	W15	W17	Single Genotype Likelihood Ratio	Single Allele Likelihood Ratio
1	1	55092379	9326036	503	TACTCATAACRAATGACTTAA	1.476	-	-	-	-	-	-	-	-	1.466	-	-	-	-	-	-	-	-	1.084	1.476
1	1	55064946	2767503	504	CATCCTTTTAGGTTCTTCAC	1.934	1.680	-	-	-	-	-	-	-	1.926	1.666	-	-	-	-	-	-	-	1.646	1.934
1	1	55086324	748284	505	AAGAAAATAAAGAGCAATTAT	2.053	0.915	0.459	-	-	-	-	-	-	2.025	1.116	0.514	-	-	-	-	-	-	1.516	2.053
1	1	55099081	2071226	506	GATAAATATGWCACATTGACT	1.877	0.570	0.577	1.544	-	-	-	-	-	1.841	0.855	0.563	1.455	-	-	-	-	-	1.531	1.877
1	1	55913045	2767499	509	TCATGTTTCCYAGAGCTTCA	2.207	1.055	2.345	2.430	0.932	-	-	-	-	2.165	1.099	2.218	2.227	0.897	-	-	-	-	2.331	2.207
1	1	55920205	1165512	510	TGCATTTCAAACTCTTTGTGG	0.365	4.021	3.491	1.445	-	-	-	-	-	0.339	4.297	3.127	1.432	-	-	-	-	-	0.375	0.365
1	1	55927431	1165504	511	CTGCCAGCTGRATCAGAAGAG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	1	55936001	17112758	512	TTCCAATTTGRGTTGACTGAC	5.121	4.636	3.268	-	-	-	-	-	-	5.097	4.690	2.920	-	-	-	-	-	-	4.636	5.121
1	1	55946108	17111768	513	AACACAGTTTYGAAAAATTAG	5.121	3.918	-	-	-	-	-	-	-	5.155	3.950	-	-	-	-	-	-	-	3.340	3.446
1	1	55981978	17112021	515	ATATTAACCTWAACTCCAAAC	3.446	-	-	-	-	-	-	-	-	3.378	-	-	-	-	-	-	-	-	-	-
1	1	56004459	4304273	594	GATCCCATGCSCTCAGTTGTG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	1	75016464	211715	695	CTTTAGAGTGSAAAGAAAAA	1.387	-	-	-	-	-	-	-	-	1.405	-	-	-	-	-	-	-	-	1.182	1.387
2	1	75032321	440826	696	TGATTTTAGGYGTAAATGAGC	1.387	1.547	-	-	-	-	-	-	-	1.393	1.554	-	-	-	-	-	-	-	1.182	1.387
2	1	75047967	7552404	521	CTAAAAGAAARCCTTCCCAAT	2.057	1.256	1.084	-	-	-	-	-	-	1.891	1.306	1.089	-	-	-	-	-	-	1.959	2.057
2	1	75065766	2066163	697	CTCTCTGACTRGAGAGCAAAC	1.907	1.373	2.735	2.255	-	-	-	-	-	1.882	1.356	2.513	2.075	-	-	-	-	-	1.959	1.907
2	1	75080530	11164037	524	GTGCCAAAGTWTAATACGGTC	1.428	1.568	1.322	2.092	2.533	-	-	-	-	1.440	1.709	1.320	1.879	2.212	-	-	-	-	1.039	1.428
2	1	75907568	1313071	525	TTTTCAAGCAGTGTGAGGATA	1.832	1.420	1.333	1.450	3.648	-	-	-	-	1.771	1.396	1.353	1.139	2.521	-	-	-	-	1.977	1.832
2	1	75922498	11161611	527	GCCGGGTCTCYGTTTTATATT	2.114	1.542	1.473	2.118	-	-	-	-	-	2.053	1.719	1.383	1.952	-	-	-	-	-	2.054	2.114
2	1	75931022	1694419	529	TTTTTAATTGRTAGGTACAGT	1.957	1.442	1.759	-	-	-	-	-	-	1.934	1.518	1.671	-	-	-	-	-	-	2.050	1.957
2	1	75947105	6664254	530	TTCTTTGAAGKGGGATCATTT	0.288	1.591	-	-	-	-	-	-	-	0.285	1.816	-	-	-	-	-	-	-	0.369	0.288
2	1	75964032	1140634	532	CAGACAGGGTRGAGTAGGGTT	1.145	-	-	-	-	-	-	-	-	1.141	-	-	-	-	-	-	-	-	0.745	1.145
3	1	201629229	2002853	699	AAGCTGGGCTSTAGACCAGGG	0.413	-	-	-	-	-	-	-	-	0.382	-	-	-	-	-	-	-	-	0.222	0.413
3	1	201659075	11810003	599	TAAGAACCCRTAGCAGAAAA	1.254	0.356	-	-	-	-	-	-	-	1.165	0.348	-	-	-	-	-	-	-	0.959	1.254
3	1	201687982	2290198	900	AAATAGCAGASTCTGAGCTAT	0.503	0.378	0.188	-	-	-	-	-	-	0.755	0.364	0.187	-	-	-	-	-	-	0.455	0.503
3	1	201720192	2794806	901	AGAAGTTCCTYTGACTAGGCT	0.235	0.534	0.799	1.781	-	-	-	-	-	0.189	0.545	0.820	1.153	-	-	-	-	-	1.412	0.235
3	1	201760031	871423	552	CTGGACAGTAMCTCTCTGGGT	1.687	1.285	0.900	1.201	1.244	-	-	-	-	1.601	1.275	0.866	1.134	1.195	-	-	-	-	1.437	1.687
3	1	201792049	13150848	555	CCCTCCAAGAYGACCAGGTTA	1.821	1.511	1.622	0.573	0.836	0.586	-	-	-	1.740	1.565	1.589	0.559	0.806	0.845	-	-	-	1.274	1.821
3	1	201824502	2095854	902	AAAGCAACAARGAAAAATNTC	0.026	1.934	2.259	3.882	3.557	0.856	0.242	-	-	0.000	1.932	2.067	0.993	0.519	0.862	1.199	-	-	0.018	0.026
3	1	201854186	13142514	903	GCAATACAARGCTCAGACTC	0.133	0.155	1.859	1.541	1.498	1.076	1.845	1.762	-	0.104	0.133	1.734	1.452	1.405	1.339	1.745	1.655	-	0.031	0.133
3	1	201884283	3051267	904	CTGAAAATGGYTAATTTATGA	0.000	0.546	0.522	2.261	2.957	2.546	1.992	2.144	2.116	0.000	0.598	0.545	2.042	2.662	2.121	1.545	2.051	2.020	0.067	0.000
3	1	201922839	7531256	570	CAACTGGACTKTAAGGCAGGG	1.251	0.659	1.486	2.065	3.268	3.332	2.814	1.440	-	1.190	0.727	1.428	1.882	3.505	2.842	2.496	1.402	-	0.754	1.251
3	1	201930152	1172161	571	GGAATGTTGGYGACCTAGGCT	1.248	1.555	2.883	3.437	2.403	3.453	3.162	-	-	1.197	1.884	2.524	2.167	2.186	2.549	2.859	-	-	0.788	1.248
3	1	201944681	1172140	572	GACAGACCCARCTGACATGTA	0.295	2.332	3.142	3.225	2.578	3.251	-	-	-	0.267	2.839	2.771	2.759	2.353	2.842	-	-	-	0.195	0.295
3	1	201953518	1172150	573	AAATTTAGCTRCTTTTTAGCG	0.220	2.931	3.727	3.337	2.886	-	-	-	-	0.183	3.566	3.104	3.215	3.586	-	-	-	-	0.138	0.220
3	1	201964104	1151704	574	AGGATTTGCAYGCCCTCAGAT	2.002	2.714	3.349	3.452	-	-	-	-	-	1.912	2.978	2.898	3.261	-	-	-	-	-	1.413	2.002
3	1	201971436	1668809	575	TCACATAACASTCATCSTTACG	2.002	1.152	2.348	-	-	-	-	-	-	1.991	1.120	2.123	-	-	-	-	-	-	1.413	2.002
3	1	201976923	9326060	576	CAACACCCARCTCCACAAGT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	1	201998099	1172125	577	CATCCTGTTGWCTAACCCAAA	1.846	1.450	-	-	-	-	-	-	-	1.751	1.567	-	-	-	-	-	-	-	1.235	1.846
3	1	201994779	17345837	578	AGGTCTTCAGRTGAGCCCGGC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	1	202003987	16055350	579	GACTCTGTCGYGGGCATTCCC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	1	202033124	1134082	905	TTAGAGCAGGYCTAGGGAACT	0.394	-	-	-	-	-	-	-	-	0.342	-	-	-	-	-	-	-	-	0.163	0.394
4	2	60121615	10206105	906	TTTACCTTCCYTGCCGTGTTT	0.743	-	-	-	-	-	-	-	-	0.682	-	-	-	-	-	-	-	-	0.378	0.743
4	2	60136951	-	586	GAAGGCAGTCSTGGTGAGCTG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	2	60145978	11125023	587	TCTTCTCATGYCCCCTCTCAT	0.831	0.198	-	-	-	-	-	-	-	0.783	0.187	-	-	-	-	-	-	-	0.477	0.831

4	2	60163473	4672341	907	GGACTATATCMAAATTATAAA	0.873	1.535	1.839	-	-	-	-	-	-	0.861	1.955	1.881	-	-	-	-	-	-	0.518	0.873
4	2	60165881	13003093	589	TGAGTGGTCAGTAGGCAAATG	3.069	2.853	2.819	1.731	-	-	-	-	-	3.077	3.154	2.439	1.610	-	-	-	-	-	2.748	3.069
4	2	60172290	2215897	589	GAATATCCAGHAGCAATGCCT	1.363	1.572	3.075	2.847	2.507	-	-	-	-	1.331	3.055	2.486	2.555	2.290	-	-	-	-	0.896	1.363
4	2	60180536	4672357	590	TTCCTCGTTTATAAAGAGTTC	0.226	0.834	1.629	3.012	-	-	-	-	-	0.195	1.103	1.497	2.217	-	-	-	-	-	0.131	0.226
4	2	60191069	994972	991	AAGGAAACTTYCTGCCTGGCC	0.330	1.229	1.682	-	-	-	-	-	-	0.309	1.216	1.798	-	-	-	-	-	-	0.152	0.330
4	2	60204672	4671308	908	CCGAGGATTGYCCATGGGATT	0.478	1.842	-	-	-	-	-	-	-	0.461	1.830	-	-	-	-	-	-	-	0.250	0.478
4	2	60217208	6725448	909	AGTGACTAAGGTGGAATTCCT	0.696	-	-	-	-	-	-	-	-	0.630	-	-	-	-	-	-	-	-	0.423	0.696
4	2	115310871	17731643	597	CCAAGTATGTRTGATTCCAAC	1.116	-	-	-	-	-	-	-	-	1.117	-	-	-	-	-	-	-	-	0.722	1.116
5	2	115353658	10182492	601	GAGCTTCCCTYAAGCTGGAT	3.112	2.811	-	-	-	-	-	-	-	3.084	3.006	-	-	-	-	-	-	-	2.696	3.112
5	2	115384367	1401755	604	TTTTTACTACNTTCTGATTTC	2.730	3.025	3.031	-	-	-	-	-	-	2.731	3.165	3.032	-	-	-	-	-	-	2.204	2.730
5	2	115421231	17043023	605	ATTTGAGAGTYAATTTTTTTA	4.723	3.565	3.193	3.272	-	-	-	-	-	4.627	3.896	2.973	2.150	-	-	-	-	-	4.108	4.723
5	2	115446541	10480426	611	AATCAAGAAAKTSTTCTATGA	5.397	3.426	2.082	2.451	2.223	-	-	-	-	5.432	3.892	1.952	2.358	2.145	-	-	-	-	4.532	5.397
5	2	115474886	7423711	614	TGTTTTACTCNAACCTGTAACA	1.013	2.975	4.781	2.371	1.746	1.680	-	-	-	1.029	3.833	3.922	2.205	1.665	1.638	-	-	-	0.554	1.013
5	2	115510683	2175175	616	CATTAGAATCKAATTAAAATA	2.992	2.107	4.433	4.662	3.019	1.675	-	-	-	2.962	2.345	3.697	3.915	1.504	1.635	-	-	-	2.308	2.992
5	2	115539945	11092506	619	GGCAGAAGCTKTTTATCTTAG	0.123	2.348	2.778	3.809	2.526	-	-	-	-	0.097	2.649	2.401	3.147	2.332	-	-	-	-	0.827	0.123
5	2	115580001	1076001	621	ACTTTGATGTRCAAATYAAAA	2.803	2.287	2.779	1.428	-	-	-	-	-	2.875	2.586	2.445	1.337	-	-	-	-	-	2.251	2.803
5	2	115598002	7579207	625	AGACCATTCARTGACTTCCTA	0.975	1.775	1.032	-	-	-	-	-	-	0.961	2.003	0.981	-	-	-	-	-	-	0.634	0.975
5	2	115639051	12711618	910	GTTTTTTGAGRTGAGTCCTGT	0.079	0.215	-	-	-	-	-	-	-	0.093	0.243	-	-	-	-	-	-	-	0.014	0.079
5	2	115694508	6730842	630	ATGACATCCTSTCACTAGAAA	0.029	-	-	-	-	-	-	-	-	0.011	-	-	-	-	-	-	-	-	0.008	0.029
6	2	103287840	12017446	911	TACCCACCTCRTAATTTTGCC	1.231	-	-	-	-	-	-	-	-	1.218	-	-	-	-	-	-	-	-	0.552	1.231
6	2	103292982	10487608	912	GTAATTTAAAARATAGAATAGT	1.137	0.175	-	-	-	-	-	-	-	1.093	0.172	-	-	-	-	-	-	-	1.072	1.137
6	2	103296973	10150521	913	CTTGCATTTCYAAATAGAGTG	0.578	0.229	0.079	-	-	-	-	-	-	0.572	0.199	0.098	-	-	-	-	-	-	1.346	0.578
6	2	103302816	6720813	914	AAGGCACATRTATTTCACTC	1.323	0.609	0.185	0.088	-	-	-	-	-	1.267	0.555	0.222	0.079	-	-	-	-	-	0.952	1.323
6	2	103306948	2141269	915	AGTGGTTCTCWATCTTGACCA	1.276	1.048	0.424	0.052	0.086	-	-	-	-	1.203	1.274	0.448	0.098	0.101	-	-	-	-	0.840	1.276
6	2	103311909	13403435	634	GGGTTTCAAARTATTTAATCTC	1.276	0.714	0.421	0.168	0.705	-	-	-	-	1.251	0.746	0.479	0.188	0.724	-	-	-	-	0.840	1.276
6	2	103315704	7900943	935	ACARTTGATAYTAAGTTCMGG	0.622	0.440	0.338	1.686	-	-	-	-	-	0.672	0.485	0.388	1.549	-	-	-	-	-	1.357	0.622
6	2	103322291	1527883	917	TTTTATTCTCTYTCCAGACTGA	0.364	0.233	1.654	-	-	-	-	-	-	0.348	0.218	1.624	-	-	-	-	-	-	0.975	0.364
6	2	103329924	6733013	918	AGCAAAAGCARCTAAAATCAT	0.422	0.544	-	-	-	-	-	-	-	0.379	0.630	-	-	-	-	-	-	-	1.328	0.422
6	2	103333552	1569103	919	AATTGCCTGGYACGCCACAGG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	2	103338232	10180264	920	TTTCCTCTAGRAAGCTGATTT	0.605	-	-	-	-	-	-	-	-	0.582	-	-	-	-	-	-	-	-	0.350	0.605
7	2	236423479	2675144	921	TGCTCTCATCYTGAGTTGTCCT	5.093	-	-	-	-	-	-	-	-	5.102	-	-	-	-	-	-	-	-	4.354	5.093
7	2	236427015	1018113	922	GTCATTTAATSGTACGTCATC	2.631	5.032	-	-	-	-	-	-	-	2.667	5.057	-	-	-	-	-	-	-	2.111	2.631
7	2	236432013	12460145	923	CTGTCGTTTGKTTTGTTTTTT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	2	236437852	2696393	924	GGCCTAATGCRCCCTCTAGTG	2.854	2.371	6.149	-	-	-	-	-	-	2.792	3.949	4.983	-	-	-	-	-	-	2.234	2.854
7	2	236440088	1720206	925	TTTTTAGGAARCTGGTGCCGA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	2	236445586	3106790	926	TGGAGCGTTCKCGCTCCTGCT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	2	236450830	1720291	927	TCTTACCCTCNTGAAGCTATA	2.206	2.603	4.845	5.475	-	-	-	-	-	2.139	3.178	3.881	5.231	-	-	-	-	-	1.696	2.206
7	2	236457602	2696387	639	CTGCGCAAATKTTTTCCTTTT	2.970	2.752	4.522	4.545	5.476	-	-	-	-	2.935	3.522	3.958	3.856	5.284	-	-	-	-	2.497	2.970
7	2	236461236	2675126	928	CGTGCAAAGAYGACGGSCGTG	2.895	3.151	4.918	4.822	5.886	5.728	-	-	-	2.864	3.748	3.552	3.922	5.087	4.759	-	-	-	2.446	2.895
7	2	236466546	2292705	929	AGTATGAASAYGCCTGTTGA	2.895	2.895	4.415	5.157	5.053	4.880	4.410	-	-	2.893	3.548	3.758	4.003	4.238	4.112	3.935	-	-	2.444	2.895
7	2	236469394	6761949	640	ATGATAGAGSYCCTGCCAGTG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	2	236473982	1070154	641	ATTTAGATGTRTGCTTAATGT	2.895	2.895	5.191	4.815	5.992	5.123	4.304	3.678	-	2.855	3.844	3.509	3.785	3.410	4.280	3.795	3.353	-	2.444	2.895
7	2	236477836	1070155	930	GCTAGCAGGGKCCGCCTCTAC	2.895	3.199	3.420	4.302	5.599	3.494	4.443	3.069	1.955	2.844	3.562	2.594	3.543	4.301	3.128	3.973	2.834	1.891	2.444	2.895
7	2	236482821	12471567	931	AACGTGGGCTYGATTTACACC	3.041	2.214	4.259	4.491	3.648	5.002	2.955	3.288	1.439	3.012	2.792	3.386	3.383	3.002	4.120	3.713	2.576	1.367	0.017	0.041
7	2	236486602	4471841	932	TATTTGACTTRTCTCTAGGAAG	1.694	2.419	4.491	3.648	4.046	3.143	3.446	0.778	1.925	1.649	2.524	3.411	2.668	3.262	2.837	3.275	0.775	1.808	1.270	1.694
7	2	236491123	6760477	642	CAAAAATGCTRCCAGCCTGTC	4.457	2.936	3.521	3.579	2.793	1.657	0.807	1.143	0.630	4.435	3.737	2.234	3.035	2.561	1.987	0.758	1.141	0.845	3.377	4.457
7	2	236494086	2917003	933	CCACACGCTTSATCAAATACA	1.751	2.753	4.215	1.668	1.598	1.085	1.017	0.982	1.289	1.715	3.353	3.143	1.759	1.528	1.380	1.035	0.973	1.287	1.359	1.751
7	2	236498843	925809	934	ATCCTTCTTGRGCAGAAGCCT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	2	236504996	967373	935	GTGAAATCCARTTCCCCTTGA	1.636	0.887	2.855	1.765	0.468	0.978	0.935	0.936	0.691	1.621	0.898	2.584	1.672	0.452	0.970	0.908	0.917	0.846	1.227	1.638
7	2	236510114	12469191	936	CCAGGAGCCAYACTTTGAGAT	1.490	2.054	1.547	1.024	0.884	0.725	1.315	1.237	1.943	1.432	2.438	1.515	1.079	0.880	0.734	1.295	1.338	1.832	1.077	1.410
7	2	236516213	883521	937	GCTTTTAAAARTAGATGGTAG	3.243	1.491	0.693	0.531	0.895	0.693	0.992	1.943	-	3.205	1.551	0.670	0.545	0.877	0.559	0.593	1.854	-	2.654	3.243
7	2	236521670	3795893	938	GCATTTGCTASTTTTATAAGA	0.490	1.231	0.655	0.547	0.364	0.678	1.431	-	-	0.444	1.247	0.677	0.577	0.353	0.881	1.357	-	-	0.428	0.490
7	2	236524101	908444	643	TCTCCTCATCRCCCCCGCCCCC	0.193	0.448	0.349	0.469	0.395	0.932	-	-	-	0.171	0.464	0.570	0.483	0.393	0.907	-	-	-	0.597	0.193
7	2	236529643	6717724	939	GCTAACGCGGRTGTACGTGAG	0.477	0.541	0.194	0.627	0.820	-	-	-	-	0.439	0.526	0.188	0.643	0.787	-	-	-	-	0.219	0.477
7	2	236534552	2176237	940	CTCACATTGTSATCAGTGTCA	0.143	0.324	0.377	0.187	-	-	-	-	-	0.131	0.376	0.366	0.194	-	-	-	-	-	0.503	0.143
7	2	236540422	908450	941	GGCCTTTCCCYGCTCCACCGT	0.132	0.382	0.275	-	-	-	-	-	-	0.109	0.345	0.309	-	-	-	-	-	-	0.381	0.132
7	2	236544304	4632286	942	GGATGTTAATRAAAATTTAAT	0.846	0.509	-	-	-	-	-	-	-	0.830	0.550	-	-	-	-	-	-	-	0.525	0.846
7	2	236549343	4663621	644	TAATGTATCCKTCTACCATGA	0.149	-	-	-	-	-	-	-	-	0.090	-	-	-	-	-	-	-	-	0.129	0.149
8	3	6425439	156109	943	GGTGAACAGAYAGAAATATAT	1.135	-	-	-	-	-	-	-	-	1.155	-	-	-	-	-	-	-	-	0.719	1.135

8	3	6433324	155514	944	CTTTTGTACARAAAGGAGCTG	1.789	0.589	-	-	-	-	-	-	-	9.769	0.584	-	-	-	-	-	-	-	1.305	1.789
8	3	6439043	4686655	945	TTGGTGAAGTGTGTAAGCCAG	1.023	2.120	1.570	-	-	-	-	-	-	5.048	2.353	1.768	-	-	-	-	-	-	0.772	1.023
8	3	6450006	156480	946	TGGTGAACAAATGAATTGCAT	3.614	2.423	2.341	1.452	-	-	-	-	-	3.553	2.566	2.391	1.435	-	-	-	-	-	3.475	3.514
8	3	6457374	267482	946	GGAAATTCTCTCTCCTCTGTAA	1.861	2.355	1.949	2.009	1.748	-	-	-	-	4.752	3.791	1.946	1.852	1.640	-	-	-	-	1.350	1.551
8	3	6465033	266418	947	CATACATCTARCTCGTGGACC	1.432	1.058	2.105	2.629	2.886	1.981	-	-	-	9.414	1.062	1.935	2.675	2.641	1.781	-	-	-	0.968	1.432
8	3	6472100	7610839	948	AGAATAATGGRCAAAGGCCAG	1.172	0.954	1.650	3.391	2.612	-	-	-	-	5.167	0.879	1.575	2.543	2.458	-	-	-	-	0.748	1.172
8	3	6479367	263717	949	ATGTTCCCTAYACACACAG	1.071	0.843	0.532	1.638	-	-	-	-	-	1.023	0.975	0.839	1.552	-	-	-	-	-	0.578	1.071
8	3	6486713	155307	950	ACTAAATATGMTTCTTTGCGT	0.244	0.970	1.639	-	-	-	-	-	-	0.237	1.018	1.709	-	-	-	-	-	-	0.224	0.244
8	3	6493040	155264	951	ATTTCCTCAGMTAAAAGGGA	0.190	0.524	-	-	-	-	-	-	-	0.182	0.505	-	-	-	-	-	-	-	0.356	0.190
8	3	6497643	9049828	952	TAACTGCAAARTAGGAACACG	0.183	-	-	-	-	-	-	-	-	0.168	-	-	-	-	-	-	-	-	0.045	0.183
9	3	76611818	6772812	953	TATTTACTGGRCAGAGCCAAA	0.918	-	-	-	-	-	-	-	-	0.837	-	-	-	-	-	-	-	-	0.544	0.918
9	3	76617447	2324541	954	GTGTGAGCCTRCTTTGAGAAA	0.016	0.442	-	-	-	-	-	-	-	0.003	0.483	-	-	-	-	-	-	-	0.335	0.016
9	3	76621782	6789109	955	ACATTAGGACRTGCCTTAAAT	0.436	0.120	0.398	-	-	-	-	-	-	0.408	0.074	0.422	-	-	-	-	-	-	0.177	0.436
9	3	76624945	6801929	955	GCACATATGCRTAGCSTTGTG	0.039	0.882	0.538	0.486	-	-	-	-	-	0.022	0.932	0.539	0.533	-	-	-	-	-	0.474	0.039
9	3	76631502	17014202	956	GGTAACTATTYGACCAAACAC	1.692	1.073	0.582	0.622	0.446	-	-	-	-	1.658	1.150	0.593	0.537	0.477	-	-	-	-	1.357	1.692
9	3	76634093	7431017	956	TACAGCTGCTMCTCCCTTCAA	1.643	1.155	0.630	0.588	-	-	-	-	-	1.593	1.355	0.518	0.577	-	-	-	-	-	1.036	1.643
9	3	76638562	13888369	957	GTTAGCGTACSCTGGCAACCT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	3	76642077	3913572	958	TATGACTGACRGTGATAACAA	0.979	0.599	0.443	-	-	-	-	-	-	0.953	0.556	0.482	-	-	-	-	-	-	0.638	0.979
9	3	76646675	13866043	959	AGATACATGAYTGCCTGAATG	1.044	0.232	-	-	-	-	-	-	-	1.021	0.220	-	-	-	-	-	-	-	0.732	1.044
9	3	76651471	3848502	957	TCACTTTTGTRTACCGGGGSA	0.981	-	-	-	-	-	-	-	-	0.957	-	-	-	-	-	-	-	-	0.589	0.981
10	4	14249397	10461009	960	CTTATATCTTYTAAGTAGAGT	1.308	-	-	-	-	-	-	-	-	1.215	-	-	-	-	-	-	-	-	1.211	1.308
10	4	14261728	2173948	961	CTGCACAAAAWCATGGGACTA	0.756	1.971	-	-	-	-	-	-	-	0.755	1.882	-	-	-	-	-	-	-	0.438	0.756
10	4	14274320	13105355	963	AGCCCACATCRTCTTCTCCTG	2.606	1.835	1.535	-	-	-	-	-	-	2.567	1.566	1.485	-	-	-	-	-	-	2.338	2.606
10	4	14281200	4303614	964	TTCTACTCCTYTTGAATTTCC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	4	14286397	16090800	965	TATAATAMGTWCAGGTCAATA	1.653	2.291	1.652	0.859	-	-	-	-	-	1.571	2.565	1.732	0.827	-	-	-	-	-	1.496	1.993
10	4	14302700	16090803	966	GGTAGGAATAYGGTGTCCTTA	3.495	2.151	1.684	0.752	-	-	-	-	-	3.473	2.362	1.787	0.773	-	-	-	-	-	2.715	3.495
10	4	14321654	4202168	962	AAAGAGCTGCRTAAAGGAGGA	0.801	2.059	1.093	-	-	-	-	-	-	0.746	2.237	1.042	-	-	-	-	-	-	0.584	0.531
10	4	14336342	764759	963	TTAGGCTGCTYCCAAGGCTCA	2.905	1.466	-	-	-	-	-	-	-	2.931	1.509	-	-	-	-	-	-	-	2.254	2.905
10	4	14349730	4690079	964	CTTCCTTATARTCCCTCAGGC	1.441	-	-	-	-	-	-	-	-	1.374	-	-	-	-	-	-	-	-	1.959	1.441
11	5	75934970	3797385	965	AATATCTGAGWTTGACCCTCC	0.247	-	-	-	-	-	-	-	-	0.211	-	-	-	-	-	-	-	-	0.385	0.247
11	5	75938289	1393098	966	ATTGGAACACRATCTAACTGT	0.052	0.384	-	-	-	-	-	-	-	0.035	0.389	-	-	-	-	-	-	-	0.042	0.052
11	5	75942821	3797390	967	GCAGAGGGAGRGAAGGAGGAG	0.183	0.077	0.177	-	-	-	-	-	-	0.165	0.269	0.177	-	-	-	-	-	-	0.259	0.183
11	5	75948599	961530	968	AATACCACTCRCAGAAATAGG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	5	75955389	2869646	969	CAAATTGAGCRGTGAATCACT	0.043	0.072	0.757	0.881	-	-	-	-	-	0.040	0.049	0.778	0.873	-	-	-	-	-	0.005	0.043
11	5	75962825	3797405	970	TAGGCAACCARGAGTTTAATG	0.286	0.757	1.225	1.048	1.555	-	-	-	-	0.348	0.756	1.202	1.015	1.044	-	-	-	-	0.164	0.286
11	5	75968473	2455229	970	TTCATTCCNTRTTAGTGGGTC	1.417	2.120	1.638	0.584	-	-	-	-	-	1.295	2.153	1.680	0.613	-	-	-	-	-	1.008	1.417
11	5	75974090	2537414	971	AGAATGCAAAWTAGACACTGA	2.112	2.096	1.397	-	-	-	-	-	-	2.095	2.178	1.309	-	-	-	-	-	-	1.493	2.112
11	5	75982084	457182	972	TTACATGTGTRTATCATAATA	0.190	1.357	-	-	-	-	-	-	-	0.194	1.493	-	-	-	-	-	-	-	0.489	0.190
11	5	75985949	462307	973	TTGAATGGTGRACATGAACAC	0.756	-	-	-	-	-	-	-	-	0.748	-	-	-	-	-	-	-	-	0.703	0.756
12	5	106906691	8079279	974	CCAAGAAATGGAAGAGGCCAG	0.556	-	-	-	-	-	-	-	-	0.535	-	-	-	-	-	-	-	-	0.503	0.556
12	5	106912962	341272	975	CATTATTTCAYTTGATCCCCA	0.185	3.224	-	-	-	-	-	-	-	0.164	3.451	-	-	-	-	-	-	-	1.214	0.185
12	5	106917770	41375	975	GCACTGTTTTWAAAATCATAC	1.915	1.751	2.751	-	-	-	-	-	-	1.911	1.884	3.318	-	-	-	-	-	-	1.370	1.915
12	5	106921902	352003	977	TTTTAAGTAAYTCGCTTTTTT	0.002	1.622	2.784	2.864	-	-	-	-	-	0.000	1.801	2.541	2.733	-	-	-	-	-	0.995	0.002
12	5	106924516	2484103	681	AGTTCTGTAAYGTGAGCECCA	2.696	1.842	0.995	1.652	1.636	-	-	-	-	2.649	2.067	1.005	1.594	1.585	-	-	-	-	2.179	2.696
12	5	106929756	184851	978	CTGCTCCTGCYCCCCTCTAAG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	5	106932896	609539	979	AGCAACTCAARTTTCCAAGTA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	5	106936032	184700	980	CGTTAAGACAMCCAAAAAGAA	0.047	1.558	1.139	0.997	-	-	-	-	-	0.035	1.574	1.147	0.999	-	-	-	-	-	0.017	0.047
12	5	106941403	597083	682	TTTTTGGAGCRCGGGGGTTCA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	5	106946111	872240	981	AGGAAGCTAARCAAGAAATGT	0.388	0.779	1.233	-	-	-	-	-	-	0.365	0.810	1.211	-	-	-	-	-	-	0.145	0.388
12	5	106950160	9329000	982	AGAGTGCTTTYACTGGATGAC	1.807	0.855	-	-	-	-	-	-	-	1.788	1.053	-	-	-	-	-	-	-	1.719	1.807
12	5	106953865	3044311	683	TATTCTAAACYCTACCTTCCA	0.542	-	-	-	-	-	-	-	-	0.503	-	-	-	-	-	-	-	-	0.452	0.542
13	6	133697011	296457	983	ACTTACAAGAGAAGACAGAAATT	1.631	-	-	-	-	-	-	-	-	1.615	-	-	-	-	-	-	-	-	1.417	1.631
13	6	133704714	439123	984	CAGCAGTCTCYCCCAAGTGACT	0.790	0.505	-	-	-	-	-	-	-	0.823	0.836	-	-	-	-	-	-	-	0.434	0.790
13	6	133716327	9309671	985	CACATAAAAARTAACTCTATA	1.113	0.529	0.638	-	-	-	-	-	-	1.037	0.667	0.593	-	-	-	-	-	-	1.353	1.113
13	6	133722330	7743259	685	TGATTTCAGCWATTCCCTTAT	0.241	0.332	0.325	0.843	-	-	-	-	-	0.201	0.319	0.361	0.844	-	-	-	-	-	0.057	0.241
13	6	133730946	212828	687	TACCATTTCTSTGTTTCTAGG	0.388	0.636	0.360	0.668	3.086	-	-	-	-	0.376	0.616	0.384	0.716	2.805	-	-	-	-	0.153	0.388
13	6	133737200	9402508	688	TCCTGGAAGGYCCTCGCATTA	0.164	0.602	0.565	1.978	1.971	-	-	-	-	0.111	0.628	0.598	1.795	1.833	-	-	-	-	0.429	0.164
13	6	133745088	2216405	986	TTACATACTARTGGAGTCCAG	0.337	0.814	2.348	1.662	-	-	-	-	-	0.303	0.893	2.081	1.643	-	-	-	-	-	0.327	0.337
13	6	133750355	211602	689	GGTTGTACAGWCCACAGAATT	0.587	2.008	2.227	-	-	-	-	-	-	0.538	2.353	2.058	-	-	-	-	-	-	0.783	0.587

13	6	133756903	3777011	690	CAATAAGCCARTGGATGAATG	1.805	0.403	-	-	-	-	-	-	-	0.533	0.408	-	-	-	-	-	-	-	0.616	1.005
13	6	133767521	266423	987	TTTGAGACTAGTAACTTCCTA	0.748	-	-	-	-	-	-	-	-	0.724	-	-	-	-	-	-	-	-	0.450	0.748
14	7	26403805	10231161	694	AAGCTATAAAGGCAAAATATT	1.683	-	-	-	-	-	-	-	-	1.625	-	-	-	-	-	-	-	-	1.252	1.683
14	7	26407530	7706450	695	TAGTGGTGTAYGGAGATCAAT	0.170	3.930	-	-	-	-	-	-	-	0.158	4.113	-	-	-	-	-	-	-	0.394	0.170
14	7	26408560	212053	696	GATTAAATCARTTGGTAAATG	2.524	2.983	5.508	-	-	-	-	-	-	2.643	3.244	4.500	-	-	-	-	-	-	2.498	2.924
14	7	26520304	2299102	698	TAGCACASAAYCTTGCCTATA	3.141	2.204	3.115	4.889	-	-	-	-	-	3.073	2.737	2.545	3.938	-	-	-	-	-	2.597	3.141
14	7	26527922	17310051	699	CCTTCTGGTAYAGTGTTAGGA	0.025	1.945	2.553	2.594	4.786	-	-	-	-	0.003	2.255	2.435	2.279	3.585	-	-	-	-	0.496	0.325
14	7	26530185	1024454	700	CTCAGTTACAYTGGGGAAGAA	3.075	2.255	2.141	2.047	-	-	-	-	-	3.042	2.601	1.900	1.877	-	-	-	-	-	2.547	3.075
14	7	26543634	12780757	701	CAGTTGGCTCYGATTRCCTAC	2.860	2.242	2.206	-	-	-	-	-	-	2.725	2.376	2.005	-	-	-	-	-	-	2.311	2.860
14	7	26550239	7010127	702	ATTACTACATYTGAACATACT	1.562	1.783	-	-	-	-	-	-	-	1.912	1.937	-	-	-	-	-	-	-	2.553	1.952
14	7	26561463	10406476	704	TAGCAAGTAAMACATTAAATT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	7	26575912	1030048	908	CACTGACCAMKTGAAGTTTAC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	7	26580870	4949466	989	TACTGCTCCAYGGATATCTCA	3.290	-	-	-	-	-	-	-	-	3.231	-	-	-	-	-	-	-	-	2.924	3.290
15	7	97460978	6465057	990	ATTGGCTTACYGTTTTATTAG	2.235	-	-	-	-	-	-	-	-	2.224	-	-	-	-	-	-	-	-	1.648	2.235
15	7	97483791	17169462	712	CACTGGAAAAYCAAACCTGTT	1.493	1.093	-	-	-	-	-	-	-	1.483	1.062	-	-	-	-	-	-	-	1.078	1.493
15	7	97507796	2291750	713	CTGGTCTCTCTCRACTTCCGAGG	1.168	0.869	0.874	-	-	-	-	-	-	1.119	0.876	0.837	-	-	-	-	-	-	0.805	1.168
15	7	97530370	12536125	714	AACAAGGATCRGGAGTAGACA	0.117	1.098	1.520	1.341	-	-	-	-	-	0.099	1.134	1.492	1.332	-	-	-	-	-	0.434	0.117
15	7	97552847	4586131	715	CGAGGGATGGGCCAGATATCC	1.972	1.967	2.623	1.654	1.341	-	-	-	-	1.969	2.146	2.341	1.552	1.375	-	-	-	-	1.636	1.972
15	7	97583094	1006169	716	TCTCTACTTCRCTGGTGTTGA	0.580	1.941	2.857	2.823	1.786	1.485	-	-	-	0.578	2.555	2.160	2.373	1.676	1.427	-	-	-	0.846	0.580
15	7	97609637	6963210	717	ATCAGATAGCRTGCAGCCTGA	3.083	2.605	3.960	2.944	2.877	0.630	-	-	-	3.010	2.625	2.602	2.373	2.600	0.826	-	-	-	2.540	3.083
15	7	97590922	10000107	718	GAGSTCCCATYGGGCACAGCT	3.083	3.083	3.636	4.453	1.630	-	-	-	-	3.016	2.852	2.816	2.757	1.386	-	-	-	-	2.540	3.083
15	7	97602273	7007455	719	CTATAGAGAAMTCAAGGACAC	3.083	2.888	4.033	1.990	-	-	-	-	-	3.016	2.728	2.700	1.678	-	-	-	-	-	2.540	3.083
15	7	97641244	4336537	991	CAGGAAACTGYCACTGCTACA	0.470	2.229	1.992	-	-	-	-	-	-	0.464	2.733	1.714	-	-	-	-	-	-	0.817	0.473
15	7	97671957	2394854	992	ATATTTGATAYTGGCTACAGA	3.083	1.719	-	-	-	-	-	-	-	3.034	1.751	-	-	-	-	-	-	-	2.540	3.083
15	7	97699155	10240312	993	ACTGAAGCGTKGGGGAGGTTA	0.889	-	-	-	-	-	-	-	-	0.863	-	-	-	-	-	-	-	-	0.543	0.889
16	7	144659113	2372095	730	TCGTGTCAATGGCAAAATGG	0.764	-	-	-	-	-	-	-	-	0.765	-	-	-	-	-	-	-	-	0.462	0.764
16	7	144662673	7709735	994	CTAAGCTTCAYTGAATGATGC	0.361	0.541	-	-	-	-	-	-	-	0.368	0.337	-	-	-	-	-	-	-	0.373	0.361
16	7	144667900	6950691	731	TAGATGATGGRGTTAACCACC	0.304	1.382	0.659	-	-	-	-	-	-	0.260	1.435	0.581	-	-	-	-	-	-	0.270	0.334
16	7	144674136	6952652	732	GGCTAATGATRTTTAAGTATG	2.014	1.210	1.125	0.652	-	-	-	-	-	1.999	1.483	1.138	0.595	-	-	-	-	-	1.557	2.014
16	7	144676012	1357623	995	GGACTTCAGGYATAAAGCAAA	1.950	1.346	1.310	1.150	0.546	-	-	-	-	1.914	1.582	1.242	1.141	0.695	-	-	-	-	1.535	1.950
16	7	144684484	1523929	996	GGTCTAAATAKCTTCAACCTA	0.284	1.638	1.858	1.316	0.868	-	-	-	-	0.247	1.837	1.617	1.225	0.805	-	-	-	-	0.211	0.284
16	7	144689049	6959021	997	TATATTAGATRATACCGTTTT	0.309	0.312	2.343	1.321	-	-	-	-	-	0.272	0.308	1.988	1.345	-	-	-	-	-	0.245	0.309
16	7	144693586	6464087	998	TAATCATAAAMCCAAGGCTCA	0.284	0.309	0.107	-	-	-	-	-	-	0.275	0.255	0.099	-	-	-	-	-	-	0.211	0.284
16	7	144700050	2056553	999	CTATAGAAACRGGGGTTATAC	0.259	0.151	-	-	-	-	-	-	-	0.225	0.190	-	-	-	-	-	-	-	0.179	0.259
16	7	144705856	6943077	735	GCACTATTGAYTTACAATTTC	0.282	-	-	-	-	-	-	-	-	0.275	-	-	-	-	-	-	-	-	0.435	0.282
17	10	20042742	2025709	736	AGTCTGGTTCYGGAGTTAGCA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	10	20057858	7100283	739	ACACAACTTAMTCAAATTAAT	1.720	-	-	-	-	-	-	-	-	1.668	-	-	-	-	-	-	-	-	1.209	1.720
17	10	20069776	7073601	740	TGGTCTGGCTYTTTTAGAGGT	1.999	0.588	-	-	-	-	-	-	-	2.002	0.568	-	-	-	-	-	-	-	1.453	1.999
17	10	20081089	11811420	1000	TGAAGGTACTRATTTATCCAG	1.061	0.731	0.750	-	-	-	-	-	-	1.093	0.703	0.744	-	-	-	-	-	-	0.817	1.061
17	10	20106875	4740610	741	CTCAGTGCTTYGATAYAGGAG	1.780	1.098	2.590	1.486	-	-	-	-	-	1.746	1.274	2.422	1.444	-	-	-	-	-	1.212	1.780
17	10	20119150	12763948	742	GGTCTTTCAAYTGAAACTAGA	0.329	1.998	3.638	2.247	0.693	-	-	-	-	0.305	2.441	2.401	2.080	0.692	-	-	-	-	0.173	0.329
17	10	20125405	12263244	743	CCTATCCCCARTTTGATACTA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	10	20131109	16918362	744	CAGCAGTCAGYAATTGAATGGT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	10	20142825	11811381	745	CACTCTGCATYGTACAGCATA	2.120	1.234	2.153	1.592	1.125	-	-	-	-	1.963	1.403	1.881	1.538	1.423	-	-	-	-	1.899	2.120
17	10	20150341	11811036	746	GACGTGTTGTYTTAGAATTCT	1.554	1.624	0.984	1.606	-	-	-	-	-	1.483	1.655	0.984	1.531	-	-	-	-	-	1.577	1.554
17	10	20159908	12263862	747	TTCAGGTCAGYATCTACCATG	1.554	1.467	3.325	-	-	-	-	-	-	1.497	1.528	2.067	-	-	-	-	-	-	1.677	1.554
17	10	20170064	333093	748	TCCTTCTCTTYAAATCTTACA	0.384	1.968	-	-	-	-	-	-	-	0.385	2.004	-	-	-	-	-	-	-	0.174	0.384
17	10	20178305	7800249	749	GTCATTATCTSTGGTGGAAAA	2.605	-	-	-	-	-	-	-	-	2.562	-	-	-	-	-	-	-	-	2.054	2.605
18	10	53970048	2403963	750	ATTGGCTCATRTATACTAAAA	0.146	-	-	-	-	-	-	-	-	0.129	-	-	-	-	-	-	-	-	0.299	0.146
18	10	53985118	11002517	1001	AATAAAAACTRTGATTACATA	0.100	0.382	-	-	-	-	-	-	-	0.093	0.381	-	-	-	-	-	-	-	0.410	0.100
18	10	53997189	7072039	752	CCGGAAATTTRGTACAGTGCT	0.919	1.169	1.084	-	-	-	-	-	-	0.829	1.243	1.072	-	-	-	-	-	-	0.715	0.919
18	10	54010434	1919781	1002	GATTGGGAGRCCTCACAGTG	2.110	2.222	1.431	1.070	-	-	-	-	-	2.072	2.315	1.293	1.058	-	-	-	-	-	1.753	2.110
18	10	54023254	4444025	1003	AGTGTAATCAYAAGACACTTT	1.150	2.217	2.251	1.100	0.746	-	-	-	-	1.153	2.448	1.996	1.073	0.763	-	-	-	-	0.928	1.150
18	10	54029767	12415510	753	AAAGTGGTCCRTGGGAAGCAA	2.384	2.574	1.997	1.662	0.986	-	-	-	-	2.318	3.173	1.769	1.412	0.554	-	-	-	-	1.710	2.384
18	10	54040012	2138092	754	TCTTTTGATGWAGGAATTCAG	3.074	2.576	2.690	1.481	-	-	-	-	-	3.092	2.986	2.261	1.345	-	-	-	-	-	2.597	3.074
18	10	54053922	11931702	755	TGTCTGTCACYGAACATATTT	2.747	2.406	2.449	-	-	-	-	-	-	2.721	2.577	2.030	-	-	-	-	-	-	2.093	2.747
18	10	54061798	1919700	756	GGATGGAWGTKGTGGCTGTCT	3.073	2.491	-	-	-	-	-	-	-	3.012	2.916	-	-	-	-	-	-	-	2.548	3.073
18	10	54075328	12210308	757	CTAAAAGTCTRAATTCCTCTG	0.151	-	-	-	-	-	-	-	-	0.155	-	-	-	-	-	-	-	-	0.428	0.151
19	11	92024911	2050706	759	GTGAGATTAAMCCAGAAGCTG	0.736	-	-	-	-	-	-	-	-	0.738	-	-	-	-	-	-	-	-	1.158	0.736

19	11	92030631	2650791	759	TCTGTTTTGGYCACTGTGCAG	0.674	3.453	-	-	-	-	-	-	-	0.632	0.441	-	-	-	-	-	-	-	0.598	0.674
19	11	92048120	2650280	760	AATTCAGATAATGTCGGTTTC	0.015	1.735	1.845	-	-	-	-	-	-	0.003	1.827	1.824	-	-	-	-	-	-	0.433	0.015
19	11	92060783	2650603	1004	CTGTATGTCAGGGACAATGT	3.222	1.712	0.729	0.837	-	-	-	-	-	3.173	1.826	0.737	0.881	-	-	-	-	-	2.479	3.222
19	11	92071877	2650766	1005	TGTCGTGGGGRTTCAACAGCA	0.751	1.483	0.575	0.441	0.311	-	-	-	-	0.719	1.627	0.594	0.448	0.294	-	-	-	-	0.457	0.751
19	11	92083170	2650610	762	TTTGGCCTTAAGGCTACAGTC	1.236	1.064	1.055	0.867	0.117	-	-	-	-	1.145	1.129	1.089	0.852	0.105	-	-	-	-	0.931	1.236
19	11	92090509	986134	763	CAGGATTGGAYCCGGAAATTA	2.221	1.014	0.893	0.232	-	-	-	-	-	2.254	1.019	0.902	0.233	-	-	-	-	-	1.921	2.221
19	11	92098749	10765835	1006	CAATTTCTGCRGTACTGATGA	1.571	1.434	0.377	-	-	-	-	-	-	1.554	1.453	0.368	-	-	-	-	-	-	1.335	1.571
19	11	92923152	4753472	765	ACCACACTAAAATCACAACTC	0.509	3.734	-	-	-	-	-	-	-	0.505	0.721	-	-	-	-	-	-	-	0.215	0.509
19	11	92934982	2399685	1007	TCAGATGATTYCATAGCATGC	0.727	-	-	-	-	-	-	-	-	0.677	-	-	-	-	-	-	-	-	0.403	0.727
20	11	120340968	546239	768	CCTACTCCTCYGTCCCCACTC	1.526	-	-	-	-	-	-	-	-	1.477	-	-	-	-	-	-	-	-	1.350	1.526
20	11	120369953	694068	1009	GATGAAACAGRAAATTATTTT	0.817	1.584	-	-	-	-	-	-	-	0.784	1.564	-	-	-	-	-	-	-	1.397	0.817
20	11	120403249	590436	1009	TAAAATACAATATCTGCCAAC	0.817	0.854	1.084	-	-	-	-	-	-	0.741	0.820	1.136	-	-	-	-	-	-	1.397	0.817
20	11	120437779	10893965	1010	TTCTGTTAAANTACTGTTACT	1.055	0.761	0.515	0.611	-	-	-	-	-	1.048	0.780	0.493	0.608	-	-	-	-	-	1.517	1.055
20	11	120466903	657820	1011	GCTTAGGACGACAACTTCAGT	0.238	0.415	0.370	0.537	0.408	-	-	-	-	0.229	0.423	0.380	0.549	0.436	-	-	-	-	0.596	0.238
20	11	120493738	674207	1012	GTAAAGATGGYAAAGATTAAG	0.979	0.367	0.642	0.423	0.462	0.388	-	-	-	0.948	0.354	0.565	0.445	0.479	0.399	-	-	-	1.539	0.979
20	11	120522190	624054	1013	ACTACACACTKTTGCAACCTA	0.898	0.486	0.407	0.818	0.403	0.395	0.629	-	-	0.874	0.463	0.439	0.864	0.429	0.407	0.627	-	-	1.657	0.898
20	11	120558210	2438791	1014	CATCTCCATTTRGCTTTTGAGA	0.565	0.315	0.150	0.374	0.469	0.446	-	-	-	0.504	0.326	0.283	0.381	0.459	0.446	-	-	-	0.279	0.565
20	11	120560286	2511818	781	TAGAACCAGAYATCAATATTA	0.567	0.429	0.448	0.668	0.496	-	-	-	-	0.542	0.375	0.452	0.647	0.509	-	-	-	-	0.284	0.567
20	11	120610465	6590309	783	ACATTTACTGKGCTCCTCCCT	0.744	0.518	0.267	0.899	-	-	-	-	-	0.692	0.589	0.259	0.896	-	-	-	-	-	0.518	0.744
20	11	120640322	4461739	1015	CGTGCATTCTRGTACCTATCA	0.892	0.453	0.652	-	-	-	-	-	-	0.879	0.467	0.583	-	-	-	-	-	-	1.975	0.598
20	11	120669973	4644651	785	CCTTAAAATCRTAGACCTGAT	0.373	0.423	-	-	-	-	-	-	-	0.373	0.458	-	-	-	-	-	-	-	0.527	0.373
20	11	120696334	10894022	1016	AGTAGCTGGCYCCCGGATGTT	0.678	-	-	-	-	-	-	-	-	0.627	-	-	-	-	-	-	-	-	0.460	0.678
21	13	60642623	3102210	1017	TAGAGCAAAAYTGAGCAAATA	0.357	-	-	-	-	-	-	-	-	0.387	-	-	-	-	-	-	-	-	3.145	3.357
21	13	60676034	3127284	792	CATAAATTTCYGCGTGGTTTA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	13	60807408	9570517	793	TCTACGTCTGKTTTTTATGCAT	0.616	2.498	-	-	-	-	-	-	-	0.565	2.860	-	-	-	-	-	-	-	0.630	0.616
21	13	60714270	7332634	794	CAGCAAATGTYAGGTAGATAA	2.277	1.137	3.213	-	-	-	-	-	-	2.172	1.388	2.715	-	-	-	-	-	-	2.523	2.277
21	13	60727881	17059540	795	ACGGGTTTTAYAGATCTTTCA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	13	6073564C	-	796	ATCATACTACYTGTACCCCCT	0.347	1.050	0.599	1.634	-	-	-	-	-	0.115	1.279	0.584	1.557	-	-	-	-	-	0.243	0.147
21	13	60744842	17059505	797	AGGTAGGTCAYTTGTCTCAGA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	13	60746124	2120881	798	TATATCTATAYGGTAACTCCA	2.282	0.813	0.415	0.540	1.884	-	-	-	-	2.251	0.790	0.456	0.557	1.808	-	-	-	-	2.545	2.282
21	13	60780705	1158312	799	CAACAATTATKTCTGATTTAT	0.825	0.887	0.375	0.488	-	-	-	-	-	0.848	0.862	0.418	0.484	-	-	-	-	-	1.002	0.825
21	13	60800670	7909120	800	CCTGTATAAAYGCATAACAAT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	13	60835413	1999246	1019	CATTATGGCASTACCAAAAAA	0.914	0.526	0.531	-	-	-	-	-	-	0.903	0.486	0.529	-	-	-	-	-	-	0.822	0.914
21	13	60865415	7904186	1019	ATCTTGGGCTYCTYCTTCCTT	0.028	0.341	-	-	-	-	-	-	-	0.012	0.326	-	-	-	-	-	-	-	0.070	0.028
21	13	60892492	9590274	1020	AAACCATCTAYGGTTAAGATC	0.316	-	-	-	-	-	-	-	-	0.315	-	-	-	-	-	-	-	-	0.536	0.316
22	13	61002818	1413564	1021	TCCAATATGCRAAATAAAAAA	0.122	-	-	-	-	-	-	-	-	0.092	-	-	-	-	-	-	-	-	0.052	0.122
22	13	61069137	878472	1022	CTCATCATGTWTGAGTCCTAG	0.208	0.419	-	-	-	-	-	-	-	0.195	0.469	-	-	-	-	-	-	-	0.149	0.238
22	13	61075009	9593570	1023	AAAGCGGATRTAAGCTTATT	0.102	0.483	0.190	-	-	-	-	-	-	0.094	0.473	0.177	-	-	-	-	-	-	0.045	0.102
22	13	61081884	9545801	1024	ATCTATTCCTYAGAGTTTCTG	0.547	0.995	0.850	0.856	-	-	-	-	-	0.511	1.016	0.823	0.915	-	-	-	-	-	0.350	0.547
22	13	61084876	17074824	814	ACTGTATAGCMAAGACCTACT	1.093	0.561	0.678	0.850	1.089	-	-	-	-	1.056	0.543	0.654	0.895	1.095	-	-	-	-	1.055	1.093
22	13	61091040	9601653	1025	TAGGTCTAGARAATTGTTAAA	0.552	0.941	0.840	0.742	1.243	1.053	-	-	-	0.777	0.997	0.847	0.760	1.221	1.047	-	-	-	0.506	0.552
22	13	61096172	12427990	1026	TTTGCATAAAWTGTTATGGTG	0.552	0.941	0.839	0.687	0.735	0.669	-	-	-	0.543	0.935	0.918	0.697	0.776	0.694	-	-	-	0.426	0.552
22	13	61099123	17074541	816	TATGCTTAGCRAGAATGAAAA	1.093	0.897	1.091	0.793	0.415	-	-	-	-	1.046	0.912	1.066	0.831	0.435	-	-	-	-	1.055	1.093
22	13	61104960	2329201	1027	TCTATCTAATKCTGATTAATC	1.207	1.210	0.793	0.823	-	-	-	-	-	1.180	1.352	0.818	0.797	-	-	-	-	-	0.879	1.207
22	13	61107909	2329204	1028	TTTACCGATAYTTATTTCAAT	1.311	1.257	0.445	-	-	-	-	-	-	1.334	1.382	0.529	-	-	-	-	-	-	0.935	1.311
22	13	61113141	9574936	818	TTGTAACAGCRAAGTGGTATC	1.171	0.775	-	-	-	-	-	-	-	1.182	0.834	-	-	-	-	-	-	-	0.854	1.171
22	13	61116860	2876773	1029	TTATATAAAAYCGTGGCTTTG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	13	61120127	9531231	819	TAAGCCTCCTRTAAGTCACAG	0.691	-	-	-	-	-	-	-	-	0.666	-	-	-	-	-	-	-	-	0.544	0.691
23	14	46954105	7150335	1030	GTAATTTCTAYATCAGTAACC	1.093	-	-	-	-	-	-	-	-	1.037	-	-	-	-	-	-	-	-	1.487	1.093
23	14	46977272	10146014	1031	AACAGCATACSGAAAGTTACT	0.877	1.588	-	-	-	-	-	-	-	0.859	1.719	-	-	-	-	-	-	-	1.318	0.877
23	14	47015329	12091944	828	CACGTAAACAHCTTAGGTTTT	3.327	2.165	1.392	-	-	-	-	-	-	3.353	2.184	1.315	-	-	-	-	-	-	2.655	3.327
23	14	47021150	11624882	829	CCCAGGATGAYAGAAATTTAG	2.217	2.267	2.079	1.480	-	-	-	-	-	2.154	2.437	1.797	1.423	-	-	-	-	-	1.841	2.217
23	14	47046266	2017732	832	GAATTATGAAYATGGTTTGTG	0.737	2.403	3.305	1.240	1.035	-	-	-	-	0.743	2.947	2.415	1.438	1.035	-	-	-	-	0.719	0.737
23	14	47059654	10220462	834	AGTCAGAAAAYAGATACAGAA	0.777	2.153	2.051	2.204	1.773	-	-	-	-	0.757	2.353	1.776	1.858	1.641	-	-	-	-	0.634	0.777
23	14	47065968	6572434	835	ACAGTATCAGYATGCTCACTG	3.253	1.703	2.099	3.462	-	-	-	-	-	3.245	1.739	1.911	2.148	-	-	-	-	-	2.630	3.253
23	14	47080731	4900780	837	TCCATTTTATMCTCTTCTTAA	2.542	1.535	2.358	-	-	-	-	-	-	2.547	2.088	2.151	-	-	-	-	-	-	2.247	2.542
23	14	47080453	10483505	839	TTGGTCTGAAYGGAATACAAA	2.015	1.538	-	-	-	-	-	-	-	2.014	1.781	-	-	-	-	-	-	-	1.817	2.015
23	14	47112076	1955793	849	GCGTAATCCCRTAACCAAATA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	14	47130425	8015902	1032	TTGGTAAAATWTTCCAACTAT	1.037	-	-	-	-	-	-	-	-	1.024	-	-	-	-	-	-	-	-	2.013	1.037

Table 4. List of significantly associated haplotypes based on the Schizophrenia Fine Mapping results using the Quebec Founder Population (QFP). Individual haplotypes with associated relative risks are presented in each row of the table; these values were extracted from the associated marker haplotype window with the most significant p value for each SNP in Table 3. The first column lists the region ID as presented in Table 1. The Haplotype column lists the specific nucleotides for the individual SNP alleles contributing to the haplotype reported. The Case and Control columns correspond to the numbers of cases and controls, respectively, containing the haplotype variant noted in the Haplotype column. The Total Case and Total Control columns list the total numbers of cases and controls for which genotype data was available for the haplotype in question. The RR_1 column gives to the relative risk for each particular haplotype. The remainder of the columns lists the SeqIDs for the SNPs contributing to the haplotype and their relative location with respect to the central marker. The Central marker (0) column lists the SeqID for the central marker on which the haplotype is based. Flanking markers are identified by minus (-) or plus (+) signs to indicate the relative location of flanking SNPs.

Region ID	Haplotype	Case	Control	Total Case	Total Control	RR	Central marker (-6)	Central marker (-5)	Central marker (-4)	Central marker (-3)	Central marker (-2)	Central marker (-1)	Central marker (0)	Central marker (+1)	Central marker (+2)	Central marker (+3)	Central marker (+4)	Central marker (+5)	Central marker (+6)
1	G	547	512	916	934	1.122							503						
1	A	369	422	916	934	0.818							503						
1	T	407	361	916	934	1.269							504						
1	G	509	573	916	934	0.788							504						
1	T	409	361	916	934	1.280							893						
1	A	507	573	916	934	0.781							893						
1	A	406	361	916	934	1.264							508						
1	T	510	573	916	934	0.791							508						
1	TTACCAG	146	69	916	934	1.823				504	893	508	509	510	512	513			
1	GATTCGA	476	542	916	934	0.782				504	893	508	509	510	512	513			
1	CCA	160	96	916	934	1.647						509	510	512					
1	TCG	651	716	916	934	0.748						509	510	512					
1	G	756	838	916	934	0.541							512						
1	A	160	96	916	934	1.847							512						
1	A	756	838	916	934	0.541							513						
1	G	160	96	916	934	1.847							513						
1	T	710	735	916	934	0.854							515						
1	A	206	149	916	934	1.529							515						
2	G	553	613	908	936	0.821							895						
2	C	355	323	908	936	1.218							895						
2	GAG	247	307	908	936	0.765						895	896	521					
2	CGA	355	323	908	936	1.218						895	896	521					
2	G	247	307	908	936	0.765							521						
2	A	661	629	908	936	1.305							521						
2	AGGTG	245	307	908	936	0.757					895	521	897	524	526				
2	GAAAC	259	227	908	936	1.245					895	521	897	524	526				
2	GAGGTGATC	243	307	908	936	0.749			895	895	521	897	524	526	527	528	530		
2	CGAAADGCC	259	226	908	936	1.254			895	895	521	897	524	526	527	528	530		
2	AGGTGATCT	213	280	908	936	0.718			896	521	897	524	526	527	528	530	532		
2	GAAACGCCC	259	225	908	936	1.261			896	521	897	524	526	527	528	530	532		
2	A	248	309	908	936	0.762							527						
2	G	660	627	908	936	1.312							527						
2	T	280	341	908	936	0.778							528						
2	C	628	595	908	936	1.285							528						
2	TCT	248	313	908	936	0.748						528	530	532					
3	GTGCCAT	17	32	924	924	0.522				898	899	900	901	552	556	902			
3	CTGTCAT	56	80	924	924	0.681				898	899	900	901	552	556	902			

5	T	264	185	926	930	1.595						608					
5	G	633	724	926	930	0.615						611					
5	T	293	205	926	930	1.627						611					
5	CGTCC	331	379	926	930	0.812				608	611	514	616	619			
5	TTAAC	214	154	926	930	1.515				608	611	614	516	619			
5	TTACC	23	11	926	930	2.120				609	611	614	616	619			
5	CGACC	112	161	926	930	0.657				608	511	614	616	619			
5	TTAACAC	186	125	926	930	1.604			608	611	514	616	619	521	625		
5	TTACCGC	23	10	926	930	2.343			608	611	614	616	519	621	625		
5	CGACCGC	105	145	926	930	0.687			609	611	614	516	619	621	625		
5	TAACACT	213	142	926	930	1.658			611	614	616	619	621	625	910		
5	GACCGCT	100	135	926	930	0.713			611	614	616	619	521	625	910		
5	G	640	705	926	930	0.714						621					
5	A	286	225	926	930	1.400						621					
5	ACT	273	211	926	930	1.425					621	625	910				
5	GCT	168	207	926	930	0.774					621	625	910				
6	T	459	423	922	936	1.202						914					
6	C	463	513	922	936	0.832						914					
6	TAT	459	423	922	936	1.202					914	915	634				
6	TATCTCG	440	404	922	936	1.202			914	915	634	916	917	918	920		
7	T	402	586	916	932	0.655						921					
7	C	434	345	916	932	1.525						921					
7	TCG	475	585	916	932	0.639					921	922	924				
7	CGA	228	181	916	932	1.375					921	922	924				
7	CCG	198	163	916	932	1.301					921	922	924				
7	TCGCG	259	303	916	932	0.818				921	922	924	927	639			
7	TCGCT	214	280	916	932	0.710				921	922	924	927	639			
7	CGATG	216	175	916	932	1.335				921	922	924	927	639			
7	CCGCG	194	151	916	932	1.393				921	922	924	927	639			
7	TCGCTGC	214	277	916	932	0.721			921	922	924	927	639	928	929		
7	CGATGAT	216	175	916	932	1.335			921	922	924	927	639	928	929		
7	CCGCGAT	194	151	916	932	1.390			921	922	924	927	639	928	929		
7	TCGCTGCGT	214	277	916	932	0.721		921	922	924	927	639	928	929	641	930	
7	CGATGATAG	216	175	916	932	1.335		921	922	924	927	639	928	929	641	930	
7	CCGCGATAG	194	151	916	932	1.390		921	922	924	927	639	928	929	641	930	
7	CGCGATAGA	248	293	916	932	0.818		922	924	927	639	928	929	641	930	931	
7	CGCTGCGTG	218	288	916	932	0.690		922	924	927	639	928	929	641	930	931	
7	CGCGATAGG	205	159	916	932	1.402		922	924	927	639	928	929	641	930	931	
7	CTGCGTG	229	292	916	932	0.731			927	639	928	929	641	930	931		
7	TGATAGA	173	143	916	932	1.285			927	639	928	929	641	930	931		
7	TGATAGG	50	33	916	932	1.573			927	639	928	929	641	930	931		
7	CGATAGG	206	159	916	932	1.411			927	639	928	929	641	930	931		
7	GCGATAGAAAG	240	290	916	932	0.785	924	927	639	928	929	641	930	931	932	642	933
7	GCTGCGTGGAC	218	284	916	932	0.713	924	927	639	928	929	641	930	931	932	642	933
7	GCGATAGGAGG	197	154	916	932	1.384	924	927	639	928	929	641	930	931	932	642	933
7	GATAGAAAG	245	291	916	932	0.804		639	928	929	641	930	931	932	642	933	
7	TGCGTGGAC	228	287	916	932	0.745		639	928	929	641	930	931	932	642	933	
7	GATAGAAGG	184	147	916	932	1.342		639	928	929	641	930	931	932	642	933	
7	GATAGGAGG	221	171	916	932	1.415		639	928	929	641	930	931	932	642	933	
7	GATAGAAAGTG	244	290	916	932	0.804	639	928	929	641	930	931	932	642	933	935	936
7	TGCGTGGACCA	227	283	916	932	0.755	639	928	929	641	930	931	932	642	933	935	936
7	GATAGAAGGTG	180	146	916	932	1.317	639	928	929	641	930	931	932	642	933	935	936
7	GATAGGAGGTG	221	171	916	932	1.415	639	928	929	641	930	931	932	642	933	935	936
7	GAAAG	245	291	916	931	0.804				930	931	932	642	933			

10	TTT	107	67	926	936	1.695				663	665	666			
10	C	598	677	926	936	0.697					666				
10	T	328	259	926	936	1.434					666				
10	CGG	580	664	926	936	0.687				666	962	963			
10	TGA	136	93	926	936	1.560				666	962	963			
10	G	710	774	926	936	0.688					963				
10	A	216	162	926	936	1.454					963				
10	G	755	797	926	936	0.779					964				
10	A	171	139	926	936	1.299					964				
11	TCTGTGT	21	8	920	936	2.710		965	966	967	969	970	676	971	
11	TGTAA	260	311	920	936	0.792			967	969	970	676	971		
11	TGTGT	63	41	920	936	1.605			967	969	970	676	971		
11	TAA	343	393	920	936	0.821				970	676	971			
11	CGA	38	62	920	936	0.607				970	676	971			
11	GTC	300	258	920	936	1.272				676	971	972			
11	GAC	93	128	920	936	0.710				676	971	972			
11	TCT	282	244	920	936	1.254				971	972	973			
12	GTT	362	317	924	938	1.262				974	975	976			
12	GTA	249	323	924	938	0.702				974	975	976			
12	CCA	10	22	924	938	0.456				974	975	976			
12	GTTCT	360	313	924	938	1.275			974	975	976	977	681		
12	GTACC	226	290	924	938	0.723			974	975	976	977	681		
12	CTATT	23	11	924	938	2.151			974	975	976	977	681		
12	CCATT	6	19	924	938	0.316			974	975	976	977	681		
12	GTTCTCG	163	132	924	938	1.308		974	975	976	977	681	980	981	
12	GTACCCG	68	112	924	938	0.585		974	975	976	977	681	980	981	
12	CCATTCA	5	15	924	938	0.335		974	975	976	977	681	980	981	
12	T	668	616	924	938	1.364					681				
12	C	256	322	924	938	0.733					681				
12	CCG	77	122	924	938	0.603				681	980	981			
12	TCGGC	280	244	924	938	1.237			681	980	981	982	683		
12	CCGAT	34	60	924	938	0.559			681	980	981	982	683		
12	G	783	755	924	938	1.346					982				
12	A	141	183	924	938	0.743					982				
13	C	453	507	924	934	0.819					983				
13	G	471	427	924	934	1.235					983				
13	CGC	450	501	924	934	0.821				983	984	985			
13	GGT	121	94	924	934	1.347				983	984	985			
13	GGTAGGA	119	93	924	934	1.337		983	984	985	686	687	688	986	
13	GGCACAA	9	22	924	934	0.408		983	984	985	686	687	688	986	
13	GGCACAATA	9	22	924	934	0.408	983	984	985	686	687	688	986	689	690
13	CGCTGGGTA	15	33	924	934	0.415	983	984	985	686	687	688	986	689	690
13	GATTGGGTA	3	11	924	934	0.273	983	984	985	686	687	688	986	689	690
13	GCTGGATAG	45	67	924	934	0.678	984	985	685	687	688	986	689	690	987
13	GCTGGGTAG	35	54	924	934	0.642	984	985	685	687	688	986	689	690	987
13	ATTGGGTAG	3	11	924	934	0.273	984	985	685	687	688	986	689	690	987
13	GGGTA	38	70	924	934	0.529			687	688	986	689	690		
13	GTA	39	70	924	934	0.544				986	689	690			
14	G	408	542	926	934	0.805					694				
14	C	438	392	926	934	1.241					694				
14	GCG	459	528	926	934	0.756				694	695	696			
14	CCA	326	278	926	934	1.282				694	695	696			
14	GCA	27	7	926	934	3.977				694	695	696			
14	GCGTG	456	528	926	934	0.746			694	695	696	696	699		

14	CCACG	326	278	926	934	1.282			694	695	696	698	699		
14	GCACG	27	7	926	934	3.977			694	695	696	698	699		
14	GCGTGCG	425	495	926	934	0.752		694	695	696	698	699	700	701	
14	CCACGTA	326	277	926	934	1.289		694	695	696	698	699	700	701	
14	GCACGTA	27	7	926	934	3.977		694	695	696	698	699	700	701	
14	GCGTGCGAA	425	495	926	934	0.752	694	695	695	696	699	700	701	702	989
14	GCACGTAGG	25	4	926	934	6.451	694	695	695	696	699	700	701	702	989
14	C	427	503	926	934	0.733					700				
14	T	499	431	926	934	1.364					700				
14	G	431	504	926	934	0.743					701				
14	A	495	430	926	934	1.346					701				
14	GAA	426	502	926	934	0.733				701	702	989			
14	AGG	363	316	926	934	1.261				701	702	989			
14	A	459	538	926	934	0.723					989				
14	G	467	396	926	934	1.382					989				
15	C	599	461	924	930	0.773					990				
15	T	525	469	924	930	1.293					990				
15	T	133	168	924	930	0.763					712				
15	C	791	762	924	930	1.311					712				
15	TATCG	62	94	924	930	0.640			712	713	714	715	716		
15	CGTGG	453	398	924	930	1.265			712	713	714	715	716		
15	TATGG	8	1	924	930	8.114			712	713	714	715	716		
15	ATCGG	123	178	924	930	0.686			713	714	715	716	717		
15	GTGGA	400	420	924	930	1.313			713	714	715	716	717		
15	ATGGA	8	1	924	930	8.114			713	714	715	716	717		
15	CGG	217	269	924	930	0.754				715	716	717			
15	GGA	495	427	924	930	1.359				715	716	717			
15	G	428	503	924	930	0.733					717				
15	A	496	427	924	930	1.365					717				
15	A	428	503	924	930	0.733					718				
15	G	496	427	924	930	1.365					718				
15	C	428	503	924	930	0.733					719				
15	A	496	427	924	930	1.365					719				
15	CAT	215	272	924	930	0.734				719	991	992			
15	AAC	495	427	924	930	1.359				719	991	992			
15	T	428	503	924	930	0.733					992				
15	C	496	427	924	930	1.365					992				
16	TAT	270	333	924	932	0.774				994	731	732			
16	TAC	201	167	924	932	1.274				994	731	732			
16	C	646	599	924	932	1.292					732				
16	T	270	333	924	932	0.774					732				
16	T	645	599	924	932	1.285					995				
16	C	279	333	924	932	0.778					995				
16	CTC	279	333	924	932	0.778				995	996	997			
16	TTC	195	159	924	932	1.300				995	996	997			
16	CTCCG	279	333	924	932	0.778			995	996	997	998	999		
16	TTCCG	195	159	924	932	1.300			995	996	997	998	999		
17	G	556	513	922	934	1.247					739				
17	T	366	421	922	934	0.802					739				
17	A	504	537	922	934	1.277					740				
17	G	338	397	922	934	0.783					740				
17	AATAA	441	398	922	934	1.235			740	1000	741	742	745		
17	AGTAG	73	47	922	934	1.623			740	1000	741	742	745		
17	GGCAA	322	376	922	934	0.795			740	1000	741	742	745		

17	AGTAA	7	20	922	934	0.350		740	1000	741	742	745	
17	TAG	78	52	922	934	1.568			741	742	745		
17	CAA	330	388	922	934	0.784			741	742	745		
17	A	840	881	922	934	0.616				745			
17	G	82	53	922	934	1.623				745			
17	AGC	840	881	922	934	0.615			745	746	747		
17	GAT	69	46	922	934	1.537			745	746	747		
17	AGCGG	343	305	922	934	1.215		745	746	747	748	749	
17	AGCGC	84	115	922	934	0.714		745	746	747	748	749	
17	GATAG	52	33	922	934	1.632		745	746	747	748	749	
17	AGCAC	237	280	922	934	0.808		745	746	747	748	749	
17	CGG	349	307	922	934	1.244			747	748	749		
17	CGC	91	121	922	934	0.738			747	748	749		
17	TAG	52	33	922	934	1.632			747	748	749		
17	CAC	238	280	922	934	0.813			747	748	749		
17	G	577	520	922	934	1.332				749			
17	C	345	414	922	934	0.751				749			
18	ATC	93	123	920	936	0.717			1001	752	1002		
18	TTC	465	407	920	936	1.328			752	1002	1003		
18	TCC	67	105	920	936	0.615			752	1002	1003		
18	TCC	481	414	920	936	1.381			1002	1003	753		
18	CCT	65	102	920	936	0.632			1002	1003	753		
18	CCT	490	423	920	936	1.382			1003	753	754		
18	CTA	65	102	920	936	0.632			1003	753	754		
18	T	490	425	920	936	1.364				754			
18	A	430	518	920	936	0.733				754			
18	G	848	822	920	936	1.633				755			
18	A	72	114	920	936	0.612				755			
18	C	489	425	920	936	1.364				756			
18	A	431	511	920	936	0.733				756			
19	CTG	316	275	926	932	1.231			759	760	1004		
19	CTA	153	195	926	932	0.748			759	760	1004		
19	G	471	400	926	932	1.377				1004			
19	A	455	532	926	932	0.725				1004			
19	GTC	399	342	926	932	1.305			1004	1005	762		
19	ATA	232	285	926	932	0.755			1004	1005	762		
19	G	514	458	926	932	1.291				763			
19	A	412	474	926	932	0.775				763			
19	C	649	695	926	932	0.794				1006			
19	T	277	236	926	932	1.259				1006			
19	CAC	158	197	926	932	0.768			1006	766	1007		
20	T	808	785	914	924	1.350				769			
20	C	106	139	914	924	0.741				769			
20	TCC	720	677	914	924	1.354			769	1008	1009		
20	CCC	106	139	914	924	0.741			769	1008	1009		
20	CCCAG	106	139	914	924	0.741		769	1008	1009	1010	1011	
20	CCAGTTA	174	213	914	924	0.785	1008	1009	1010	1011	1012	1013	1014
20	TATGTAG	136	178	914	924	0.733	1013	1014	781	783	1015	785	1016
21	C	291	367	924	934	0.710				1017			
21	T	633	567	924	934	1.408				1017			
21	CGT	267	349	924	934	0.681			1017	793	794		
21	TGC	507	455	924	934	1.280			1017	793	794		
21	CGTTT	228	301	924	934	0.689		1017	793	794	796	798	
21	TGCTC	504	452	924	934	1.280		1017	793	794	796	798	

21	CGTTTTC	150	208	924	934	0.676		1017	793	794	796	798	799	1018	
21	CGCTCGG	12	2	924	934	6.132		1017	793	794	796	798	799	1018	
21	T	371	435	924	934	0.779					798				
21	C	553	499	924	934	1.298					798				
21	TTC	235	281	924	934	0.793				798	799	1018			
21	GAC	49	31	924	934	1.631				1018	1019	1020			
22	GTCGC	1	9	922	930	0.111			1021	1022	1023	1024	814		
22	CGC	1	9	922	930	0.111				1023	1024	814			
22	GTCGCTTTT	1	9	922	930	0.111	1021	1022	1023	1024	814	1025	1026	816	1027
22	TCGCTTTTA	1	9	922	930	0.111	1022	1023	1024	814	1025	1026	816	1027	1028
23	AGG	289	357	922	936	0.740				1030	1031	828			
23	GCT	221	178	922	936	1.343				1030	1031	828			
23	G	404	486	922	936	0.722					828				
23	T	518	450	922	936	1.385					828				
23	GTT	396	479	922	936	0.718				828	829	832			
23	TCT	243	197	922	936	1.343				828	829	832			
23	TTT	396	479	922	936	0.718				829	832	834			
23	CTT	243	197	922	936	1.343				829	832	834			
23	TTA	396	478	922	936	0.721				832	834	835			
23	TTG	243	198	922	936	1.334				832	834	835			
23	A	404	485	922	936	0.725					835				
23	G	518	451	922	936	1.379					835				
23	T	289	355	922	936	0.747					837				
23	G	633	581	922	936	1.338					837				
23	G	682	740	922	936	0.753					838				
23	A	240	195	922	936	1.329					838				
24	CATGT	38	61	920	936	0.584			1034	1035	1036	1037	1038		
24	TGA	54	34	920	936	1.654				846	1039	1040			
25	T	839	875	924	936	1.740					852				
25	C	35	60	924	936	0.575					852				
25	CCA	31	55	924	936	0.556				852	853	854			
25	CAC	31	55	924	936	0.556				853	854	855			
25	G	656	836	924	936	1.506					855				
25	C	68	100	924	936	0.664					855				
25	T	839	876	924	936	1.740					856				
25	C	35	60	924	936	0.575					856				
25	G	638	873	924	936	1.780					860				
25	A	38	63	924	936	0.562					860				
25	AGA	34	59	924	936	0.568				860	1043	1044			
26	GCA	310	355	617	625	0.763				1045	1046	1047			
26	ACG	147	114	617	625	1.402				1045	1046	1047			
26	AAG	3	27	617	625	0.108				1045	1046	1047			
26	GCAGT	282	328	617	625	0.762			1045	1046	1047	1048	1049		
26	AAGAT	3	11	617	625	0.273			1045	1046	1047	1048	1049		
26	GCAGTAT	68	93	617	625	0.709		1045	1046	1047	1048	1049	869	1050	
26	ACGGCGC	57	36	617	625	1.665		1045	1046	1047	1048	1049	869	1050	
26	AAGATAC	3	11	617	625	0.273		1045	1046	1047	1048	1049	869	1050	
26	GAGGCGC	25	12	617	625	2.157		1045	1046	1047	1048	1049	869	1050	
26	ACGGCGCTA	48	30	617	625	1.673	1045	1046	1047	1048	1049	869	1050	1051	1052
26	GAGGCGCTA	19	6	617	625	3.170	1045	1045	1047	1048	1049	869	1050	1051	1052
26	GAGGTATCA	5	14	617	625	0.357	1045	1045	1047	1048	1049	869	1050	1051	1052
26	GCAGTATTA	6	19	617	625	0.313	1045	1046	1047	1048	1049	869	1050	1051	1052
26	CAGTACTGT	12	27	617	625	0.439	1046	1047	1048	1049	869	1050	1051	1052	874
26	CGGCGCTAT	36	20	617	625	1.874	1046	1047	1048	1049	869	1050	1051	1052	874

26	AGGCGCTAT	10	6	617	625	3.100	1046	1047	1048	1049	869	1050	1051	1052	874
26	AGGTATCAT	5	16	617	625	0.311	1046	1047	1048	1049	869	1050	1051	1052	874
26	CAGTATTAT	3	13	617	625	0.230	1046	1047	1048	1049	869	1050	1051	1052	874
26	TACTG	16	31	617	625	0.510			1049	869	1050	1051	1052		
26	CGCTA	79	46	617	625	1.849			1049	869	1050	1051	1052		
26	TATTA	13	29	617	625	0.442			1049	869	1050	1051	1052		
26	CTA	272	219	617	625	1.462				1050	1051	1052			
26	TTA	21	38	617	625	0.544				1050	1051	1052			
26	TAT	207	173	617	625	1.319				1051	1052	874			
27	T	363	331	616	627	1.283					1053				
27	A	253	295	616	627	0.779					1053				
27	C	335	295	616	627	1.342					1054				
27	T	231	332	616	627	0.745					1054				
27	C	363	332	616	627	1.275					879				
27	T	253	295	616	627	0.734					879				
27	CCCAA	121	82	616	627	1.625			1054	879	881	882	885		
27	A	312	288	616	627	1.272					885				
27	C	304	347	616	627	0.785					885				
27	AGA	273	231	616	627	1.364				885	886	887			
27	CGG	303	346	616	627	0.785				885	886	887			
27	A	306	273	616	627	1.280					887				
27	G	310	354	616	627	0.781					887				
27	T	294	251	616	627	1.368					1055				
27	C	322	375	616	627	0.731					1055				

Table 5. List of Schizophrenia candidate genes from the regions identified from the genome wide and fine mapping association analysis. The first column corresponds to the region identifier provided in Table 1. The second and third columns correspond to the chromosome and cytogenetic band, respectively. The fourth and fifth columns corresponds to the chromosomal start coordinates of the NCBI genome assembly derived from build 35 (B35) and the end coordinates (the start and end position relate to the + orientation of the NCBI assembly and don't necessarily correspond to the orientation of the gene). The sixth and seventh columns correspond to the official gene symbol and gene name, respectively, and were obtained from the NCBI Entrez Gene database. The eighth column corresponds to the NCBI Entrez Gene Identifier (GeneID). The ninth and tenth columns correspond to the Sequence IDs from nucleotide (cDNA) and protein entries in the Sequence Listing.

Region ID	Chromosome	Cytogenetic Band	Start Position B35	End position B35	Gene Symbol	Gene Name	Entrez Gene ID	Nucleotide Seq ID	Protein Seq ID
1	1	1p32.3	55758731	55912695	FLJ45337	FLJ45337 protein	400754	1	2
2	1	1p31	75902303	75941204	ACADM	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	34	3	4
2	1	1p31	75919705	75922715	DLSTP	dihydrolipoamide S-succinyltransferase pseudogene (E2 component of 2-oxo-glutarate complex)	1744	-	-
2	1	1p31	75952607	75972785	RABGGTB	Rab geranylgeranyltransferase, beta subunit	5876	5	6
3	1	1q32.1	201645935	201719153	NFASC	neurofascin homolog (chicken)	23114	7	8
3	1	1q32.1	201743997	201776795	CNTN2	contactin 2 (axonal)	6900	9	10
3	1	1q32.1	201733915	201765245	TMEM81	transmembrane protein 81	338730	11	12
3	1	1q32	201737995	201822765	RBBP5	retinoblastoma binding protein 5	5929	13	14
3	1	1q32.1	201843230	201912347	RIPK5	receptor (interacting protein) kinase 5	25778	15,17	16,18
3	1	1q32.1	201947405	201972939	TMCC2	transmembrane and coiled-coil domain family 2	9911	19	20
3	1	1q32.1	202008265	202022540	NUAK2	NUAK family, SNF1-like kinase, 2	81788	21	22
3	1	1q32.1	202037335	202367536	KLHDC8A	kelch domain containing 8A	55220	23	24
7	2	2p24.3-p24.1	23618477	23681987	CENTG2	centaurin, gamma 2	116987	25,27	26,28
11	5	5q13.3	75734505	76035711	IQGAP2	IQ motif containing GTPase activating protein 2	10788	29	30
11	5	5q13	76947064	76954996	F2RL2	coagulation factor II (thrombin) receptor-like 2	2151	31	32
12	5	5q21	106744250	107034435	EFNA5	ephrin-A5	1946	33	34
13	5	5q23	133634236	133852353	EYA4	eyes absent homolog 4 (Drosophila)	2070	35,37,39	36,38,40
14	7	7p21-p15	26479930	26577531	SCAP2	src family associated phosphoprotein 2	8935	41	42
15	7	7q21.3	97383649	97430252	LMTK2	lemur tyrosine kinase 2	22853	43	44
15	7	7q21.3	97486255	97496842	BHLHB9	basic helix-loop-helix domain containing, class B, 9	168620	45	46
15	7	7q21.3	97492687	97526119	DKFZP434E0335	DKFZP434E0335 protein	25851	47	48
15	7	7q21.3	97555663	97556487	BRI3	brain protein I3	25796	49	50
15	7	7q21.3	97557207	97575031	BAIAP2L1	BAI1-associated protein 2-like 1	55971	51	52
17	10	10p12.33-p12.32	19432550	20145269	C10orf112	chromosome 10 open reading frame 112	343695	53	54
17	10	10p12.32-p12.31	23145378	23539121	PLXDC2	plexin domain containing 2	84898	55	56
19	11	11q13.3-q23.3	92851266	92872164	C11orf75	chromosome 11 open reading frame 75	56935	57	58
20	11	11q24-q25	128343052	128359222	RICS	Rho GTPase-activating protein	9743	59	60
21	13	13q21	60631622	60857232	FCDH20	protocadherin 20	64881	61	62
23	14	14q21.3	46983039	46993452	RPS15AP3	ribosomal protein S15a pseudogene 3	325317	-	-
25	20	20p11.1	25543002	25552614	NANP	N-acetylneuraminic acid phosphatase	140838	63	64
25	20	20p11.21	25602652	25625459	ZNF337	zinc finger protein 337	25152	65	66
26	X	Xp22.32	8306651	8509563	KAL1	Kallmann syndrome 1 sequence	3730	67	68
27	X	Xq21.3-q22	92732079	92734753	NAP1L3	nucleosome assembly protein 1-like 3	4675	69	70
27	X	Xq21.32	92735350	92773406	RP1-32F7.2	hypothetical protein FLJ37855	286499	71	72

Table 6. List of additional Schizophrenia candidate genes from the genome wide and Fine Mapping association analyses derived from B36. In order to identify genes not placed in the regions from Table 1 according to Build 36, the region coordinates were converted to Build 36 using the UCSC (University of California Santa Cruz) online program LiftOver. Only new genes that were mapped to this version of the genome assembly are included in this table. The first column corresponds to the region identifier provided in Table 1. The second and third columns correspond to the chromosome and cytogenetic band, respectively. The fourth and fifth columns correspond to the chromosomal start and end coordinates of the NCBI genome assembly derived from build 36 (the start and end position relate to the + orientation of the NCBI assembly and do not necessarily correspond to the orientation of the gene). The sixth and seventh columns correspond to the official gene symbol and gene name, respectively, and were obtained from the NCBI Entrez Gene database. The eighth column corresponds to the NCBI Entrez Gene Identifier (GeneID).
The ninth and tenth columns correspond to the Sequence IDs from nucleotide (cDNA) and protein entries in the Sequence Listing.

Region ID	Chromosome	Cytogenetic Band	Start Position B36	End position B36	Gene Symbol	Gene Name	Entrez Gene ID	Nucleotide Seq ID	Protein Seq ID
2	1	1p31.1	78875445	79661303	LOC646425	hypothetical protein LOC646425	646425	73	74
5	2	2q14.1	118636271	119316408	DPP10	dipeptidyl-peptidase 10	57628	75,77	76,78
9	3	3p12.3	76642959	76643026	LOC643345	hypothetical protein LOC643345	643345	79	80
15	7	7q21.3	97775139	97777695	FLJ30064	hypothetical protein LOC644975	644975	81	82
15	7	7q21.3	97841346	97853865	LOC644972	similar to ribosomal protein BSa	644972	83	-
25	20	20p11.21	25623212	25681395	LOC646356	similar to nuclear receptor co-repressor 1	646356	84	85
25	20	20p11.21-p11	25692084	25796780	RP13-401N9	hypothetical gene supported by BC042812	388388	86,88,90,92	87,89,91,93
26	X	Xp22	8382871	8394647	VCX-D	variably charged X-D	426054	94	95

Table 7 List of candidate genes based on EST clustering from the regions identified from the various Fine Mapping analyses. The first column corresponds to the region identifier provided in Table 1. The second column corresponds to the chromosome number. The third and fourth columns correspond to the chromosomal start and end coordinates of the NCBI genome assemblies derived from build 35 (B35). The fifth column corresponds to the ECGene Identifier, corresponding to the ECGene track of UCSC. These ECGene entries were determined by their overlap with the regions from Table 1, based on the start and end coordinates of both Region and ECGene identifiers.The sixth and seventh columns correspond to the Sequence IDs from nucleotide and protein entries in the Sequence Listing.

Region ID	Chromosome	Start Position Build35	End Position Build35	Name	Nucleotide Seq ID	Protein Seq ID
1	1	56001480	56010260	H1C9375.1	96	97
1	1	56001500	56045938	H1C9375.2	98	99
2	1	75806660	75820154	H1C11099.1	100	101
2	1	75887711	75901717	H1C11103.1	102	103
2	1	75902080	75941222	H1C11105.1	104	105
2	1	75902080	75941385	H1C11105.2	106	107
2	1	75902410	75941222	H1C11105.3	108	109
2	1	75903517	75904821	H1C11107.1	110	111
2	1	75910348	75910945	H1C11105.4	112	113
2	1	75922104	75924758	H1C11110.1	114	115
2	1	75963902	75972539	H1C11113.1	116	117
2	1	75963902	75972795	H1C11113.2	118	119
2	1	75963903	75972795	H1C11113.3	120	121
2	1	75963906	75967819	H1C11113.4	122	123
2	1	75963981	75972539	H1C11113.5	124	125
3	1	201529434	201711687	H1C23767.1	126	127
3	1	201529472	201677931	H1C23767.2	128	129
3	1	201529472	201719153	H1C23767.3	130	131
3	1	201529472	201719530	H1C23767.4	132	133
3	1	201529472	201720355	H1C23767.5	134	135
3	1	201571387	201645813	H1C23767.6	136	137
3	1	201687459	201719530	H1C23767.7	138	139
3	1	201687459	201720355	H1C23767.8	140	141
3	1	201697628	201719530	H1C23767.9	142	143
3	1	201743996	201778801	H1C23797.1	144	145
3	1	201759389	201778801	H1C23797.2	-146	147
3	1	201774480	201778801	H1C23802.1	148	149
3	1	201779134	201782970	H1C23804.1	150	151
3	1	201783912	201785361	H1C23805.1	152	153
3	1	201786928	201789603	H1C23807.1	154	155
3	1	201786928	201822800	H1C23808.1	158	159
3	1	201786928	201822800	H1C23808.2	156	157
3	1	201786933	201788072	H1C23809.1	160	161
3	1	201786935	201787629	H1C23810.1	162	163
3	1	201843280	201851530	H1C23817.1	164	165
3	1	201843280	201864539	H1C23817.2	166	167
3	1	201843280	201912347	H1C23817.3	170	171
3	1	201843280	201912347	H1C23817.4	168	169
3	1	201843280	201912384	H1C23817.5	172	173
3	1	201860703	201875300	H1C23817.6	174	175
3	1	201882156	201912384	H1C23817.7	176	177
3	1	201928960	201942607	H1C23833.1	178	179
3	1	201930516	201974128	H1C23833.2	180	181
3	1	201934600	201937652	H1C23836.1	182	183
3	1	201947404	201974128	H1C23833.3	184	185

3	1	201956985	201974128	H1C23833.4	188	189
3	1	201956985	201974128	H1C23833.5	186	187
3	1	201958015	201974128	H1C23833.6	190	191
3	1	201968248	201974128	H1C23833.7	192	193
3	1	202002843	202022540	H1C23842.1	194	195
3	1	202036849	202040646	H1C23853.1	196	197
3	1	202036849	202045687	H1C23851.1	198	199
3	1	202036849	202057662	H1C23852.1	200	201
3	1	202036849	202057696	H1C23853.2	202	203
3	1	202036849	202057823	H1C23853.3	204	205
3	1	202037304	202040646	H1C23853.4	206	207
3	1	202037304	202057662	H1C23852.2	208	209
3	1	202037304	202057696	H1C23853.5	210	211
3	1	202037304	202057823	H1C23853.6	212	213
3	1	202042342	202046781	H1C23856.1	214	215
5	2	115302239	115307355	H2C12594.1	216	217
5	2	115604785	115634998	H2C12603.1	218	219
5	2	115617854	115635582	H2C12603.2	220	221
5	2	115617854	115657324	H2C12603.3	222	223
5	2	115635961	116000192	H2C12598.1	224	225
5	2	115635997	116317136	H2C12598.2	226	227
5	2	115635997	116318556	H2C12598.3	228	229
7	2	236184715	236543845	H2C24136.1	230	231
7	2	236184715	236815899	H2C24136.2	232	233
7	2	236184715	236817410	H2C24136.3	234	235
7	2	236418208	236421124	H2C24227.1	236	237
7	2	236464496	236466438	H2C24246.1	238	239
7	2	236464496	236474031	H2C24246.2	240	241
10	4	14250039	14566043	H4C1875.1	242	243
10	4	14306653	14394488	H4C1878.1	244	245
11	5	75734904	76039713	H5C6651.1	246	247
11	5	75938193	75940551	H5C6697.1	248	249
11	5	75940694	76038990	H5C6651.2	250	251
11	5	75946993	75954996	H5C6700.1	252	253
11	5	75947063	75954996	H5C6700.2	254	255
12	5	106744112	107034495	H5C9152.2	256	257
13	6	133604187	133894951	H6C13079.1	260	261
13	6	133604187	133894951	H6C13079.2	258	259
13	6	133604187	133894951	H6C13079.3	262	263
13	6	133604187	133894951	H6C13079.4	264	265
13	6	133723035	133809585	H6C13079.6	266	267
14	7	26479920	26677604	H7C2856.1	268	269
14	7	26553224	26677497	H7C2856.2	270	271
14	7	26571128	26571874	H7C2880.1	272	273
15	7	97380847	97479698	H7C10256.1	274	275
15	7	97380847	97483601	H7C10256.2	276	277
15	7	97485428	97489419	H7C10289.1	278	279
15	7	97486254	97486842	H7C10290.1	280	281
15	7	97489405	97490716	H7C10292.1	282	283
15	7	97489405	97496924	H7C10293.1	284	285
15	7	97489405	97526162	H7C10293.2	286	287
15	7	97489479	97496924	H7C10293.3	288	289
15	7	97489479	97507805	H7C10293.4	290	291
15	7	97489479	97526162	H7C10293.5	292	293
15	7	97500554	97526181	H7C10293.6	294	295
15	7	97514696	97515258	H7C10293.7	296	297
15	7	97524876	97526138	H7C10293.8	298	299

15	7	97526327	97565490	H7C10297.1	300	301
15	7	97526362	97526991	H7C10298.1	302	303
15	7	97527397	97539455	H7C10300.1	304	305
15	7	97555629	97565490	H7C10297.2	306	307
15	7	97555629	97565490	H7C10297.3	308	309
15	7	97555629	97567558	H7C10297.4	310	311
15	7	97555680	97556682	H7C10297.5	312	313
15	7	97555680	97565487	H7C10304.1	314	315
15	7	97555680	97565490	H7C10297.6	316	317
15	7	97557545	97565490	H7C10297.7	318	319
15	7	97565612	97566988	H7C10297.8	320	321
15	7	97565612	97578937	H7C10309.1	322	323
15	7	97565612	97675068	H7C10309.2	324	325
15	7	97565612	97675078	H7C10309.3	326	327
15	7	97565648	97566931	H7C10310.1	328	329
15	7	97567201	97675078	H7C10309.4	330	331
15	7	97567202	97578937	H7C10309.5	332	333
15	7	97567202	97675068	H7C10309.6	334	335
15	7	97575994	97577898	H7C10316.1	336	337
15	7	97581475	97584404	H7C10319.1	338	339
15	7	97659691	97660570	H7C10337.1	344	345
15	7	97659691	97660570	H7C10337.2	342	343
15	7	97659691	97660570	H7C10337.3	346	347
15	7	97659691	97660570	H7C10338.1	340	341
15	7	97659710	97660571	H7C10337.4	348	349
15	7	97659712	97660570	H7C10337.5	350	351
17	10	19963498	20063413	H10C2475.1	352	353
17	10	20021244	20119336	H10C2475.2	354	355
17	10	20021413	20063413	H10C2475.3	356	357
17	10	20039267	20057485	H10C2479.1	358	359
17	10	20050522	20057457	H10C2479.2	360	361
17	10	20050522	20057485	H10C2479.3	362	363
17	10	20062708	20064004	H10C2480.1	364	365
17	10	20145352	20609127	H10C2485.1	368	369
17	10	20145352	20609127	H10C2485.2	366	367
17	10	20145352	20609288	H10C2485.3	370	371
17	10	20145352	20610607	H10C2485.4	372	373
18	10	53986346	54185170	H10C5984.1	374	375
18	10	53986746	54185170	H10C5984.2	376	377
19	11	92851286	92854311	H11C10888.1	378	379
19	11	92851286	92896181	H11C10888.2	380	381
19	11	92851286	92910729	H11C10888.3	382	383
19	11	92851286	92916322	H11C10888.4	384	385
19	11	92877687	92896867	H11C10894.1	386	387
20	11	128340169	128345810	H11C14777.1	388	389
20	11	128340169	128399298	H11C14778.1	390	391
20	11	128342923	128343838	H11C14781.1	392	393
20	11	128343868	128345810	H11C14777.2	394	395
20	11	128343868	128539450	H11C14778.2	396	397
20	11	128360984	128441918	H11C14778.3	398	399
20	11	128468538	128539533	H11C14778.4	400	401
20	11	128608882	128612812	H11C14805.1	402	403
20	11	128608882	128658269	H11C14804.1	404	405
21	13	60881679	60887288	H13C4225.1	406	407
21	13	60881679	60887294	H13C4225.2	408	409
21	13	60881993	60887294	H13C4225.3	410	411
21	13	60881993	60899983	H13C4225.4	412	413

23	14	46379746	47213703	H14C2609.4	414	415
23	14	47029284	47076104	H14C2632.1	416	417
25	20	25541574	25552621	H20C2759.1	418	419
25	20	25543001	25552615	H20C2759.2	420	421
25	20	25545326	25552607	H20C2759.3	422	423
25	20	25552671	25563460	H20C2762.1	424	425
25	20	25552680	25606706	H20C2762.2	426	427
25	20	25552688	25606184	H20C2762.3	428	429
25	20	25552694	25605787	H20C2762.4	430	431
25	20	25552716	25568455	H20C2762.5	432	433
25	20	25552792	25574283	H20C2762.6	434	435
25	20	25552860	25599487	H20C2762.7	436	437
25	20	25560793	25563551	H20C2762.8	438	439
25	20	25602762	25614417	H20C2768.1	440	441
25	20	25602762	25625231	H20C2768.2	442	443
25	20	25602762	25625451	H20C2768.3	444	445
25	20	25602762	25625477	H20C2768.4	446	447
25	20	25602851	25614417	H20C2768.5	448	449
25	20	25602851	25615593	H20C2768.6	450	451
25	20	25602851	25625231	H20C2768.7	452	453
25	20	25602851	25625451	H20C2768.8	454	455
25	20	25602851	25625477	H20C2768.9	456	457
25	20	25603333	25604164	H20C2762.9	458	459
25	20	25625768	25663608	H20C2776.1	460	461
25	20	25625799	25663845	H20C2777.1	462	463
25	20	25681195	25681490	H20C2781.1	464	465
25	20	25692090	25724878	H20C2782.1	466	467
25	20	25692090	25796786	H20C2782.2	468	469
25	20	25698233	25702079	H20C2783.1	470	471
25	20	25698233	25729886	H20C2782.3	472	473
25	20	25702200	25703906	H20C2782.4	474	475
25	20	25702200	25715459	H20C2782.5	476	477
25	20	25702654	25791837	H20C2782.6	478	479
25	20	25703397	25790213	H20C2782.7	480	481
25	20	25762777	25769694	H20C2797.1	482	483
25	20	25773302	25782657	H20C2782.8	484	485
25	20	25774180	25777579	H20C2782.9	486	487
26	X	8243155	8244283	HXC512.1	488	489
26	X	8306650	8509964	HXC517.1	490	491
27	X	92732070	92734697	HXC5192.1	492	493
27	X	92732070	92734712	HXC5192.2	494	495
27	X	92732070	92734793	HXC5192.3	496	497
27	X	92732078	92734802	HXC5192.4	498	499
27	X	92735349	92773418	HXC5193.1	500	501

WE CLAIM:

1. A method of constructing a schizophrenia disorder GeneMap in a human population, comprising screening for the expression level of or presence or absence of at least one allele of at least one gene from Tables 5-7 in at least one sample.

2. A method of claim 1, wherein said population is a general population.

3. A method of claim 1, wherein said population is a founder population.

4. A method of claim 3, wherein said founder population is the population of Quebec.

5. A method of constructing a GeneMap in a human population, comprising screening for the presence or absence of an allele of at least one single nucleotide polymorphisms (SNPs) from Tables 2-4 in at least one sample.

6. A method of claim 5, wherein said population is a general population.

7. A method of claim 5, wherein said population is a founder population.

8. A method of claim 7, wherein said founder population is the population of Quebec.

9. A method of genetic mapping for detecting the association of at least one marker for schizophrenia comprising: a) obtaining biological samples from at least one patient; b) screening for the presence or absence of an allele of at least one SNP or a group of SNPs from Tables 2-4 within each biological sample; and c) evaluating whether said SNP or a group of SNPs shows a statistically significant skewed genotype distribution between a group of cases compared to a group of control.

10. A method of claim 9, wherein said biological samples are hair, fluid, serum, tissue or buccal swabs, saliva, mucus, urine, stools, spermatozoids, vaginal secretions, lymph, amniotic fluid, pleural liquid or tears.

11. A method of claim 9, wherein said cases and controls are from a human population.

12. A method of claim 11, wherein said population is the general population.

13. A method of claim 11, wherein said population is a founder population.

14. A method of claim 13, wherein said founder population is the population of Quebec.

15. A method of claim 9, wherein said cases and controls are recruited independently according to specific phenotypic criteria.

16. A method of claim 9, wherein said cases and controls are recruited in the form of trios, comprising two parents and one child or one parent and two children, or in the form of unrelated individuals.

17. A method of claim 9, wherein said screening is performed by a method selected from the group consisting of an allele-specific hybridization assay, an oligonucleotide ligation assay, an allele-specific elongation/ligation assay, an allele-specific amplification assay, a single-base extension assay, a molecular inversion probe assay, an invasive cleavage assay and a selective termination assay. Other assay: RFLP, sequencing assay, SSCP, mismatch-cleaving assay, denaturing gradient gel electrophoresis.

18. A method of claim 9, wherein said screening is carried out on each individual of a cohort at each of at least one SNP or a group of SNPs from Tables 2-3.

19. A method of claim 9, wherein said screening is carried out on pools of cases and pools of controls.

20. A method of claim 9, wherein the genotype distribution is compared one SNP at a time.

21. A method of claim 9, wherein a genotype distribution is compared with a group of markers from Tables 2-4 forming a haplotype.

22. A method of claim 9, wherein a genotype distribution is compared using the allelic frequencies between the cases pools and control pools.

23. A set of genetic markers comprising at least two SNPs of Tables 2-4.

24. A set of nucleic acid probes that specifically detect said SNPs of claim 23.

25. A solid support or collection of solid supports comprising the nucleic acid probes of claim 24.

26. A solid support of claim 25, wherein the support is selected from the group consisting of at least one microarray and a set of beads.

27. A method for predicting the efficacy of a drug for treating an schizophrenia in a human patient, comprising: a) obtaining a sample of cells from the patient; b) obtaining a gene expression profile from the sample in the absence and presence of *in vitro* modulation of the cells with specific mediators; the gene expression profile comprising one or more genes from Tables 5-7; and c) comparing the gene expression profile of the sample with a reference gene expression profile, wherein similarity between the sample expression profile and the reference expression profile predicts the efficacy of the drug for treating schizophrenia in the patient.

28. The method of claim 27, further comprising exposing the sample to the drug for treating schizophrenia prior to obtaining the gene expression profile of the sample.

29. The method of claim 27, wherein the sample of cells is derived from a tissue selected from the group consisting of: any parts of the body such as the scalp, blood, dermis, brain, colon, epidermis and other skin cells, cutaneous surfaces, intertrigious areas, genitalia, vessels and endothelium.

30. The method of claim 29, wherein the cells are selected from the group consisting of: red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

31. The method of claim 27, wherein the sample is obtained via biopsy.

32. The method of claim 27, wherein the gene expression profile comprises expression values for at least two of the genes listed in Tables 5-7.

33. The method of claim 32, wherein the gene expression profile of the sample is obtained by detecting the protein products of said genes.

34. The method of claim 27, wherein the gene expression profile of the sample is obtained using a hybridization assay to oligonucleotides contained in a microarray.

35. The method of claim 34, wherein the oligonucleotides comprises nucleic acid molecules at least 95% identical to SEQ ID from Tables 2-7.

36. The method of claim 27, wherein the reference expression profile is that of cells derived from patients that have schizophrenia.

37. The method of claim 27, wherein the drug is selected from the group consisting of symptom relievers and drugs for schizophrenia.

38. A method for inducing a schizophrenia-like state in a resident tissue or cell, comprising contacting the tissue or cell with at least one gene from Tables 5-7 that induces a schizophrenia-like state.

39. The method of claim 38, wherein the resident tissue cell is selected from the group consisting of red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

40. A method for screening drug candidates for treating schizophrenias, comprising: a) contacting a resident cell induced by the method of claim 38 with a drug candidate for treating schizophrenia; and b) assaying for a schizophrenia-like state, such that an absence of the schizophrenia-like state is indicative of the drug candidate being effective in treating schizophrenia.

41. A method for inducing a resident tissue cell to mimic schizophrenia, comprising modulating the expression of at least one gene from Tables 5-7 in the cells.

42. The method of claim 41, wherein the resident tissue cell is selected from the group consisting of red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

43. A method for screening drug candidates for treating schizophrenia, comprising: a) contacting the resident tissue cell induced according to the method of claim 41 with a drug candidate for treating schizophrenia; and b) assaying for a schizophrenia-like state, such that an absence of the schizophrenia-like state is indicative of the drug candidate being effective in treating schizophrenia.

44. A method for treating an animal having schizophrenia, or related phenotype, comprising administering a drug identified by the method of claim 43.

45. A drug screening assay comprising: a) administering a test compound to an animal having an schizophrenia, or related phenotype, or a cell composition isolated therefrom; and b) comparing the level of gene expression of at least one gene from Tables 5-7 in the presence of the test compound with one or both of the level of said gene expression in the absence of the test compound or in normal cells; wherein test compounds which cause the level of expression of one or more genes from Tables 5-7 to approach normal are candidates for drugs to treat schizophrenia.

46. A method for treating an animal having schizophrenia, or related phenotype, comprising administering a compound identified by the assay of claim 45.

47. A pharmaceutical preparation for treating an animal having schizophrenia, or related phenotype, comprising a compound identified by the assay of claim 45 and a pharmaceutically acceptable excipient.

48. A method for identifying a gene that regulates drug response in an schizophrenia, comprising: a) obtaining a gene expression profile for at least one gene from Tables 5-7 in a resident tissue cell induced for a schizophrenia-like state in the presence of the candidate drug; and b) comparing the expression profile of said gene to a reference expression profile for said gene in a cell induced for the schizophrenia-like state in the absence of the candidate drug, wherein genes whose expression relative to the reference expression profile is altered by the drug may identifies the gene as a gene that regulates drug response in the schizophrenia disorder.

49. An expression profile indicative of the presence of schizophrenia disorder in a patient, comprising the level of expression of at least one gene of Tables 5-7.

50. A microarray comprising probes that hybridize to one or more genes of Tables 5-7.

51. A method of diagnosing susceptibility to schizophrenia in an individual, comprising screening for an at-risk haplotype of at least one gene or gene region from Tables 5-7, or at least one SNP from Table 2-4, that is more frequently present in an individual susceptible to schizophrenia compared to a control individual, wherein the at-risk haplotype increases risk of schizophrenia.

52. The method of claim 51 wherein the risk increase is at least about 20%.

53. A method of diagnosing susceptibility to schizophrenia in an individual, comprising screening for an at-risk haplotype of at least one gene from Tables 5-7, or comprising at least one SNP from Tables 2-4, that is more frequently present in an individual susceptible to schizophrenia, compared to the frequency of its presence in a control individual, wherein the presence of the at-risk haplotype is indicative of a susceptibility to schizophrenia.

54. The method of claim 53 wherein the at-risk haplotype is characterized by the presence of at least one single nucleotide polymorphism from Tables 2-4.

55. The method of claim 53 wherein screening for the presence of an at-risk haplotype in at least one gene from Tables 5-7, or comprising at least one SNP from

Table 1, comprises enzymatic amplification of nucleic acid from said individual or amplification using universal oligos on elongation/ligation products.

56. The method of claim 55 wherein the nucleic acid is DNA.

57. The method of claim 55 wherein the DNA is human DNA.

58. The method of claim 53 wherein screening for the presence of an at-risk haplotype in at least one gene from Tables 5-7 or comprising at least one SNP from Tables 2-4 comprises: a) obtaining material containing nucleic acid from the individual; (b) amplifying said nucleic acid as for claim 55; and c) determining the presence or absence of an at-risk haplotype in said amplified nucleic acid.

59. The method of claim 58 wherein determining the presence of an at-risk haplotype is performed by electrophoretic analysis.

60. The method of claim 58 wherein determining the presence of an at-risk haplotype is performed by restriction length polymorphism analysis.

61. The method of claim 58 wherein determining the presence of an at-risk haplotype is performed by sequence analysis.

62. The method of claim 58 wherein determining the presence of an at-risk haplotype is performed by hybridization analysis.

63. A kit for diagnosing susceptibility to schizophrenia in an individual comprising: primers for nucleic acid amplification of a region of at least one gene from Tables 5-7 or a region comprising at least one SNP from Tables 2-4.

64. The kit of claim 63 wherein the primers comprise a segment of nucleic acids of length suitable for nucleic acid amplification of a target sequence, selected from the group consisting of: single nucleotide polymorphism from Tables 2-4, primers flanking a SNP from Tables 2-4, and combinations thereof.

65. A method of diagnosing a susceptibility to schizophrenia, comprising detecting an alteration in the expression or composition of a polypeptide encoded by at least one gene from Tables 5-7 in a test sample, in comparison with the expression or composition of a polypeptide encoded by said gene in a control sample, wherein the

presence of an alteration in expression or composition of the polypeptide in the test sample is indicative of a susceptibility to schizophrenia.

66. The method of claim 65, wherein the alteration in the expression or composition of a polypeptide encoded by said gene comprises expression of a splicing variant polypeptide in a test sample that differs from a splicing variant polypeptide expressed in a control sample.

67. A method of treating schizophrenia in a patient in need thereof, comprising expressing *in vivo* at least one gene from Tables 5-7 (wild type/non-disorder associated allele) in an amount sufficient to treat the disorder.

68. A method of claim 67, comprising: a) administering to a patient a vector comprising a gene selected from Tables 5-7 that encodes the protein; and b) allowing said protein to be expressed from said gene in said patient in an amount sufficient to treat the disorder.

69. A method of claim 68, wherein said vector is selected from the group consisting of an adenoviral vector, and a lentiviral vector.

70. A method of claim 68, wherein said vector is administered by a route selected from the group consisting of: topical administration, intraocular administration, parenteral administration, intranasal administration, intratracheal administration, intrabronchial administration and subcutaneous administration.

71. A method of claim 68, wherein said vector is a replication-defective viral vector.

72. A method of claim 68, wherein said gene encodes a human protein.

73. A method of treating an schizophrenia in a patient in need thereof, comprising administering an agent that regulates the expression, activity or physical state of at least one gene or its encoded RNA from Tables 5-7 in the patient.

74. A method of claim 73, wherein the encoded protein from said gene comprises an alteration.

75. A method of claim 73, wherein said gene comprises an associated allele, a particular allele of a polymorphic locus, or the like that modulates the expression of the encoded protein.

76. A method of claim 73, wherein said agent is selected from the group consisting of chemical compounds, oligonucleotides, peptides and antibodies.

77. A method of claim 73, wherein said agent is an antisense molecule or interfering RNA.

78. A method of claim 73, wherein said agent is an expression modulator.

79. A method of claim 78, wherein said modulator is an activator.

80. A method of claim 78, wherein said modulator is a repressor.

81. A method of claim 73, wherein said gene comprises an associated allele, a particular allele of a polymorphic locus, or the like that modifies at least one property or function of the encoded protein.

82. A method of claim 81, wherein the agent modulates at least one property or function of said gene or a polymorphism wherein at least one allele of said polymorphism modifies at least one property or function of the encoded protein.

83. A method for preventing the occurrence of schizophrenia in an individual in need thereof, comprising regulating the level of at least one gene from Tables 5-7 compared to a control.

84. The method of claim 83 wherein said level is regulated by regulating expression of at least one gene from Tables 5-7 using a binding agent, a receptor to said gene, a peptidomimetic, a fusion protein, a prodrug, an antibody or a ribozyme.

85. The method of claim 83 wherein said level is controlled by genetically altering the expression level of at least one gene from Tables 5-7, whereby the regulated level of said gene mimics the level in a control individual.

86. A method for monitoring the effectiveness of treatment on the regulation of expression of one or more genes from Tables 5-7 at the RNA or protein level, or its

enzymatic activity by measuring RNA, protein or enzymatic activity in a sample of peripheral blood or cells derived thereof.

87. A method of diagnosing an schizophrenia, the predisposition to schizophrenia, or the progression of schizophrenia, comprising the steps of a) obtaining a biological sample of mammalian body fluid or tissue to be diagnosed; b) determining the amount and/or concentration of at least one polypeptide from Tables 5-7 and/or nucleic acids encoding the polypeptide present in said biological sample; and c) comparing the amount and/or concentration of said polypeptide determine in said biological sample with the amount and/or concentration of said polypeptide as determined in a control sample and/or comparing the amount and/or concentration of nucleic acids encoding said polypeptide determined in said biological sample with the amount and/or concentration of nucleic acids encoding said polypeptides measured in a control sample, wherein the difference in the amount of said polypeptides and/or nucleic acids encoding the polypeptides is indicative of schizophrenia or the stage of schizophrenia.

88. A method according to claim 87, wherein a nucleic acid probe is used for determining the amount and/or concentration of at least one nucleic acid from Tables 5-7 encoding the polypeptide.

89. A method according to claim 88, wherein said nucleic acid probe is derived from the nucleic acid sequence depicted in SEQ ID NO: 1 to 1056.

90. A method according to claim 88, wherein said nucleic acid probe comprises nucleic acids hybridizing to the nucleic acid sequence depicted in SEQ ID NO: 1 to 1056, and/or fragments thereof.

91. A method according to claim 88, wherein a PCR technique is employed.

92. A method according to claim 87, wherein a specific antibody is used for determining the amount and/or concentration of at least one polypeptide from Tables 5-7.

93. A method according to claim 92, wherein said antibody is selected from the group comprising polyclonal antiserum, polyclonal antibody, monoclonal antibody, antibody fragments, single chain antibodies and diabodies.

94. A method of claim 87, wherein at least five polypeptides or nucleic acids are determined.

95. Use of a method according to claim 87, wherein the diagnosis serves as a basis for prevention and/or monitoring of schizophrenia disorder.

96. A method of treatment of schizophrenia in a mammal in need thereof, comprising the steps of a) performing steps a) to c) according to claim 87; and b) treating the mammal in need of said treatment; wherein said medical treatment is based on the stage of the disorder.

97. A method for determining the phenotype of a cell comprising detecting the differential expression, relative to a normal cell, of at least one gene from Tables 5-7.

98. The method of claim 97, wherein said difference in the level of expression of said gene, is of at least a factor of about two.

99. The method of claim 98, including the further step of cloning said genes which are up- or down-regulated.

100. The method of claim 98, including the further step of generating nucleic acid probes for detecting the level of expression of said genes which are up- or down-regulated.

101. A kit for assessing a patient's risk of having or developing schizophrenia, comprising: a) detection means for detecting the differential expression, relative to a normal cell, of at least one gene shown in Tables 5-7 or the gene product thereof; and b) instructions for correlating the differential expression of said gene or gene product with a patient's risk of having or developing schizophrenia.

102. The kit of claim 101, wherein the detection means includes nucleic acid probes for detecting the level of mRNA of said genes.

103. A kit for assessing a patients risk of having or developing schizophrenia, comprising: at least one means for amplifying or detecting a sequence of (a) at least one gene in Tables 5-7, or at least one sequence comprising a SNP in Tables 2-4, wherein the detection means includes nucleic acid probes or primers for detecting the presence or absence of an associated allele, a particular allele of a polymorphic locus, or the like or changes to at least one sequence of Tables 5-7 or Tables 2-4, and (b) instructions for correlating the presence or absence of at least one sequence of Tables 5-7 or Tables 2-4 with a patient's risk of having schizophrenia.

104. The kit of claim 101, wherein the detection means includes an immunoassay for detecting the level of at least one gene product from Tables 5-7.

105. A method of assessing a patient's risk of having or developing schizophrenia, comprising: a) determining the level of expression of at least one gene from Tables 5-7 or gene products thereof, and comparing the level of expression to a normal cell; and b) assessing a patient's risk of having or developing schizophrenia, if any, by determining the correlation between the differential expression of said genes or gene products with known changes in expression of said genes measured in at least one parent suffering from schizophrenia.

106. A nucleic acid array comprising a solid support comprising nucleic acid probes which selectively hybridize to at least 5 different genes from Tables 5-7 or at least 5 different SNPs of Tables 2-4.

107. The array of claim 106, wherein the solid support is selected from the group consisting of paper, membranes, filters, chips, pins, and glass.

108. A method of diagnosing schizophrenia in a patient, comprising detecting a nucleic acid molecule encoding at least one protein from Tables 5-7 in a fluid or tissue sample from the patient.

109. A method of claim 108, wherein the detecting comprises detecting at least one associated allele, particular allele of a polymorphic locus, or the like in the nucleic acid molecule encoding said protein.

110. A method of claim 109, wherein said method comprises hybridizing a probe to said patient's sample of DNA or RNA under stringent conditions which allow hybridization of said probe to nucleic acid comprising said associated allele, a particular allele of a polymorphic locus, or the like, wherein the presence of a hybridization signal indicates the presence of said associated allele, particular allele of a polymorphic locus, or the like, in at least one gene from Tables 5-7.

111. A method of claim 110, wherein the patient's DNA or RNA has been amplified and said amplified DNA or RNA is hybridized.

112. A method of claim 111, wherein said method comprises using a single-stranded conformation polymorphism technique to assay for said associated allele, particular allele of a polymorphic locus, or the like.

113. A method of claim 111, wherein said method comprises sequencing at least one gene from Tables 5-7 in a sample of DNA from a patient.

114. A method of claim 27, wherein said patient's sample of DNA has been amplified or cloned.

115. A method of claim 111, wherein said method comprises sequencing at least one gene from Tables 5-7 in a sample of RNA or DNA from a patient.

116. A method of claim 111, wherein said method comprises determining the sequence of at least one gene from Tables 5-7 by preparing cDNA from RNA taken from said patient and sequencing said cDNA to determine the presence or absence of an associated allele, a particular allele of a polymorphic locus, or the like.

117. A method of claim 110, wherein said method comprises performing an RNAse assay.

118. A method of claim 110, wherein said probe is attached to a microarray or a bead.

119. A method of claim 110, wherein said probes are oligonucleotides.

120. A method of claim 108, wherein said sample is selected from the group consisting of blood, normal tissue and tumor tissue.

121. A method of claim 109, wherein the associated allele, particular allele of a polymorphic locus, or the like is selected from the group consisting of at least one of the SNPs from Tables 2-4, alone or in combination.

122. A method for assaying the presence of a nucleic acid associated with resistance or susceptibility to schizophrenia in a sample, comprising: contacting said sample with a nucleic acid recited in claim 1 or claim 5 under stringent hybridization conditions; and detecting a presence of a hybridization complex.

123. A method for diagnosing schizophrenia or the predisposition to schizophrenia, or the progression of schizophrenia, comprising obtaining a sample from a patient; contacting the sample with a nucleic acid of Tables 2-4; and detecting the presence or absence of a hybridization complex, wherein the presence or absence of a hybridization complex is a diagnosis of schizophrenia.

124. A method for assaying the presence or amount of a polypeptide encoded by a gene of Tables 5-7 for use in diagnostics, prognostics, prevention, treatment, or study of schizophrenia, comprising: contacting a sample with an antibody that specifically binds to a gene of Tables 5-7 under conditions appropriate for binding; and assessing the sample for the presence or amount of binding of the antibody to the polypeptide.

125. A method for diagnosing or prognosticating schizophrenia comprising comparing the level of expression or activity of a polypeptide encoded by a gene of Tables 5-7 in a test sample from a patient with the level of expression or activity of the same polypeptide in a control sample wherein a difference in the level of expression or activity between the test sample and control sample is indicative of schizophrenia.

126. A method for identifying an agent that can alter the level of activity or expression of a polypeptide encoded by a gene of Tables 5-7 for use in diagnostics, prognostics, prevention, treatment, or study of schizophrenia, comprising: contacting a cell, cell lysate, or the polypeptide, with an agent to be tested; assessing a level of activity or expression of the polypeptide; and comparing the level of activity or expression of the polypeptide with a control sample in an absence of the agent, wherein if the level of activity or expression of the polypeptide in the presence of the

agent differs by an amount that is statistically significant from the level in the absence of the agent then the agent alters the activity or expression of the polypeptide.

127. A cosmetic composition for inhibiting schizophrenia in a patient, said cosmetic composition comprising a compound that modulates as an example amount of delusions, hallucinations, paranoia, psychosis, disorganized thinking, agitation, lack of drive or initiative, social withdrawal, apathy, impaired attention, cognitive impairments and emotional unresponsiveness.

128. The cosmetic composition of claim 127, wherein said composition is in a form selected from the group consisting of shampoos, conditioners, lotions, sprays, ointments, oils, and gels.

129. A method for predicting the efficacy of a drug for treating schizophrenia in a human patient, comprising: a) obtaining a sample of cells from the patient; b) obtaining a set of genotypes from the sample, wherein the set of genotypes comprises genotypes of one or more polymorphic loci from Tables 2-4; and c) comparing the set of genotypes of the sample with a set of genotypes associated with efficacy of the drug, wherein similarity between the set of genotypes of the sample and the set of genotypes associated with efficacy of the drug predicts the efficacy of the drug for treating schizophrenia in the patient.

130. The method of claim 129, wherein the sample of cells is derived from a tissue selected from the group consisting of: any parts of the body such as the scalp, brain, colon, blood, dermis, epidermis and other skin cells, cutaneous surfaces, intertrigious areas, genitalia, vessels and endothelium.

131. The method of claim 130, wherein the cells are selected from the group consisting of: red blood cells, muscle cells, heart cells, nerve cells, insulin-producing cells, pancreatic cells, brain cells, germ cells, keratinocytes, monocytes, neutrophils, langerhans cells, CD4+ and CD8+ T cells, B and T lymphocytes, leukocytes, hormonal cells, bone marrow cells, skin cells, buccal cells, spinal cord cells, bone cells, adipose cells, cartilage cells, dendritic cells, intestinal cells, hepatic cells, mucous cells, olfactory cells, retinal cells, somatic cells and arterial cells.

132. The method of claim 129, wherein the sample is obtained via biopsy.

133. The method of claim 129, wherein the set of genotypes from the sample comprises genotypes of at least two of the polymorphic loci listed in Tables 2-4.

134. The method of claim 129 wherein the set of genotypes from the sample is obtained by hybridization to allele-specific oligonucleotides complementary to the polymorphic loci from Tables 2-4, wherein said allele-specific oligonucleotides are contained on a microarray.

135. The method of claim 134, wherein the oligonucleotides comprise nucleic acid molecules at least 95% identical to SEQ ID from Tables 2-7.

136. The method of claim 129 wherein the set of genotypes from the sample is obtained by sequencing said polymorphic loci in said sample.

137. The method of claim 129, wherein the drug is selected from the group consisting of symptom relievers and drugs for schizophrenia.

138. A method of treating schizophrenia in a patient in need thereof, comprising administering an agent that regulates the expression, activity or physical state of at least one polypeptide encoded by a gene from Tables 5-7 in the patient.

139. A method of claim 138, wherein the encoded protein from said gene comprises an alteration, wherein said alteration is encoded by a polymorphic locus in said gene.

140. A method of claim 138, wherein said gene comprises an associated allele, a particular allele of a polymorphic locus, or the like that modulates the expression of the encoded protein.

141. A method of claim 138, wherein said agent is selected from the group consisting of chemical compounds, oligonucleotides, peptides and antibodies.

142. A method of claim 138, wherein said agent is an antisense molecule or interfering RNA.

143. A method of claim 138, wherein said agent is an expression modulator.

144. A method of claim 143, wherein said modulator is an activator.

145. A method of claim 143, wherein said modulator is a repressor.

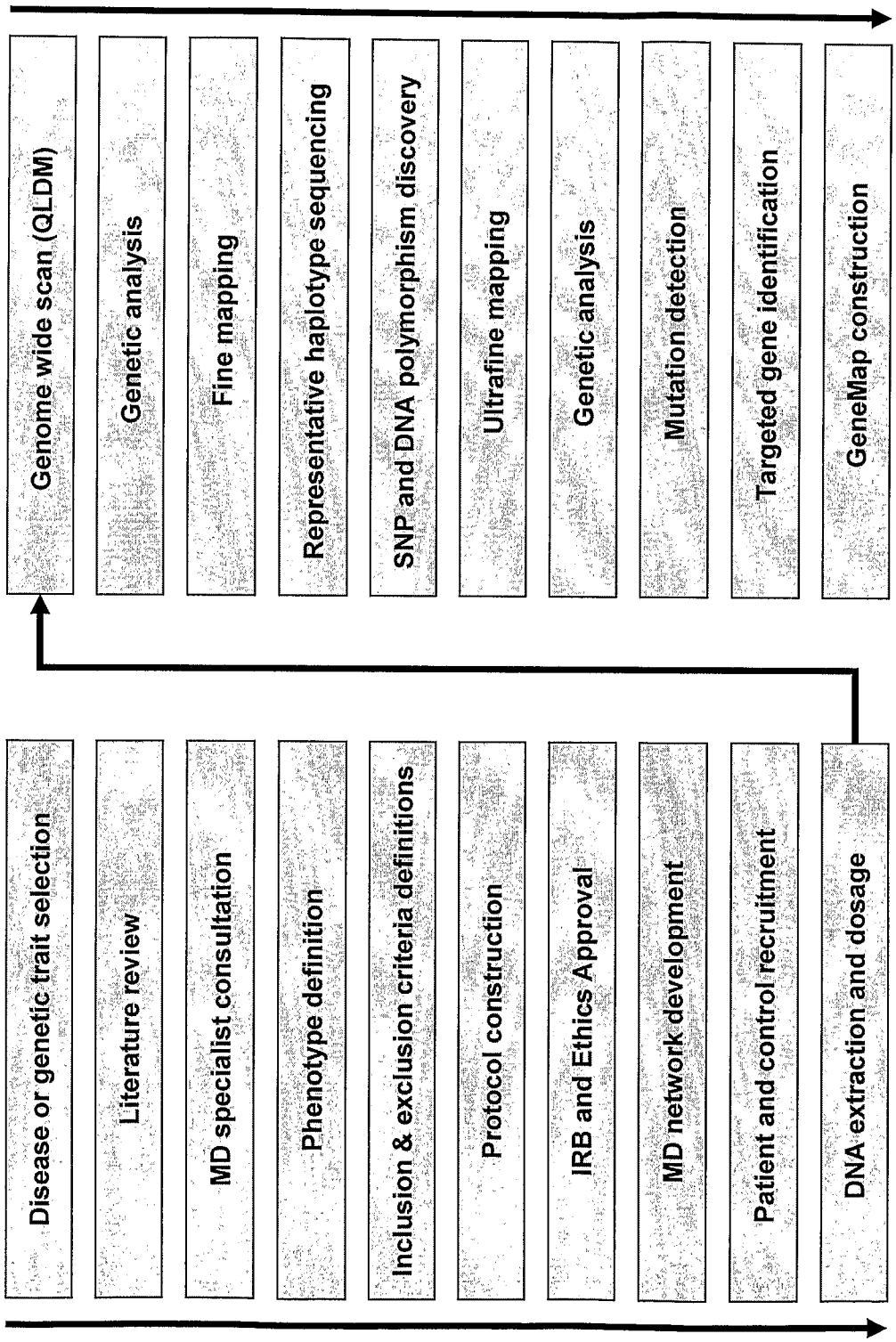
146. A method of claim 138, wherein said gene comprises an associated allele, a particular allele of a polymorphic locus, or the like that modifies at least one property or function of the encoded protein.

147. A kit for assessing a patient's risk of having or developing schizophrenia, comprising: a) detection means for detecting the genotype of at least one polymorphic locus shown in Tables 2-7; and b) instructions for correlating the genotype of said at least one polymorphic locus with a patient's risk of having or developing schizophrenia.

148. The kit of claim 147, wherein the detection means includes nucleic acid probes for detecting the genotype of said at least one polymorphic locus.

Figure 1a

From Trait Selection to GeneMap



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US06/33147

A. CLASSIFICATION OF SUBJECT MATTER
IPC: **C12Q 1/68**(2006.01);**A61K 49/00**(2006.01)

USPC: 424/9.1;435/6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 424/9.1;435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST, caplus, scisearch, embase, wpids, medline

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 6,020,135 A (LEVINE et al) 01 February 2000 (01.02.00), claim 1, example.	1-2 ---------- 1-4
Y	MORRISETTE, J. et al Genome-Wide search for Linkage of bipolar affective disorders in very large pedigree derived from a homogeneous population in Quebec points to a lovuls of major effect on chromosome 12q23-q24. Am. J. Medical Genetics. 1997, Vol. 88, pages 567-587, abstract and p 569.	3-4

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 September 2007 (30.09.2007)	13 NOV 2007
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201	Authorized officer Christopher M. Gross Telephone No. (571) 272-1600

Form PCT/ISA/210 (second sheet) (April 2005)

BOX III. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This International Search Authority has found 25 inventions claimed in the International Application covered by the claims indicated below:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group	Claims	Invention
I	1-4	a method of constructing GeneMap comprising screening for expression or presence of at least one allele from tables 5-7
II	5-8,122	a method of constructing GeneMap comprising screening for at least one allele from tables 2-4
III	9-22	a method of genetic mapping for detecting the association of at least one marker for schizophrenia
IV	23-26	a set of genetic markers comprising two SNPs from tables 2-4
V	27-37,114-115	a method of predicting drug efficacy
VI	38-39,41-42	a method for inducing a schizophrenia-like state
VII	40,43-44,45-47,126	a method for screening drug candidates
VIII	48	a method for identifying a gene that regulates a drug response
IX	49,50,105-107	an expression profile
X	51-62	a method of diagnosing schizophrenia using a SNP from tables 2-4
XI	63-64, 147-148	a kit for diagnosing susceptibility to schizophrenia using genes from tables 2-7
XII	65-66,105,125	a method of diagnosing a susceptibility to schizophrenia using one gene from tables 5-7
XIII	67-72	a method of treating schizophrenia by administering a vector comprising a gene from table 5-7
XIV	73-82	a method of treating schizophrenia comprising administering an agent that regulates the expression or activity of at least one gene set forth in tables 5-7
XV	83-85	a method for preventing the occurrence of schizophrenia
XVI	86	a method for monitoring the efficacy or treatment on the regulation of expression of one or more genes from tables 5-7.
XVII	87-96	a method of diagnosing schizophrenia using the polypeptide(s) set forth in tables 5-7
XVIII	97-100	a method for determining the phenotype of a cell comprising the genes set forth in tables 5-7
XIX	101-104	a kit for assessing patient risk of developing schizophrenia
XX	108-113,115-121	a method of diagnosing schizophrenia using at least one protiein set forth in tables 5-7

XXI	123	a method of assaying for the presence of a nucleic acid set forth in tables 2-4
XXII	124	a method of assaying for the presence of a polypeptide encoded by the genes set forth in tables 5-7
XXIII	127-128	a cosmetic composition
XXIV	129-137	a method for predicting the efficacy of a drug in treating schizophrenia
XXV	138-146	a method of treating schizophrenia comprising administering an agent which regulates the expression, etc. of a gene set forth in tables 5-7

The technical feature of groups I-XXV appears to be related to the DNA sequences set forth in tables 2-7 of the present disclosure

However, the technical feature appears lacking from all the inventions of groups I-XXV *a priori* because some claims are drawn to full gene(s), such as set forth tables 5-7 whereas other claims are drawn various single nucleotide polymorphisms (SNPs), set forth in tables 2-4.

Notably, unlike full genes, SNPs in of themselves do not provide an appropriate open reading frame. Thus, said full gene(s) would be required for expression, such as required by claim 41, yet are not used for hybridization with the microarray of claim 50. Said SNPs and full genes are not interchangeable. Therefore the inventions are not so linked as to form a single general inventive concept as defined by PCT Rule 13.1

Alternatively, the technical feature might be considered genetic analysis associated with schizophrenia, however such methods were known in the art at the time the invention was made (Morrisette et al 1999 Am. J. Med. Genetics 88:567-587). Therein, the technical feature linking the inventions of groups I-XXV does not constitute a *special* technical feature as defined by PCT rule 13.2, as it does not define a contribution over the prior art.

Groups IV,IX,XI, XIX, XXIII (each independently) are drawn to materially and structurally different compositions of matter. For example, Group XXIII is drawn to a cosmetic composition, including shampoo, which would not be required for use of the expression profile pertaining to invention IX, the kits for assessing patient risk for developing schizophrenia of inventions XI,XIX, or the set of markers of invention IV

Groups I-III,V-VIII,X,XII-XVIII,XX-XXII,XXIV-XXV (each independently) are drawn to various methods requiring materially different method steps with technical features highlighted as follows. Inventions I and II are drawn to different manners or preparing a GeneMap, using full genes or SNPs, which represent patentably distinct inventions for the reasons mentioned above. Invention III is drawn to assessing SNPs for a possible correlation with schizophrenia. Invention V is concerned with drug efficacy. Invention VI is drawn to developing biochemical models of schizophrenia. Invention VII is drawn to finding drug candidates. Invention VIII is drawn to elucidating a gene involved in a pharmacological response. Inventions X and XII are drawn to different diagnostics, using full genes or SNPs, which represent patentably distinct inventions for the reasons mentioned above. Invention XIII is drawn to a method of treatment using gene therapy. Invention XIV is drawn to drawn to a method of treatment using a small molecule, etc. Invention XV is drawn to a method of preventing schizophrenia. Invention XVI is drawn to a assessing prognosis associated with treating schizophrenia. Invention XVII is drawn to treating schizophrenia using a polypeptide. Invention XVIII is drawn to assessing a cell phenotype. Invention XX is drawn to a schizophrenia diagnostic using a polypeptide. Invention XXI is drawn to a drawn to a nucleic acid based diagnostic. Invention XXII is drawn to an ELISA type protein assay. Invention XXIV is drawn to a method of predicting the efficacy of a drug in treating schizophrenia. Invention XXV includes a method of treating schizophrenia using a siRNA directed against a gene set forth in tables 5-7 of the present disclosure.

Each composition of matter is patently distinct, as is its method of use. Each product is not used in each method. For example, the shampoo, set forth as part of invention XXIII is not required for the assessing a cell phenotype set forth as invention XVIII. As another example, the kit for assessing patient risk for developing schizophrenia of invention XIX would not employ the gene therapy method of invention VIII. Therefore, group III,IX, XIX, XIII methods lack unity with the product of groups I-II,V-VII,X-VXIII,XX-XXII,XXIV-XXV.

Accordingly, Groups I-XXV are not so linked by the same or a corresponding special technical feature as to form a single inventive concept.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US06/33147

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of any additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-4

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.