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

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2744424 A1 2012/03/14

(21) **2 744 424**

(12) **DEMANDE DE BREVET CANADIEN**
CANADIAN PATENT APPLICATION

(13) **A1**

(22) Date de dépôt/Filing Date: 2011/06/09

(41) Mise à la disp. pub./Open to Public Insp.: 2012/03/14

(30) Priorité/Priority: 2010/09/14 (US61/382,834)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01)

(71) Demandeur/Applicant:
THE HOSPITAL FOR SICK CHILDREN, CA

(72) Inventeurs/Inventors:
SCHERER, STEPHEN W., CA;
VINCENT, JOHN B., CA

(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : BIOMARQUEURS POUR LES TROUBLES DU SPECTRE AUTISTIQUE
(54) Title: BIOMARKERS FOR AUTISM SPECTRUM DISORDERS

(57) Abrégé/Abstract:
Methods of determining the risk of ASD or ID in an individual are provided which comprise identifying the presence of one or more specific genomic mutations in, upstream of, or comprising the PTCHD1 gene. Additionally provided are methods of determining the risk of ASD or ID in an individual comprising analyzing genomic mutations in PTCHD1AS1 and/or PTCHD1AS2 and/or PTCHD1AS3.

ABSTRACT

Methods of determining the risk of ASD or ID in an individual are provided which comprise identifying the presence of one or more specific genomic mutations in, upstream of, or comprising the PTCHD1 gene. Additionally provided are methods of determining the risk of ASD or ID in an individual comprising analyzing genomic mutations in PTCHD1AS1 and/or PTCHD1AS2 and/or PTCHD1AS3.

WAT_LAW\ 503807\3

BIOMARKERS FOR AUTISM SPECTRUM DISORDERS

FIELD OF THE INVENTION

[0001] The present invention relates to genetic markers for Autism Spectrum Disorders (ASD), and methods of determining risk of ASD in an individual.

BACKGROUND OF THE INVENTION

[0002] Autism (MIM 209850) is a severe, lifelong neurodevelopmental disorder characterized by impairments in communication and socialization, and by repetitive behavior. Autism is not a distinct categorical disorder but is the prototype of a group of conditions defined as Pervasive Developmental Disorders (PDDs) or Autism Spectrum Disorders (ASD), which include Asperger's Disorder, Childhood Disintegrative Disorder, Pervasive developmental disorder-not otherwise specified (PDD-NOS) and Rett Syndrome. ASD is diagnosed in families of all racial, ethnic and social-economic backgrounds with incidence roughly four times higher in males compared to females. Data from several epidemiological twin and family studies provide substantial evidence that autism has a significant and complex genetic etiology. The concordance rate in monozygotic twins is 60-90%, and the recurrence rate in siblings of affected probands has been reported to be between 5-10% representing a 50 fold increase in risk compared to the general population. Although autism spectrum disorders are among the most heritable complex disorders, the genetic risk is clearly not conferred in simple Mendelian fashion.

[0003] Recent studies of sub-microscopic genomic copy number variation (CNV) have identified several loci associated with Autism Spectrum Disorder (ASD; MIM 209850). *De novo* CNVs associated with ASD have been reported in ~7% of simplex families and ~2% of multiplex families. CNV studies have also led to the identification of autism candidate genes such as *SHANK3* (MIM 606230) and *NRXN1* (MIM 600565). Intellectual disability (ID) is frequently associated with autism (in up to ~30% of cases for ASD, and ~67% for autism). Moreover, mutations in several X-

linked ID (XLID) genes (e.g. *NLGN4* and *IL1RAPL1*) have been shown to result in an autistic phenotype, which suggests that autism and ID may often share a common genetic etiology. Currently available data suggest substantial genetic heterogeneity, with the most likely cause of non-syndromic idiopathic ASD involving multiple epistatically-interacting loci. The identification of large scale copy number variants (CNVs) represents a considerable source of genetic variation in the human genome that contributes to phenotypic variation and disease susceptibility found in small inherited deletions in autistic kindreds, suggesting possible susceptibility loci.

[0004] It would thus be desirable to characterize putative susceptibility loci to identify genetic markers of ASD, as well as to understand the role of candidate genes for ASD in order to facilitate determination of the risk of ASD in an individual, and to assist in the diagnosis of ASD.

SUMMARY OF THE INVENTION

[0005] Systematic screening at *PTCHD1* and 5'-flanking regions, suggests involvement of this locus in ~1% of autism spectrum disorder (ASD) and intellectual disability (ID) individuals. Provided herein are mutations in the X-chromosome *PTCHD1* (*patched*-related) locus, which are useful in assessing the risk of ASD and/or the risk of ID in an individual, as well as being useful to diagnose carrier status of an individual, or other condition(s). Provided markers are useful both individually and in the form of a microarray to screen individuals for risk of ASD and/or ID or for carrier status for risk of ASD and/or ID.

[0006] Thus, in one aspect of the present invention, a method of determining the risk of ASD in an individual is provided, comprising analyzing a nucleic acid-containing sample obtained from the individual for the presence or absence of a genomic sequence mutation at the *PTCHD1* locus, wherein the mutation comprises a deletion of a region upstream to the *PTCHD1* gene (e.g., a deletion as set forth in Table 2), a disruption of a non-coding RNA (ncRNA) selected from *PTCHD1AS1*, *PTCHD1AS2*, or *PTCHD1AS3*, or splice variants of these ncRNAs, or a disruption of

other regulatory elements upstream of the PTCHD1 coding region. Presence of the mutations has been found to be indicative of ASD.

[0007] These and other aspects of the present invention are described by reference to the following figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Figure 1 depicts the cDNA sequence (SEQ ID No:1) of a *PTCHD1* (A) and the amino acid sequence (SEQ ID No: 2) of the protein it encodes (B).

[0009] Figure 2 depicts detailed genomic organization of the *PTCHD1* locus.

[0010] Figure 3 depicts pedigrees of families. (A) Pedigrees showing *PTCHD1* mutations. (B) Pedigrees showing deletions at the *PTCHD1/PTCHD1AS1-3* locus.

[0011] Figure 4 depicts PTCHD1 missense variants. Electropherograms indicate the nucleotide substitutions within PTCHD1 in unrelated ASD families and ID families.

[0012] Figure 5 depicts PTCHD1 domain structure (A) and protein sequence conservation (B).

[0013] Figure 6 depicts the consensus sequence for non-coding RNA of *PTCHD1AS1* (SEQ ID No:11).

[0014] Figure 7 depicts the consensus sequence for non-coding RNA of *PTCHD1A2* (SEQ ID No:12).

[0015] Figure 8 depicts the consensus sequence for non-coding RNA of *PTCHD1A3* (SEQ ID No:13).

DETAILED DESCRIPTION OF THE INVENTION

[0016] A method of determining the risk of an autism spectrum disorder (ASD) in an individual, or carrier status of an individual, is provided comprising screening a biological sample obtained from the individual for a mutation that may modulate the expression of *PTCHD1*.

[0017] The term "an autism spectrum disorder" or "an ASD" is used herein to refer to at least one condition that results in developmental delay of an individual such as autism, Asperger's Disorder, Childhood Disintegrative Disorder, Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) and Rett Syndrome (APA DSM-IV 2000).

[0018] The term "intellectual disability" or "ID" refers to a disability originating before age 18, characterized by significant limitations in both intellectual functioning and adaptive behavior as expressed in conceptual, social, and practical adaptive skills.

[0019] Microdeletions that directly disrupt the *PTCHD1* gene have been identified in males in families affected with ASD, ID or learning disability. Identified deletions are maternally inherited and were not observed in more than 10,000 controls, indicating that these alterations are associated with ASD and ID. Maternally inherited missense mutations in *PTCHD1* in male probands have also been reported.

[0020] *PTCHD1* encodes a Patched-related protein with 12 transmembrane domains and a sterol-sensing domain, structurally similar to the Hh receptors PTCH1 and PTCH2, as well as the Niemann-Pick Type C1 protein (NPC1) and several others. Many Patched-related genes have been found in various organisms, from nematodes to humans, and they appear to play diverse biological functions, including cytokinesis, growth and pattern formation (Zugasti, O. et al., *Genome Res.* **15**, 1402-1410 (2005)). For instance, there are just seven patched-related genes in humans (*PTCH1*, *PTCH2*, *PTCHD1*, *PTCHD2*, *PTCHD3*, *NPC1* and *c6orf138*, whereas in *C. elegans* there are at least 26 patched-related genes, with diverse roles in development in addition to Hh

signaling, including cytokinesis, growth and pattern formation (Zugasti, O. et al., *Genome Res.* **15**, 1402-1410 (2005)). We have found in 10T1/2 cells, an inhibitory effect of PTCHD1 was demonstrated on Gli-dependent transcription. Although these results suggest that PTCHD1 exhibits biochemical activity in Hh-dependent processes similar to that of PTCH1 and 2, other functions or roles for PTCHD1 cannot be excluded at this point.

[0021] We have further characterized the *PTCHD1* locus and found variants identified in *PTCHD1* were not seen in more than 500 controls, further supporting a role of *PTCHD1* in autism and ID. As used herein, the term "*PTCHD1* locus" refers to the region in the X chromosome which extends from about the distal-most exon of mRNA clone DA355362 at the distal end to a proximal boundary which at least includes the coordinate according to the UCSC 2006/hg18 build ChrX:23,329,120 and which may extend to BX115199 as illustrated in Figure 2. As will be appreciated by one of skill in the art, the *PTCHD1* locus may encompass *PTCHD1* corresponding to Figure 1 or isoforms thereof.

[0022] Furthermore, 10 deletions were found that map to regions upstream of the coding region of *PTCHD1*. The region 5' and distal to *PTCHD1* is relatively gene poor. Within this upstream region, a coding gene, *DDX53*, encoding DEAD Box 53, lies ~335 Kb 5' to *PTCHD1*. Five of the 10 upstream deletions span *DDX53*. However, based on the function of the DDX53 protein and the expression pattern of this gene (which is restricted mainly to testis and tumor cells (Cho, B. et al., *Biochem.Biophys. Res.Commun.* **292**, 715-726 (2002)), it is unlikely to contribute to the ASD or ID phenotype. Additionally, within the gene-poor region between *PTCHD1* and *DDX53*, there is a putative pseudogene of *FAM3C*, *FAM3C2,* which is disrupted by five of the 10 upstream deletions. *FAM3C*, a cytokine-like gene on 7q31.31, consists of 10 exons (Zhu, Y. et al.,*Genomics* **80**, 144-150 (2002)) whereas *FAM3C2*, although 99% identical, has no intron/exon structure and is interrupted by a short interspersed nuclear element (SINE). It appears to have inserted on Xp22 after human/chimp evolutionary divergence. Since no mRNA or EST matches exactly to *FAM3C2*, it is most likely an untranscribed processed pseudogene.

[0023] The region just distal to *PTCHD1* was examined in detail and a number of putative enhancer and promoter sequences were identified, as well as conserved (and putative regulatory) elements (Figure 2). Several overlapping spliced long (>200nt) non-coding (nc) RNAs (*PTCHD1AS1* (from cDNA clone IMAGE:1560626; BX115199) and *PTCHD1AS2* (from cDNA clone BRSTN2000219; DA355362)), were identified, which map to the opposite strand and distal to *PTCHD1* (see Figure 2). 5'RACE (Rapid Amplification of cDNA Ends) shows that a number of splice variants of these transcripts originate at the CpG island just upstream of *PTCHD1*, encompassing its putative promoter. Similar antisense transcripts are present at syntenic loci in other mammalian species, at least two exons of which appear to be conserved between rat, mouse and humans (see Figure 2)

[0024] Although the ncRNAs do not appear to encode protein, they may serve as regulators for other coding genes, particularly for *PTCHD1*, since the 5' exons are adjacent on opposite strands. Such ncRNAs may regulate expression of a coding transcript on the opposite strand through a number of mechanisms, including modification of chromatin, transcriptional regulation and post-transcriptional modification (Mercer, T.R. et al., *Nat.Rev.Genet.* **10**, 155-159 (2009); Kleinjan, D.A et al., *Am.J.Hum.Genet.* **76**, 8-32 (2005)).

[0025] All of the upstream deletions identified, as well as *PTCHD1* deletions (e.g., Family 1) disrupt conserved (and putative regulatory) sequences and/or exons of ncRNAs (see Figure 2). Deletions were not inherited by a subset of the affected family members; also, missense variants do not segregate with disease in all families (e.g., Family 6) (Figure 3). These findings are similar to other previously reported major affect ASD loci such as 16p11.2 (Weiss, L.A. et al., *N.Engl.J.Med.* **358**, 667-675 (2008)) and are also consistent with the complex, non-Mendelian inheritance believed to control the etiology of autism. A recently proposed threshold model of relative contribution in ASD has been described (Cook, Jr., E.H. et al., *Nature* **455**, 919-923 (2008).), whereby it is anticipated that multiple common and rare variants may act in concert to generate the phenotype. For instance, under this model, some *de novo* CNVs may be solely sufficient to cause ASD. Conversely, other *de novo* CNVs may have

weaker effects, requiring contributions from additional loci (for example additional risk haplotypes, or other CNVs), or environmental risk factors, for the burden of contributory factors to cross a risk threshold and result in an ASD phenotype. In families that carry putative *PTCHD1* missense mutations (e.g., Families 9 and 10), other CNVs involving genes that may also contribute to the phenotype were identified. In Family 9, in addition to the I173V substitution,a *de novo* ~1.1 Mb loss was found at 1p21.3 resulting in deletion of the entire *DPYD* gene (MIM 274270), encoding dihydropyrimidine dehydrogenase (DPD) (Marshall, C.R. et al., *Am.J.Hum.Genet.* **82**, 477-488 (2008)). Complete DPD deficiency results in highly variable clinical outcomes, with convulsive disorders, motor retardation, and mental retardation being the most frequent manifestations, and autistic features occasionally reported (van Kuilenburg, A.B. et al., *Hum.Genet.* **104**, 1-9 (1999)). In this family, a balanced translocation, t(19;21)(p13.2; q22.12) is also present in the proband, but is inherited from the unaffected mother and shared with an unaffected sister. In Family 10, which shows the V195I substitution in *PTCHD1*, a 66 Kb *de novo* loss at 7q36.2 was previously reported that results in deletion of the third exon of *DPP6* (MIM 126141) – previously reported as a positional and functional candidate gene for autism (Marshall, C.R. et al., *Am.J.Hum.Genet.* **82**, 477-488 (2008)).

[0026] Thus, in ASD individuals there is evidence for the possible involvement of more than one locus in the disease, and these findings may support the threshold model of relative contribution in ASD and polygenic inheritance in autism. As such, some *de novo* CNVs may be highly penetrant in causing ASD susceptibility (e.g. disruption of *PTCHD1* in Family 1). Conversely, other *de novo* CNVs (e.g. *DPP6* and *DPYD* deletions) may have more subtle effects, requiring contributions of additional loci (e.g. *PTCHD1* missense mutations in the case of Families 9 & 10) for ASD to be phenotypically evident. This scenario may also apply to the ID families with *PTCHD1* mutations.

[0027] Cerebellar abnormalities have frequently been linked to autism, including recent magnetic resonance imaging (MRI) studies showing significant decrease in cerebellar grey matter (Courchesne, E. et al., *Neurology* **57**, 245-254

(2001); Toal, F. et al., *Br.J.Psychiatry* **194**, 418-425 (2009)), and decreased cerebellar connectivity and activity (Mostofsky, S.H. et al., *Brain* **132**, 2413-2425 (2009)).

[0028] In the present methods, it is possible to determine ASD risk in an individual, as well as to determine carrier status of an individual (e.g., testing of females for the presence of mutations associated with ASD, to determine whether they are carriers). In the methods, a biological sample obtained from the individual is utilized. A suitable biological sample may include, for example, a nucleic acid-containing sample or a protein-containing sample. Examples of suitable biological samples include saliva, urine, semen, other bodily fluids or secretions, epithelial cells, cheek cells, hair and the like. Although such non-invasively obtained biological samples are preferred for use in the present method, one of skill in the art will appreciate that invasively-obtained biological samples, may also be used in the method, including for example, blood, serum, bone marrow, cerebrospinal fluid (CSF) and tissue biopsies such as tissue from the cerebellum, spinal cord, prostate, stomach, uterus, small intestine and mammary gland samples. Techniques for the invasive process of obtaining such samples are known to those of skill in the art. The present method may also be utilized in prenatal testing for the risk of ASD using an appropriate biological sample such as amniotic fluid and chorionic villus.

[0029] In one aspect, the biological sample is screened for nucleic acid encoding selected genes in order to detect mutations associated with an ASD. It may be necessary, or preferable, to extract the nucleic acid from the biological sample prior to screening the sample. Methods of nucleic acid extraction are well-known to those of skill in the art and include chemical extraction techniques utilizing phenol-chloroform (Sambrook et al., 1989), guanidine-containing solutions, or CTAB-containing buffers. As well, as a matter of convenience, commercial DNA extraction kits are also widely available from laboratory reagent supply companies, including for example, the QIAamp DNA Blood Minikit available from QIAGEN (Chatsworth, CA), or the Extract-N-Amp blood kit available from Sigma (St. Louis, MO).

[0030] Once an appropriate nucleic acid sample is obtained, it is subjected to well-established methods of screening, such as those described in the specific examples that follow, to detect genetic mutations indicative of ASD, i.e. ASD-linked mutations. Representative methods of screening include straight sequencing; use of arrays as described herein; as well as quantitative PCR (qPCR) and multiplex ligation-dependent probe amplification (MLPA). For example, various platforms can be used: affymetrix 500k SNP arrays; Illumina 1M BeadChips; NimbleGen 385K arrays; Affymetrix 6.0 arrays; Illumina 550X arrays; and other platforms.

[0031] Mutations, including sequence mutations in coding and/or regulatory regions of a gene, as well as in flanking regions of a gene, have been found to be indicative of ASC. Representative mutations include, for example, genomic copy number variations (CNVs), which include gains and deletions of segments of DNA (e.g., segments of DNA greater than about 1kb, such as DNA segments over about 50 kb, such as between 50 and 300 kb, or between about 300 and 500 kb); as well as base pair mutations such as nonsense, missense and splice site mutations.

[0032] Genomic sequence variations of various types in different genes have been identified as indicative of ASD. As described herein, deletions in the 5' flanking region of PTCHD1 that disrupted a complex non-coding RNA (e.g., PTCHD1AS1, PTCHD1AS2, PTCHD1AS3), and potential regulatory element(s) in the *PTCHD1* locus have been associated with ASD. In one embodiment, genomic sequence variations that alter the expression of PTCHD1 have been linked to ASD. The terminology "alter expression" refers broadly to sequence variations that may alter (e.g., inhibit, or at least reduce) *any one* of transcription and/or translation of the coding nucleic acid sequence of PTCHD1, as well as the activity of the PTCHD1 protein.

[0033] Genomic sequence variations other than CNVs have also been found to be indicative of ASD, including, for example, missense mutations which result in amino acid changes in a protein that may also affect protein expression. In one embodiment, missense mutations in the PTCHD1 gene have been identified which are indicative of ASD. In certain embodiments, a missense change is associated with a further genetic

mutation and the presence of the combination of the missense change and the deletion is associated with ASD.

[0034] In another embodiment, sequence variations associated with ASD include deletions in the region that is within the 5' region upstream of the PTCHD1 gene (e.g., in whole or in part, or a portion or more of the upstream region thereof). In certain embodiments, mutations include deletions (e.g., deletions described in Table 2). The term "upstream region," as used herein, refers to a region that is distal to the PTCHD1 gene within approximately 1.2 mbp. For example, in one embodiment, the region comprises cDNA clone BRSTN2000219 (DA355362) (see Figure **2**). In another embodiment, the region comprises the 5' RACE and RT-PCR region as shown in Figure **2**. In additional embodiments, the region comprises any of the regions comprising non-coding mRNA regions of PTCHD1AS1, PTCHD1AS2, and/or PTCHD1AS3 or splice variants thereof. Upstream regions can be of varying sizes, from under 1kbp to over 1mbp. Representative upstream regions include regions varying in size from approximately 50kbp and approximately 1mbp; from approximately 60kbp and approximately 500kbp; from approximately 100kbp and approximately 400kbp; from approximately 100kbp to 300kbp. In certain embodiments, representative upstream regions comprise one or more of the breakpoint deletions, for example, those identified in Table 2. In certain embodiments, representative upstream regions comprise chrX:22,200,000-23,260,000, chrX:22,300,000-23,260,000, chrX:22,670,000-23,260,000, chrX:22,900,000-23,260,000 or chrX:22,900,000-23,050,000.

[0035] To determine risk of ASD in an individual, it may be advantageous to screen for multiple genomic mutations, including CNVs and/or mutations as indicated above applying array technology. In this regard, genomic sequencing and profiling, using well-established techniques as exemplified herein in the specific examples, may be conducted for an individual to be assessed with respect to ASD risk/diagnosis using a suitable biological sample obtained from the individual. Identification of one or more mutations associated with ASD would be indicative of a risk of ASD, or may be indicative of a diagnosis of ASD. This analysis may be conducted in combination with

an evaluation of other characteristics of the individual being assessed, including for example, phenotypic characteristics.

[0036] In view of the determination of gene mutations which are linked to ASD, a method for determining risk of ASD in an individual is also provided in which the expression or activity of a product of an ASD-linked gene mutation is determined in a biological protein-containing sample obtained from the individual. Abnormal levels of the gene product or abnormal levels of the activity thereof, i.e. reduced or elevated levels, in comparison with levels that exist in healthy non-ASD individuals, are indicative of a risk of ASD, or may be indicative of ASD. Thus, a determination of the level and/or activity of the gene product of PTCHD1, may be used to determine the risk of ASD in an individual, or to diagnose ASD. Further, a determination of the level and/or activity of the gene product of PTCHD1AS1, PTCHD1AS2, and/or PTCHD1AS3 or splice variants thereof, may be used to determine the risk of ASD in an individual, or to diagnose ASD. As one of skill in the art will appreciate, standard assays may be used to identify and quantify the presence and/or activity of a selected gene product.

[0037] Embodiments of the invention are described by reference to the following specific exemplification which is not to be construed as limiting.

EXEMPLIFICATION

METHODS

[0038] *Subjects:* CNVs at the PTCHD1 locus were initially assessed in 427 ASD patients as described (Marshall, C.R. et al., *Am.J.Hum.Genet.* **82**, 477-488 (2008)). DNA samples from 900 individuals diagnosed with ASD were sequenced for PTCHD1 mutations, and compared to a reference nucleic acid sequence to identify mutations. In this regard, Figure 1 illustrates the cDNA sequence (A) of the *PTCHD1* gene and the corresponding amino acid sequence (B).

[0039] Among the samples assessed, 400 samples were collected at three sites, namely The Hospital for Sick Children (HSC) in Toronto and child diagnostic centers in

Hamilton, Ontario and St, John's, Newfoundland. Details of these samples are published elsewhere (Moessner, R. et al., *Am.J.Hum.Genet.* **81**, 1289-1297 (2007)). 420 ASD cases were recruited at Montreal, details of these samples are published elsewhere (Gauthier, J. et al., *Mol.Psychiatry* **11**, 206-213 (2006)). Another 80 ASD probands from the Autism Genetic Resource Exchange (AGRE) were also included. The second cohort of 996 autism probands was recruited at different sites as a part of the Autism Genome Project (AGP); ascertainment is described elsewhere (Pinto, D. et al., *Nature* **466**, 368-372 (2010)). 246 male patients with intellectual disability were recruited from the UK, United States, Australia, Europe and South Africa as the IGOLD study. A subset of 225 from this cohort were also used for sequence analysis of *PTCHD1*. Details of these samples are published elsewhere (Tarpey, P.S. et al., *Nat.Genet.* **41**, 535-543 (2009)). 167 unrelated patients diagnosed with ADHD were recruited through the Department of Psychiatry at the Hospital for Sick Children, Toronto. Microarray data from controls included 1,123 (M=623, F=500) controls recruited from northern Germany as a part of the PopGen project, 1,234 (M=586, F=648) healthy controls of European origin recruited from the province of Ontario, Canada, 1,287 (M=383, F=904) controls from the Study of Addiction: Genetics and Environment (SAGE), 1,320 (M=589, F=1320) controls from Children's Hospital of Philadelphia (CHOP), 4783 (M=2460, F=2323) controls were recruited by the Wellcome Trust Case Control Consortium, 440 (M=158, F=282) controls were recruited by The Centre of Addiction and Mental Health (CAMH) and GlaxoSmithKline (GSK), and 59 (M=30, F=29) from the Centre d'Etude Polymorphisme Humaine (CEPH) HapMap controls (total *N*=5,023). More than 650 Ontario controls were obtained from The Centre for Applied Genomics (TCAG) and The Centre for Addiction and Mental Health (CAMH) and sequenced. Institutional ethical review board approval (CAMH, HSC, CHOP and all other collaborating institutions) was obtained for the study, and informed written consent was obtained for each family. Details of the clinical findings in families with *PTCHD1* mutations or CNVs are summarized in Table 1.

Table 1: Clinical description of cases with disruptions at the *PTCHD1* locus on Xp22.11

Family ID	Genes; Mutation	#Chromosomes Tested in Controls	Clinical Details in Proband‡	Family Segregation Comments
Family 1 (1-0186)	PTCHD1, PTCHD1AS2/3 167 Kb del	15,663 (M=4,829 F=10,834)	Proband (deletion) = Autism (based on ADI & ADOS-Module 1) & ADHD. Leiter-R brief IQ: 97 (42%)†; PLS-3: 86 (18%); VABS: COM=88 (21%); DLS=79 (8%), SOC=80 (9%), MOT=75 (5%), ABC=74 (4%).	Simplex family. Proband's brother DZ twin (deletion) = ASD features and Learning Disability. WASI: Non-Verbal IQ=67 (1%), Verbal IQ=86 (18%); VABS: COM=84 (14%), DLS=95(37%), SOC=104 (61%), ABC=92 (30%) Proband's sister (heterozygous deletion) = non-ASD
Family 3 (S01407)	PTCHD1 I173V	1101 (M=613 F=488)	Proband (mutation) = Autism (based on ADI & ADOS-Module 1). Non-Verbal IQ=95, Verbal IQ=85.	Simplex family. No other siblings.
Family 4 (S01433)	PTCHD1 ML336-7II	1193* (M=643 F=550)	Proband (mutation) = Autism (based on ADI & ADOS-Module 1). Some traits were observed that might be related to schizophrenia.	Simplex family. No other siblings.
Family 5 (S01355)	PTCHD1 E479G	869 (M=531 F=338)	Proband (mutation) = High Functioning Autism	Simplex family. Proband's brother (no genotype data) = non-ASD
Family 6 (AU0501)	PTCHD1 L73F	869 (M=531 F=338)	Proband (mutation) = Autism	Multiplex family. Proband's brother #1 (no mutation) = ASD Proband's brother #2 (mutation) = phenotype is currently unclear.
Family 9 (1-0215)	PTCHD1 I173V and *de novo* ~1.1 Mb loss at DPYD	1101 (M=613 F=488)	Proband (mutation) = Autism (based on ADI & ADOS-Module 1), intellectual disability, hyperactive, poor motor coordination. Leiter-R Brief IQ = 38. OWLS = 40 (<1%). VABS: COM=36 (<1%); DLS=<20 (<1%), SOC=31 (<1%), ABC=26 (<1%).	Simplex family. Proband's sister (mutation) = non-ASD
Family 10 (3-0002)	PTCHD1 V195I and 66 Kb *de novo* loss at DPP6	1101 (M=613 F=488)	Proband (mutation) = Autism (based on ADI & ADOS-Module 1). Severe expressive/receptive language delay. CT head=Normal.	Simplex family. No other siblings
Family 11 (5298)	PTCHD1AS1-3, DDX53 125 Kb del	15,663 (M=4,829 F=10,834)	Proband (deletion) = Autism (based on ADI-R & ADOS-Module1), ID, speech delay, apraxia. Uses single words. Leiter Brief IQ: 42 (<1%). PPVT-4: 20 (<1%). VABS: COM=<20 (<1%); DLS=47 (<1%), SOC=44 (<1%), ABC=34 (<1%).	Simplex family. Proband's sister (heterozygous deletion) = non-ASD.
Family 12 (5065)	PTCHD1AS1 65 Kb del	15,663 (M=4,829 F=10,834)	Proband (deletion) = Autism (based on ADI-R & ADOS-Module 4). Verbally fluent. Leiter IQ: 71 (3%). VABS: COM=68 (2%), DLS=45 (<1%), SOC=58 (<1%), ABC=52 (<1%).	Mulitplex family. Paternal family history of ASD. Proband's brother (no deletion) = Autism (based on ADI & ADOS-Module 4). Verbally Fluent. VABS: COM=71 (3%), DLS=38 (<1%), SOC=51 (<1%), ABC=49 (<1%).
Family 13 (3424)	104 Kb del	15,663 (M=4,829 F=10,834)	Proband (deletion) = Autism (based on ADI & ADOS). WISC-R: Non-Verbal IQ=58, Verbal IQ=50, Total IQ=50	Simplex family. Proband's brother (no deletion) =non-ASD

Family 14 (5111)	PTCHD1AS1 59 Kb del	15,663 (M=4,829 F=10,834)	Proband (deletion) = Autism (based on ADI-R & ADOS-Module1). Uses single words. MRI = normal. Leiter IQ: 46 (<1%). VABS: COM=37 (<1%), DLS=31 (<1%), SOC=52 (<1%), ABC=37 (<1%).	Mulitplex family. Paternal family history of ASD & ADHD. Proband's brother (no deletion) = Autism (based on ADI & ADOS-Module 3). Verbally fluent. Leiter IQ: 105 (63%). VABS: COM=108 (70%), DLS=62 (1%), SOC=92 (30%), ABC=83 (13%). Proband's sister (heterozygous deletion) = non-ASD, Bassen-Kornzweig syndrome. Proband's father (no deletion) = non-ASD, OCD.
Family 15 (3253)	PTCHD1AS1 54 Kb del	15,663 (M=4,829 F=10,834)	Proband (deletion) = Autism (based on ADI & ADOS) Non-Verbal IQ=75, Verbal IQ=56	Multiplex family. Proband's brother (no deletion) = ASD Proband's sister (no deletion) = non-ASD.
Family 16 (13047)	PTCHD1AS1-3, DDX53 389 Kb del	15,663 (M=4,829 F=10,834)	Proband (deletion) = Autism (based on ADI & ADOS). No epilepsy, history of language delay followed by a rapid language learning progression. Average to above average Non-Verbal and Verbal IQ.	Multiplex family. Proband's brother #1 (no deletion) = Autism (based on ADI & ADOS), IQ=average to above average Proband's brother #2 (no deletion) = ASD Proband's sister (no CNV data) = non-ASD, semantic-pragmatic language disorder.
Family 17 (8273)	101 Kb del	15,663 (M=4,829 F=10,834)	Proband (no deletion) = ASD WISC III IQ: Non-verbal=120, Verbal=130	Multiplex family. Proband's brother (deletion) = ASD Proband's sister #1 (deletion) = ASD Proband's sister #2 (deletion) = ASD
Family 18 (8013)	PTCHD1AS1 65 Kb del	15,663 (M=4,829 F=10,834)	Proband (no deletion) = Autism (based on ADI & ADOS-Module 3). WISC-III: Non-Verbal IQ=139 (>99%), Verbal IQ=89 (23%). VABS: SOC=76 (5%).	Multiplex family. Proband's brother #1(deletion) = Autism (based on ADI & ADOS-Module 3). WISC III: Total IQ=44 (1%). Proband's brother #2 (deletion) = non-ASD. WPPSI-R: Verbal IQ=89 (23%), non-verbal=100 (50%).
Family 19 (3387)	PTCHD1AS1-3, DDX53 213 Kb del	15,663 (M=4,829 F=10,834)	Proband (no deletion) = ASD	Multiplex family. Proband's father (deletion) = Broad Autism Phenotype Proband's brother (no deletion) = ASD Proband's sister (deletion) = non-ASD.
Family 20 1-27075	PTCHD1AS1-3, DDX53 388 Kb del	15,663 (M=4,829 F=10,834)	Proband (deletion) = ADHD, NVLD Verbal IQ =131, Performance IQ =113. Proband has some ASD spectrum features (disinterest in social relationships, preference for being alone, difficulty with change and over-adherence to structure and rules, difficulty with reading nonverbal cues resulting in social difficulties) but no evidence of	Simplex family. Proband's sister#1 (genotype unknown) = non-ASD Proband's sister#2 (genotype unknown) = non-ASD

			restricted, repetitive, or stereotyped behaviour.	

§All probands are male and are of European ancestry except for those in family 9 (Mixed European), family 4 (East Asian), and families 6 and 7 (Not available). The referring diagnosis for all probands is Autism Spectrum Disorder (ASD) except for Families 2, 7, 8 (intellectual disability; ID) and Family 20 (ADHD)
‡Abbreviations used: ADHD: Attention-Deficit Hyperactivity Disorder; BAP: Broad Autism Phenotype; NVLD: Non-verbal Learning Disability; ADOS: Autism Diagnostic Observation Schedule; ADI(-R): Autism Diagnostic Interview(-Revised); Leiter-R: Leiter International Performance Scale-Revised (non-verbal); WISC-(R or III): Wechsler Intelligence Scale for Children –(Revised or 3rd Edition); WPPSI-R: Wechsler Preschool and Primary Scale of Intelligence – Revised; VABS: Vineland Adaptive Behaviour Scale – consists of the following domains: COM-Communication, DLS-Daily Living Scales, SOC-Socialization, MOT-Motor Skills, ABC-Adaptive Behaviour Composite; PLS-3: Preschool Language Scale-3; OWLS: Oral and Written Language Scale; PPVT-4: Peabody Picture Vocabulary Test (4th Edition).
†Standard Score 100 ±15 (percentile)
*Controls included N=92 of Asian ancestry

[0040] *Copy Number Variation Analysis:* Affymetrix 500K SNP arrays were used to assess CNVs in a cohort of 427 ASD cases. Details on the methods of copy number analysis and complete results are published elsewhere (Marshall, C.R. et al., *Am.J.Hum.Genet.* **82**, 477-488 (2008)). Only the CNV result at *PTCHD1* is described here. Another cohort of 996 autism probands was analyzed on 1M BeadChips (Illumina) (Pinto, D. et al., *Nature* **466**, 368-372 (2010)). 246 male patients with ID were analyzed on a custom designed NimbleGen 385K array. Genomic DNA samples were sent to NimbleGen for the hybridizations to be performed. Each patient sample (Cy5-labelled) was co-hybridised with DNA from the reference sample NA10851 (Cy3-labelled; obtained from Coriell Cell Repository). After data normalisation, the ADM-1 algorithm (CGH Analytics 3.4, Agilent) was used for CNV discovery. The ADHD cohort was analyzed on Affymetrix 6.0 arrays. Three algorithms (Birdsuite, iPattern and Affymetrix Genotyping console (GTC)) were used to infer CNVs. The CEPH, PopGen and Ontario controls were analyzed on Affymetrix 6.0 arrays, SAGE controls were analyzed using 1M BeadChips (Illumina), and Illumina 550K arrays were used for the CHOP and CAMH\GSK controls. Similar methods were used to infer CNVs in controls. Fisher's Exact Test was used to calculate the two-tailed p value.

[0041] *DNA Sequencing and Mutation Screening:* PCR primers were designed with Primer 3 (v. 0.3.0) to amplify all three exons and intron-exon boundaries. PCR were performed under standard conditions, and products were purified and sequenced directly with the BigDye Terminator v3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems).

[0042] *X-Inactivation Studies:* X Chromosome Inactivation assays were performed on genomic DNA extracted from peripheral blood as described (Allen, R.C. et al., *Am.J.Hum.Genet.* **51**, 1229-1239 (1992)). Briefly, X Chromosome Inactivation was measured by the analysis of the (CAG)n repeat in the androgen receptor gene at Xq11-q12 before and after digestion with methylation sensitive restriction enzymes *Hha*I and *Hpa*II. Quantitative PCR amplification of androgen receptor gene repeat

alleles was compared, with and without restriction digestion, to determine the ratio of X-active/inactive alleles.

[0043] *Expression Analysis and Protein Localization:* Expression analysis and tissue distribution for *PTCHD1*, *PTCHD1AS1* and *PTCHD1AS2* was performed by RT-PCR, with a multiple tissue panel of first strand cDNA. The housekeeping gene *G3PDH* was used as a control. Origene human adult brain tissue panel was used to check the expression of *PTCHD* mRNA in different regions of the brain. qRT-PCR was performed with TaqMan Gene Expression assay Hs00288486, and samples were pre-normalized to *GAPDH* expression. Northern blot analysis was performed with a six tissue mRNA blot (BioChain). The BioChain FastHyb solution was used to hybridize the probe according to manufacturer's instructions. RNA *in situ* hybridization was performed on paraffin sections and whole-mounted fetal mouse and adult mouse brain using a 411 bp (chrX:152,008,934-152,009,344, UCSC Mouse July, 2007 (UCSC Genome Browser)) digoxigenin-labeled mouse antisense probe (and sense probe as negative control), using standard methods. To examine cellular localization of PTCHD1 protein, full-length human fetal brain *PTCHD1* cDNA was PCR amplified and cloned into the pcDNA3.1/CT-GFP-TOPO expression vector (Invitrogen). After confirming sequence and orientation of the insert,COS-7 and SK-N-SH cells were transiently infected with 2 μg of purified construct DNA with SuperFect (Qiagen). 24 hours after transfection, the PTCHD1-GFP fusion protein was visualized in transfected cells using a Zeiss Axioplan 2 imaging microscope, equipped with the LSM510 array confocal laser scanning system, and the Zeiss LSM510 version 3.2 SP2 software package.

[0044] *Luciferase Assays:* A luciferase assay was performed to compare the effect of PTCH1, PTCH2 and PTCHD1 on Gli-dependent transcription with a previously described method (Nieuwenhuis, E. et al., *Mol.Cell Biol.* **26**, 6609-6622 (2006)). Briefly, the 10T1/2 cells were transiently transfected with mixtures containing 0.1 μg β-galactosidase to normalize for transfection efficiency, 1 μg reporter plasmid (8xGlipro) encoding multimerized Gli binding sites fused to the luciferase gene and up to 1 μg of Gli2, PTCH1 or PTCH2 or PTCHD1. Gli-dependent transcription was measured and normalized by β-galactosidase. Data were replicated in independent

experiments performed in triplicates. In another assay, 10T1/2 cells were transiently transfected with mixtures containing 0.1 μg β-galactosidase, 1 μg 8xGlipro reporter plasmid and purmorphamine, PTCH1 or PTCH2 or PTCHD1. The effect of PTCH1, PTCH2 and PTCHD1 on the endogenous Gli-dependent transcription was measured. Statistical significance was calculated as p below 0.05, using the Student's t-test.

[0045] *Cytogenetic and CNV analysis of proband from Family 9:* Localization of translocation breakpoints was performed by fluorescence *in situ* hybridization (FISH; performed in accordance with standard procedures) initially using bacterial artificial chromosome (BAC) clones across the suspected breakpoint regions, and then narrowing the search using fosmid clones. BAC clones were obtained from the RP11 human genomic library, and fosmid clones from the Whitehead fosmid library WIBR2. For the chromosome 19 locus, the clone G248P85500F11 was translocated, and thus distal to the breakpoint, while clone G248P85559B4 was not translocated, and thus proximal to the breakpoint. The breakpoint therefore lies within a 32 Kb region between these two clones (UCSC March 2006: Chr19: 7,843,511-7,874,724. This region encompasses just two genes: *FLJ22184*, *LRRC8E*. At the chromosome 21 translocation site, fosmid clone G248P87249E2 was translocated, and G248P89542E9 was not translocated, and the breakpoint thus lies within a ~14.5 Kb region between these two clones, within an intron of the *RUNX1* gene.

[0046] Whole-genome SNP analysis was performed using the Affymetrix 260K *Nsp*I SNP microarray. Analysis using the dCHIP and CNAG programs indicated a loss of heterozygosity from SNPs rs10875047 at Chr1:97,367,581 and rs822559 at Chr1:98,424,675 (inclusive; UCSC March 2006). This apparent deletion spans from intron 20 of the gene *DPYD* to include the first 20 *DPYD* exons, as well as two proximal putative genes, AK094607 and AX747691.

RESULTS

[0047] *CNV Analysis of PTCHD1:* Precise breakpoints of the 167 Kb deletion at *PTCHD1* identified in the male proband from Family 1 were characterized. This CNV also disrupts long, spliced non-coding RNAs (ncRNAs) on the opposite strand

that codes for PTCHD1, however, no other coding genes were interrupted. See Figure **2** which depicts a detailed genomic organization of the *PTCHD1* locus. Known genes, predicted CpG islands (>300 bp), predicted promoters (ElDorado Suite from Genomatix) and conserved sequences (>75% identity with chicken, >90% identity with opossum or 100% identity with dog or horse) are shown.

[0048] The 167 kb deletion was validated in the family using both PCR and SYBR-Green I-based real-time quantitative PCR (qPCR) and was found to be transmitted from a heterozygous unaffected mother to two affected dizygotic twin sons, also to an unaffected daughter (Figure **3**). X-chromosome inactivation (XCI) analysis of the mother, carrier of the *PTCHD1* deletion, revealed a highly skewed allelic ratio of 94:6. The third male in Family 18 was assessed at age 4 and had speech and language problems, but was not available for further assessment. The father in Family 19 has a broader autism phenotype (BAP) (Pinto, D. et al., *Nature* **466**, 368-372 (2010)). The proband in Family 20 (hatched) has ADHD plus BAP. A diamond symbol represents siblings who were not tested as part of the study, and with gender not indicated.

[0049] *Mutation Screening of PTCHD1:* In order to identify additional cases with *PTCHD1* mutations, the coding regions in 900 (M=723; F=177) unrelated ASD cases and 225 unrelated male ID cases were sequenced. Missense changes were identified in unrelated ASD probands and ID probands (Figure **3**; Figure **4**; see also Table 1, above). In Figure **5**, the protein structure of the transmembrane protein PTCHD1 is illustrated. In 5A, twelve transmembrane domains (cylinders) and Patched-domain (line) were identified using the SMART tool (http://smart.embl-heidelberg.de/) with the Pfam domain option selected. In addition, the locations of missense sequence variants discovered among ASD and ID probands are shown. 5A shows the position of missense mutations among ASD and ID probands. Amino acid positions given are relative to the human PTCHD1 sequence (NP_775766). Other sequences used include mouse (NP_001087219), opossum (XP_001366520), platypus (XP_001512040), chicken (XP_425565), zebrafish (XP_690754), sea urchin (XP_001199849) and nematode (*C. elegans*) (NP_499380). **5B/C** depicts PTCH1, showing missense mutations reported for holoprosencephaly[14,15], and includes sequences from human PTCH1 (NP_000255), mouse (NP_032983), opossum

(XP_001368370), chicken (NP_990291), *Xenopus laevis* (NP_001082082), zebrafish (XP_001922161), fruitfly (NP_523661) and nematode (*C. elegans*; NP_495662).

[0050] All of these variants, which resulted in the substitution of highly conserved amino acids, were inherited from unaffected carrier mothers (Figure **4**). In six of the eight families the missense variants appear to segregate with the phenotype, however in Family 6 L73F did not segregate, (see Figure **4** and Table 1 for details).

[0051] The entire coding region of *PTCHD1* was sequenced in 700 control individuals (M=531 F=169), and none of the missense changes identified from among the ASD and ID patient cohorts has been detected. Only two missense changes have been identified: P252L from amongst the controls, and N497K reported in the SNP database (rs35880456, in 1 out of 39 screened; NCBI), both in females who were heterozygotes. Altogether, absence of *PTCHD1* missense variants indicates that these variants are significantly enriched in the males with ASD (6/723 male ASD versus 0/531 male control: Fisher's exact test: p =0.042) and may contribute to the phenotype.

[0052] Additional controls were sequenced for the exons in which missense mutations were identified. Control chromosomes were tested for the sequence underlying the I173V and V195I mutations (N=1101 chromosomes), the ML336_337II mutation (N=1193), and the L73F and E479G mutations (N=869) and detected none of these variants.

[0053] *CNVs upstream of PTCHD1 (PTCHD1AS1/PTCHD1AS2 locus):* Copy number variations were also identified upstream of the coding region for PTCHD1. A study of 996 ASD families examined with the Illumina 1M BeadChip (Pinto, D. et al., *Nature* **466**, 368-372 (2010)) identified deletions in probands or affected siblings, and in a father with a diagnosis of Broad Autism Phenotype (BAP) (Hurley R.S. et al., *J.Autism Dev.Disord.* **37**, 1679-1690 (2007); Constantino, J.N. et al., *Biol.Psychiatry* **57**, 655-660 (2005)). All of the upstream CNVs occurred 5' of *PTCHD1*, and overlapping with an anti-sense non-coding RNA, *PTCHD1AS1/PTCHD1AS2*. A tenth deletion at this upstream locus was identified in a patient from a CNV study of 167 unrelated attention deficit-hyperactivity disorder (ADHD) patients. The ADHD proband

with the deletion also has a BAP diagnosis. See Figure **2**. Putative non-coding RNA transcripts *PTCHD1AS1* (from cDNA clone IMAGE:1560626; BX115199) and *PTCHD1AS2* (cDNA clone BRSTN2000219; DA355362) from human, mouse and rat genomes are also shown, with transcripts assembled from RT-PCR and 5' RACE (*PTCHD1AS3*) results. The dotted line between the two exons in transcript *PTCHD1AS1* indicates that this is a putative exon, identified through clone sequencing. This exon is putative because, although this location represents its best genomic hit, it only partially matches the 5' end of the clone sequence. The consensus sequences for noncoding RNA of PTCHD1AS1, PTCHD1AS2 and PTCHD1AS3 are shown in Figures 6, 7 and 8, respectively.

[0054] In Figure **2**, Black boxes within the spliced transcripts indicate homologous exons between the sequences. White bars with black borders indicate CNV losses within this locus that have been identified in patients with ASD and controls. Cross-hatched or grey bars indicate CNV losses identified in patients with ADHD and ID, respectively. Lines within these bars indicate overlap with exons of known transcripts or ncRNA.

[0055] The breakpoints of the deletions for all families that are reported here were mapped by sequencing the junction. Breakpoints for all CNVs in controls were mapped by using the physical positions of microarray probe fragments. Deletions were validated with qPCR and exact breakpoints at the PTCHD1 locus were mapped (See Table 2). Additional CNV data for the individuals in other regions is included in Table 3.

Table 2: Breakpoint of deletions at the *PTCHD1* locus:

Family	**Breakpoints***	**Deletion size (bp)**	**Method used to map the breakpoints**
Family 1(5240)	chrX:23,114,179-23,281,723	167,543	Sequencing of junction fragment.
Family 11 (5298)	chrX:22,890,415-23,015,667	125,253	Sequencing of junction fragment.
Family 12 (5065)	chrX:22,859,294-22,924,136	64,843	Sequencing of junction fragment.
Family 13 (3424)	chrX:23,011,719-23,116,212	104,494	Sequencing of junction fragment.
Family 14 (5111)	chrX:22,841,534-22,900,490	58,957	Sequencing of junction fragment.
Family 15 (3253)	chrX:22,853,977-22,908,345	54,367	Sequencing of junction fragment.
Family 16 (13047)	chrX:22,826,477-23,215,032	388,556	Sequencing of junction fragment.
Family 17 (8273)	chrX: 22,989,332-23,091,080	101,749	Sequencing of junction fragment.
Family 18 (8013)	chrX:22,859,294-22,924,136	64,843	Sequencing of junction fragment.
Family 19 (3387)	chrX:22,824,496-23,037,508	213,013	Sequencing of junction fragment.
Family 20 (1-27075)	chrX: 22,678,814–23,066,819	388,006	Sequencing of junction fragment.

* refers to genome assembly HG18

Table 3: Additional CNVs in 9 subjects with upstream deletions:

Family	Gender	Inheritance	Physical Position	Size (bp)	CNV	Cytoband	Genes
Family 1 (5240)	M	Maternal	2:236932539_236990050	57,512	3	2q37.2	IQCA1
Family 11 (5298)	M	Paternal	14:43889940_44003766	113,827	3	14q21.3	No gene.
	M	Maternal	16:16225138_16726778	501,641	3	16p12.3,16p13.11	ABCC6,NOMO3
	M		16:18153166_18699648	546,483	3	16p12.3	ABCC6P1,NOMO2,LOC339047,RPS15A
Family 12 (5065)	M	Maternal	1:17079505_17140083	60,579	1	1p36.13	CROCC
	M	paternal	3:1719782_1786952	67,171	3	3p26.3	No gene.
	M	Maternal	3:17494057_17542224	48,168	1	3p24.3	TBC1D5
	M	Maternal	3:197219312_197527449	308,138	3	3q29	PCYT1A,TCTEX1D2,TFRC,ZDHHC19,OSTalpha
	M	Maternal	4:22488002_22620537	132,536	3	4p15.31	No gene.
	M	Maternal	10:68138586_68227559	88,974	1	10q21.3	CTNNA3
	M	paternal	11:61516315_61632187	115,873	3	11q12.3	No gene.
	M	Maternal	16:21506626_21647775	141,150	3	16p12.2	METTL9,IGSF6,OTOA
Family 13 (3424)	M	paternal	5:98798044_98836932	38,889	1	5q21.1	No gene.
	M	Maternal	7:149089061_149159195	70,135	3	7q36.1	SSPO,ZNF467
Family 14	M	Maternal	18:66315754_66382003	66,250	1	18q22.2	No gene.

(5111)							
Family 15 (3253)	M	NA	5:20975886_21105120	129,235	1	5p14.3	No gene.
	M	NA	7:109552072_109593909	41,838	1	7q31.1	No gene.
	M	NA	9:11936421_12032535	96,115	1	9p23	No gene.
Family 16 (13047)	M	Maternal	1:244036261_245191978	1,160,000	1	1q44	AHCTF1,TFB2M,LOC149134,SCCPDH,SMYD3,C1orf71
	M	Maternal	9:24652558_24705098	52,541	1	9p21.3	No gene.
	M	Maternal	18:67894269_67931021	36,753	1	18q22.3	No gene.
Family 17 (8273)	NA						
Family 18 (8013)	NA						
Family 19 (3387)	NA						
Family 20 (1-27075)	NA						

[0056] SNP microarray data was analyzed from 10,246 control individuals (4,829 male; 5,417 female), for CNVs at *PTCHD1* and the upstream region. In a 1.4-Mb region spanning from *PTCHD1* to adjacent genes *PRDX4* (proximal) and *ZNF645* (proximal), 15 CNVs were identified (7 duplications and 8 deletions); however, it is notable that only 1 male control with a deletion was identified, which was 20.6 Kb in length and did not disrupt any known exons of any genes or non-coding RNAs, or any of the identified conserved or putative regulatory sequences. The remaining 7 deletions were all identified among female controls, consistent with the X-linked recessive inheritance observed for the *PTCHD1* mutations. Thus, *PTCHD1* and upstream deletions were not observed in 4,829 male controls, or in the Database of Genomic Variants (Iafrate, A.J. et al., *Nat.Genet.* **36**, 949-951 (2004)), which suggests that the CNV directly disrupting *PTCHD1* and the 6 CNVs located just upstream in unrelated ASD probands are associated with autism (male ASD cases N=7, out of 1,185; male controls N=0 out of 4,829; Fisher's exact test: $p = 1.2\times10^{-5}$).

[0057] *Expression and Functional Studies of PTCHD1:* Expression analysis for the *PTCHD1* and the ncRNA transcripts suggests that they are transcribed in brain regions, notably the cerebellum, as well as in other tissues (data not shown). RNA *in situ* hybridization of *Ptchd1* in mouse showed widespread expression in the developing brain from E9.5/10.5 to P1 (data not shown), as well as broad expression in the adult mouse brain (6 months), with highest density in the cerebellum (see Allen brain atlas online (Allen Institute mouse brain atlas *in situ* hybridization data for Ptchd1: http://mouse.brain-map.org/brain/Ptchd1.html)).

[0058] Gene expression and genes co-expressed with *PTCHD1* were also analyzed, from gene Affymetrix gene expression microarray analysis from BioGPS (Gene Atlas U133A, gcrma; http://biogps.gnf.org); UCLA Gene Expression Tool (UGET: http://genome.ucla.edu/~jdong/GeneCorr.html; using human HG-U133_Plus_2 microarrays (*2*), and correlation with mouse Ptchd1 using UGET and Mouse430_2 microarrays. These algorithms correlate expression based on banked Affymetrix gene microarray data, and is not tissue specific. Ranking counts multiple probes as single hits, and excludes hypothetical proteins.*PTCHD1* gene expression showed high

correlation with expression of other cerebellar genes such as *ZIC1*, *CADPS2*, *EN2*, *CBLN1*, and with synaptic genes such as *PCLO*, *NRXN3*, *SNAP25*, *SYT2*, *DPP6* and *DPP10* (see Table 4).

[0059] To investigate its function, the sub-cellular localization of PTCHD1 was studied. It was found that a PTCHD1-GFP fusion protein predominantly localizes to the cell membrane (data not shown). It was further hypothesized that PTCHD1 may function in the Hh-signaling pathway and have similar functional attributes as PTCH1 and PTCH2. A Gli-dependent transcription assay was performed in Hh-responsive 10T1/2 cells to test whether PTCHD1 could interfere with Hh signaling. In 10T1/2 cells, overexpression of PTCH1 or PTCH2 inhibits transcription from a Gli-luciferase reporter containing multiple copies of the Gli protein-binding site in the presence of Smoothened agonist purmorphamine (Sinha, S. and J.K. Chen, *Nat.Chem.Biol.* **2**, 29-30 (2006)) or Gli2 (data not shown). Similar to PTCH proteins, PTCHD1 also exerted a statistically significant inhibitory effect in these assays suggesting that PTCHD1 functions in the Hedgehog signalling pathway.

Table 4. Genes co-expressed with *PTCHD1*

Gene Name	Correlation	Rank#	OMIM #	Comments
A. BioGPS co-expression data for PTCHD1 from Gene Atlas, U133A				
PTCHD1	1	1		
ZIC1	0.7564	2	600470	Zinc finger protein in cerebellum; homologue of Gli
GABRD	0.7064	12	137163	Receptor subunit (delta) for GABA neurotransmitter
MAB21L1	0.6916	17	601280	Autism susceptibility locus, AUTS3, candidate gene
CBLN1	0.6832	21	600432	Precerebellin 1
CADPS2	0.6827	22	609978	Cerebellar gene; involved in vesicular trafficking; autism candidate gene
CACNA1A	0.6801	23	601011	Gene for spinocerebellar ataxia 6
CALN1	0.6675	26	607176	Calneurin 1; cerebellar homologue of calmodulin
NRXN3	0.6041	42	600567	Neurexin 3; synaptic adhesion and presynaptic voltage-gated Ca^{2+} signalling
EN2	0.5799	50	131310	Engrailed 2; candidate gene at autism locus, AUTS10
SYT2	0.5782	51	600104	Synaptotagmin 2; synaptic vesicle associated protein, CA^{2+} sensor
GRM1	0.5747	52	604473	Metabotropic glutamate neurotransmitter receptor
GABRA6	0.5171	77	137143	Receptor subunit (alpha-6) for GABA neurotransmitter
SNAP25	0.5034	87	600322	Synaptosomal-associated protein
B. UGET co-expression data for PTCHD1 from HG-U133_Plus_2 platform				
PTCHD1	0.85455	1		
SNAP25	0.5389	7	600322	Synaptosomal-associated protein
CACNA1A	0.52815	10	601011	Gene for spinocerebellar ataxia 6
NRXN3	0.514	13	600567	Neurexin 3; synaptic adhesion and presynaptic voltage-gated Ca^{2+} signalling
GABRA6	0.50935	15	137143	Receptor subunit (alpha-6) for GABA neurotransmitter
GRM1	0.50555	19	604473	Metabotropic glutamate neurotransmitter receptor
GABRD	0.4958	24	137163	Receptor subunit (delta) for GABA neurotransmitter
KCNC1	0.4935	25	176258	Voltage-gated K+ channel, Shaw-related, Kv3.1
SYT4	0.4934	26	600103	Synaptotagmin 4; synaptic vesicle associated protein, CA^{2+} sensor
CBLN3	0.4867	32	612978	Precerebellin 3
DPP6	0.4771	45	126141	Dipeptidyl peptidase 6: forms complex with Kv4.2 channels at synapse

CADPS2	0.4699	54	609978	Cerebellar gene; involved in vesicular trafficking; autism candidate gene
C. UGET co-expression data for mouse Ptchd1 from Mouse430_2 platform				
Ptchd1	0.7053	1		
Olfm3	0.4714	2	607567	Olfactomedin 3
Gria4	0.4397	3	138246	Glutamate receptor (AMPA); L-glutamate-gated ion channel
Pclo	0.4235	5	604918	Piccolo; presynaptic cytoskeletal matrix component
Dpp10	0.4165	9	608209	Dipeptidyl peptidase 10; forms complex with Kv4.2 channels at synapse
Cadps2	0.39	19	609978	Cerebellar gene; involved in vesicular trafficking; autism candidate gene
Nrxn3	0.3879	21	600567	Neurexin 3; synaptic adhesion and presynaptic voltage-gated Ca^{2+} signalling
En2	0.3816	30	131310	Engrailed 2; candidate gene at autism locus, AUTS10

Gene Affymetrix gene expression microarray analysis from **A.** BioGPS (Gene Atlas U133A, gcrma; http://biogps.gnf.org); **B.** UCLA Gene Expression Tool (UGET: http://genome.ucla.edu/~jdong/GeneCorr.html; using human HG-U133_Plus_2 microarrays (*2*), and **C.** correlation with mouse Ptchd1 using UGET and Mouse430_2 microarrays. These algorithms correlate expression based on banked Affymetrix gene microarray data, and is not tissue specific. Ranking counts multiple probes as single hits, and excludes hypothetical proteins.

[0060] *RT-PCR failed to find evidence for a shortened 3' PTCHD1 transcript from individual with PTCHD1 exon 1 deletion:* It was speculated that the difference in phenotype between the *PTCHD1* deletion families, could be explained by residual PTCHD1 protein function in relevant brain regions in Family 1 due to downstream transcription and translation of a shorter isoform, possibly driven by a secondary promoter just upstream of exon 2, resulting in the milder ASD symptoms, rather than the severer ID with the full deletion. However, RT-PCR did not detect any evidence of shorter downstream transcripts.

[0061] RT-PCR and 5' RACE (Rapid Amplification of cDNA Ends) analysis of the ncRNAs, PTCHD1AS1 and PTCHD1AS2 and the PTCHD1 gene: By RT-PCR, the annotated exons of PTCHD1AS1 and PTCHD1AS2 were amplified from human cerebellum cDNA. Sequencing of RT-PCR product confirmed the current annotation of the ncRNAs. Additionally, the annotation of PTCHD1AS1 was verified by re-sequencing of the IMAGE clone 1560626.

[0062] It was attempted to identify additional 5' sequence of the ncRNAs and *PTCHD1* by 5' RACE analysis using the Clontech Marathon-Ready™ fetal brain cDNA (Cat. No. 639300). According to the manufacturer instructions the gene specific primers were designed for *PTCHD1AS1*, *PTCHD1AS2* and *PTCHD1* and RT-PCR was performed. The PCR products were cloned into the Promega pGEM®-T Easy Vector and the clones were sequenced using standard methods. No additional upstream sequence for *PTCHD1* could be found; however, for the *PTCHD1AS1* at least two additional exons were identified. One of these exons completely overlaps with the *PTCHD1AS2* exon 2 (chrX:23,198,089-23,198,215), while the second exon mapped further upstream at chrX:23,261,313-23,261,767 (UCSC 2006). RT-PCR also identified another splice variant with an initial exon at ChrX:23,262,967-23,262,009, which skips to exon 2 in the *current annot*ation of *PTCHD1AS1*. It is possible that the extremely GC-rich nature of the 5' region of *PTCHD1* prevented the finding of additional upstream sequence.

[0063] Alternative 5' exons for *PTCHD1AS1*, identified by 5'RACE, are shown

NCRNA	Exon	Size (bp)	Coordinates	Comments
PTCHD1AS1	1^{I}	126	chrX:23,198,089-23,198,214	This exon is alternatively spliced and completely overlaps with the exon 2 of the NCRNA355362.
PTCHD1AS1	1^{II}	455	chrX:23,261,313-23,261,767	Starts 1.1Kb upstream of *PTCHD1* and overlaps with the exon 1 of mouse transcript AK028243 and the *PTCHD1* CpG island.
PTCHD1AS1	1^{III}	43	chrX:23,261,967-23,262,009	Starts ~900 bp upstream of *PTCHD1* and overlaps with the *PTCHD1* CpG island. The transcript starting from this exon skips the Exon 1^{II}, 1^{I} and exon 1.

in Table 5 below.

Table 5 Alternative 5' exons for *PTCHD1AS1*, identified by 5'RACE

[0064] The relevant sequences are as follows:

Sequence of exon $1^{I:}$
CAATTGGTAGACATCTGGGTAGCTTCCACTTTTCCTGAACCAACTTTTAC
TGCAATTTGACAGCTAGTTGTCCACGTTCTGTGTTCTCCTCTCCAGGACT
CCAACTTCCTAAGTGGCTGTGGGTGC (SEQ ID No: 14)

Sequence of exon $1^{II;}$
ACCTGTGCGTGGCCGTTCCCGCCGCCGCCGCAGGTCTATCCCGGGGCCGA
AGCCGGCGCCCGCCTTCTCGGGGAATTCTCCGGAGGGGGAGTGCGAGGGG
AACCACGGTGACTGCCTGCTAGCTCACGGCTGGCGCGCACACGCACACGC
CCAACTTTGCCAAGCCGTCGGCGCCCCGCGGGCTCCCCCGCGCCCCCTGC
GGCTCAACACGCTCGGAGACCTGTATCTCTCCTGCTCTGAGATAAGGTTC
CCTCCACTCTCACACCTTCGCATGTAGGGGAGGAGAGGGCGGAGTGAGGC
AGAGAAGGGGGTTAATGCTACTGACTCCCTGGCCAGCCTTTCTCAAACAC
TCTACGCCCGCAGGGGCGCCCGCGCCAGCCACGCCGCACCAGGTCCCCCA
GACCTGCTGGTGACGACAGAGAGAGGAGGAGGAAGAGAAGGCAGGGCGAA
GAACC (SEQ ID No: 15)

Sequence of exon $1^{III:}$
CTTTTGAGTGGACGTGCTCCAGACACACACCCGGACCCCGTGG (SEQ ID No: 16)

[0065] *Putative promoter and enhancer sequences in intergenic region between DDX53 and PTCHD1:* The identification of predicted promoter sequences may indicate the presence of an alternative upstream transcription start site for *PTCHD1* (or possibly another unknown gene), that may be disrupted by the CNVs identified

upstream of *PTCHD1* in ASD families (see *supra*). The Genomatix ElDorado suite was used to predict promoter sequences. The promoter sequence for DDX53 is (hg18/UCSC March 2006 build):

TCTACACAAACCAGATGAACCNTCCAATCTCCTGCCTCGAGTATTGAAGCCTGGCTACTGTGACTGTGGG
GAAGGGATTAATGGTCTCAGCATTCAGCCAACAACAATACCTGCTCACTATAAGCATTCAGAAAACAGAA
AAGTTTCAAGAAGCAGGAAGAAAAGACTCACCTATGATCCCAACACCCAGAGATAAGAGTCCTGAAGCTC
AGATGACACAGCTGATAACAGGGAAGCCAGGACAGAATCTCATTGTTTTGAACACCAAAACCCGTTCCCT
TGACAACTTGGCTATACTACACTATTCGAATGTTGCAGATACTGTGGTCACATTTCAAAGGCCAGATCTT
TCCCAGGGCTTAAGCTGTTCCTTGGATACTTTTGGTAAGTCATTTATCCACTAATCATTTAGTAATCGTC
TCTGACATGCCAAACACCCTGCTCAGGGCTGGAAATGCAGAACCTGGGAAGCCACTGGCCTTGTCCTCAA
GATCTCTCTCTGGCTCCCTTTGAATTTGCTAATTCAGACTTTCACATTTCCCCCAGGAAAAATCATAAGG
ACCAAATCATATCCGTTTTCTCAAATGGCTTCAAAGACCCATGTCATCGTTTGGCATCATGTAATTCTTT
ACTGATGTACTTTAAGAGTCACGTTTTATTCTCTTTATGCAGCTGTCAAGGACAGACACAAAGAGGGGGG
GGGNGGNCTTCCTCACTAAATACTTTTCCCACAACA (SEQ ID No: 17)

[0066] In addition to promoter sequences at the 5' ends of DDX53 and PTCHD1, on the plus strand a putative promoter sequence was identified in the intergenic region, from ChrX:22,927,508-22,928,108 (hg18/UCSC March 2006 build):

AATGATGAATTTATCCTGACAAAGTACTGTATTCACTCCAAAAGAAATTT
ACCAAAATAAATGAACACACGAATATATAAATAAATAGTTTTACTTTAAA
TGCATTATTTTTTTCTCTTAGGGAAATAACTGGCTTATATAAAGGACAAT
GTGTATATGGTGTGTATGTTTAAGGCGTGCTTCAAGGTTGCTCTCAAGCT
GAGCCAGAACTATCACGAGAAGAGTGAAAGGAGCACCCGGGACGCAGAAG
TTAAGGAGGCAGTTACTCCTAGGGTCCTGTAAGTGCTGGCAGGGTCAGCC
CGTGAGAGTGAGTGCCTCTTTAAATTTGCGTCACAGACGCCTGCTTACCT
CACCCCAGTCCAAGCCCTGTGATTGGTCAGGCCATCAAAGCCTCGCCCCC
TACACGACCCGGAATTCGACGCCAACACTGGTTTCTGGGGCAACTTCTGC
GTAGCTATGTGACTAGCACCCGGAAATAATTGCCACCGCCATCTTTTGGT
GCAGAAGGTGACGGGAAACAGGCCGCAGACCTGAACTTCCAACCGTATGT
AGGCGAGAAGCCGGTGCCGATACTCCCACTATCCCACAATGTCCCACTGG
G (SEQ ID No: 18)

[0067] This putative promoter lies ahead of ENSEMBL predicted non-coding transcript ENST00000407873. On the minus strand a putative promoter sequence was identified in the intergenic region, from ChrX: chrX:23,022,123-23,022,723, which lies just ahead of ENSEMBL predicted non-coding transcript ENST00000356867 and an EST clone (AU118198) (hg18/UCSC March 2006 build):

ATTTTTAAAAAATATGCTGAATTTGAAGTTTCTTTCAAAGTACAGTGTTT
CAATGGGGGGAGTCCAATTTTTGTAAAATTTTACAAAAACTGTATTGCCC
TAAAGGCAGCCTACTGCACACAAGGATCACAGTGACTTTTACTTGTTATT
CTACATGATTACTTAAAATTTTTCTGATTTTTTTACCCTCATCTATCTTC
TAACTTGTCTAGTTAACTCTTAAGAATTTCAAATTTTCTTTGAAAGATGA
TAGGCAATATGAGATGAGAGATAATCTACAAAAGTTACAGATGCTCACAT
GTATAAAACAGTCAAAATATCACAGGTCAATGACATAAACTGCATTAAAT
AAATTATGTTTATAGGCATCAGTAGTTGAAAATGCTCAATAATTCTGGGC
TCCTTCCCCAAAATGTAAGACTTAAGTACTTCAAAGGCATTATTCTTTAC
TCATGAGGATCAGTGGCTTCATTTAGTAAAAGAAAAAGGAATGGACCCAG
GATCCCAGTAAATAATTACTAACTGATCGCAACGCTCTTTTATCTAATGA
ACAACCAACAACCAACAGAAAACCCTTGATTCACAGAGGAGCAAGTCCTA
G (SEQ ID No: 19)

[0068] The ElDorado Suite from Genomatix, as well as the FPROM algorithm from the Softberry suite, was also used to predict promoter/enhancer sequences just upstream of the *FAM3C2* predicted pseudogene.

[0069] Comparative sequence analysis indicated a number of regions located in the gene desert upstream of *PTCHD1* and between *DDX53* where nucleotide sequence conservation is relatively high through vertebrate evolution or through mammalian evolution. Such conserved regions may represent functional regions, possibly cis-regulatory sequences for *PTCHD1*. Regions were selected through the Vertebrate Multiz Alignment & PhastCons Conservation (28 Species) track on the UCSC (March 2006 build) browser. Results are shown in Table 1 and indicate which conserved elements overlap with CNV losses upstream of *PTCHD1*.

[0070] *eQTL at PTCHD1 locus*: The SNP rs7878766, located within *PTCHD1* intron 1, has been reported as a quantitative trait locus for expression of mRNA levels of *MAP8KIP2* in control brain cortex (http//eqtl.uchicago.edu), with a QTL score of 5.3. RefSeq Summary reports this to encode a scaffold protein involved in the c-Jun N-terminal kinase signaling pathway, and is thus thought to act as a regulator of signal transduction. Using mRNA by SNP Browser 1.0.1, other SNPs at the *PTCHD1* locus that showed as suggestive QTLs for mRNAs included rs5925800 (*ACSM2A*; LOD= 5.039, p=1.5 $\times 10^{-6}$; *GALNT4*, LOD=5.095, p=1.3 X 10^{-6}; *PIK3C2G*, LOD= 5.27, p=8.4 x 10^{-7}), rs868659 (*DLEU2*, LOD= 5.427, p=5.8 x 10^{-7}), and rs6526278 (*SGCG*, LOD= 5.248, p=8.8 x 10^{-7}).

[0071] In summary, the data indicate that mutations at the *PTCHD1* locus are highly penetrant and strongly associated with ASD (including BAP) and ID in ~1.1% and ~1.3% of the individuals analyzed, respectively (based on probands for whom comprehensive mutation screening, for both CNVs and sequence variants, has been performed (4 out of 353 ASD, and 3 out of 225 ID). As one of skill in the art will appreciate, mutations indicative of ASD and ID may vary from the exact CNVs identified (e.g. in Table 2 or other mutations), but will include at least a portion of one or more of the identified CNVs.

[0072] Overall, the findings are reminiscent of genetic findings for several other X chromosome genes, including *NLGN4* (Jamain, S. et al., *Nat.Genet.* **34**, 27-29 (2003); Laumonnier, F. et al., *Am.J.Hum.Genet.* **74**, 552-557 (2004)) and *IL1RAPL1* (Bhat, S.S. et al., *Clin.Genet.* **73**, 94-96 (2008); Piton, A. et al., *Hum.Mol.Genet.* 17, 3965-3974 (2008); Carrie, A. et al., *Nat.Genet.* **23**, 25-31 (1999)), in that mutations can apparently cause either ASD or ID (or both), and thus *PTCHD1* may be a gene for both. *IL1RAPL1*, for example, was initially reported as a gene for non-syndromic X-linked ID (Carrie, A. et al., *Nat.Genet.* **23**, 25-31 (1999)), and then subsequently was also found to harbor mutations in ASD pedigrees (Bhat, S.S. et al., *Clin.Genet.* **73**, 94-96 (2008); Piton, A. et al., *Hum.Mol.Genet.* 17, 3965-3974 (2008)). Families have also been identified in whom at least two loci may be contributing to the pathogenesis of ASD, and other families bearing upstream microdeletions that disrupt a complex non-coding RNA, providing possible genetic explanations for the clinical heterogeneity of these disorders. Finally, the results raise the possibility that Hh signaling may be perturbed in these conditions.

CLAIMS

What is claimed is:

1. A method of determining the risk of ASD in an individual comprising:
analyzing a nucleic acid-containing sample obtained from the individual for the presence or absence of a genomic sequence mutation at the *PTCHD1* locus wherein the mutation comprises a deletion of a region upstream to the *PTCHD1* gene, a disruption of a non-coding RNA selected from *PTCHD1AS1*, *PTCHD1AS2*, or *PTCHD1AS3*, or splice variants of these ncRNAs, or a disruption of other regulatory elements upstream of the PTCHD1 coding region, and wherein the presence of the mutation is indicative of a risk of ASD.

2. The method as defined in claim 1, wherein the mutation comprises a deletion of a region upstream to the *PTCHD1* gene.

3. The method as defined in claim 2, wherein the deletion comprises at least a portion of a region of the X chromosome selected from the regions: 23,114,179-23,281,723, 22,890,415-23,015,667, 22,859,294-22,924,136, 22,859,294-22,924,136, 22,841,534-22,900,490, 22,853,977-22,908,345, 22,826,477-23,215,032, 22,989,332-23,091,080, 22,859,294-22,924,136, 22,824,496-23,037,508 and 22,678,814–23,066,819.

4. The method as defined in claim 1, wherein the mutation comprises a disruption of a non-coding RNA selected from *PTCHD1AS1*, *PTCHD1AS2*, or *PTCHD1AS3*, or splice variants of these ncRNAs.

5. The method as defined in claim 4, wherein the mutation comprises a disruption of a non-coding RNA *PTCHD1AS1*, or splice variants thereof.

6. The method as defined in claim 4, wherein the mutation comprises a disruption of a non-coding RNA *PTCHD1AS2* or a splice variant thereof.

7. The method as defined in claim 4, wherein the mutation comprises a disruption of a non-coding RNA *PTCHD1AS3* or a splice variant thereof.

8. The method as defined in claim 1, wherein the mutation comprises a disruption of regulatory elements upstream of the PTCHD1 coding region.

9. The method of claim 8, wherein the mutation comprises a disruption of at least a portion of a promoter sequence in the intergenic region, from ChrX:22,927,508-22,928,108 or a promoter sequence in the intergenic region, from ChrX: chrX:23,022,123-23,022,723.

10. The method of claim 8, wherein the mutation comprises a disruption of cis-regulatory sequences for *PTCHD1*.

GCTCTAGGAT GCTGCGGCAG GTTCTGCACA GGGGCTTGAG GACGTGTTTC 50
TCCCGGCTCG GCCACTTCAT TGCCAGTCAC CCTGTCTTCT TCGCCTCGGC 100
GCCGGTGCTC ATCTCCATCC TGCTCGGCGC CAGCTTCAGC CGCTACCAGG 150
TCGAGGAGAG CGTGGAGCAC CTGCTGGCGC CCCAGCACAG CCTGGCCAAG 200
ATCGAGCGCA ACCTCGTTAA CAGCCTCTTC CCGGTCAACC GCTCCAAGCA 250
CCGTCTCTAC TCGGACCTGC AGACCCCCGG GCGCTACGGC CGGGTCATCG 300
TCACCTCCTT CCAGAAAGCC AACATGCTGG ACCAGCATCA CACCGACCTG 350
ATCTTAAAGT TGCATGCTGC TGTCACCAAG ATCCAGGTTC CAAGGCCTGG 400
TTTTAATTAC ACGTTTGCCC ATATATGTAT CCTGAATAAT GATAAGACTT 450
GCATCGTGGA TGACATAGTG CACGTCCTGG AAGAGCTAAA GAATGCTCGG 500
GCCACCAATC GGACCAATTT TGCTATCACA TACCCAATCA CTCACTTAAA 550
GGACGGGAGG GCTGTGTACA ATGGGCACCA GCTTGGGGGC GTCACTGTGC 600
ACAGCAAAGA CCGGGTGAAA TCTGCAGAGG CCATCCAGCT CACCTACTAC 650
CTGCAGTCAA TCAACAGTCT CAATGACATG GTGGCTGAGA GGTGGGAGTC 700
CAGCTTCTGC GACACTGTCA GACTGTTTCA GAAATCCAAC AGCAAAGTCA 750
AAATGTACCC TTACACGTCC TCCTCACTGA GGGAAGATTT CCAGAAGACC 800
AGCCGCGTAT CAGAACGTTA CCTGGTCACC AGCCTGATTC TGGTGGTTAC 850
CATGGCCATC CTGTGTTGCT CTATGCAGGA CTGCGTCCGC AGCAAACCCT 900
GGCTAGGCCT GCTCGGATTG GTGACCATAA GCCTGGCCAC TCTCACTGCA 950
GCCGGGATCA TCAATCTTAC TGGTGGGAAA TATAATTCCA CCTTCCTGGG 1000
AGTCCCTTTC GTCATGCTAG GTCATGGATT ATATGGGACT TTTGAAATGT 1050
TATCCTCCTG GAGGAAAACT AGAGAAGACC AACATGTTAA AGAGAGAACT 1100
GCAGCAGTCT ATGCAGACTC CATGCTCTCC TTTTCTCTCA CCACTGCCAT 1150
GTACCTGGTC ACCTTTGGCA TAGGGGCCAG CCCTTTCACG AACATTGAGG 1200
CAGCCAGGAT TTTCTGCTGC AATTCCTGTA TTGCAATCTT CTTCAACTAC 1250
CTCTATGTAC TCTCGTTTTA TGGTTCCAGC CTAGTGTTCA CTGGCTACAT 1300
AGAAAACAAT TACCAGCATA GTATCTTCTG TAGAAAAGTC CCAAAGCCTG 1350
AGGCATTGCA GGAGAAGCCG GCATGGTACA GGTTTCTCCT GACGGCCAGA 1400
TTCAGTGAGG ACACAGCTGA AGGCGAGGAA GCGAACACTT ACGAGAGTCA 1450
CCTATTGGTA TGTTTCCTCA AACGCTATTA CTGTGACTGG ATAACCAACA 1500
CCTATGTCAA GCCTTTTGTA GTTCTCTTTT ACCTTATTTA TATTTCCTTT 1550
GCCTTAATGG GCTATCTGCA GGTCAGTGAA GGGTCAGACC TTAGTAACAT 1600
TGTAGCAACC GCGACACAAA CCATTGAGTA CACTACTGCC CAGCAAAAGT 1650
ACTTCAGCAA CTACAGTCCT GTGATTGGGT TTTACATATA TGAGTCTATA 1700
GAATACTGGA ACACTAGTGT CCAAGAAGAT GTTCTAGAAT ACACCAAGGG 1750
GTTTGTGCGG ATATCCTGGT TTGAGAGCTA TTTAAATTAC CTTCGGAAAC 1800
TCAATGTATC CACTGGCTTG CCTAAGAAAA ATTTCACAGA CATGTTGAGG 1850
AATTCCTTTC TGAAAGCCCC TCAATTTTCA CATTTTCAAG AGGACATCAT 1900
CTTCTCTAAA AAATACAATG ATGAGGTCGA TGTAGTGGCC TCCAGAATGT 1950
TTTTGGTGGC CAAGACCATG GAAACAAACA GAGAAGAACT CTATGATCTC 2000
TTGGAAACCC TGAGGAGACT TTCTGTCACC TCCAAGGTGA AGTTCATCGT 2050
CTTCAATCCG TCCTTTGTAT ACATGGATCG ATATGCCTCC TCTCTGGGAG 2100
CCCCCCTGCA CAACTCCTGC ATCAGTGCTT TGTTCCTGCT CTTCTTCTCG 2150
GCATTCCTGG TGGCAGATTC ACTGATTAAC GTCTGGATCA CTCTCACAGT 2200
TGTGTCCGTG GAGTTTGGAG TGATAGGTTT CATGACATTA TGGAAAGTAG 2250
AACTGGACTG CATTTCTGTG CTATGCTTAA TTTATGGAAT TAATTACACA 2300
ATTGACAATT GTGCTCCAAT GTTATCCACA TTTGTTCTGG GCAAGGATTT 2350
CACAAGAACT AAATGGGTAA AAAATGCCCT GGAAGTGCAT GGGGTAGCTA 2400
TTTTACAGAG TTACCTCTGC TATATTGTTG GTCTGATTCC TCTTGCAGCT 2450
GTGCCTTCAA ATCTGACCTG TACACTGTTC AGGTGCTTGT TTTTAATAGC 2500
ATTTGTCACC TTCTTTCACT GCTTTGCCAT TTTACCTGTG ATACTGACTT 2550
TCCTGCCACC CTCTAAGAAA AAAAGGAAAG AGAAGAAAAA TCCTGAGAAC 2600
CGGGAGGAAA TTGAGTGTGT AGAAATGGTA GATATCGATA GTACCCGTGT 2650
GGTTGACCAA ATTACAACAG TGTGATAATG TCTGCTTGGC ATATTTTCAC 2700

Figure 1A

```
CTTAGGTCTT ATCAAGACCA AAGAGATTAT GTTAATGAAA CAATTAAATT  2750
CAAAGTTCTT CCCTTTTTTA AAGATAGGAA ACAGGCATTG CCAAAAAAAA  2800
AAAAAAAAAA AAAAGGAAAG GACAGTGGGG AGAAATGGGC CTGGCATATT  2850
TTCAGTCTTT AAAACAAAGG AGTTGTTATG AGAATTCACA CACACATAGA  2900
CACACACACA CACACACACA CACACACACA CACACACACA CCCTGGGAGA  2950
CCTATAGTCT CTTAAACTAA GATCAAGTAG AAGAAAGCTT ATTAACAAGC  3000
AGGATCCTGC CTTATCCAAA CTGCAGATGT TGCTGGCATT GTGACAAAAC  3050
CCACTGATTG AAAGGTCAAC TGCCAAGGCA GAAACACCTT TAAGCATTGT  3100
TCAAACAATA AGGCTTCCAG AACTTCTGTA GAGCAGTAGC TCCAGTCATG  3150
GTCTGTGGTT TGAGGTTTTA GCTGTCTCAC CTAGCTCCCT AACACTGAAG  3200
GAGATACTTG TGAAAGTTCT GACCAGCAAA AGCAAGCCAG AGCCTTGGAA  3250
ACTGATATGT GGTAGAGTGG CCATCACTCA TGGACTAAAA TTGATTCACC  3300
GCTAAATTTA CCCAGGTGAA GCAGTTTCGT TGTCTAGAAT GAAATTATCA  3350
TATTCCGCCA TTGGTATGCC TTTAACATTT GTATAGTTTG GTTTGCTTAA  3400
AACACCTTAA AACCAATGAC AGCTCCAGCA CTGCAGAATT GGTGTGATTC  3450
TACTTTGGAA TAGCTTGTCA CTTGTCACCA AATGGGTCTG CTTTATTAGT  3500
TACAGCTCTT GGCAGGAGGA TCCAGGGACC CAAAACCACA GGGCCAAACC  3550
CAAATACCTG GCATGATGGA GCAAAAGCAG GTGTCTACTT GGACCCAGAT  3600
ATAGTGTCTC CATTTTAACA ACAACAACAA AATAGCCAGC TGGTACAGCT  3650
GTTTGCATTG GCCCTACATG CATTTTTTGC ATGGATATCC AGAAACATCT  3700
GCCCACACAA AACTGCGGGG AAAAAAAATG AACACTGAAA TAGTTATTTG  3750
CTGTTGCTTC CAACTTGTAG TGCCAGTCTG CCTTTGCTGT GAAACACACC  3800
TGCTCAGAGA CAGAGAGGGG AAGAAGATCT TTGGTAAGTC TAAGTCCTGA  3850
CGCTGAGAAG CTTTGTAAAA GTGCAGGGAG ATAAAGGGCC AAAAGGGAGA  3900
TAGATGGAAA ACACTGGAAA AAGTATTCAC TGATACAAAT CTATCAATGA  3950
TGGCAGTCCA ATTCTCTTGC TAAAGTGGCT GCACCTCACC TTGCTGGTCC  4000
CCCCCACACC TTTTTTGATG TCCTTCTGCG TCATCATAGC AAGGCCCTTC  4050
TGTAAATTAA CAAGCCTAGA TATTTATACT CTTGACTTCC AGTATCTACA  4100
GAAGAATGGT TCATAGATCT AAACAGAAAT GGTTTAGATC TAAAAAGGCT  4150
GTATACGTTG CCCAGGCCCC TGCATTTCTT TAAATTTATA AAAATGAAGC  4200
TAAAACCTGG TTACATTTGA AGCAAATATC TACAGTATTT TTCCCTTTTA  4250
GAGATGTAGC TTCCTTAGAC ATCTGTAGTG GTAAGCATTT CCCAAAAGCA  4300
TCTTACCTTT CTGAACCTTA GCAGACATAC TGTGCAGCTT ACCTATCTTC  4350
TGCAGAGGAG GAAACTGAGA CCTAGGAGAA TAAAGTGACT CACTCAGGTC  4400
ACACCACTAA AGGGTTTTCA TCATTTCAGC ATACCTAAGA CAGGGCAGTC  4450
CAATTTTCAG TATTCTCATA AGATGGCTAT TACTCCTCTC AAAATGCATT  4500
TCCAAAGTAG GAACATAGGA CTTCGTTGGC CACAGGGCAG ACATTTTTTT  4550
AGTGTCTGGA ATTAAAATGT TTGAGGTTTA GGTTTGCCAT TGTCTTTCCA  4600
AAAGGCCAAA TAATTCAGAT GTAACCACAC CAAGTGCAAA CCTGTGCTTT  4650
CTATTTCACG TACTGTTGTC CATACAGTTC TAAATACATG TGCAGGGGAT  4700
TGTAGCTAAT GCATTACACA GTCGTTCAGT CTTCTCTGCA GACACACTAA  4750
GTGATCATAC CAACGTGTTA TACACTCAAC TAGAAGATAA TAAGCTTTAA  4800
TCTGAGGGCA AGTACAGTCC TGACAAAAGG GCAAGTTTGC ATAATAGATC  4850
TTCGATCAAT TCTCTCTCCA AGGGGCCCGC AACTAGGCTA TTATTCATAA  4900
AACACAACTG AAGAGGGGAT TGGTTTTACT GTTAAATCAT GTGTTGCTAA  4950
ATCATTTTCT GAACAGTGTG TTCTAAATCA GTCATTGATT TAGTGTCAGC  5000
CACGTGGAGC ACCTCGGCTT AAAGCAGCTC CACAAAACCT GACACAACAC  5050
ACACACCAAT TAAATGGATT TTGTTGAGAA TTTAATCATT CAATTTGGTC  5100
AACCAGAATG ACTTCCTGTG GAACTCTGTT TTATGACAGA TAATAGTTTT  5150
CCAACTTGAT TGAGTCTCTG TATACCCTGG GATATTGTAT TTTTTAATGA  5200
AGGGCATTTT CAAACTTGTC AACTTCTCTT TTCAGCACTT GAAATGAAGG  5250
CTTATGGAAT TCTGACTGTG AAATGAATTT TTCTATTGGG AAAAAAAAAA  5300
AAAAA (SEQ ID No:1)
```

Figure 1A cont’d

MLRQVLHRGLRTCFSRLGHFIASHPVFFASAPVLISILLGASFSRYQVEE
SVEHLLAPQHSLAKIERNLVNSLFPVNRSKHRLYSDLQTPGRYGRVIVTS
FQKANMLDQHHTDLILKLHAAVTKIQVPRPGFNYTFAHICILNNDKTCIV
DDIVHVLEELKNARATNRTNFAITYPITHLKDGRAVYNGHQLGGVTVHSK
DRVKSAEAIQLTYYLQSINSLNDMVAERWESSFCDTVRLFQKSNSKVKMY
PYTSSSLREDFQKTSRVSERYLVTSLILVVTMAILCCSMQDCVRSKPWLG
LLGLVTISLATLTAAGIINLTGGKYNSTFLGVPFVMLGHGLYGTFEMLSS
WRKTREDQHVKERTAAVYADSMLSFSLTTAMYLVTFGIGASPFTNIEAAR
IFCCNSCIAIFFNYLYVLSFYGSSLVFTGYIENNYQHSIFCRKVPKPEAL
QEKPAWYRFLLTARFSEDTAEGEEANTYESHLLVCFLKRYYCDWITNTYV
KPFVVLFYLIYISFALMGYLQVSEGSDLSNIVATATQTIEYTTAQQKYFS
NYSPVIGFYIYESIEYWNTSVQEDVLEYTKGFVRISWFESYLNYLRKLNV
STGLPKKNFTDMLRNSFLKAPQFSHFQEDIIFSKKYNDEVDVVASRMFLV
AKTMETNREELYDLLETLRRLSVTSKVKFIVFNPSFVYMDRYASSLGAPL
HNSCISALFLLFFSAFLVADSLINVWITLTVVSVEFGVIGFMTLWKVELD
CISVLCLIYGINYTIDNCAPMLSTFVLGKDFTRTKWVKNALEVHGVAILQ
SYLCYIVGLIPLAAVPSNLTCTLFRCLFLIAFVTFFHCFAILPVILTFLP
PSKKKRKEKKNPENREEIECVEMVDIDSTRVVDQITTV(SEQ ID NO:2)

Figure 1B

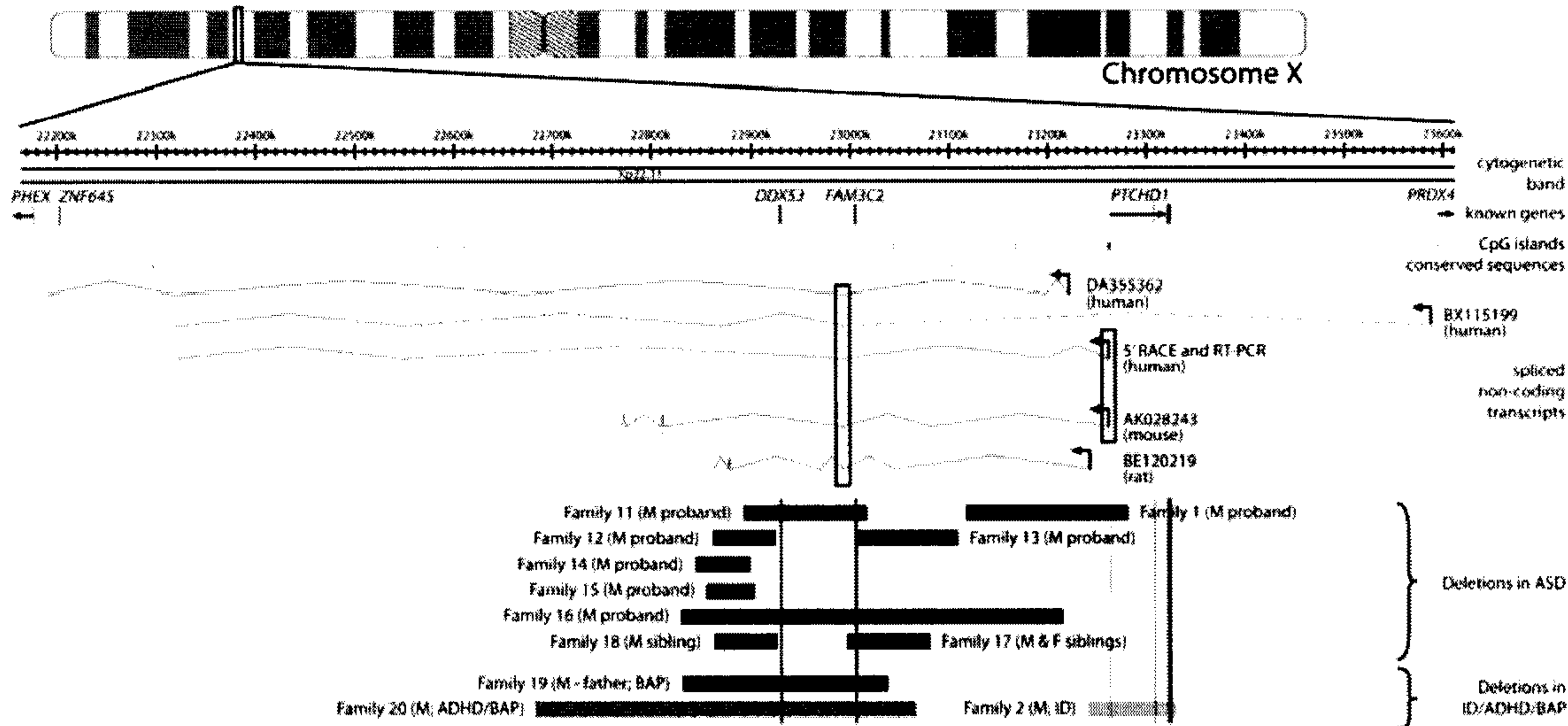
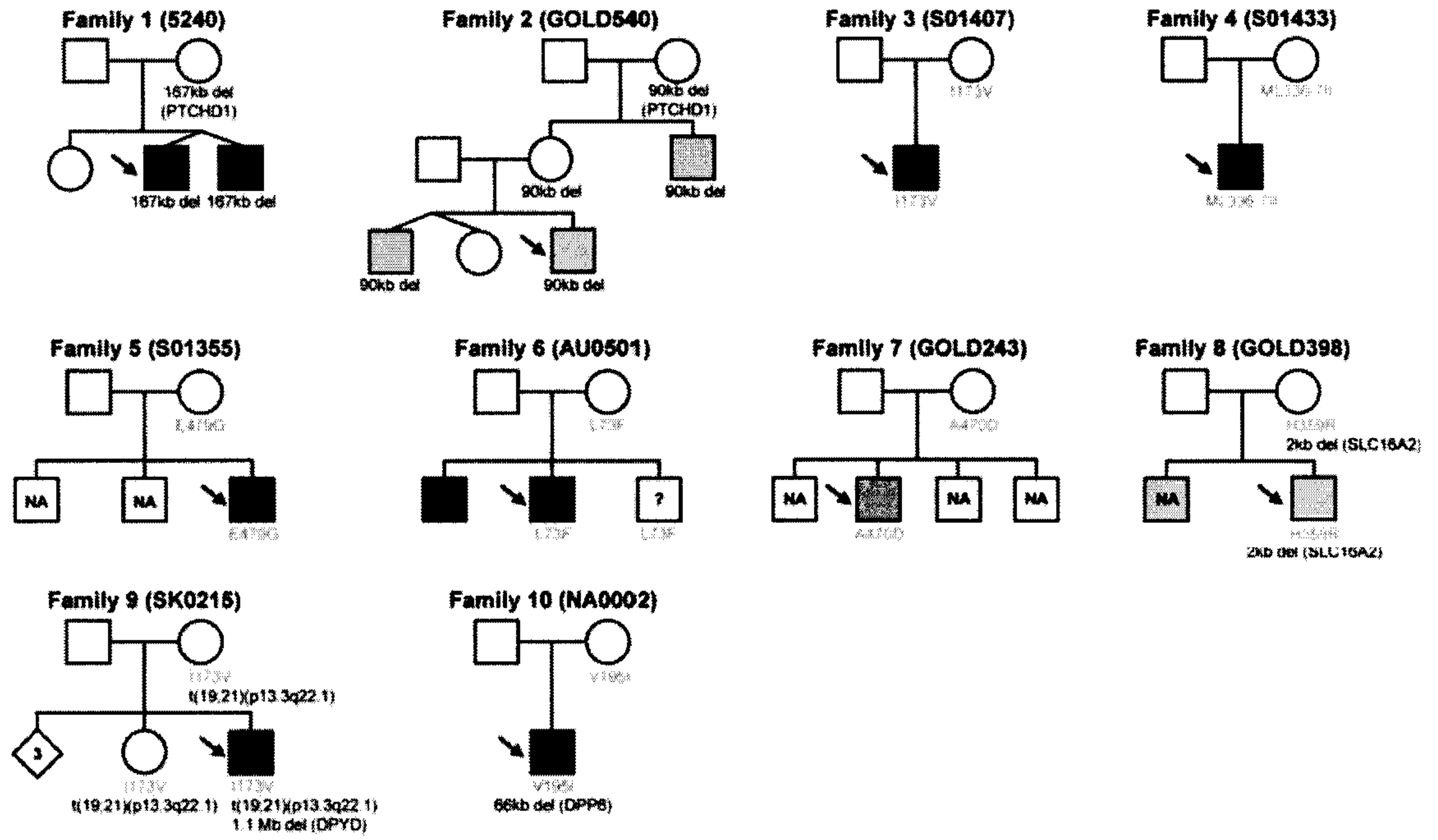
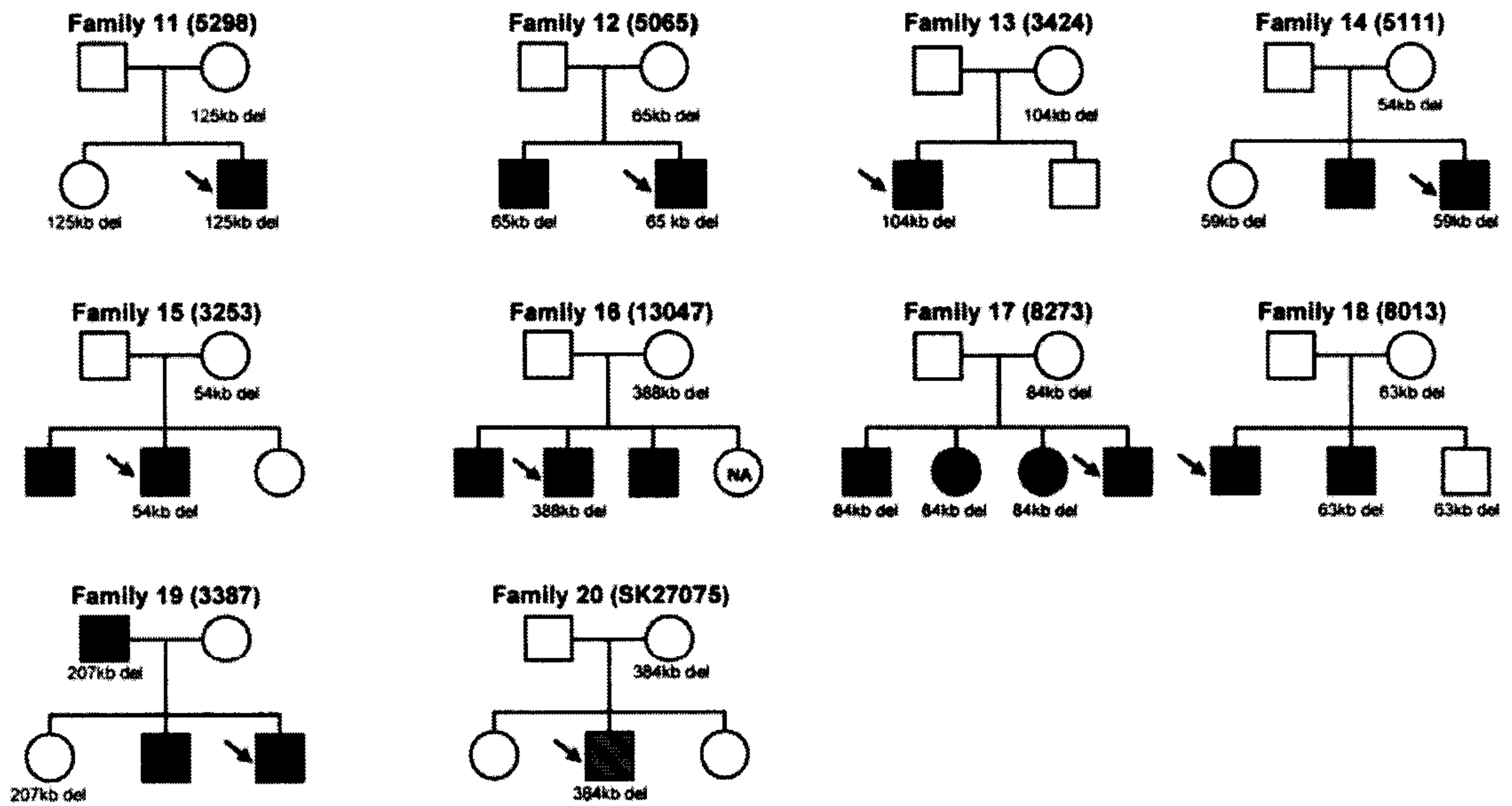


Figure 2

a) PTCHD1 mutations



b) ncRNA355362/ncRNA115199/PTCHD1 locus



ID ■ ASD ■ ADHD/BAP ↘ Proband ? – Phenotype unclear NA – Genotype/Phenotype not available

Figure 3

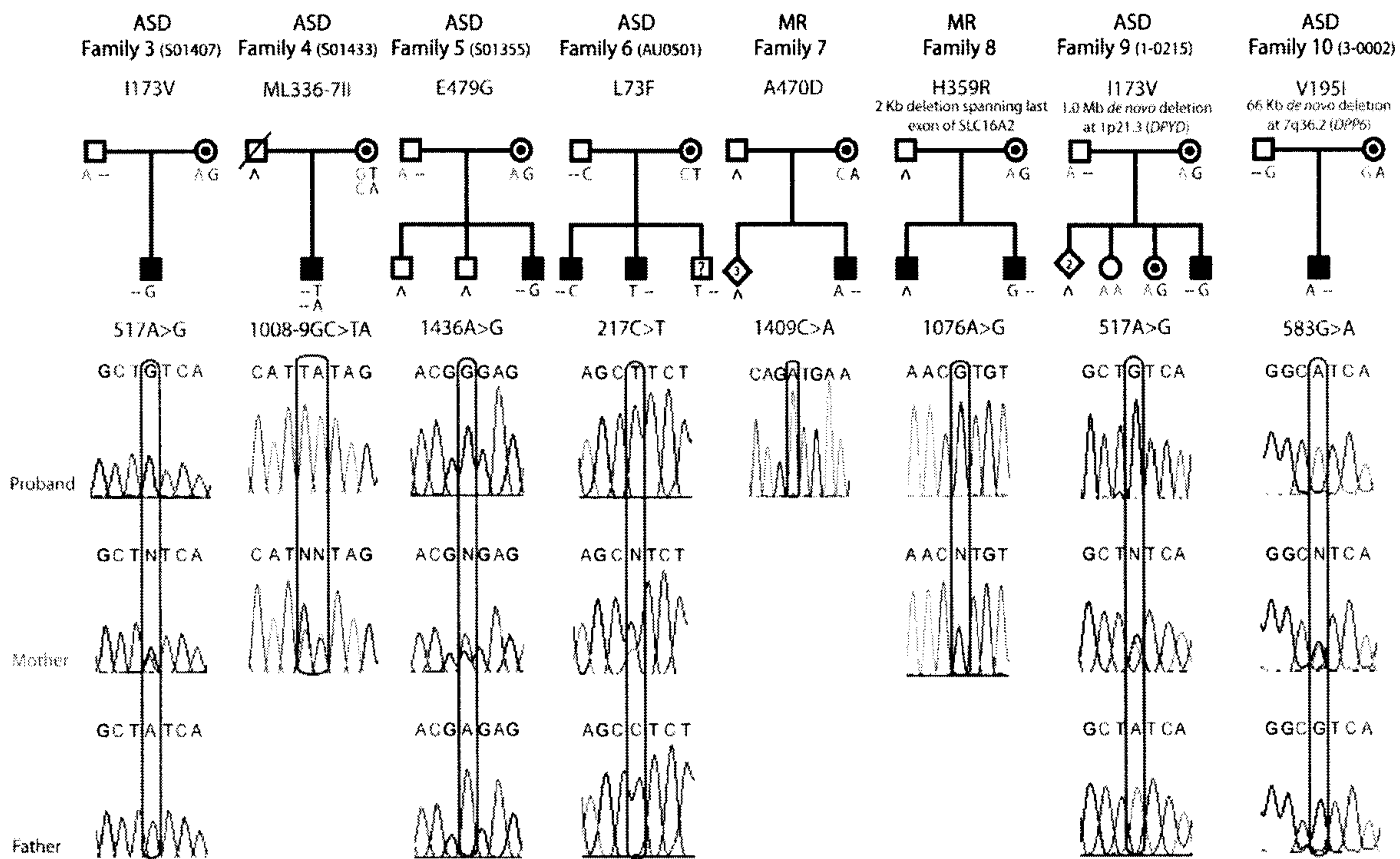
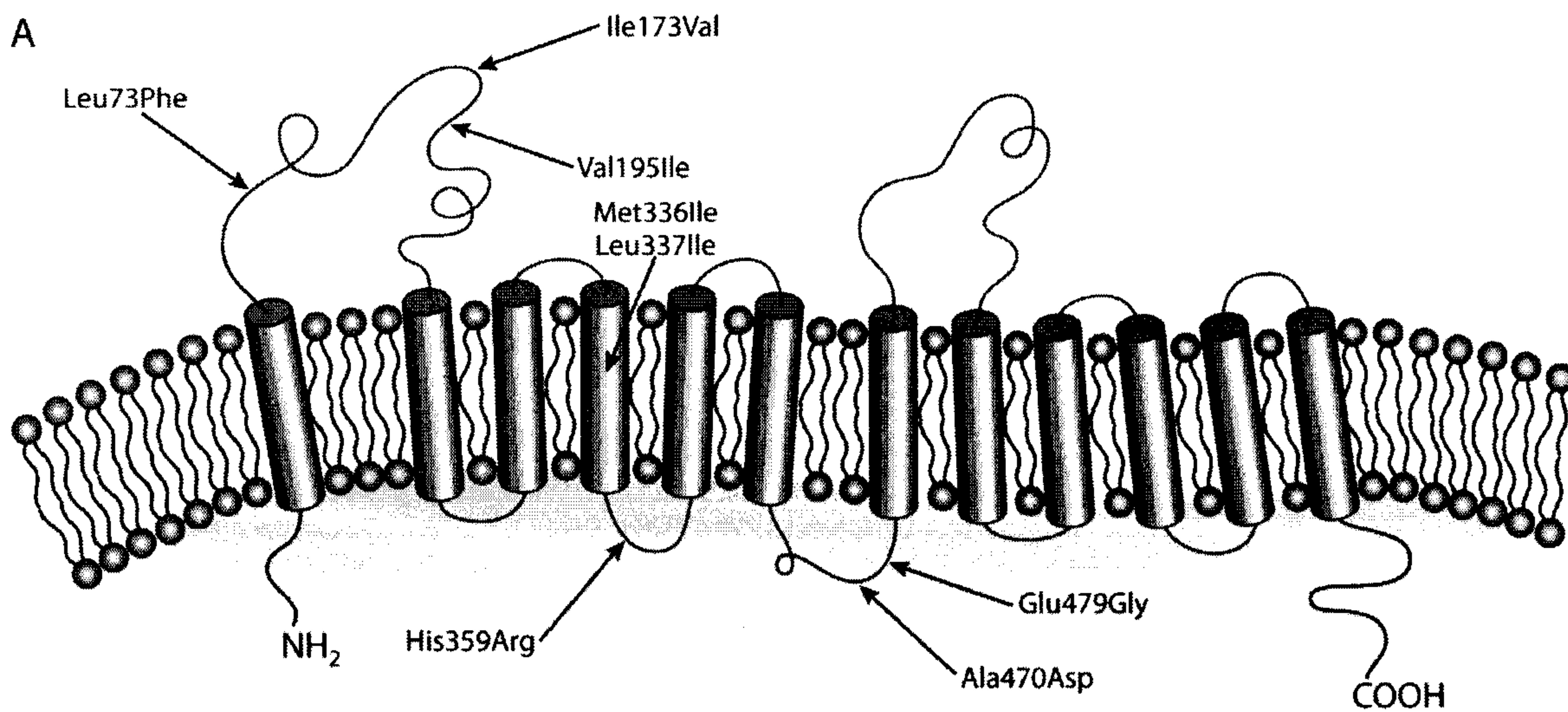


Figure 4

A



B

```
                 73            173                      195        336               359             470      479
human       KIERNLVNSL     NFA—ITYPITHLKDGRAVYNGHQLGGVTVHS   FLGVPFVMLGHGLYG   TREDQHVKERT     TAEGEEANTYESHL
mouse       KIERNLVNSL     NFA—ITYPITHLKDGRAVYNGHQLGGVTVHS   FLGVPFVMLGHGLYG   TREDQHVKERT     TAEGEEANTYESHL
opossum     KIERNLVNSL     NFA—ITYPITHLKDGREVYNGHQLGGVTVHS   FLGVPFVMLGHGLYG   TREDQHVKERT     TTDAEEANTYESHL
chicken     -----MVDVL     NFA—ITYPITHLKDGREVYNGHQLGGVTVHS   FLGIPFVMLGHGLYG   TREDQHVKERT     --------------
zebrafish   KIEGNLVDSL     VPP—LRYPITKLKDGREAYIGHQLGGVLASG   YLGIPFVMLGHGLFG   TREDQHVKERV     TTDSEETNTYESHL
platypus    KIERSLAGSL     AGGQVNYPNAKLKDGRSSFIGHQLGGVLETP   LLGVPFFAMGHGTKG   TRETLPFKDRV     QTSHHETNPYQNHF
sea_urchin  VGSNSIPVSL     FP---FLPLPPPHEGR-VFVGSQLGGVDLYP   VSLMPFLIIGVGVDN   LSIYLPVHERM     DSKCRKPEGHIIHP
nematode    RKELSQLDHL     IDEMTLSQISDAIQFDSGGMTHLLGGVTLDD   AYSMPFIVFSVGVDN   TSSTETLEHRM     IAAQGDRSFEKNTI

                 393            443             728             751              827             908
human       WNEDKAAAILE    VSVIRVASGYLLML   LEPPCTKWTLSS    KPKAKVVVIFLF     HRSFSNVKYVM     QRLVDADG (SEQ ID No: 3)
mouse       WNEDRAAAILE    VSVIRVASGYLLML   LEPPCTKWTLSS    KPKAKVVVILLF     HRSFSNVKYVM     QRLVDADG (SEQ ID No: 4)
opossum     WNEDKAAAILE    VSVIRVASGYLLML   LEPPCTKWTLSS    KPKAKVVVILLF     HKSFSSVKYVM     QRLVDADG (SEQ ID No: 5)
chicken     WNEDKAAAILE    VSVIRVASGYLLML   LEPPCTKWTLST    KPKAKVVVILLF     HRSFSNVTYVL     QRLVDADG (SEQ ID No: 6)
Xenopus     WNEDKAAAILE    VSVIRVASGYLLML   QCTPDSKWTLSS    KPKTKVAVILGF     HKSFVGVRYVL     QRLVDADG (SEQ ID No: 7)
zebrafish   WNEDKAAAILE    VSVIRVASGYLLML   LDSPYSRWTFAS    QSTTKVVVIFLF     HQRFGSVKYIL     QRLVSADG (SEQ ID No: 8)
fruitfly    WTQEKAAEVLN    PSALSIVIGVAVTV   -------FSLAT    RSWVKFLTVMGF     HDSFVRVPHVI     NRLVNSDG (SEQ ID No: 9)
nematode    WNETAAEQVLQ    FNYTIILAGYALML   -------WSLHS    KFASKVAIIVGC     RQSIGSSKYVI     IRLVDASG (SEQ ID No: 10)
```

Figure 5

TCTACACAAA CCAGATGAAC CTTCCAATCT CCTGCCTCGA GTATTGAAGC
CTGGCTACTG TGACTGTGGG GAAGGGATTA ATGGTCTCAG CATTCAGCCA
ACAACAATAC CTGCTCACTA TAAGCATTCA GAAAACAGAA AAGTTTCAAG
AAGCAGGAAG AAAAGACTCA CCTATGATCC CAACACCCAG AGATAAGAGT
CCTGAAGCTC AGATGACACA GCTGATAACA GGGAAGCCAG GACAGAATCT
CATTGTTTTG AACACCAAAA CCCGTTCCCT TGACAACTTG GCTATACTAC
ACTATTCGAA TGTTGCAGAT ACTGTGGTCA CATTTCAAAG GCCAGATCTT
TCCCAGGGCT TAAGCTGTTC CTTGGATACT TTTGGTAAGT CATTTATCCA
CTAATCATTT AGTAATCGTC TCTGACATGC CAAACACCCT GCTCAGGGCT
GGAAATGCAG AACCTGGGAA GCCACTGGCC TTGTCCTCAA GATCTCTCTC
TGGCTCCCTT TGAATTTGCT AATTCAGACT TTCACATTTC CCCCAGGAAA
AATCATAAGG ACCAAATCAT ATCCGTTTTC TCAAATGGCT TCAAAGACCC
ATGTCATCGT TTGGCATCAT GTAATTCTTT ACTGATGTAC TTTAAGAGTC
ACGTTTTATT CTCTTTATGC AGCTGTCAAG GACAGACACA AAGAGGGGGG
GGGNGGCCTT CCTCACTAAA TACTTTTCCC ACAACA (SEQ ID No:11)

FIGURE 6

ACAACTGCAG CGAGAGAAGA GGCTGGCAGC ATGGGTGGCA GGAGGCTTGG
CAGCCTCACA GGATGCCTGC AAATACCTTT CACTTATGCA GTTTGGCAGT
GCAGTGGTGC ATGGAGACAG CGTCTTGGGC CTGGCACCCA CAGtCACTTA
GGAAGTTGGA GTCCTGGAGA GGAGAACACA GAACGTGGAC AACTAGCTGT
CAAATTGCAG TAAAAGTTGG TTCAGGAAAA GTGGAAGCTA CCCAGATGTC
TACCAATTGA GATGAACCAT CCAATCTCCT GCCTCGAGTA TTGAAGCCTG
GCTACTGTGA CTGTGGGGAA GGGATTAATG GTCTCAGCAT TCAAAGCTTC
TATTCTGGAA TAGAACAGCT AGCATACTAC CAAGACTTTT CAAGGAGCAA
GAATGGAGCT CCCTGGAGAA CTGACTGAAC ATGGCTTCAG AGGCAGTATC
CATGTCACAT TTCAAAGGCC AGATCTTTCC CAGGGCTTAA GCTGTTCCTT
GGATACTTTT GCTGATGGTT TACACATCTT CTTCCCACAT TATATTGTAA
CTTTCTT (SEQ ID No:12)

FIGURE 7

ATTTTTAAAA AATATGCTGA ATTTGAAGTT TCTTTCAAAG TACAGTGTTT
CAATGGGGGG AGTCCAATTT TTGTAAAATT TTACAAAAAC TGTATTGCCC
TAAAGGCAGC CTACTGCACA CAAGGATCAC AGTGACTTTT ACTTGTTATT
CTACATGATT ACTTAAAATT TTTCTGATTT TTTTACCCTC ATCTATCTTC
TAACTTGTCT AGTTAACTCT TAAGAATTTC AAATTTTCTT TGAAAGATGA
TAGGCAATAT GAGATGAGAG ATAATCTACA AAAGTTACAG ATGCTCACAT
GTATAAAACA GTCAAAATAT CACAGGTCAA TGACATAAAC TGCATTAAAT
AAATTATGTT TATAGGCATC AGTAGTTGAA AATGCTCAAT AATTCTGGGC
TCCTTCCCCA AAATGTAAGA CTTAAGTACT TCAAAGGCAT TATTCTTTAC
TCATGAGGAT CAGTGGCTTC ATTTAGTAAA AGAAAAAGGA ATGGACCCAG
GATCCCAGTA AATAATTACT AACTGATCGC AACGCTCTTT TATCTAATGA
ACAACCAACA ACCAACAGAA AACCCTTGAT TCACAGAGGA GCAAGTCCTA
G (SEQ ID No:13)

FIGURE 8