American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 129B:1–9 (2004)

ProtocadherinX/Y, a Candidate Gene-Pair for Schizophrenia and Schizoaffective Disorder: A DHPLC Investigation of Genomic Sequence

Maria Giouzeli,[1]* Nic A. Williams,[1] Lorne J. Lonie,[2] Lynn E. DeLisi,[3] and Timothy J. Crow[1]

[1]*Department of Psychiatry, POWIC SANE Research Centre, University of Oxford, Warneford Hospital, Oxford, United Kingdom*
[2]*Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, United Kingdom*
[3]*New York University School of Medicine, New York, New York*

Protocadherin X and Protocadherin Y (PCDHX and PCDHY) are cell-surface adhesion molecules expressed predominantly in the brain. The *PCDHX/Y* gene-pair was generated by an X–Y translocation approximately 3 million years ago (MYA) that gave rise to the *Homo sapiens*-specific region of Xq21.3 and Yp11.2 homology. Genes within this region are expected to code for sexually dimorphic human characteristics, including, for example, cerebral asymmetry a dimension of variation that has been suggested is relevant to psychosis. We examined differences in patients with schizophrenic or schizoaffective psychosis in the genomic sequence of *PCDHX* and *PCDHY* in coding and adjacent intronic sequences using denaturing high performance liquid chromatography (DHPLC). Three coding variants were detected in *PCDHX* and two in *PCDHY*. However, neither the coding variants nor the intronic polymorphisms could be related to psychosis within families. Low sequence variation suggests selective pressure against sequence change in modern humans in contrast to the structural chromosomal and sequence changes including fixed X–Y differences that occurred in this region earlier in hominid evolution. Our findings exclude sequence variation in *PCDHX/Y* as relevant to the aetiology of psychosis. However, we note the unusual status of this region with respect to X-inactivation. Further investigation of the epigenetic control of *PCDHX/Y* in relation to psychosis is warranted.

KEY WORDS: psychosis; hominidae; laterality; language; sex chromosomes

INTRODUCTION

Schizophrenic illnesses occur in all populations with approximately 1% lifetime prevalence with a characteristic and sex-dependent distribution of ages of onset. This universality entails that the genetic variation accounting for the predisposition to the disorder is at least as old as the diaspora of modern *Homo sapiens* [Crow, 1999b]. The illness is associated with a fecundity disadvantage, persistence of which implies that there is a balancing advantage common to all human populations; it has been suggested [Crow, 1997] that "schizophrenia is the price *Homo sapiens* pays for language."

According to this view schizophrenia is a component of *sapiens*-specific variation that arose in the transition (the "speciation event") from a prior hominid species, for example, *Homo antecessor* [deCastro et al., 1997]. If the "speciation event" gave rise to language, then the genetic transition is predicted to be associated with a critical anatomical change in the brain. During the course of hominid evolution the brain became lateralized to allow some component of language (a component associated with word form) to localize to one hemisphere (90% of the population has language lateralized to the left). It appears that amongst extant species this cerebral asymmetry is *Homo sapiens*-specific, since lateralization and handedness at a population level is absent in our closest relative the chimpanzee [McGrew and Marchant, 1997; Buxhoeveden and Casanova, 2000].

It has been suggested [Crow, 1993] that the gene for cerebral asymmetry is on the X and Y chromosomes. This is supported by: a sexual dimorphism in asymmetry [Bear et al., 1986], greater relative hand skill (right > left) and verbal ability in females than in males [Crow et al., 1998] and non-dominant hemispheric (spatial) deficits in XO individuals and dominant hemispheric (verbal) deficits in XXY or XXX individuals. According to the rule that deviations in sex chromosome aneuploidies relate to genes that are also present on the Y chromosome (and therefore are protected from X-inactivation) these latter findings suggest that a gene for asymmetry is present in a region of X–Y homology. Consistent with this hypothesis is an association within families between handedness and sex [Corballis et al., 1996].

Some mammalian blocks of homology between X and Y chromosomes have arisen from reduplicative transpositions from the X to the Y in the course of mammalian evolution [Lambson et al., 1992]. After the separation of the lineages leading to the chimpanzees and hominids, a block from Xq21.3 was replicated on Yp11. This block was split by a subsequent paracentric inversion, with the consequence that there are two blocks of homologous sequences on Yp11 that are specific to *Homo sapiens*. These blocks, therefore, contain genes that have changed in the course of hominid evolution and these changes are candidates for speciation and cerebral asymmetry. The successive chromosomal changes presumably have been subject to species-specific selection on some gene or genes within the region. Only three genes in the Xq21.3/Yp11 block of homology have been characterized so far; *TGIFLX/Y*, which has so far only been detected in testis, *PABPC5*, which has been deleted from the Y chromosome and *ProtocadherinX/Y* (*PCDHX/Y*), which stands out because it is predominantly expressed in brain [Blanco et al., 2000] and has a coding

*Correspondence to: Maria Giouzeli, Department of Psychiatry, POWIC SANE Research Centre, University of Oxford, Warneford Hospital, Oxford, OX3 7JX, UK.
E-mail: maria.giouzeli@psychiatry.ox.ac.uk

Received 15 July 2003; Accepted 7 January 2004
DOI 10.1002/ajmg.b.30036

sequence that has been maintained (although subject to unusual selective pressures, Williams and Crow, in preparation) on both the X and the Y chromosomes. It must code for a protein product that has a function that has changed in the transition from a Great Ape precursor to *Homo sapiens*.

Protocadherins constitute a conserved subclass of the cadherin (calcium-dependent cell adhesion) superfamily, widely expressed in the brain and it has been suggested [Sano et al., 1993] that they are involved in cell–cell interactions in neural network formation. They are transmembrane proteins that can form Ca^{2+}-dependent homophilic interactions with high specificity [Hirano et al., 1999]. These homophilic interactions are weak, however, suggesting that protocadherins may not have a primary role in cell–cell adhesion [Sano et al., 1993] and instead determine the specificity of cell–cell interactions and signal transduction [Suzuki, 2000].

PCDHX and PCDHY consist of a C-terminus cytoplasmic domain, a transmembrane domain and an N-terminus extracellular segment comprising seven ectodomain modules (ED) with conserved calcium binding motifs [Nollet et al., 2000]. Alternative splicing of *PCDHX/Y* yields at least six *PCDHX* and three *PCDHY* transcripts. The *PCDHX* locus extends over approximately 844 kb and consists of at least 11 exons (10,071 bp), while that of *PCDHY* is approximately 738 kb long and consists of at least nine exons (9,960 bp) as it lacks exons 7 (81 bp) and 8 (30 bp) when compared to *PCDHX* and *PcdhX* from other Great Apes (Fig. 1). This results from a loss of 106 kb of sequence from the Y chromosome, which shortens the cytoplasmic domain of PCDHY and might, therefore, affect intracellular interactions. Size, positions and GenBank accession and GI numbers of the exons of *PCDHX* and *PCDHY* are shown in Tables I and II, respectively. Here we report a DHPLC investigation of the structure of *PCDHX/Y* in relation to psychosis.

MATERIALS AND METHODS

Genomic DNA from 99 families was used for the present study. These families were selected based on the availability of parental DNA from a larger cohort of families with multiple affected siblings previously recruited to participate in a series of international studies on the genetics of schizophrenia. Details of the recruitment and diagnostic procedures have been detailed in previous publications [DeLisi et al., 2002a,b]. Thus, the current cohort consisted of 71 USA, 14 UK, 12 Italian, and two Costa Rican families. These included 214 affected individuals (62 females and 152 males; 84 sibling-pairs: 44 male–male, nine female–female, and 34 mixed-sex; 11 sibling-trios: two all male, two all female, and seven mixed-sex; three sibling-quartets: two all male and one mixed-sex; one all male sibling-quintet) diagnosed with DSM-III-R schizophrenia or schizoaffective disorder and their parents. DNA from these 214 individuals was isolated from blood and amplified for the desired *PCDHX/Y* sequence by polymerase chain reaction (PCR). The PCR fragments were then analyzed by denaturing high performance liquid chromatography (DHPLC) and only those shown to have single nucleotide polymorphisms (SNPs) were sequenced in order to determine the position and type of SNP. The DNA from parents was reserved for further studies of transmission of alleles in families where sibling-pairs were concordant for a specific mutation.

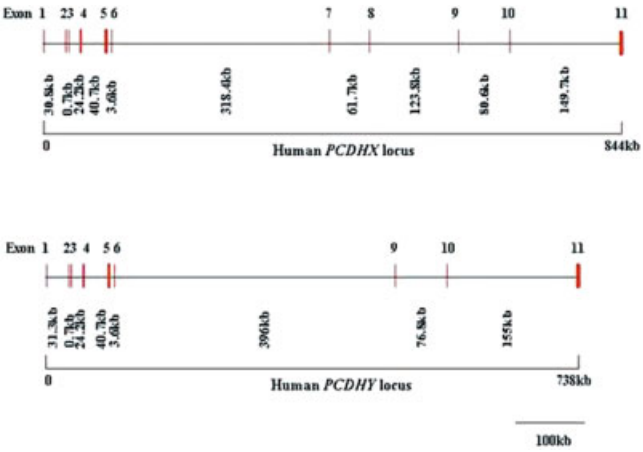

Fig. 1. Gene structure of Protocadherin X and Protocadherin Y (*PCDHX* and *PCDHY*). Exon positions are indicated and approximate intron sizes are shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Primer Design

Primers (Table III) were designed in order to yield amplicons 350–750 bp long and to amplify flanking sequences for full examination of the gene sequence. Considering that *PCDHX* and *Y* have known positions of sequence differences and that DHPLC detects any such positions of heteroduplex formation (as explained in the DHPLC section below), the primers had to be designed in order to specifically target either the *PCDHX* or *Y* sequence. This was achieved by designing the 3′-end of each primer at a position that shows an X–Y difference. A reference DNA sample was then amplified by X or Y specific primers at a PCR-gradient of various annealing temperatures (AT) in order to find the optimal temperature at which the maximum specificity and yield of either *PCDHX* or *Y* would be produced (Table III).

PCR

Genomic DNA was amplified by touchdown PCR using a combination of 9:1 AmpliTaq Gold (Applied Biosystems) and *Pfu* Turbo (Stratagene) to ensure the best available proof reading required for DHPLC. The PCR protocol comprised of initial denaturation at 95°C for 10 min to activate AmpliTaq Gold, 14 cycles of denaturation at 95°C for 30 sec, annealing for 30 sec at 7.5°C above AT with 0.5°C decrements per cycle and synthesis at 72°C for 30 sec, followed by 25 cycles at 95°C for 30 sec, AT for 30 sec, 72°C for 30 sec and a final extension for 7 min. Each 50 μl reaction contained 30 ng genomic DNA, 0.1 U/μl polymerase mix (AmpliTaq Gold and *Pfu* Turbo), 10 pM each of forward/reverse primers, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, and 0.2 mM dNTP mix. PCR yields were determined semi-quantitatively on ethidium bromide stained 1% agarose gels.

Sequencing

PCR fragments of the reference and test DNA sample were purified with QIAquick spin columns (Qiagen), mixed with primer and sequenced on a Perkin-Elmer ABI PRISM 3700 DANN Analyzer with Big Dye terminator. The sequences of each reference and test fragment were then compared to published sequences from the Human Genome Project and those available from GenBank.

DHPLC

PCR fragments were mixed with an equivalent fragment generated from the reference sample at a 3:1 ratio. The mix was then heated to 94°C for 4 min to melt the existing homoduplexes and then gradually re-annealed for 42 cycles starting from 95°C with 1.6°C decrements for 1 min each cycle resulting in homo- and hetero-duplexes. Each sample was

TABLE I. Size, Position, and GenBank Accession Numbers of *PCDHX* Exons

	Exon	Size (bp)	BAC accession number	GI number	Exon position in BAC
1	5′-UTR	380	AC004388	3046271	61318–61697
2	5′-UTR	106	AC004388	3046271	92486–92591
3	5′-UTR	59	AC004388	3046271	93250–93308
4	5′-UTR + coding	44 + 540 = 584	AC004388	3046271	117474–118057
5	Coding	2,493	AC004388	3046271	158794–161286
6	Coding + 3′-UTR	45 + 1,056 = 1,101	AC004388	3046271	164920–166020
7	Coding	81	AL133321	6933828	16010–16090
8	Coding	30	AL133321	6933828	77749–77778
9	Coding	198	AL133274	8217428	31665–31862
10	Coding	24	AL121869	7406494	16266–16289
11	Coding + 3′-UTR	678 + 4,288 = 4,966	AL121869	7406494	166231–166908

then run on the WAVE DNA Fragment Analysis System (Transgenomic). PCR products were injected by the autosampler at a constant flow rate of 0.9 ml/min and were eluted from the solid phase, which consisted of a DNASep analytical column (Transgenomic) with a binary gradient of 0.1 M triethylamonium acetate (TEAA) and 0.1 M TEAA/25% acetonitrile. Elution gradients and temperatures were automatically predicted by WAVEMaker software (Transgenomic). Partial denaturing analysis was conducted at the temperatures indicated in Table III. Sequences with SNPs (heteroduplexes) were detected in this way, the location and nature of which were then determined by automated fluorescent DNA sequencing of both strands.

Restriction Digestions

The restriction enzyme *Tsp*509 I (New England BioLabs) cuts at position 5′-AATT-3′ and we used this to confirm the presence of the two SNPs (G → T, ss5608044 and T → G, ss5608045, see Results) found in exon 5 of *PCDHY*. A 20 μl reaction consisting of 10 μl amplified DNA, 1 μl *Tsp*509 I, 2 μl 10× buffer (100 mM Tris Propane-HCl, 100 mM $MgCl_2$, 10 mM dithiothreitol, pH 7.0) and 7 μl H_2O was incubated at 65°C for 1 hr. The restriction digests were run on ethidium bromide stained 2% 3:1 Nusieve:agarose gels and bands were determined semi-quantitatively with HyperLadder V (Bioline). The absence or presence of the two SNPs produced different numbers and sizes of bands.

Statistical Analysis

Tajima proposed that neutrality of mutations could be tested by the difference 'D' between the estimates of θ based on the number of segregating sites (S) and the average number of pairwise differences (π) between the sequences in a population sample. Tajima's D value is predicted to be 0 under neutrality, while a negative value would suggest a selective sweep and a positive value a population subdivision.

The following assumptions were made for the neutral model: (a) a large constant diploid population (b) random mating (c) non-overlapping generations (d) no recombination and (e) an infinite sites, constant rate mutation where an offspring differs from its parent by a number of mutations [Simonsen et al., 1995]. Tajima's D test was applied for the coding regions of *PCDHX* and *PCDHY* for 214 and 152 sequences, respectively. As the population used is in equilibrium, Tajima's D-test need not be applied on the non-coding region of the gene [Tajima, 1989b]. Molecular evolutionary analysis was conducted using MEGA version 2.1 [Kumar et al., 2001].

RESULTS

DHPLC analysis of 4,192 bp coding and 3,544 bp flanking non-coding sequence of *PCDHX* and 4,035 bp coding and 3,050 bp flanking non-coding sequence of *PCDHY* gave the following results.

TABLE II. Size, Position, and GenBank Accession Numbers of *PCDHY* Exons (N/A = Not Applicable)

	Exon	Size (bp)	BAC accession number	GI number	Exon position in BAC
1	5′-UTR	380	AC010722	6587939	11187–11566
2	5′-UTR	106	AC010722	6587939	42867–42972
3	5′-UTR + coding	27 + 32 = 59	AC010722	6587939	43631–43689
4	Coding	571	AC010722	6587939	67850–68420
5	Coding	2,493	AC010722	6587939	109176–111668
6	Coding + 3′-UTR	18 + 1,083 = 1,101	AC010722	6587939	115305–116405
7	N/A	N/A	N/A	N/A	N/A
8	N/A	N/A	N/A	N/A	N/A
9	Coding	198	AC012067	7656648	147448–147645
10	Coding	24	AC012667	10801466	63858–63881
11	Coding + 3′-UTR	672 + 4,231 = 4,903	AC010081	7684533	144551–145222

TABLE III. Details of Primers Used in This Study [Primer Sequence (F, Forward; R, Reverse), Amplicon Size, Annealing (AT), and DHPLC Temperatures (DHPLC-T) of Each PCR Fragment]

Exon	PCR fragment	Sequence (5′–3′)	Size (bp)	AT (°C)	DHPLC-T (°C)
3	3X (F)	TTAGCAGACAAATAAGGATAGCAGAGAC	421	59	56, 58
	3X (R)	TTCAAACATGAAAAATTGGAGTTCAGAC			
	3Y (F)	TTAGCAGACAAATAAGGATAGCAGAGAG	421	59	56, 58
	3Y (R)	CTCAAACATGAAAAATTGGAGTTCAGAT			
4	41X (F)	CATAATATTTGATTCAGAACAA	410	51	56, 59, 61
	41X (R)	ACTAGCTTGAACTGCATAGC			
	41Y (F)	CATAATATTTGATTCAGAACAG	397	50	56, 58, 61
	41Y (R)	ACTAGCTTGAACTGCATAGT			
	42X (F)	AGAAAACGTCCTGATAGGCG	629	55	54, 57, 59
	42X (R)	ACCATGAAAATTCATAATCCACC			
	42Y (F)	AGAAAACGTCCTGATAGGCA	628	53	54, 57, 59
	42Y (R)	CCATGAAAATTCATAATCCACA			
5	51X (F)	AAATATCAAGAAAAGTAAAGTAACTTGT	448	52	51, 55, 58
	51X (R)	GTTGTCATTTGTATCAGTAACACTC			
	51Y (F)	AAATATCAAGAAAAGTAAAGTAACTTGC	448	54	51, 55, 58
	51Y (R)	GTTGTCATTTGTATCAGTAACACTT			
	52X (F)	GATAGGGAAGAGAAGGATACCTAC	662	55	58
	52X (R)	GACTCATAGTCAAGATATGCTGCAG			
	52Ya (F)	AGATAGGGAAGAGAAGGATACCTAT	611	57	57, 59
	52Ya (R)	CTGGCCTTAATCTGAAAGGA			
	52Yb (F)	CCAGTACTGCTATTTTGCAAGTA	596	57	58
	52Yb (R)	GGACTCATAGTCAAGATATGCTGCAT			
	53X (F)	CATGCTTCACAGATCATGAAATC	462	55	57, 59
	53X (R)	TTAAGGGTGGTACCCCG			
	53Y (F)	ATGCTTCACAGATCATGAAATT	461	50	57, 59
	53Y (R)	TTAAGGGTGGTACCCCA			
	54X (F)	ACCTGAATTCAGCCTGGATT	528	58	57
	54X (R)	AGAACCAATTCATAAGAACAGTTGG			
	54Y (F)	ACCTGAATTCAGCCTGGATC	530	57	57
	54Y (R)	GTAGAACCAATTCATAAGAATAGTTGT			
	55X (F)	CATTGTCCCTCCTTCCAACTG	369	59	59
	55X (R)	GGGGTCACTGGTGCTTCAG			
	55Y (F)	CAGTTTTCATTGTCCCTCCTTA	553	55	59
	55Y (R)	GCCCATTCAGAATTCTGCA			
	56X (F)	ATCTGTTCGTGAATGAGTCG	501	58	58, 60
	56X (R)	CCAAATCAGGGCTGTCG			
	56Y (F)	TCAATCTGTTCGTGAATGAGTCA	505	58	58, 60
	56Y (R)	GCCAAATCAGGGCTGTCA			
	57X (F)	AAGGCTGCTCAGAAAAACAA	489	55	58, 61
	57X (R)	CACAGGTACCTCGAAGGTC			
	57Y (F)	TTAAGGCTGCTCAGAAAAACAT	491	54	58, 60
	57Y (R)	CACAGGTACCTCGAAGGTT			
	58X (F)	GCCTGAAACTCCCCTGAATTC	532	55	55, 60
	58X (R)	CCATGACTATTCTGATCCTCATGCA			
	58Y (F)	CCTGAAACTCCCCTGAATTT	529	60	55
	58Y (R)	AGACTATTCTGATCCTCATGCC			
6	61X (F)	ACTATTGTAAAATTCAGTAACTTTGAC	717	50	54, 56
	61X (R)	TTTTCTCAGTAATATTATGTTAGTG			
	61Y (F)	GCTATTGTAAAATTCAGTAACTTTGAT	717	50	54, 56
	61Y (R)	TTTTCTCAGCAATATTATGTTAGTC			
7	7X (F)	CCCAGGCACTATTTTAGCTACC	318	60	55, 59
	7X (R)	CCAGCACAAAATCACCTGAA			
8	8X (F)	CATTGCACTGGGAATGTTGA	321	60	55
	8X (R)	TGAAAACAGCTCAGCTTGACTC			
9	9X (F)	CTGGCTGTGCCTTGTTTC	542	57	57, 62
	9X (R)	ATGTGTCACTATTCACAATAACATAATG			
	9Y(F)	CTGGCTGTGCCTTGTTTT	542	57	57, 62
	9Y (R)	ATGTGTCACTATTCACAATAACATAATA			
10	10X (F)	AAGGAGCAGCAGTCCTAGTG	311	61	54, 56
	10X (R)	GCAAACAGTTGATGTTTTATGAGA			
	10Y (F)	TGAAGGAGCAGCAGTCCTACTA	314	59	54, 56
	10Y (R)	AGCAAACAGTTGATGTTTTATGAGT			
11	11-1X (F)	CTTTTTTGTCACAATGATATTTAAGTG	598	60	60, 63
	11-1X (R)	ATCAACAGAGCATAGCCATC			
	11-1Y (F)	CTTTTTTGTCACAATGATATTTAAGTA	599	58	60, 63
	11-1Y (R)	GATCAACAGAGCATAGTCCATT			
	11-2X (F)	CAATCAGCCACAGCTCTC	572	60	55, 60

TABLE III. (*Continued*)

Exon	PCR fragment	Sequence (5′–3′)	Size (bp)	AT (°C)	DHPLC-T (°C)
	11-2X (R)	AAAAAATACTGTACATCAGAAATAGTG			
	11-2Y (F)	CAATCAGCCACAGCTCTT	573	54	55, 60
	11-2Y (R)	CAAAAAATACTGTACATCAGAAATAGTA			

DHPLC, denaturing high performance liquid chromatography.

PCDHX

In total, we detected four intronic SNPs, three SNPs in the *PCDHX* coding region that resulted in amino acid changes and one SNP in a 3′-UTR of a short transcript. The individuals with these changes are as follows. Three females (1 USA, 1 UK, and 1 Italy) and 18 males (12 USA and six Italy) had an intronic G → C change 131 bp 5′ to exon 3 (NCBI Assay ID: ss5608035). One female (UK) had an A → G change in exon 4 (NCBI Assay ID: ss5608036), which results in an isoleucine → valine amino acid change. Two females (USA) and four males (Italy) had an intronic G → C change 33 bp 5′ to exon 6 (NCBI Assay ID: ss5608037). Two females (USA) had a T → C change in the 3′-UTR in exon 6 (NCBI Assay ID: ss5608038). Two males (UK) had a G → A change in exon 7 (NCBI Assay ID: ss5608039), which results in an arginine → glutamine amino acid change. One female (UK) and four males (one USA, two UK, and one Italy) had an intronic T → G change 4 bp 3′ to exon 7 (NCBI Assay ID: ss5608040) and 15 females (seven USA, four UK, three Costa Rica, and one Italy) and 30 males (22 USA, three UK, and five Italy) had an intronic C → A change 13 bp 3′ to exon 7 (NCBI Assay ID: ss5608041) (Figs. 2 and 3: DHPLC and sequencing chromatograms). Six females (USA) and four males (USA) had a T → C change in exon 10 (NCBI Assay ID: ss5608042), which results in a leucine → proline amino acid change (Figs. 4 and 5: DHPLC and sequencing chromatograms).

PCDHY

We detected four intronic SNPs and two SNPs in the *PCDHY* coding region that resulted in amino acid changes. Four males (two USA and two Costa Rica) had an intronic G → T change 137 bp 3′ to exon 4 (NCBI Assay ID: ss5608043). Four males (two USA and two Italy) had an intronic T → G change 47 bp 3′ to exon 5 (NCBI Assay ID: ss5608046) and two males (USA) had an intronic C → T change 70 bp 3′ to exon 5 (NCBI Assay ID: ss5608047). Two males (USA) had an intronic A → G change 91 bp 5′ to exon 9 (NCBI Assay ID: ss5608048). DHPLC for fragment 57Y showed profiles for 69 males (54 USA, six UK, two Costa Rica, and seven Italy) (Fig. 6: DHPLC chromatogram). Twenty males were sequenced in order to determine the position, number and type of SNP/s within the fragment. Two SNPs were detected, 288 bp apart, in exon 5: a G → T (NCBI Assay ID: ss5608044) and a T → G change (NCBI Assay ID: ss5608045) (Figs. 7 and 8: sequencing chromatograms), which result in a valine → phenylalanine and an asparagine → lysine amino acid change, respectively, in the cytoplasmic region of *PCDHY*. Restriction digest of the PCR product from fragment 57Y from the remaining 49 males with *Tsp*509 I showed that they all had the above two SNPs, resulting in two *PCDHY* alleles: one with and one without the two SNPs. All the above results are summarized in Table IV.

Statistical Analysis

We further investigated the number and distribution of these newly identified SNPs using Tajima's D-test, which provides an indication of neutral evolution. We found three segregating sites in 214 *PCDHX* coding sequences and two segregating sites in 152 *PCDHY* coding sequences, which results in D values for the coding region of *PCDHX* and *PCDHY* of $D = -1.218$ ($\pi = 0.29 \times 10^{-4}$, $\theta = 1.25 \times 10^{-4}$, $SE = 0.32$) and $D = 2.559$ ($\pi = 2.49 \times 10^{-4}$, $\theta = 0.89 \times 10^{-4}$, $SE = 0.25$), respectively.

DISCUSSION

We examined sequence variation detected by DHPLC in the entire coding region and flanking intronic sequences of the

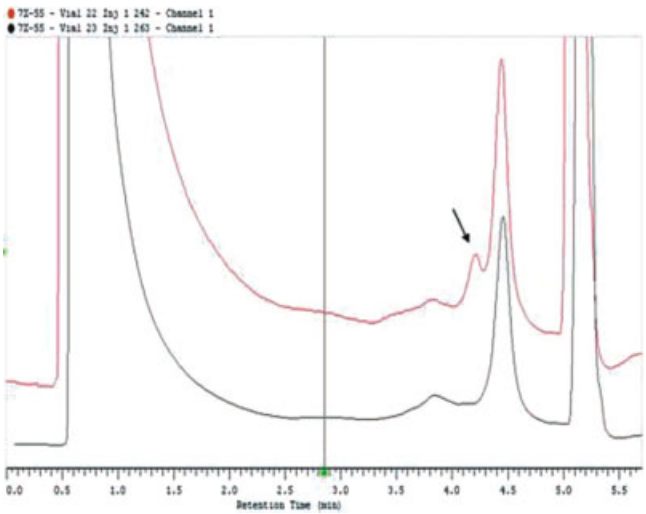

Fig. 2. Denaturing high performance liquid chromatography (DHPLC) chromatogram of fragment 7X. An arrow indicates the extra peak caused by the single nucleotide polymorphism (SNP) in the sample sequence (shown in red). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

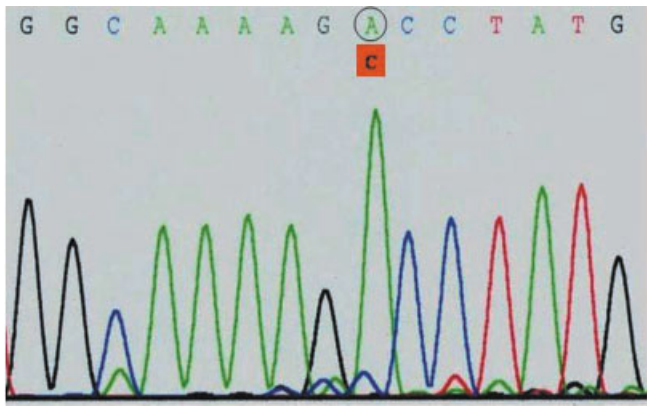

Fig. 3. C → A (ss5608041) change. Sequencing chromatogram of the C → A (ss5608041) change 3′ to exon 7 in *PCDHX* shown in Figure 2. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

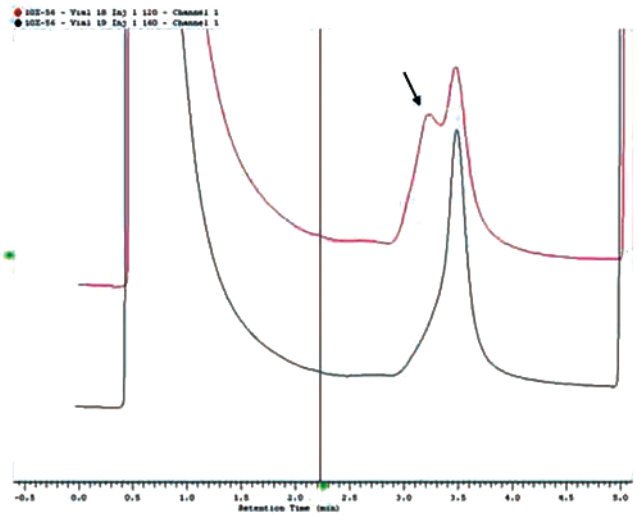

Fig. 4. DHPLC chromatogram of fragment 10X. An arrow indicates the extra peak caused by the SNP in the sample sequence (shown in red). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

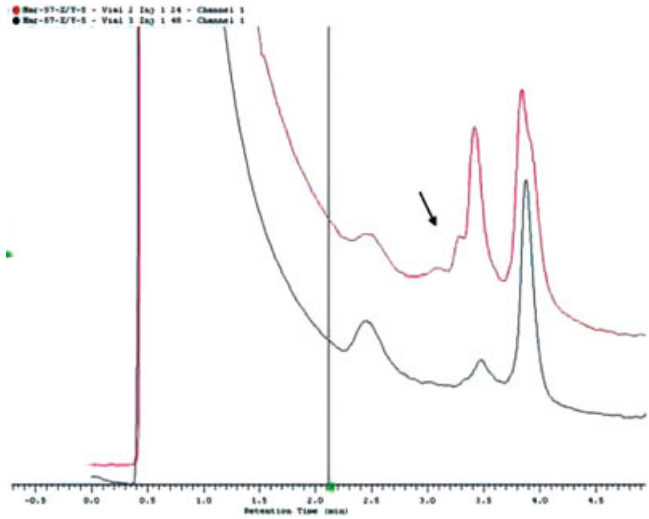

Fig. 6. DHPLC chromatogram of fragment 57Y. An arrow indicates the extra peak/s caused by the two SNPs in the sample sequence (shown in red). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

PCDHX and *PCDHY* genes, as summarized in Table IV. Eight sites of sequence variation were found in our study of *PCDHX*, four were from intronic sequences, one from 3′-UTR of a short transcript and the remaining three were located in the coding region and would result in an amino acid substitution (isoleucine → valine; arginine → glutamine; leucine → proproline). Six changes were detected in *PCDHY*, four intronic and two from the coding region that would also cause an amino acid substitution (valine → phenylalanine; asparagine → lysine).

Intronic base-pair changes can be of interest because they may influence gene expression and/or function. A significant proportion (10–15%) of disease-related mutations in human genes affect pre-mRNA splicing [Nissim-Rafinia and Kerem, 2002]. Such mutations can produce a pathogenic form of RNA splicing, resulting in exon skipping, intron retention, exon insertion, or partial exclusion of normal exons. For example, the ss5608041 (C → A) *PCDHX* SNP (found in one female and four males) located just 13 bp 3′ to exon 7 could disrupt exon splicing leading to aberrant splicing, frameshifts in the sequence, or an unstable RNA transcript leading to a nonfunctional polypeptide. The other intronic *PCDHX* and *PCDHY* sequence changes were located more distantly to known exons and are unlikely to affect exon splicing directly. Little is known of the elements controlling *PCDHX/Y* expression, however, so it is possible that these sequence variants and the 3′-UTR variant coincide with enhancer elements which affect gene expression.

We note that the remaining five SNPs detected in *PCDHX/Y* coding regions all occurred at non-synonymous sites thereby changing the encoded amino acid. Such sequence changes may have a variety of effects. Many proteins contain regulatory (allosteric) sites, which control their binding to other molecules and change their catalytic rate. Allosteric interactions are mediated by structural transitions that enable non-contiguous sites to communicate with each other and changes to such sites may block the protein conformational changes that are required for interaction with other proteins. Protein–protein interactions may also be disrupted by loss of a particular sequence motif required for recognition and binding.

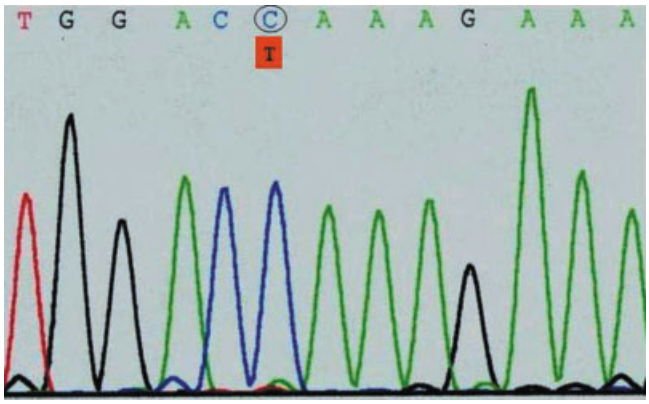

Fig. 5. T → C (ss5608042) change. Sequencing chromatogram of the T → C (ss5608042) change in exon 10 in *PCDHX* shown in Figure 4. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

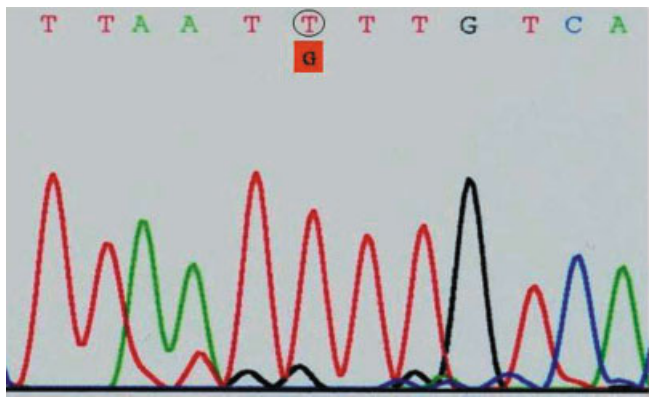

Fig. 7. G → T (ss5608044) change. Sequencing chromatogram of the G → T (ss5608044) change in exon 5 in *PCDHY* shown in Figure 6. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

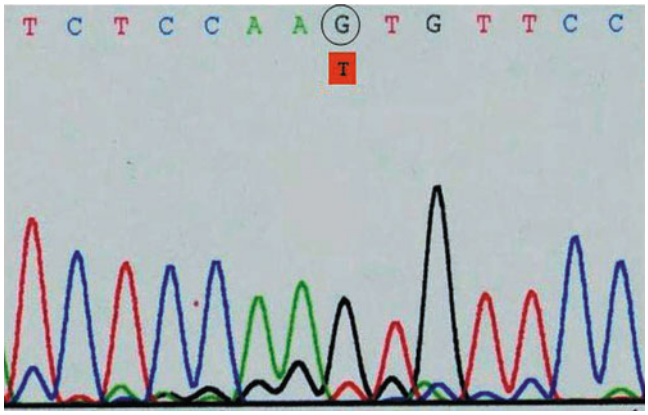

Fig. 8. T → G (ss5608045) change. Sequencing chromatogram of the T → G (ss5608045) change in exon 5 in *PCDHY* shown in Figure 6. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

One of the three coding changes in *PCDHX* is of particular interest in terms of the biochemical changes it could cause to the protein structure; the ss5608042 (T → C) *PCDHX* SNP (found in four males and six females including a male–male and a female–female sib pair and a mixed sex sibling pair) results in a radical amino acid change from leucine, an aliphatic non-polar hydrophobic amino acid that forms α helixes, to proline, an aliphatic non-polar hydrophobic amino acid that forms β turns and is a strong α helix breaker. Proline has no free amino group, resulting in a cyclic structure that influences the protein's architecture. Such a change may also disrupt a protein binding site located in this region of the cytoplasmic domain. The remaining two *PCDHX* coding SNPs (ss5608036 and ss5608039) cause conservative amino acid change and are unlikely to significantly alter protein structure although they may affect protein binding, either to an extracellular ligand or a cytoplasmic target, respectively. The first amino acid change was found in just a single female while the second was detected in males of a single sib-pair.

The two *PCDHY* coding SNPs that occur in the cytoplasmic domain encoded by exon 5 are located 287 bp apart and always occur together. The first of these SNPs, ss5608044 (G → T), results in an amino acid change from valine, an aliphatic non-polar hydrophobic amino acid that forms β strands, to phenylalanine, an aromatic non-polar hydrophobic amino acid that also forms β strands. This change from an aliphatic to an aromatic amino acid may affect the tertiary structure of the protein. The second SNP, ss5608045 (T → G), results in an amino acid change from asparagine, an amide polar hydrophilic amino acid that forms β turns, to lysine, a basic polar hydrophilic amino acid that forms α helixes. This change from an amide to a basic amino acid could affect the salt bridges that form between basic and acidic side chains. The combination of these two changes may, therefore, result in protein conformational change, affecting spatial availability of protein binding sites as well as changing the potential binding sites themselves. Sixty-nine males from our sample of 152 males possessed the above two amino acid changes, which always showed concordance between affected brothers. However, many of these males had affected sisters, none of which had a polymorphism at either of these sites (just eight sisters possessed other SNPs) in their *PCDHX* sequences, evidence against familial concordance for these variants with psychosis.

In summary, although we find mutations in *PCDHX* in males and females and *PCDHY* in males with schizophrenic or schizoaffective psychosis, minimal concordance was found between the variations and illness in siblings, leading us to

TABLE IV. Number of Individuals With SNPs and Position of SNPs in BACs (E/I, Exon/Overlapping Introns)

Gene	E/I	Base change	Amino acid change	Position at BAC	NCBI Assay ID	F	M	*
PCDHX	3	G → C	N/A	93120	ss5608035	3	18	i
	4	A → G	Ile → Val	117977	ss5608036	1	—	
	5	—	—	—	—	—	—	
	6	G → C	N/A	164887	ss5608037	2	4	ii
		T → C	N/A	165011	ss5608038	2	—	
	7	G → A	Arg → Gln	16029	ss5608039	—	2	iii
		T → G	N/A	16094	ss5608040	1	4	iv
		C → A	N/A	16103	ss5608041	15	30	v
	8	—	—	—	—	—	—	
	9	—	—	—	—	—	—	
	10	T → C	Leu → Pro	16282	ss5608042	6	4	vi
	11	—	—	—	—	—	—	
PCDHY	3	—	—	—	—	—	—	
	4	G → T	N/A	68557	ss5608043	—	4	
	5	G → T	Val → Phe	111288	ss5608044	—	69	
		T → G	Asn → Lys	111575	ss5608045			
		T → G	N/A	111714	ss5608046	—	4	
		C → T	N/A	111737	ss5608047	—	2	
	6	—	—	—	—	—	—	
	9	A → G	N/A	147358	ss5608048	—	2	
	10	—	—	—	—	—	—	
	11	—	—	—	—	—	—	

SNP, single nucleotide polymorphisms.
F, number of female individuals; M, number of male individuals, N/A, not applicable-intronic sequence.
*Most individuals showing a particular SNP for *PCDHX* were unrelated and only in the following cases was familial concordance observed; (i) two male–male sibling-pairs; (ii) two male–male sibling-pairs; (iii) one male–male sibling-pair; (iv) one male–male sibling-pair; (v) one female–female sibling-pair, one mixed-sex sibling-trio (two males and one female), seven male–male sibling-pairs, and one all male sibling-trio with two brothers having the SNP and one not; (vi) one female–female sibling-pair, one mixed-sex sibling-pair and one male–male sibling-pair. All SNPs for *PCDHY* are by necessity concordant in the above brother–brother pairs.

conclude that such sequence variation is not directly related to illness in our sample.

The relatively low number of sequence polymorphism in 4,192 bp of coding *PCDHX* and 4,035 bp of coding *PCDHY* sequence (three and two SNPs, respectively) suggests that *PCDHX/Y* is conserved across populations and is, therefore, under purifying selective pressure. The relative lack of variation across modern human populations contrasts with what we assume has been the case earlier in hominid evolution. Following the original translocation that occurred between 2 and 3 million years ago (MYA) a total of 26 changes (9 synonymous and 17 non-synonymous including a new stop codon) have accumulated in the Y coding sequence and 10 changes (four synonymous and six non-synonymous) in the X coding sequence (Williams and Crow, in preparation). We assume that these changes reflect differential selective pressures (that may have included a phase of sexual selection) preceding the origin of modern *Homo sapiens* that were fixed by selective sweeps at or before this event. In the same period (between the translocation and the origin of current populations) a number of local chromosomal rearrangements have taken place on the Y chromosome including a paracentric inversion and the deletion of DXS214 and exons 7 and 8 of *PCDHY*. The relationship between these rearrangements and the *PCDHX/Y* sequence differences is at present unclear.

To further investigate the evolution of the observed polymorphisms we employed Tajima's D-test. If the above polymorphisms are the result of mutations that occurred after a selective sweep, they arise on a background of reduced variation and their frequency will increase with increasing time after the sweep. Because π (the average number of pairwise differences) will be more affected by the sweep than S (the number of segregating sites), Tajima's D will be reduced below its neutral expectation of 0. Stronger selection results in a strong reduction of S, π, and D (Simonsen et al., 1995). The low number of S and π and the resulting negative D value in the coding region of *PCDHX* ($D = -1.218$ for a sample size of 214) are consistent with the prediction of a selective sweep, however, this D value is not significant at a confidence limit of 90% as a value between -1.539 and -1.752 is expected for a sample size of 200–250 [Tajima, 1989a]. Although this does not rule out the possibility of selective sweep/s (too ancient or more recent sweeps will not be identified using this technique), the observed *PCDHX* DNA polymorphism level in extant populations may also be explained by neutral mutation.

In contrast we calculated a positive D value of 2.559 for *PCDHY* using Tajima's test, which suggests that the population was subdivided resulting in a higher value of π relative to S than would be expected under neutrality. The more ancient the population the greater the Tajima's D value becomes. This population subdivision could explain the establishment of the two *PCDHY* alleles that result from the ss5608044 and ss5608045 SNPs.

Although, we found no evidence that variations in the gene sequence are related to psychosis we note that this gene has special status with respect to X-inactivation. Genes with homologues on the Y are protected from such inactivation. Therefore, while in other Great Ape species one copy is inactivated in females, this situation has changed with the establishment of *PCDHY* in the course of hominid evolution. The mechanism by which this protection is brought about is obscure and if as has been suggested [Burgoyne and McLaren, 1985] it is dependent on X–Y pairing in male meiosis then the final direction of the sequence in Yp and therefore, the paracentric inversion that succeeded the original translocation and determined this direction are important. On other grounds it has been suggested that the genetic predisposition to psychosis is conferred by modifications to the sequence that are "epigenetic" rather than intrinsic to the sequence itself [Crow, 1999a, 2002]. On account of its special status with respect to hominid evolution and its changing relationship to the X-inactivation process we suggest that the epigenetic status of *PCDHX/Y* deserves particular consideration with respect to psychosis and variations in cerebral asymmetry.

ELECTRONIC-DATABASE INFORMATION

http://www.ncbi.nlm.nih.gov/SNP/; NCBI Assay IDs: ss5608035, ss5608036, ss5608037, ss5608038, ss5608039, ss5608040, ss5608041, ss5608042, ss5608043, ss5608044, ss5608045, ss5608046, ss5608047, ss5608048.

ACKNOWLEDGMENTS

We thank Dr. Rekha Wadekar for her scientific advice, Simon Field for his help with the DHPLC analysis, and the T.J.Crow Psychosis Research Trust that supported this work.

REFERENCES

Bear D, Schiff D, Saver J, Greenberg M, Freeman R. 1986. Quantitative-analysis of cerebral asymmetries-fronto-occipital correlation, sexual dimorphism, and association with handedness. Arch Neurol 43:598–603.

Blanco P, Sargent CA, Boucher CA, Mitchell M, Affara NA. 2000. Conservation of PCDHX in mammals; expression of human *X/Y* genes predominantly in brain. Mamm Genome 11:906–914.

Burgoyne PS, McLaren A. 1985. Does X–Y pairing during male meiosis protect the paired region of the X-chromosome from subsequent X-inactivation? Hum Genet 70:82–83.

Buxhoeveden D, Casanova M. 2000. Comparative lateralisation patterns in the language area of human, chimpanzee, and rhesus monkey brains. Laterality 5:315–330.

Corballis MC, Lee K, McManus IC, Crow TJ. 1996. Location of the handedness gene on the X and Y chromosomes. Am J Med Genet 67: 50–52.

Crow TJ. 1993. Sexual selection, machiavellian intelligence, and the origins of psychosis. Lancet 342:594–598.

Crow TJ. 1997. Is schizophrenia the price that homo sapiens pays for language? Schizophr Res 28:127–141.

Crow TJ. 1999a. Schizophrenia as an epigenetic puzzle—in what sense were Gottesman and Shields correct? Mol Psychiatry 4:S18.

Crow TJ. 1999b. The case for an Xq21.3/Yp homologous locus in the evolution of language and the origins of psychosis. Acta Neuropsychiatrica 11: 54–56.

Crow TJ. 2002. Epigenetics of cerebral asymmetry as the determinant of disease expression. Schizophr Res 53:25.

Crow TJ, Crow LR, Done DJ, Leask S. 1998. Relative hand skill predicts academic ability: Global deficits at the point of hemispheric indecision. Neuropsychologia 36:1275–1282.

deCastro JMB, Arsuaga JL, Carbonell E, Rosas A, Martinez I, Mosquera M. 1997. A hominid from the lower Pleistocene of Atapuerca, Spain: Possible ancestor to Neanderthals and modern humans. Science 276:1392–1395.

DeLisi LE, Mesen A, Rodriguez C, Bertheau A, LaPrade B, Llach M, Riondet S, Razi K, Relja M, Byerley W, Sherrington R. 2002. Genome-wide scan for linkage to schizophrenia in a Spanish-origin cohort from Costa Rica. Am J Med Genet 114:497–508.

DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW, Wellman N, Loftus J, Nanthakumar B, Razi K, Stewart J, Comazzi M, Vita A, Heffner T, Sherrington R. 2002. A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. Am J Psychiatry 159:803–812.

Hirano S, Yan Q, Suzuki ST. 1999. Expression of a novel protocadherin, OL-protocadherin, in a subset of functional systems of the developing mouse brain. J Neurosci 19:995–1005.

Kumar S, Tamura K, Jakobsen IB, Nei M. 2001. MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245.

Lambson B, Affara NA, Mitchell M, Ferguson-Smith MA. 1992. Evolution of DNA-sequence homologies between the sex-chromosomes in primate species. Genomics 14:1032–1040.

McGrew WC, Marchant LF. 1997. On the other hand: Current issues in and meta-analysis of the behavioral laterality of hand function in nonhuman primates. Yearbook Phys Anthropol 40:201–232.

Nissim-Rafinia M, Kerem B. 2002. Splicing regulation as a potential genetic modifier. Trends Genet 18:123–127.

Nollet F, Kools P, van Roy F. 2000. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. J Mol Biol 299:551–572.

Sano K, Tanihara H, Heimark RL, Obata S, Davidson M, Stjohn T, Taketani S, Suzuki S. 1993. Protocadherins—a large family of cadherin-related molecules in central-nervous-system. EMBO J 12:2249–2256.

Simonsen KL, Churchill GA, Aquadro CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141:413–429.

Suzuki ST. 2000. Recent progress in protocadherin research. Exp Cell Res 261:13–18.

Tajima F. 1989a. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tajima F. 1989b. The effect of change in population-size on DNA polymorphism. Genetics 123:597–601.