Do the Hominid-Specific Regions of X–Y Homology Contain Candidate Genes Potentially Involved in a Critical Event Linked to Speciation?

CAROLE A. SARGENT, PATRICIA BLANCO & NABEEL A. AFFARA

Summary. It has been postulated that the critical events leading to major differences between humans and the great apes (such as language and lateralisation of the brain) are associated with major changes on sex chromosomes. Regions of homology between the human sex chromosomes have arisen at different points during mammalian evolution. The two largest blocks are specific to hominids, having appeared at some time after the divergence of humans and chimpanzees. These are the second pairing region found at the telomeres of the sex chromosome long arms and a region of homology between Xq21.3 (X chromosome long arm) and Yp11 (Y chromosome short arm). Questions arise as to whether (1) these regions of the sex chromosomes contain functional genes and (2) these genes might be candidates for the differences in cognitive function that distinguish modern humans from their ancestors. Furthermore, divergence between functional sequences on the X and the Y, and the alteration of the immediate environment around a locus by rearrangements, may allow for the acquisition and evolution of genes conferring more subtle sexually dimorphic variation.

INTRODUCTION

THE MAMMALIAN SEX chromosomes have evolved from an ancestral pair of autosomes (Ohno, 1967). The dominant male-determining gene evolved on the proto-Y, and became genetically sequestered from the rest of the genome through the suppression of recombination over most of the length of the proto-Y with the proto-X. In all mammals, a small region of strict X–Y

Proceedings of the British Academy, **106**, 231–250, © The British Academy 2002.

Copyright © British Academy 2002 – all rights reserved

homology has been maintained in order to permit pairing and correct segregation of the sex chromosomes during male gametogenesis. This segment is known as the pseudoautosomal region (PAR). However, the PAR shows considerable variation in gene content between different groups of species. The evolution of the sex chromosomes and the establishment of distinct PARs is believed to have occurred both by the acquisition of regions from autosomes, and the loss of material from the male-determining Y chromosome: the addition–attrition model of sex chromosome evolution (Graves, 1995). Within the non-recombining portions of the sex chromosomes, several other blocks of homology and X–Y homologous genes have been described. By studying the patterns of homology in different species, either by Southern hybridisation or fluorescence *in situ* hybridisation (FISH), the events leading to discrete blocks of sequence conservation can be reconstructed (Lambson *et al.*, 1992; Affara & Ferguson-Smith, 1994; Vogt *et al.*, 1997; Perry *et al.*, 1998; Graves *et al.*, 1998; Glaser *et al.*, 1999; Lahn & Page, 1999). Consequently, regions of X–Y homology defined on the modern human sex chromosomes may represent either the ancient remnants of the ancestral pair of autosomes, or reflect more recent exchanges of material. Figure 1 summarises our present understanding of the patterns of homology between the human sex chromosomes.

One of the consequences of a heterogamous method of sex determination is that females with an XX karyotype have twice the number of copies of X-linked genes as males with an XY karyotype. Hence those genes outside the PAR on the X chromosome are subject to inactivation on one of the X chromosomes in each female cell. The exceptions to this rule are genes found to be homologous between the X and Y chromosomes, where both copies are functional. These genes escape X inactivation and are, therefore, expressed in diploid dose in both males and females. Deficits of genes in this category lead to features of Turner's syndrome (XO females, and females and males with partial deletions of the X and Y), suggesting that they are required in diploid dose in both males and females (Ferguson-Smith, 1965). The association of cognitive deficits and male/female differences in cognitive function with sex chromosome aneuploidies has given rise to the idea that a gene or genes in the X–Y homologous category may be implicated in these aspects of brain phenotype (Crow, 1993). This proposition can be tested by searching X–Y homologous sequences for genes that are likely to have a function(s) within the brain.

There are two major homology blocks on the human sex chromosomes that must represent recent additions to the Y chromosome through transposition events, because they are not found on the Y in the great apes. These regions are the Xq21.3/Yp11 homology and part of the strictly homologous PAR 2 located at the tips of the X and Y chromosome long arms. The PAR 2 is known to contain a gene related to synaptobrevin (*SYBL1*) that may have a function in the

Copyright © British Academy 2002 – all rights reserved

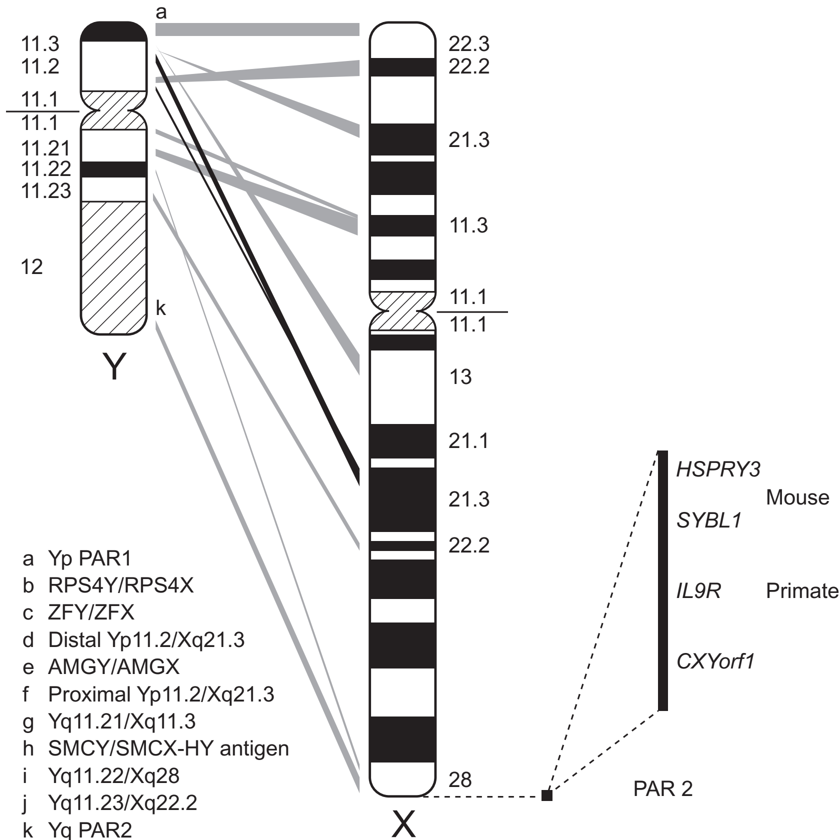

Figure 1. Homologies between the human X and Y chromosomes. Homologies between the sex chromosomes are highlighted and assigned according to the extent of conservation data available. Numerous single genes are shared on the X and Y chromosomes: for some of these genes the location within the non-recombining portion of the sex chromosomes is unique to humans, although these genes are X–Y homologous within the PAR 1 of other primate and non-primate species. An expansion of the PAR 2 is shown to the right. Homologues of *HSPRY3* and *SYBL1* are found in the mouse, *IL9R* has been added more recently in primate evolution, and *CXYorf1* is human specific. The diagram shown here is simplified, and a more detailed version can be found in Vogt *et al*. (1997).

Copyright © British Academy 2002 – all rights reserved

brain (Ciccodicola *et al.*, 2000). However, it is also worth noting that the gene is ubiquitous in its expression (D'Esposito *et al.*, 1996) and may have a role in other tissues. The synaptobrevin class of proteins is involved in synaptic vesicle docking and membrane transport and may also play a role in dorsoventral patterning in the nervous system. It has been shown that *SYBL1* is inactivated on both the inactive X and the Y chromosome (D'Esposito *et al.*, 1996), but its presence in the ancestral primate PAR 2 makes it a less attractive candidate gene. It is possible that mutations leading to reactivation of either gene may have an impact on brain phenotype as a result of inappropriate expression. The organisation and gene content of the PAR 2 is shown as an expanded inset in Figure 1.

The Xq21.3/Yp11 region is approximately 3.5 Mb of DNA and almost 10-fold larger than the PAR 2; it may therefore contain many more genes. In this chapter we discuss the structure and evolution of this block. The overall structure is well conserved between the human X and Y chromosomes, and the X chromosomes from different primates. Although the sequence data reveal high homology for the human X and Y, there are regions of significant divergence, presumably as the consequence of multiple rearrangements during evolution.

STRUCTURE OF THE Xq21.3/ Yp11 HOMOLOGY BLOCK

Previous work has established extensive yeast artificial chromosome (YAC) contigs from both the X and Y homology blocks, and has shown that the gross order of DNA markers is highly conserved (Vollrath *et al.*, 1992; Jones *et al.*, 1994; Sargent *et al.*, 1996; Mumm *et al.*, 1997). The major difference in the organisation of the sequences on the sex chromosomes is the isolation on the Y chromosome of the interval defined by DXYS34 and DXYS1 by means of a LINE (long interspersed nuclear element)-mediated recombination (Schwartz *et al.*, 1998). This results in a paracentric inversion leading to the juxtaposition of this block with the *AMELY* locus, creating a small (<300 kb) proximal Yp/Xq homology block and a larger (3 Mb) distal Yp/Xq block (DXYS42–DXYS31). Figure 2 compares the organisation of the homologous blocks on the X and the Y chromosomes.

Through the isolation of bacterial artificial chromosome (BAC), P1 artificial chromosome (PAC) and cosmid genomic clones from human male and human female DNA libraries, X-specific and Y-specific clone sets have been developed for sequencing by the genome centres as part of the Human Genome Project. Analysis of these sequence data allows a more detailed study of the level of identity between the chromosomes, and the identification of candidate genes. Comparison of non-repetitive DNA sequence from homologous

Copyright © British Academy 2002 – all rights reserved

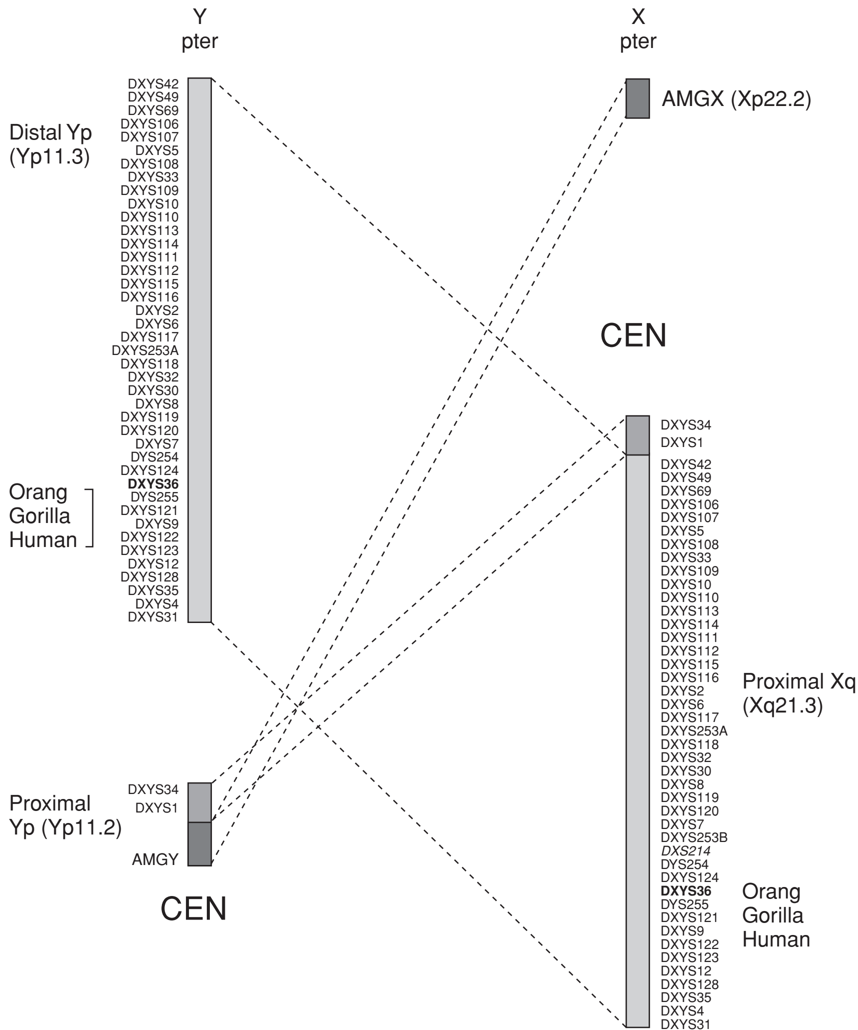

Figure 2. Transposition of the Xq21.3 block to Yp11. Diagram to show the transfer of the Xq21.3 block to the Y chromosome has occurred as a single event. Subsequent inversion of the region containing markers DXYS1 and DXYS34 have split the block on Yp11 into two. The proximal portion is sited alongside *AMGY*, a gene with an X homologue in Xp22.2. The centromeric to telomeric orientation of the Xq21.3/Yp block is reversed as a result of the transposition. DXS214 is shown in italics, and is conserved on the ancestral primate X chromosome. The marker DXYS36, in a bold text, is found on both the X and Y chromosomes of orang-utan, gorilla and human.

Copyright © British Academy 2002 – all rights reserved

segments of the X and Y gives a level of identity of between 98.8% and 99.1%. The G + C content of the Xq21.3/Yp11 block is low, with a range of 35–37%. This figure is similar for both the X and the Y chromosome, and does not currently show significant local variation, although this can only be fully assessed once the sequences have been fully completed and assembled. In addition, the percentage of repetitive DNA elements detected is high, with a range of 57–87% for different stretches of the homologous blocks. The majority of repeats fall into the LINE and/or LTR (long terminal repeat) category, with very few SINE (short interspersed nuclear element) elements. This high LINE and low Alu content, combined with the paucity of G + C, suggests that the X–Y homology block belongs to the GC-poorest subfraction of bulk genomic DNA, the L isochore. Generally, this class of DNA has a lower gene content than G + C-enriched isochores, and very few CpG islands (associated with the transcriptional start sites of many genes). In addition, most genes found in regions of low G + C content have a larger than average size, and tend to be tissue-specific genes expressed at low abundance rather than ubiquitous transcripts (reviewed by Bernardi, 2000).

The general composition of the block is reflected in the analysis of coding potential. Although expressed sequence tag (EST; sequences derived from cloned fragments of RNA molecules) matches have been defined through BLAST (basic local alignment search tool) searches, it is unclear how many of these represent fragments of true X–Y homologous genes, and how many are pseudogenes. This is also true for the gene structures predicted from computer algorithms. Of the confirmed pseudogenes identified the vast majority are uninterrupted, reminiscent of retroposition. To date, only three regions corresponding to clusters of cDNA clone sequences have been identified that potentially represent functional genes. One of these (a protocadherin gene expressed in the brain) has been analysed in detail (see below) and is a promising candidate gene for brain lateralisation. Figure 3 summarises the current information on the gene content of this homology block.

From a comparison of the sequence derived from the X and Y chromosomes, there appears to be one major block of divergence covering some 200 kb of DNA. This is characterised by the marker DXS214 on the X, and NR11 on the Y. It was unclear from initial experiments whether the differences between the human sex chromosomes were either the result of rearrangement between the flanking X–Y homologous markers or the consequence of sequence divergence. More detailed analysis of the interval shows that the pattern of X–Y homology/ divergence observed is complex, and probably reflects more than one rearrangement during evolution. However, through FISH and other molecular techniques such as polymerase chain reaction (PCR) amplification of DNA, we have found that different ancestral Y chromosome types show the same marker distribution: i.e. we would predict that there are no major

Copyright © British Academy 2002 – all rights reserved

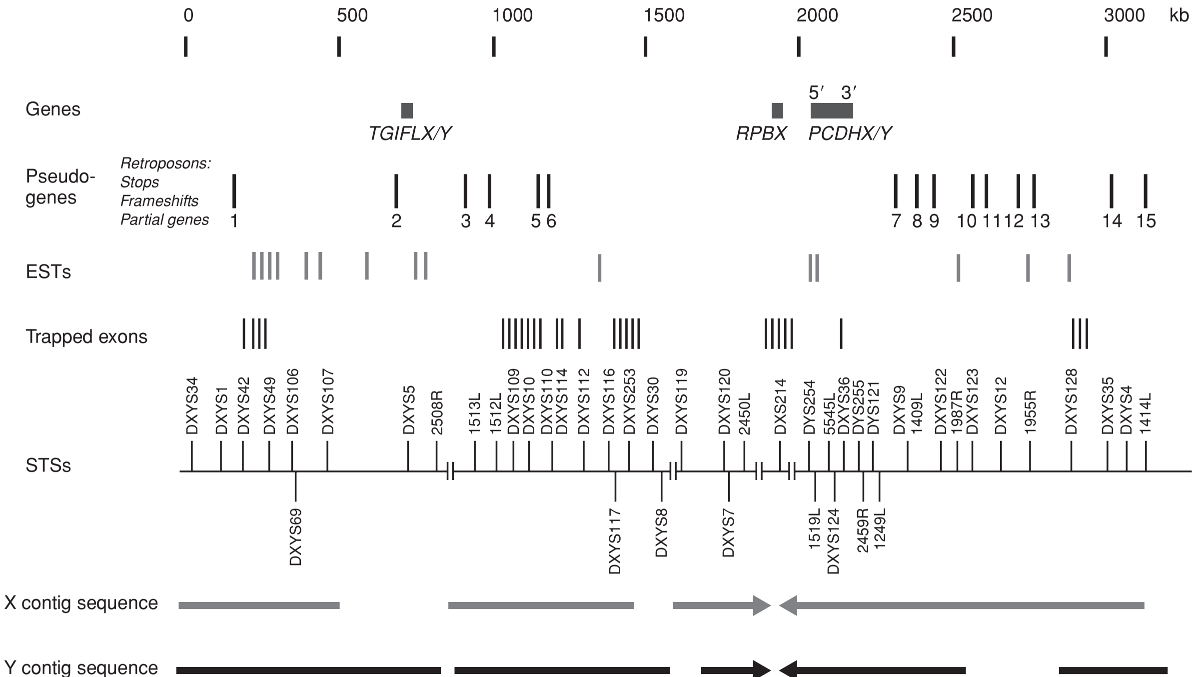

Figure 3. Coding potential of the Xq21.3/Yp11 homology block. This figure summarises the data from exon trapping experiments, and sequence analysis using the NIX program at HGMP. The extent of available X and Y genomic clones in a near completed state is shown below the STS (marker) content: the arrows indicate that the sequence extends into regions of X or Y unique sequence. Pseudogenes are 1, Sorcin; 2, Polypro dUTPase; 3, Staufen; 4, *UBH1*; 5, *CROC1B*; 6, XY body protein; 7, *VDAC1*; 8, *SAP62*; 9, *EIF4A1*; 10, *ZNF127*; 11, Transposon 10; 12, Keratin 18; 13, *SNX3*; 14, *RPL26*; 15, *FUS1*. The only EST matches that appear to be representative of genes to date are those for *TGIF*-like (*TGIFLX*/*Y*), an X-linked ribosomal binding domain protein (*RBPX*), and the protocadherin genes (*PCDHX*/*Y*).

Copyright © British Academy 2002 – all rights reserved

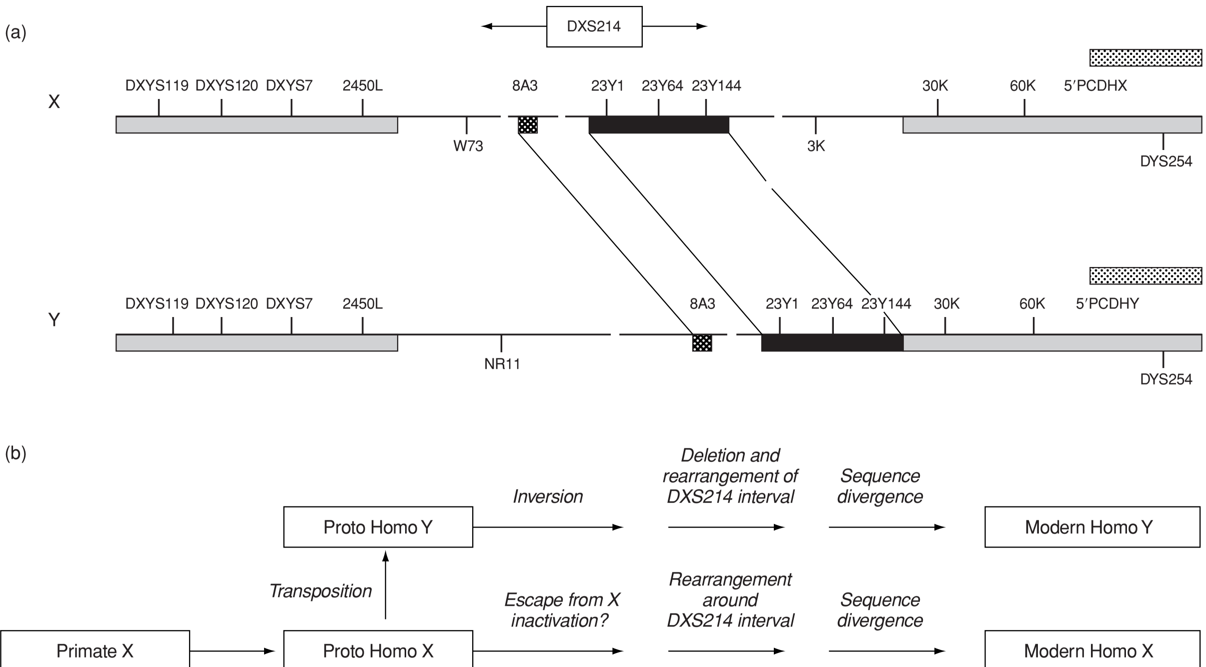

Figure 4. The human sex chromosome specific interval between DXYS7 and DYS254. The diagram in (a) shows our understanding of the X and Y chromosome intervals, with the new STS markers used to define regions of homology or specificity. The previously determined regions of homology are defined by the shaded boxes; new homology is defined by the solid box and the single exon 8A3 (white dots on black); the location of the 5′ ends of the protocadherin genes is shown by the black dots on white. The exact order of sequences in the central portion is not known (broken lines). DXS214 lies on the X chromosome between W73 and 3K. In (b), a proposed set of events leading to the current mosaic structure of this interval is outlined. The precise order and date of each event is unknown.

Copyright © British Academy 2002 – all rights reserved

variations of the interval in modern *Homo sapiens* populations. Figure 4a shows the organisation of this segment and demonstrates the mosaic interspersion of X–Y homologous and X- or Y-specific sequences.

A second region of potential sequence divergence lies between the markers DXYS5 and DXYS109, as defined by the divergence of the DNA sequence DXYS108 between the human sex chromosomes (Mumm *et al.*, 1997). As yet, the comparative sequence data for both the X and Y are not available for a more detailed commentary on possible candidate genes.

CONSERVATION OF THE HUMAN XQ21.3 INTERVAL IN THE GREAT APES

Previous failure to find homology to the probe DXS214 in primates by Southern blotting led to the supposition that this marker was representative of an insertion on the X chromosome that occurred after the transposition of the homologous block to the human Y chromosome. However, as we (Vogt *et al.*, 1997), and others (Vacca *et al.*, 1999), have demonstrated by FISH, these X-specific sequences occur on the ancestral primate X chromosome. Thus substantial homology to this DNA segment has been lost from the human Y either through chromosomal rearrangements or by sequence divergence.

For a more detailed study, genomic clones spanning the Xq21.3 interval were selected for FISH analysis of male and female normal human, male chimpanzee (*Pan troglodytes*), female gorilla (*Gorilla gorilla*), male orang-utan (*Pongo pygmaeus*), female gibbon (*Hylobates lar*) and male and female lemur (*Eulemur macaco macaco*). In the human male signals were observed on the long arm of the human X chromosome, and the short arm of the Y chromosome for all the clones, with the exception of a clone containing DXS214. The signal for this clone was restricted to the human X chromosome, as expected. In the primates and the lemur lines, all hybridisation signals were observed on the long arms of the X chromosome. With a two-colour FISH approach the results suggest that the region from DXYS42 to DXYS31, including the DXS214 interval, has a very similar gross structural arrangement between the great apes.

Additional results from Southern restriction fragment length polymorphism (RFLP) (Lambson *et al.*, 1992) and sequencing of genomic DNA allow estimation of the level of identity with the human sequence as *c*. 97% for gorilla and *c.* 98% for chimpanzee. This is concordant with other data for the evolutionary distance between these species, and supports previous claims that the transfer of the Xq21 block to the Y chromosome in hominid evolution occurred after the divergence of the great apes (Page *et al.*, 1984; Lambson *et al.*, 1992).

A schematic view of possible events leading to the arrangement of the sex chromosomes in modern humans is shown in Figure 4b.

Copyright © British Academy 2002 – all rights reserved

GENE CONTENT OF THE Xq21.3/Yp11 BLOCK

As noted above, only three potential candidate genes for brain lateralisation have been identified within this region of the X chromosome: (1) a homeodomain-related transcribed sequence (*TGIF*-like), (2) a transcribed sequence containing four RNA binding motifs that is expressed predominantly in brain and testis, and (3) a gene (*PCDHX*) belonging to the protocadherin family, expressed predominantly in the brain (Yoshida & Sugano, 1999; Blanco *et al.*, 2000). For the *TGIF*-like transcript, the functional status is unclear, as analysis is incomplete. The RNA-binding domain transcript does not have a functional Y homologue in humans, but encodes a potentially functional protein from the X locus. The absence of a Y homologue makes this gene a less favourable candidate, despite expression within the brain. The third gene appears to be a promising candidate in several respects. The X homologue has an open reading frame; it has a closely related Y homologue also with an open reading frame; both genes are predominantly expressed within different regions of fetal and adult brain, but are virtually undetectable in mRNA from other tissues. This is consistent with the general observation that protocadherins are predominantly expressed in the brain (Sano *et al.*, 1993). Redies & Takeichi (1996) have argued forcefully that these genes play a fundamental role in the cell–cell recognition essential for the segmental development and function of the central nervous system. Thus the *PCDHX* and *PCDHY* genes potentially have a role not only in the establishment of neuronal networks in the developing brain, but also in the maintenance of functions in adult structures.

THE PCDHX AND PCDHY GENES

General structure and classification of the genes

Figure 5 illustrates the predicted structures of the X- and Y-linked protocadherin genes and their products, as determined from nucleotide sequence analysis. The proteins conform closely to the general organisation of other members of the protocadherin subfamily: a signal peptide to permit secretion of the protein; seven extracellular cadherin motifs that mediate cell interactions; a transmembrane domain achieving insertion into the cell membrane; and a cytoplasmic domain that interacts with intracellular proteins. In addition to the arrangement of the protein domains, the gene organisation of *PCDHX*/*PCDHY* strongly supports their inclusion in the protocadherin gene subfamily. Typically, the protocadherin subfamily genes consist of four to five exons, with most of the coding region contained in a single large exon (Wu & Maniatis, 1999). Similarly, the majority of the open reading frame of *PCDHX*/*PCDHY* is encoded in exon 5.

Copyright © British Academy 2002 – all rights reserved

Three further features of *PCDHX*/*PCDHY* indicate that they belong to a specific subgroup of the protocadherins known as the CNR (cadherin-related neuronal receptor) genes. First, the cytoplasmic region of *PCDHX*/*PCDHY* contains a lysine-rich segment that is characteristic of the CNR genes (Kohmura *et al*., 1998; Wu & Maniatis, 1999) and may be important for interactions with other proteins within the cell. Secondly, the sequence of the cadherin domains of *PCDHX* and *PCDHY* is most similar to those found in the mouse CNR genes encoded on chromosome 18, and the cluster of human CNR genes on chromosome 5q31 (Kohmura *et al*., 1998; Wu & Maniatis, 1999; Sugino *et al*., 2000). Thirdly, like the CNR genes, the first cadherin domain of *PCDHX* and *PCDHY* possesses a variant (TGD) RGD (putative integrin binding domain) motif. Proteins of the CNR protocadherin subgroup show enriched or specific expression in the brain and subpopulations of synapses. In the mouse, the extracellular domain of CNR proteins has been found to associate with secreted Reelin proteins (Senzaki *et al*., 1999), while the cytoplasmic region has been shown to interact with Fyn, a non-receptor tyrosine kinase (Kohmura *et al*., 1998). It is thought that these molecular associations may act as a positive selection mechanism during development of the cortex, and by triggering secondary signalling pathways may also regulate the morphogenesis of the neuronal cells. It is interesting to note that mice carrying mutations of Fyn show impaired long-term potentiation, spatial learning and hippocampal development (Grant *et al*., 1992).

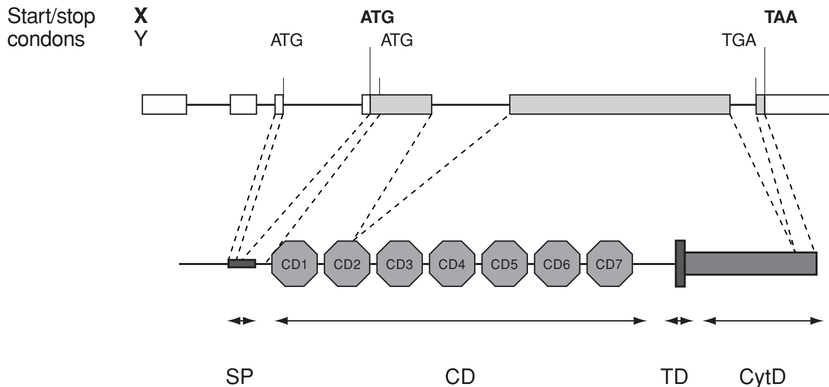

Figure 5. Comparison of the protocadherin genes. The genomic structure of *PCDHX* and *PCDHY* consists of at least six exons; only three of them contain the coding sequence (shaded). The different start and stop codons of both genes are highlighted above. The corresponding protein structure is depicted below. SP, signal peptide; CD, cadherin domains; TD, transmembrane domain; CytD, cytoplasmic domain.

Copyright © British Academy 2002 – all rights reserved

Comparison of the X and Y gene sequences

Detailed sequence analysis of the X and Y genes has shown that they share some 98% identity at the nucleotide level over their respective open reading frames (Table 1). This suggests that they may possess closely related functions. However, examination of the coding regions reveals two notable differences between the X and Y transcripts. (1) The Y gene has a 13-base pair deletion that removes the region equivalent to the initiator methionine of the X gene and the first amino acid (aspartic acid) of the putative signal peptide. A single base change in an upstream exon of the Y gene creates a new initiator methionine in-frame with the remaining open reading frame, and retains an enlarged signal peptide sequence. This initiator methionine has a weaker match than the X gene initiator to the Kozak consensus. There is a second initiator methionine with a better Kozak consensus further downstream, but if this were used there would be no signal peptide in the resulting protein, and it would not be correctly transported to the intracellular apparatus required for secretion and insertion into the cell membrane. However, it is possible that both initiator codons are used in the translation of Y transcripts and may represent a means of regulating the level of Y protein that is capable of being secreted. (2) The Y gene encodes a cytoplasmic domain that stops nine amino acids before the termination of the *PCDHX* polypeptide. This nine-amino acid difference may be very significant in terms of differentiation of function of the X and Y genes. Currently, we do not know if the respective cytoplasmic domains can interact with different cytoplasmic proteins, and whether the nine residues represent part or the whole of a specific binding domain. If these genes are involved in inter- or intracellular signalling (as are other CNR family genes), then it is possible that unique X-only or Y-only interactions will influence distinct pathways.

Expression of *PCDHX/PCDHY*

The expression of *PCDHX* and *PCDHY* was examined using reverse transcriptase (RT)-PCR with a range of tissue-specific mRNAs. The PCR primers were designed to amplify across exons 2–4 and produce a fragment from both the X and Y transcripts. From the tissue data, expression is predominantly in brain mRNA and varies from one region of the brain to the other. The protocadherin transcripts are found in fetal brain, adult cortex amygdala, hippocampus, caudate nucleus, corpus callosum, substantia nigra and thalamus and are barely detectable in the adult cerebellum. Transcription cannot be detected in mRNAs from other adult tissues (heart, liver, kidney and skeletal muscle) except at a low level in testis.

Copyright © British Academy 2002 – all rights reserved

Table 1. *PCDHX* and *PCDHY* gene structure

	Size (bp)		
	X	Y	Identity (%)
EXON 1	380	380	99
Intron 1	30,788	31,300	96.7
EXON 2	106	106	97
Intron 2	658	658	98.9
EXON 3	59	59	96
Intron 3	24,165	24,160	96.5
EXON 4	584	571	98.8
Intron 4	40,736	40,755	97.1
EXON 5	2493	2493	99.1
Intron 5	3633	3636	98.9
EXON 6	1100	1100	98.8

Using a two-step nested PCR approach that allows detection of mRNA present at very low levels, the X or Y origin of the transcripts was investigated in all the tissues. The PCR primers were designed to amplify across a coding region that contains a restriction enzyme site difference between the X- and Y-linked sequences. This allows an estimate of the relative transcription levels from the X- and Y-linked genes. The mRNA samples used in the study comprised pools from male and female individuals, with the exceptions of whole brain, heart, liver and testis (all male only). Both X and Y transcripts are evident in the subregions of the brain, except cerebellum (predominantly *PCDHX*) and the heart, whereas kidney, liver, muscle and testis are predominantly *PCDHY* (Table 2).

Additional studies into the regulation of *PCDHX*/*PCDHY* gene expression were carried out using the testis tumour-derived pluripotential cell line NTERA. This cell line can be induced with retinoic acid to differentiate along the spermatogenic pathway and into neuronal cells. The NTERA cells were cultured either in the presence of retinoic acid, hydrocortisone or DMSO, or left in an untreated state. The hydrocortisone is used to mimic the effect of steroid hormones upon the cells and, as the retinoic acid is prepared in DMSO solution, DMSO-treated cells act as an experimental control. The nested set of primers was used to amplify the relatively low level of gene expression in NTERA cells as described above. In the uninduced state and with the DMSO-treated cells, only the X-linked transcripts are detectable, whereas hydrocortisone-treated cells contain both X and Y transcripts at similar levels. In contrast, treatment with retinoic acid increases the level of Y-linked transcripts and appears to depress the X-linked mRNA. These data seem to indicate that the Y-linked gene is regulated differently to the X-linked gene at the mRNA

Copyright © British Academy 2002 – all rights reserved

Table 2. Summary of expression data for *PCDHX* and *PCDHY*

	PCDHX	*PCDHY*
Amygdala	+	+
Caudate nucleus	+	+
Cerebellum	+	−
Corpus callosum	+	+
Hypocampus	+	+
Substantia nigra	+	+
Thalamus	+	+
Total brain	+	+
Fetal brain	+	+
Heart	+	+
Liver	−	+
Kidney	−	+
Muscle	−	+
Testis	+	+

level, and although both of the promoters may contain retinoic acid response elements the effects are modified by other transcriptional control elements.

Currently the promoters of these genes have not been isolated or characterised. However, the 5′ ends of both the X and Y genes lie close to the boundary between the region of strict X–Y homology and the X or Y chromosome-specific regions. The sequences that have been inserted or deleted upstream of the translational start sites are likely to have an influence upon the promoter structures, and the position of long-range elements such as enhancers.

Conservation of *PCDHX*

cDNA probes for *PCDHX* were hybridised to Southern blot filters of male and female pairs of DNA from a range of different species. No bands were detected in kangaroo (representative of metatherian mammals) or in the armadillo, mouse or rat. However, sex-linked dosage was seen for rabbit, cow, goat, sheep, pig, horse, lemur, a New World monkey (squirrel monkey) and two Old World monkeys (rhesus macaque and stump-tailed macaque), suggesting that this gene is only on the X chromosome of these species (Table 3). Previous blotting and hybridisation studies used a set of probes from across the Xq21/Yp11 homology block (Lambson *et al.*, 1992). The observation that all of these markers, with the exception of DXYS36 (contained within an intron of the *PCDHX*/*PCDHY* genes), are confined to the X chromosome in primates and monkeys but appear absent from rodents is entirely consistent with the pattern seen for the *PCDHX* cDNA. Probe DXYS36, in contrast, is X–Y homologous

Copyright © British Academy 2002 – all rights reserved

in higher primates (orang-utan, gorilla and human) but not in lower primates or chimpanzee. This raises the question as to whether there is a functional *PCDHY* gene copy on the Y chromosome of these additional species of great apes. So far, there is no conclusive evidence of a Y gene from PCR and sequence analysis of non-human male primate DNA: sequences do not differ from those of female DNA. Either the transfer of material encompassing DXYS36 does not include a complete protocadherin gene, or the X and Y copies are identical, or the Y copy has diverged significantly from the X copy such that it no longer amplifies with the primers used in these experiments. Additional studies are in progress to resolve this issue.

Does *PCDHX/PCDHY* fit the criteria for a gene involved in brain lateralisation and/or language development?

The current hypothesis under investigation is that a gene (or genes) on the human sex chromosomes is important for brain lateralisation, and that this is fundamental for language development. It is also predicted that relative hand skill is influenced in part by the dominant hemisphere. The evidence for this is summarised as follows.

1. Brain asymmetry and directional handedness are features not observed in chimpanzees (Marchant & McGrew, 1996; Buxhoeveden & Casanova, 2000)
2. The cognitive deficits seen in individuals with sex chromosome aneuploidies can be correlated with deficits in brain lateralisation, and

Table 3. Conservation of *PCDHX/Y*

		PCDHX	*PCDHY*
Hominoid	Human	+	+
Old World	Rhesus monkey	+	–
	Stamp-tailed macaque	+	–
New World	Squirrel monkey	+	–
	Marmoset	+	–
Prosimian	Lemur	+	–
Carnivore	Dog	+	–
Artiodactyl	Cow	+	–
	Goat	+	–
	Sheep	+	–
	Pig	+	–
	Horse	+	–
	Rabbit	+	–
Rodent	Mouse	–	–
	Rat	–	–
Marsupial	Kangaroo	–	–

Copyright © British Academy 2002 – all rights reserved

suggest that a gene for relative hemispheric development is located on the X chromosome. Thus, XO patients have a relative deficit of the non-dominant hemisphere, and patients with an excess of the X chromosome (XXX, XXY) have a relative deficit of the dominant hemisphere. Males do not have hemispheric deficits equivalent to those in XO patients and, therefore, the gene must fall into the X–Y homologous category (Crow, 1993).

3 Although the gene(s) is X–Y homologous, there are sex-related differences in brain development and verbal ability, with females maturing faster than males. Individuals of either sex who show equal hand skill are delayed in developing language skills, including reading (Crow *et al.*, 1998).

To fulfil the requirements demanded of the above hypothesis, any potential candidate gene must therefore: (1) be found on the human X chromosome; (2) have a homologue on the human Y chromosome; (3) show appropriate spatial and temporal patterns of expression during embryonic and post-natal development to support a role in lateralisation; (4) have acquired sufficient differences in the nucleotide sequence of the open reading frame, or in the mode of transcriptional control, to distinguish it from the homologous gene in other great apes.

The criteria defined by the first two points are met by the protocadherin genes, but also by *SYBL1* and by the partially characterised *TGIF*-like locus. However, *SYBL1* does not have an expressed Y homologue (see below). So far, in relation to expression patterns, we know that *PCDHX* and *PCDHY* transcripts are both expressed in the fetal brain during the second half of pregnancy (mRNA sample pool 21–30 weeks). Currently, we do not have detailed information about the precise onset of gene expression, or the cell types involved. The *TGIF*-like gene is not expressed in any of the brain samples investigated to date, and thus would appear to be a less attractive candidate.

The presence of a candidate cDNA at the appropriate point in development must be validated with protein data, showing that the transcripts are actually translated to give rise to functional proteins. These proteins need to have a proven role in the initiation of asymmetric development. Recently, experiments designed to dissect the pathways that establish left–right asymmetry in the developing chick embryo identified another gene of the cadherin superfamily, N-cadherin, as a key early but transiently expressed molecule required for normal morphogenesis (Garcia-Castro *et al.*, 2000).

With respect to those changes that define the unique status of *H. sapiens*, it would be useful to compare the expression patterns of candidate genes in modern humans with those of our closest extant relatives, such as the chimpanzee. Differences between the species could then be related either to variations in the

Copyright © British Academy 2002 – all rights reserved

potential protein product encoded by the genes, or to the alterations in the sequences involved in the control of mRNA production both at the temporal and spatial levels. To this end, it is of added importance to define the promoter regions of the human *PCDHX* and *PCDHY* genes, as there are already sufficient data from the expression analysis to support the hypothesis that the transcripts are regulated by distinctive promoter and/or enhancer elements.

Other predictions arise from the criteria listed. As the brains of XY individuals are not phenotypically like those of XO individuals, either the Y gene is functionally equivalent to the X gene, or there is a mechanism in males (such as differential expression levels in response to hormonal differences) to compensate for X haploinsufficiency. From this it also follows that if the X and Y genes are equivalent for critical stages in development, the X gene must escape from X-inactivation. For the protocadherin gene, where we should expect the ancestral primate homologue to be subject to X inactivation, such an escape from the established controls would represent an evolutionary significant event. Subtle differences in male/ female verbal performance could relate to secondary functions of the diverged regions of the *PCDHY* protein. Additionally, this sex chromosome-related dimorphism might help to explain some of the gender bias observed in sets of patients with learning difficulties related to language, such as dyslexia. Traditionally, these disorders are classified as autosomal recessive with reduced female penetrance. Variants of the *PCDHX/PCDHY* genes make attractive candidates for modifying effects.

The above characteristics of a gene are all testable with the appropriate materials and experimental design. Patient, anthropological and other primate samples can be used to collect data either to strengthen the case for or to exclude a given candidate. The greatest problem for molecular biology is to try to date the events leading to the genomic organisation observed in modern humans. As methodology for dating based upon the observed rates of nucleotide variation on the X and Y chromosomes improves, it may become possible to predict more precisely the date of the transposition event that placed the copy of the Xq21.3 block onto the Y chromosome. The dating of subsequent independent rearrangements of the sex chromosomes will prove a more formidable challenge in the absence of an archaeological DNA database.

NOTE ADDED IN PROOF

The protocadherin genes on the X and Y chromosomes are also known as *PCDH22* (*PCDHX*) and *PCDH11* (*PCDHY*). Alternative transcript analysis shows that each gene is much larger than originally estimated, and currently there are 17 defined exons spanning 850 kb. *RBPX* has now received the official gene symbol *PABPC5*. Additional details are published in Blanco *et al.* (2001).

Copyright © British Academy 2002 – all rights reserved

ACKNOWLEDGEMENTS

We wish to thank Dr R.A. Furlong and Dr C.A. Boucher for their critical reading of the manuscript. P. Blanco is funded by the Fundacion Pedro Barrie de la Maza and the British Council, and C. Sargent is currently funded by PIC and The Isaac Newton Trust. Part of this work was supported by an MRC programme grant.

DISCUSSION

Questioner: Is this protcadherin gene actually absent in primates and mice?

Affara: It's present on the X chromosome in primates but not on the Y and we have not been able to detect cross-hybridisation with mouse genomic DNA.

Questioner: Have you found that the gene is inactivated on the X chromosome?

Affara: No we haven't. This is something obvious that we want to do. But its not such an easy thing to do for a gene that is expressed specifically in nervous tissue because a lot of the experiments looking at inactivation of genes on the X are done on cell lines in culture. If we can get a polymorphism for one of those neuroepithelial cell lines from a female then we can do that.

Questioner: Was that expression in human or non-human tissue?

Affara: It was human tissue. We haven't looked at non-human tissue.

Questioner: Were they adult brains?

Affara: Most were adult brain except some were whole fetal brain.

References

Affara, N.A. & Ferguson-Smith, M.A. (1994) DNA sequence homology between the human sex chromosomes. In: *Molecular Genetics of Sex Determination* (ed. S. Wachtel, pp. 225–66. San Diego: Academic Press.

Bernardi, G. (2000) Isochores and the evolutionary genomics of vertebrates. *Gene*, **241**, 3–17.

Blanco, P., Sargent, C.A., Boucher, C.A. & Affara, N.A. (2000) Conservation of PCDHX in mammals: expression of human X–Y genes predominantly in the brain. *Mammalian Genome*, **11**, 906–14.

Blanco, P., Sargent, C.A., Boucher, C.A., Howell, G., Ross, M., Affara, N.A. (2001) A novel poly(A)-binding protein gene (PABPC5) maps to an X-specific subinterval in the Xq21.3/Yp11.2 homology block of the human sex chromosomes. *Genomics*, **74**, 1–11.

Copyright © British Academy 2002 – all rights reserved

Buxhoeveden, D. & Casanova, M. (2000) Comparative lateralization patterns in the language area of normal human, chimpanzee, and rhesus monkey brain. *Laterality*, **5**, 315–30.

Ciccodicola, A., D'Esposito, M., Esposito, T., Gianfrancesco, F., Migliaccio, C., Miano, M.G., Matarazzo, M.R., Vacca, M., Franze, A., Cuccurese, M., Cocchia, M., Curci, A., Terracciano, A., Torino, A., Cocchia, S., Mercadante, G., Pannone, E., Archidiacono, N., Rocchi, M., Schlessinger, D. & D'Urso, M. (2000) Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. *Human Molecular Genetics*, **9**, 395–4012.

Crow, T.J. (1993) Sexual selection, Machiavellian intelligence and the origins of psychosis. *Lancet*, **342**, 594–8.

Crow, T.J., Crow, L.R., Done, D.J. & Leask, S. (1998) Relative hand skill predicts academic ability: global deficits at the point of hemispheric indecision. *Neuropsychologia*, **36**, 1275–82.

D'Esposito, M., Ciccodicola, A., Gianfrancesco, F., Esposito, T., Flagiello, L., Mazzarella, R., Schlessinger, D. & D'Urso, M. (1996) A synaptobrevin-like gene in the Xq28 pseudoautosomal region undergoes X inactivation. *Nature Genetics*, **13**, 227–9.

Ferguson-Smith, M.A. (1965) Karyotype–phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. *Journal of Medical Genetics*, **2**, 142–55.

Garcia-Castro, M.I., Vielmetter, E. & Bronner-Fraser, M. (2000) N-Cadherin, a cell adhesion molecule involved in establishment of embryonic left–right asymmetry. *Science*, **288**, 1047–51.

Glaser, B., Myrtek, D., Rumpler, Y., Schiebel, K., Hauwy, M., Rappold, G.A. & Schempp, W. (1999) Transposition of SRY into the ancestral pseudoautosomal region creates a new pseudoautosomal boundary in a progenitor of simian primates. *Human Molecular Genetics*, **8**, 2071–8.

Grant, S.G., O'Dell, T.J., Karl, K.A., Stein, P.L., Soriano, P. & Kandel, E.R. (1992) Impaired long-term potentiation, spatial learning, and hippocampal development in fyn mutant mice. *Science*, **258**, 1903–10.

Graves, J.A. (1995) The origin and function of the mammalian Y chromosome and Y-borne genes: an evolving understanding. *Bioessays*, **17**, 311–20.

Graves, J.A.M., Wakefield, M.J. & Toder, R. (1998) The origin and evolution of the pseudoautosomal regions of the human sex chromosomes. *Human Molecular Genetics*, **7**, 1991–6.

Jones, M.H., Khwaja, O.S., Briggs, H., Lambson, B., Davey, P.M., Chalmers, J., Zhou, C.Y., Walker, E.M., Zhang, Y., Todd, C., Ferguson-Smith, M.A. & Affara, N.A. (1994) A set of ninety-seven overlapping yeast artificial chromosome clones spanning the human Y chromosome euchromatin. *Genomics*, **24**, 266–75.

Kohmura, N., Senzaki, K., Hamada, S., Kai, N., Yasuda, R., Watanabe, M., Ishii, H., Yasuda, M., Mishina, M. & Yagi, T. (1998) Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron*, **20**, 1137–51.

Lahn, B.T. & Page, D.C. (1999) Four evolutionary strata on the human X chromosome. *Science*, **286**, 964–7.

Lambson, B., Affara, N.A., Mitchell, M. & Ferguson-Smith, M.A. (1992) Evolution of DNA sequence homologies between the sex chromosomes in primate species. *Genomics*, **14**, 1032–40.

Copyright © British Academy 2002 – all rights reserved

Marchant, L.F. & McGrew, W.C. (1996) Laterality of limb function in wild chimpanzees of Gombe National Park: comprehensive study of spontaneous activities. *Journal of Human Evolution*, **30**, 427–43.

Mumm, S., Molini, B., Terrell, J., Srivastava, A. & Schlessinger, D. (1997) Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 60-kb resolution. *Genome Research*, **7**, 307–14.

Ohno, S. (1967) Sex chromosomes and sex linked genes. In: *Monographs on Endocrinology* (eds A. Londhardt, L. Samuels & J. Zander). New York: Springer-Verlag.

Page, D.C., Harper, M.E., Love, J. & Botstein, D. (1984) Occurrence of a transposition from the X chromosome long arm to the Y chromosome short arm during human evolution. *Nature*, **311**, 119–23.

Perry, J., Feather, S., Smith, A., Palmer, S. & Ashworth, A. (1998) The human FXY gene is located within Xp22.3: implications for evolution of the mammalian X chromosome. *Human Molecular Genetics*, **7**, 299–305.

Redies, C. & Takeichi, M. (1996) Cadherins in the developing central nervous system: an adhesive code for segmental and functional subdivisions. *Developmental Biology*, **180**, 413–23.

Sano, K., Tanihara, H., Heimark, R.L., Obata, S., Davidson, M., St John, T., Taketani, S. & Suzuki, S. (1993) Protocadherins: a large family of cadherin-related molecules in the central nervous system. *EMBO Journal*, **12**, 2249–56.

Sargent, C.A., Briggs, H., Chalmers, I.J., Lambson, B., Walker, E. & Affara, N.A. (1996) The sequence organization of Yp/proximal Xq homologous regions of the human sex chromosomes is highly conserved. *Genomics*, **32**, 200–9.

Schwartz, A., Chan, D.C., Brown, L.G., Alagappan, R., Pettay, D., Disteche, C., McGillivray, B., de la Chapelle, A. & Page, D.C. (1998) Reconstructing hominid Y evolution; X-homologous block, created by X–Y transposition, was disrupted by Yp inversion through LINE–LINE recombination. *Human Molecular Genetics*, **7**, 1–11.

Senzaki, K., Ogawa, M. & Yagi, T. (1999) Proteins of the CNR family are multiple receptors for Reelin. *Cell*, **99**, 635–47.

Sugino, H., Hamada, S., Yasuda, R., Tuji, A., Matsuda, Y., Fujita, M. & Yagi, T. (2000) Genomic organization of the family of CNR cadherin genes in mice and humans. *Genomics*, **63**, 75–87.

Vacca, M., Matarazzo, M.R., Jones, J., Spalluto, C., Archidiacono, N., Ma, P., Rocchi, M., D'Urso, M., Chen, E.Y., D'Esposito, M. & Mumm, S. (1999) Evolution of the X-specific block embedded in the human Xq21.3/Yp11.1 homology region. *Genomics*, **62**, 293–6.

Vogt, P.H., Affara, N., Davey, P., Hammer, M., Jobling, M.A., Lau, Y.F., Mitchell, M., Schempp, W., Tyler-Smith, C., Williams, G., Yen, P. & Rappold, G.A. (1997) Report of the Third International Workshop on Y Chromosome Mapping 1997. Heidelberg, Germany, April 13–16, 1997. *Cytogenetics and Cell Genetics*, **79**, 1–20.

Vollrath, D., Foote, S., Hilton, A., Brown, L.G., Beer-Romero, P., Bogan, J.S. & Page, D.C. (1992) The human Y chromosome: a 43 interval map based on naturally occurring deletions. *Science*, **258**, 52–9.

Wu, Q. & Maniatis, T. (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell*, **97**, 779–90.

Yoshida, K. & Sugano, S. (1999) Identification of a novel protocadherin gene (*PCDH11)* on the human X–Y homology region in Xq 21.3. *Genomics*, **62**, 540–3.

Copyright © British Academy 2002 – all rights reserved