

# Disruption of two novel genes by a translocation co-segregating with schizophrenia

J. Kirsty Millar<sup>1,+</sup>, Julie C. Wilson-Annan<sup>1,§</sup>, Susan Anderson<sup>1</sup>, Sheila Christie<sup>1</sup>, Martin S. Taylor<sup>1</sup>, Colin A. M. Semple<sup>1</sup>, Rebecca S. Devon<sup>1</sup>, David M. St Clair<sup>2</sup>, Walter J. Muir<sup>1,3</sup>, Douglas H. R. Blackwood<sup>1,3</sup> and David J. Porteous<sup>1</sup>

<sup>1</sup>Medical Genetics Section, Department of Medical Sciences, The University of Edinburgh, Molecular Medicine Centre and MRC Human Genetics Unit, both at Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK,

<sup>2</sup>The Department of Mental Health, University Medical Building, The University of Aberdeen, Foresterhill, Aberdeen AB9 2ZD, UK and <sup>3</sup>Department of Psychiatry, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK

Received 1 February 2000; Revised and Accepted 27 March 2000

DDBJ/EMBL/GenBank accession nos AF222980–AF222987

**A balanced (1;11)(q42.1;q14.3) translocation segregates with schizophrenia and related psychiatric disorders in a large Scottish family (maximum LOD = 6.0). We hypothesize that the translocation is the causative event and that it directly disrupts gene function. We previously reported a dearth of genes in the breakpoint region of chromosome 11 and it is therefore unlikely that the expression of any genes on this chromosome has been affected by the translocation. By contrast, the corresponding region on chromosome 1 is gene dense and, not one, but two novel genes are directly disrupted by the translocation. These genes have been provisionally named Disrupted-In-Schizophrenia 1 and 2 (*DISC1* and *DISC2*). *DISC1* encodes a large protein with no significant sequence homology to other known proteins. It is predicted to consist of a globular N-terminal domain(s) and helical C-terminal domain which has the potential to form a coiled-coil by interaction with another, as yet, unidentified protein(s). Similar structures are thought to be present in a variety of unrelated proteins that are known to function in the nervous system. The putative structure of the protein encoded by *DISC1* is therefore compatible with a role in the nervous system. *DISC2* apparently specifies a non-coding RNA molecule that is antisense to *DISC1*, an arrangement that has been observed at other loci where it is thought that the antisense RNA is involved in regulating expression of the sense gene. Altogether, these observations indicate that *DISC1* and *DISC2* should be considered formal candidate genes for susceptibility to psychiatric illness.**

## INTRODUCTION

Schizophrenia is a serious and debilitating disease affecting ~1% of the population worldwide. There is compelling

evidence from family, twin and adoption studies for a significant genetic basis to the disease (1). This has initiated searches directed at identification of the genetic component using methods such as linkage analysis, association studies of candidate genes and mapping of cytogenetic abnormalities in psychiatric patients, procedures which have been applied successfully to monogenic disorders. Psychiatric illnesses are more complex, however, and apparently result from the combined effects of multiple genes, with inheritance complicated by environmental factors (1). Consequently no genes involved in the aetiology of such illnesses have yet been definitively identified.

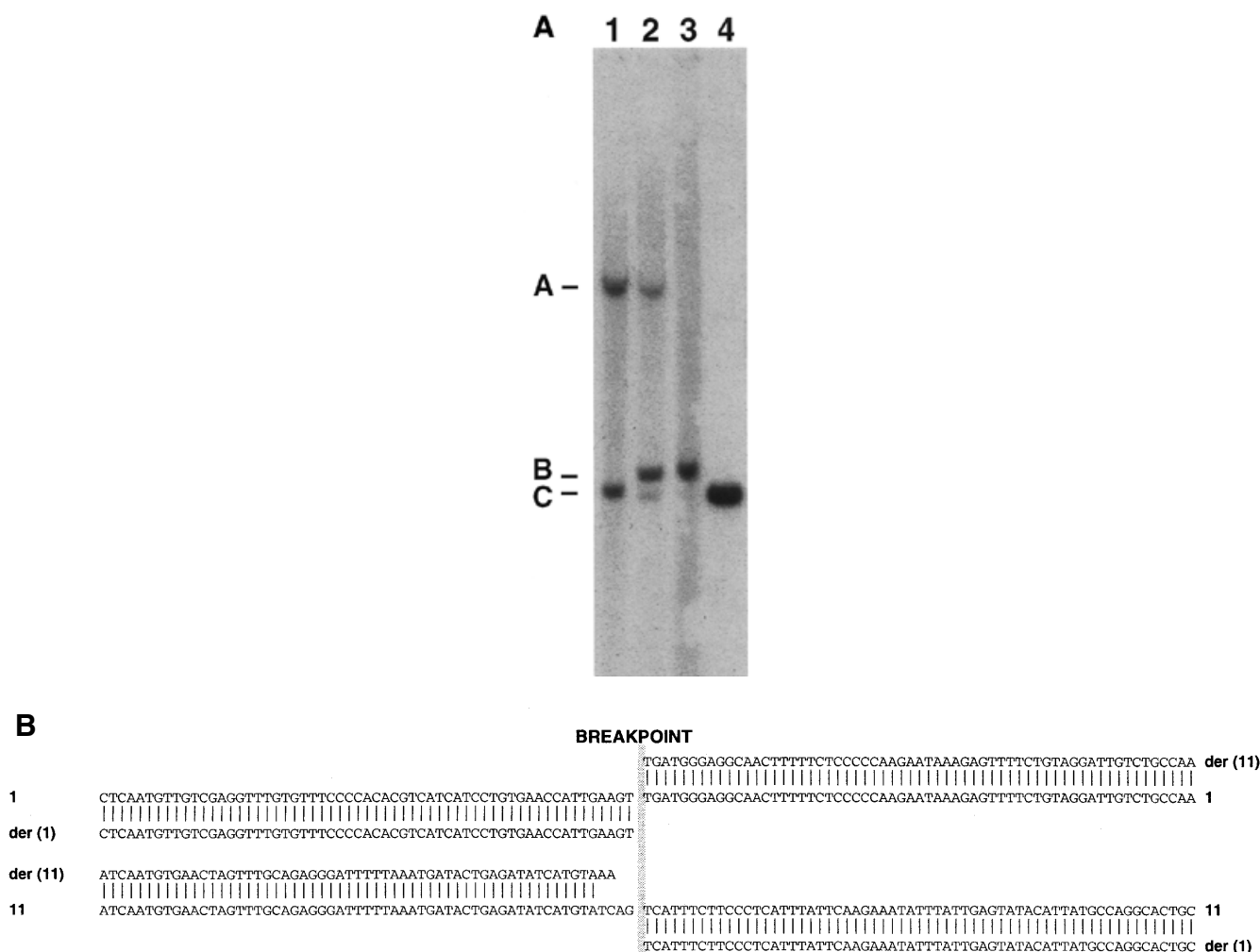
We are studying a large Scottish family (2) in which a balanced translocation segregates with major mental illness (maximum LOD = 6.0; D.H.R. Blackwood, A. Fordyce, M. Walker, E. Drysdale, J.K. Millar, D.M. St Clair, D.J. Porteous and W.J. Muir, manuscript in preparation), based on the hypothesis that the rearrangement has directly disrupted gene function, leading to psychosis. This family may be atypical due to the wide spectrum of disorders present (schizophrenia, schizoaffective disorder, bipolar affective disorder, unipolar affective disorder and adolescent conduct disorder). However, it is likely that identification of the genetic factor(s) involved in the aetiology of disease in these patients would facilitate discovery and understanding of the underlying genetic defects in unrelated psychotic individuals.

## RESULTS

To clone the chromosome 1 breakpoint, a 2.5 kb *EcoRI* fragment from chromosome 11, containing the site of the breakpoint (3), was used to screen an *EcoRI* genomic library constructed from a translocation cell line. A 2.7 kb *EcoRI* fragment, corresponding to the translocation fragment from the derived 1 chromosome was obtained (Fig. 1A). This fragment was used to rescreen the same library, yielding a 7.3 kb clone, containing the site of the chromosome 1 breakpoint (Fig. 1A). The derived 11 fragment was obtained using the polymerase

<sup>+</sup>To whom correspondence should be addressed at: Medical Genetics Section, Department of Medical Sciences, The University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, Scotland, UK. Tel: +44 131 650 1000; Fax: +44 651 1059; Email: kirsty.millar@ed.ac.uk

<sup>§</sup>Present address: Molecular Genetics of Cancer Division, WEHI, Royal Melbourne Hospital, Victoria 3050, Australia



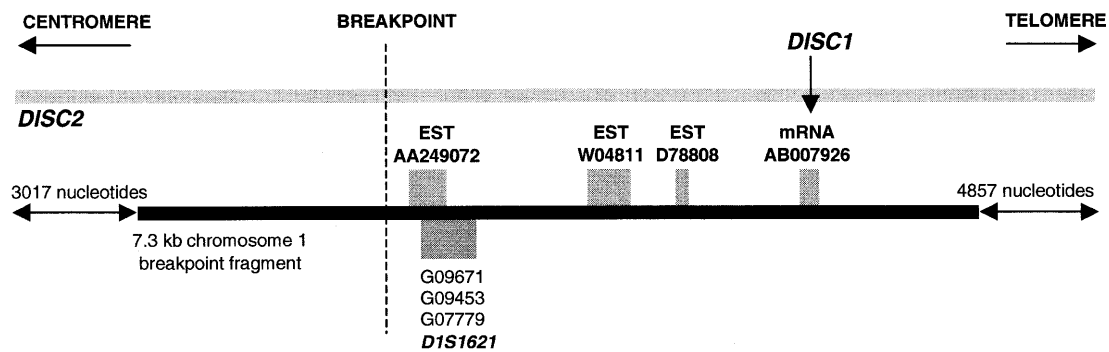
**Figure 1.** (A) Southern blot of *Eco*RI digested DNA probed with the 2.7 kb *Eco*RI derived 1 fragment. Lane 1, control human DNA; lane 2, genomic DNA from the translocation cell line MAFLI; lane 3, genomic DNA from the somatic cell hybrid MIS7.4, carrying the derived 1 translocation chromosome; lane 4: ICRF yeast artificial chromosome (YAC) y901 D0485 DNA. This YAC spans the chromosome 11 breakpoint of the translocation (28). Arrows indicate the positions of the 7.3 kb normal chromosome 1, 2.7 kb derived 1 and 2.5 kb normal chromosome 11 hybridization signals (A, B and C, respectively). (B) Alignment of sequence immediately flanking the breakpoints from the normal chromosome 1, derived 1, derived 11 and normal chromosome 11 [1, der (1), der (11) and 11, respectively]. The breakpoint sequence and minor rearrangement were confirmed by genomic sequence analysis of two other translocation carriers (data not shown). The position of the breakpoint was also confirmed by PCR and DNA sequence analysis on genomic DNA from MIS7.4 and MIS39, cell lines carrying the derived 1 and derived 11 chromosomes respectively (data not shown).

chain reaction (PCR). Alignment of breakpoint sequence from all four fragments (chromosome 11, chromosome 1, derived 11, derived 1, GenBank accession nos AF222984, AF222983, AF222986 and AF222985, respectively) (Fig. 1B) shows that the translocation resulted in replacement of TCAG with AA on the derived 11 chromosome, and that, consequently, no major rearrangement of sequence accompanied the translocation event.

Sequence analysis provided no evidence for the presence of any transcripts within the sequence surrounding the chromosome 11 breakpoint. However, the chromosome 1 breakpoint fragment sequence includes three expressed sequence tags (ESTs: AA249072, W04811 and D78808) and an exon from a messenger RNA, AB007926 (Fig. 2).

We have provisionally named the anonymous mRNA AB007926 *DISC1* (Disrupted-In-Schizophrenia 1), because analysis of the chromosome 1 breakpoint sequence indicates that the gene is directly disrupted by the translocation, which takes place within an intron (Fig. 2). Furthermore, this analysis indicates that the direction of transcription is proximal to distal.

Using a combination of cDNA library screening and RACE (rapid amplification of cDNA ends), we obtained 6913 nt of cDNA sequence transcribed from the *DISC1* gene (accession no. AF222980), which matched AB007926 (99.985%) over 6833 nt, and extended 14 nt further 5'. Four different 5' RACE products, our longest cDNA clone and AB007926 all terminate within 14 nt, suggesting that this is the true 5' end of the transcript. Furthermore, Nucleotide Identify X (NIX, <http://menu.hgmp.mrc.ac.uk/>)



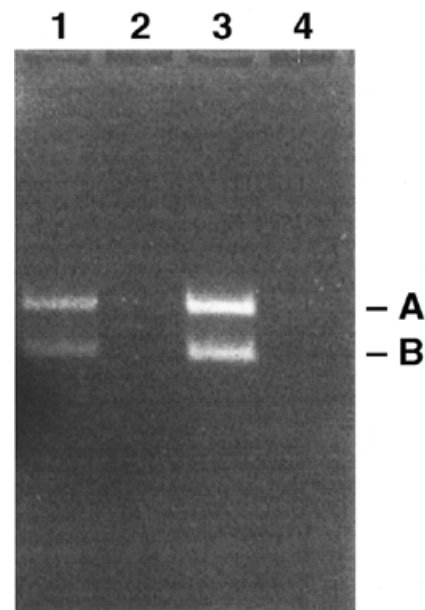
**Figure 2.** Map of the 7.3 kb *EcoRI* chromosome 1 breakpoint fragment. Breakpoint sequence was analysed using BLASTN (27) at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) and the suite of gene recognition and analysis programmes encompassed by NIX. Only 103 nt of the total 350 in EST D78808 are contained within the chromosome 1 breakpoint sequence. The remaining sequence is identical to several other ESTs (UniGene cluster Hs.31446, <http://www.ncbi.nlm.nih.gov/UniGene/index.html>), none of which contain any chromosome 1 sequence or are even present on chromosome 1, as judged by a lack of hybridization to genomic DNA from the chromosome 1 human/mouse hybrid cell line A9(Neo-1)-4 (data not shown). The messenger RNA sequence in AB007926 consists of 6833 nt of a brain-expressed transcript (29) positioned on chromosome 1 by Gene Map '98 (<http://www.ncbi.nlm.nih.gov/genemap/>). 191 nt of this transcript are contained within the 7.3 kb of chromosome 1 breakpoint sequence. NIX identified one putative exon with consensus splice sites on the forward strand of the chromosome 1 breakpoint sequence. This exon contains 189 nt of the sequence match to mRNA AB007926. The sequence of a marker, *DIS1621*, is also contained within the chromosome 1 breakpoint fragment. The extent to which *DISC2* continues centromeric and telomeric from the chromosome 1 breakpoint fragment is indicated.

menu-bin/Nix/Nix.pl) analysis of 2.4 kb of genomic sequence encompassing the 5' end of *DISC1* identified a 758 nt putative CpG island (70% GC) containing 120 nt from the 5' end of the transcript, and two potential promoters 55 and 359 nt upstream. The first ATG in *DISC1* is located at position 54 and is not part of a strong translation initiation consensus (4). The predicted coding sequence, starting at the first ATG, consists of 2565 nt, with a stop codon at position 2616. The 3' untranslated region (UTR) is 4294 nt extending to a poly(A) tail. A consensus polyadenylation signal is located at position 6892, 16 nt upstream of the poly(A) tail. The gene is tagged at the 3' end by UniGene cluster Hs.26985.

Sixty-six nucleotides (2295–2360) are not contained within the putative coding sequence of AB007926. This deletion corresponds to a common alternative splicing event (Fig. 3, Table 1). *DISC1* is present as a major transcript of ~7.5 kb (Fig. 4A and B) in all adult tissues examined. The size discrepancy between the sequence we have obtained and the transcripts detectable on northern blots may be due to polyadenylation. *DISC1* was not detected on northern blots of fetal tissues (data not shown), although reverse transcription (RT)-PCR experiments indicate that fetal transcripts do exist (Table 1).

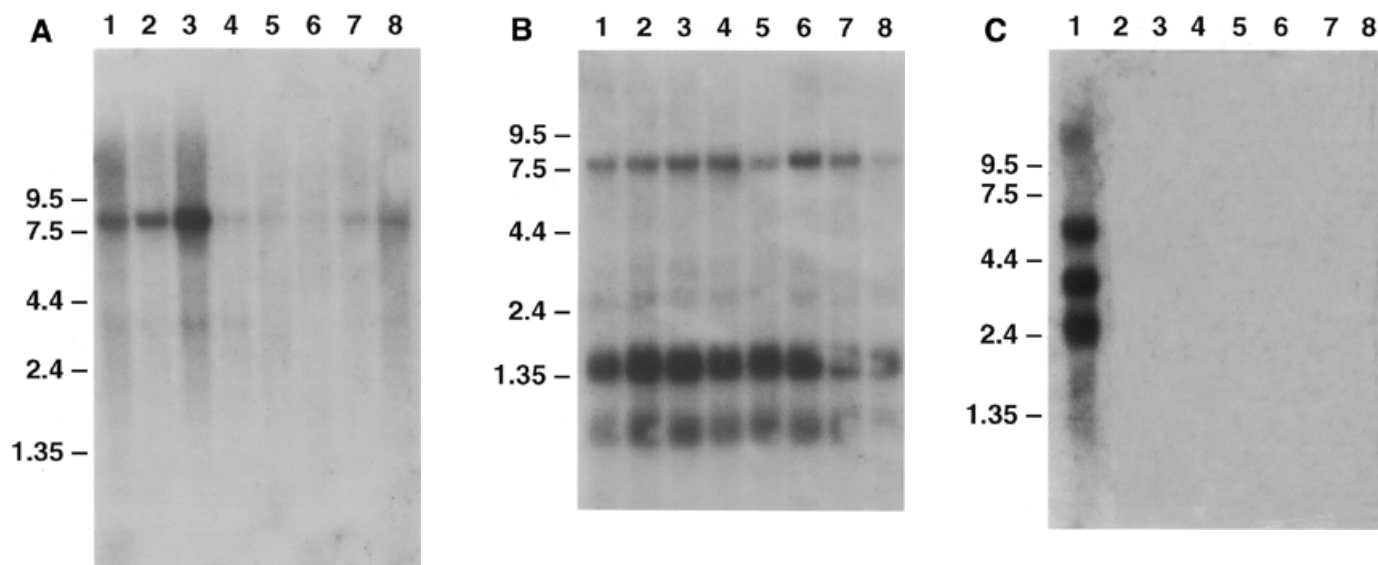
The open reading frame (ORF) in *DISC1* encodes a putative protein of 854 amino acids. Protein structure prediction programmes (<http://dodo.cpmc.columbia.edu/predictprotein/>) suggest that *DISC1* can be divided into two distinct regions of secondary structure (Fig. 5A). The N-terminal region (amino acids 1–347) is predicted to consist of one or more globular domains. The C-terminal region is predicted to consist entirely of  $\alpha$ -helix interspersed with several short loops, and contains regions with the potential to form coiled-coils, structures that arise when the helical stretches of separate proteins interact. Alternative splicing introduces an additional loop into the C-terminal domain.

BLAST 2.0 and FASTA 3 searches of the SWall database (SwissProt plus TREMBL) at the European Bioinformatics



**Figure 3.** RT-PCR analysis of *DISC1* in human fetal brain using the distal primer pair (Table 1). Position 2295, at which the 66 bp deletion starts, corresponds approximately to a splice donor site within the cDNA (AGgtattg instead of the consensus AGgtaagt). Position 2360, where the deletion ends, is known to be a splice donor (J.K. Millar, S. Christie, D. Lawson, D. Hsiao-Wei Loh, B. Arveiler and D.J. Porteus, manuscript in preparation). A product of 270 bp indicates upstream donor site usage. A product of 336 bp indicates downstream donor site usage. Lane 1, 8.3 weeks; lane 2, 10.3 weeks. Arrows indicate the 336 and 270 bp alternatively spliced products (A and B, respectively).

Institute (<http://www2.ebi.ac.uk/>) reveal matches (~21% identity, 41% similarity across 500 amino acids) between the  $\alpha$ -helical region of *DISC1*, particularly around the stretches of coiled-coil, and the known or predicted coiled-coil domains of



**Figure 4.** Clontech human multiple tissue northern blots probed with *DISC1* (nucleotides 958–1983) and *DISC2* (nucleotides 8488–9300). Each lane contains ~2 µg of poly(A)<sup>+</sup> RNA. Positions of size markers are indicated. (A) Adult human multiple tissue northern probed with *DISC1*. Lanes 1–8: heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. (B) Adult human brain multiple tissue northern probed with *DISC1*. Lanes 1–8: amygdala, caudate nucleus, corpus callosum, hippocampus, whole brain, substantia nigra, subthalamic nucleus and thalamus. An additional filter with RNA from cerebellum, cerebral cortex, medulla, spinal cord, occipital pole, frontal lobe, temporal lobe and putamen was probed and produced similar results (data not shown). (C) Adult human multiple tissue northern probed with *DISC2*. Lanes 1–8: heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. All four heart transcripts are apparently derived from *DISC2* because no other sites in the human genome to which *DISC2*-derived sequences hybridize have been detected (data not shown).

**Table 1.** RT-PCR analysis of *DISC1* and *DISC2*

Sample	Age (weeks)	<i>DISC1</i>		<i>DISC2</i>	
		proximal	distal	proximal	distal
Brain	8.3	+	+(2)	+	+
	10.3	+	+(2)	+	+
	13.3	+	+(2)	+	+
Heart	8.8	+	+(2)	+	+
	9.1	+	+(2)	+	+
	9.3	+	+(2)	+	+
Kidney	10.0	+	+(2)	–	+
Spleen	14.8	+	+(2)	+	–
Limb	10.3	+	+(2)	–	–

Approximate ages of gestation are given in weeks. 2, two bands obtained using one primer pair indicating alternative splicing; +, transcript detected; –, transcript not detected.

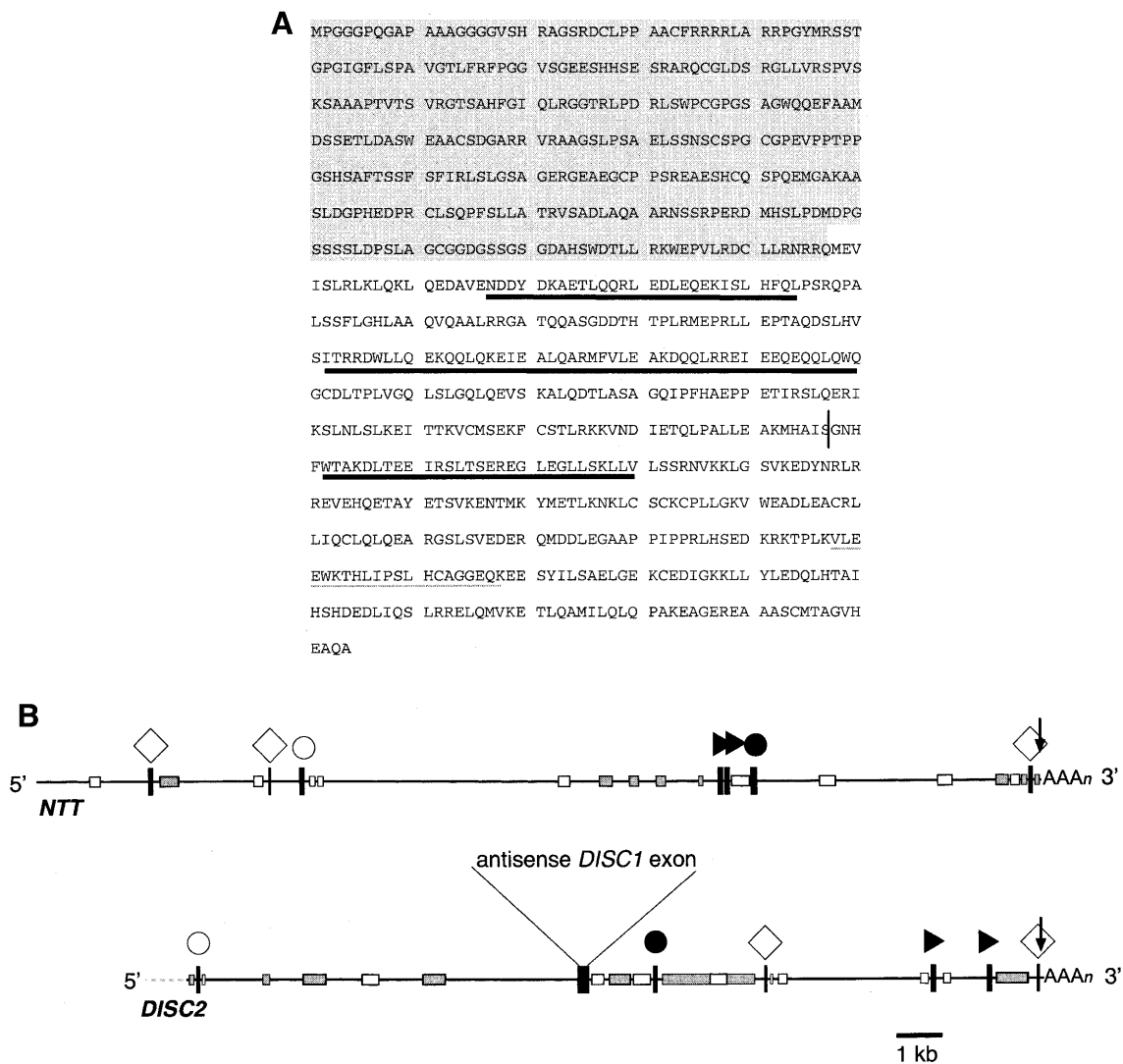
several other proteins. These similarities are likely to result from biased sequence composition imposed by coiled-coil structure and therefore probably reflect structural resemblances between the proteins.

*DISC2* (accession no. AF222981) was identified from the ESTs at the chromosome 1 breakpoint (Fig. 2) and extended by cDNA library screening, RACE and RT-PCR. This identified 15 178 nt of contiguous genomic sequence which is known to be transcribed. The transcript 3' end is tagged by Unigene cluster Hs.96883. There are two consensus polyadenylation

signals at positions 15 072 and 15 161, 107 and 18 nt upstream of the poly(A) tail.

The 5' end of the transcript has not yet been located, but *DISC2* so far consists of a single large exon encompassing the translocation breakpoint and the 189 nt exon of *DISC1*. Sequence analysis indicates that the direction of transcription is distal to proximal. *DISC2* therefore overlaps with *DISC1*, but is transcribed in the opposite orientation. *DISC2* transcripts are most abundant in heart where species of >9.5 kb, and of ~6, 3 and 2.5 kb are present (Fig. 4C). RT-PCR indicates that *DISC2* is also transcribed in several fetal tissues (Table 1) although transcripts were not detectable on northern blots (data not shown).

No significant ORF has been identified within 15 178 nt of the *DISC2* transcript. The longest ORF deduced to be present encodes 57 amino acids, while the start codon of this ORF is not in a good context (4), suggesting that the sequence lacks any protein coding potential. Furthermore, a survey of the EMBL database (<ftp://ftp.ebi.ac.uk/pub/databases/embl/release/>) indicates that the longest known 3' UTR in a human gene is 9280 nt (doublecortin), while the average length is 2131.83 nt ± 1368.09 (based on 861 full-length 3' UTR sequences). Therefore, if *DISC2* is a protein-coding gene, it possesses a 3' UTR in excess of 15 kb, substantially outside the size range of other known 3' UTRs. In addition, *DISC2* possesses certain similarities to the 17 572 nt non-coding mRNA-like transcript *NTT* (5), as summarized (Fig. 5B). These observations suggest that *DISC2* is a non-coding structural RNA gene, although final confirmation awaits identification of the transcript 5' end.



**Figure 5.** (A) Predicted protein sequence of DISC1 translated from the first ATG. Light underlining, alternatively spliced amino acids; bold underlining, stretches with coiled-coil forming potential; light shaded box, putative globular domain(s); vertical line, position of the translocation breakpoint. (B) Feature-based comparison of DISC1 and NTT transcripts. Simple repeat and interspersed repeat features were annotated based on RepeatMasker (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>) analysis of each transcript. The direct repeat features were identified using Dotmatrix analysis (30). Both transcripts and the annotated features are drawn to scale. The tetranucleotide repeat of DISC1 is (TGGA)<sub>n</sub> and of NTT is (TAGA)<sub>n</sub>. Although both transcripts possess comparable features, the linear organization of those elements is not highly similar. Open circle, tetranucleotide repeat; open diamond, AT- or GT-rich low complexity region; closed circle, polypurine tract; closed triangle, transcript-specific repeat element; vertical arrow, utilized poly(A) consensus; shaded box, LINE; open box, SINE; closed box, annotated feature.

## DISCUSSION

We have cloned and sequenced the breakpoints of a (1;11)(q42.1;q14.3) translocation linked to schizophrenia and related psychosis (2), and identified two novel genes, DISC1 and DISC2, both disrupted by the translocation. One or both of the disrupted alleles may be responsible for the psychiatric disorders suffered by carriers of the translocation. Furthermore, these genes may also be involved in the mental illness of patients unrelated to the family segregating the translocation. However, no independent evidence for a locus in this region of the genome has yet been presented, although recent reports of suggestive linkage to 1q32 and 1q32-41 in patients suffering from bipolar disorder and schizophrenia, respectively (6,7), are intriguing.

DISC1 protein matches are essentially restricted to structural similarities to myosins, structural proteins and proteins that are involved in motility and/or transport (particularly microtubule binding proteins). In the context of psychiatric illness, it is interesting to note that many of these proteins (dynactin, D-CLIP-190, citron, post-synaptic density proteins, FEZ1 and hyaluronan receptor, for example) are implicated in processes such as axon guidance, synaptogenesis, functioning of the synapse and intracellular transport along axons and dendrites (8-15). It is an intriguing possibility that the function of DISC1 is similar, suggesting a role in development of the nervous system and/or neuronal activity, and therefore adding further to the evidence pointing towards involvement in the aetiology of mental illness.

By analogy to many other examples of mammalian genes with endogenous antisense RNA transcripts (16–18), *DISC2* presents an attractive mechanism by which *DISC1* expression may be regulated. There is evidence that antisense RNAs affect expression of the sense gene (i) in the nucleus at the levels of transcription, RNA processing or export from the nucleus, or (ii) in the cytoplasm by influencing RNA stability or translation. Similarly, *DISC2* might act at one of these levels to influence expression of *DISC1*.

In translocation carriers, transcription of both *DISC1* and *DISC2* is predicted to occur from their endogenous promoters on the derived chromosomes, unless unidentified transcription signals of major effect are removed or inactivated by the rearrangement. This contention is supported by preliminary experiments which detect transcription of *DISC1* from the derived 1 chromosome (data not shown). It is unlikely that corresponding truncated transcripts lacking a 5' end would be produced as well, because the breakpoint region of chromosome 11 is apparently transcriptionally inactive (3,19).

Several scenarios resulting from production of 3' truncated *DISC1* and *DISC2* transcripts could be envisaged, but three are most obvious. First, truncated *DISC1* protein may be produced. It would lack the C-terminal 257 amino acids, including one of the regions of most strongly predicted coiled-coil forming potential. This would be expected to reduce the overall coiled-coil forming potential and stability as a multimer, while retaining the unidentified function of the globular N-terminus. Production of a partially active protein could conceivably result in a dominant-negative effect. Second, truncated *DISC2* transcripts would retain complementarity to the normal full-length *DISC1* transcripts. However, *DISC2* truncation could affect events following interaction between truncated *DISC2* and normal *DISC1* transcripts. Therefore, if *DISC2* does regulate expression of *DISC1* it is possible that this mechanism would be negatively affected, resulting in dysregulation of both the non-translocated and translocated alleles, and a pleiotropy of dominant-negative effects. Third, irrespective of the putative regulatory role of *DISC2* on *DISC1*, the gene product may have independent (regulatory) functions affected by truncation.

We propose that alteration of *DISC1* and/or *DISC2* activity, by truncation and/or by abnormal regulation of expression, is causally linked to the psychiatric illness in translocation carriers. Dysregulation of the novel functions of *DISC1* and *DISC2* in the absence of a translocation event may play a more general role in susceptibility to psychiatric illness.

## MATERIALS AND METHODS

### Cell culture

The lymphoblastoid cell line MAFLI from an individual bearing the t(1;11)(q42.1;q14.3) translocation, somatic cell hybrids MIS7.4 and MIS39 bearing the derived 1 or derived 11 translocation chromosomes respectively, and their culture conditions, have been described previously (20). On the derived chromosome 1, DNA has been lost from 1q42.1-qter and replaced with chromosome 11 material from 11q14.3-qter. The derived 11 chromosome is the reciprocal translocated chromosome. The cell line A9(Neo-1)-4, a mouse A9 hybrid cell line carrying human chromosome 1, and its culture requirements, have been reported previously (21).

### DNA preparation

Human and  $\lambda$  genomic DNA was prepared by standard methods (22). Plasmid and cosmid DNA was prepared using Qiagen (Crawley, UK) plasmid midi kits.

### RNA extraction and cDNA synthesis

Human fetal tissues were obtained from the United Kingdom Medical Research Council Tissue Bank. Total RNA was extracted using RNazol B (Biogenesis, Poole, UK) according to the manufacturer's instructions. First strand cDNA synthesis was carried out on DNase I-treated RNA using the random hexamer primer from the SUPERSCRIP<sup>T</sup> Preamplification System (Gibco BRL, Paisley, UK) according to the manufacturer's instructions. One microlitre of the resulting cDNA was used in standard PCRs.

### Genomic library construction and screening

Genomic DNA from the translocation cell line MAFLI was digested with *EcoRI*, ligated into *EcoRI*-digested and dephosphorylated  $\lambda$ ZAPII (Stratagene, Cambridge, UK), and packaged using Gigapack Gold II packaging extract (Stratagene) according to the manufacturer's instructions. Bacteriophage were plated using *Escherichia coli* XL1-Blue MRF' and the library of clones screened using standard methods. Excision of clones from the  $\lambda$  vector was carried out as advised by the manufacturer, releasing genomic fragments cloned into pBluescript SK(-). The library was screened using a 2.15 kb repeat-free *HindIII-EcoRI* sub-fragment of the 2.5 kb *EcoRI* fragment containing the chromosome 11 breakpoint (3), followed by the 2.7 kb derived 1 fragment.

### cDNA library screening

5'-STRETCH PLUS cDNA libraries of 20–26 week fetal brain and 20–25 week fetal heart, constructed in bacteriophage  $\lambda$ gt10 and gt11, respectively, were obtained from Clontech (Basingstoke, UK) and screened according to the manufacturer's instructions. Inserts were obtained from pure clones using two methods. First, cDNAs were amplified by PCR, turbocloned (23) and sequenced. In order to amplify cDNA inserts from  $\lambda$  vectors, a single plaque was picked into 25  $\mu$ l of distilled water, 1–5  $\mu$ l were then added to a PCR and the cDNA insert amplified using vector-based primers.  $\lambda$ gt10-specific primers, agcaagttcagcctggttaagt and gggaccttcttatgagtatt (annealing at 68°C) and  $\lambda$ gt11-specific primers gaagcaccat-ggctgaatcgcagcgttc and gacaccagaccaactggtaatggtagcgcac (annealing at 56°C) were used to amplify inserts from the fetal brain and fetal heart cDNA libraries, respectively. Due to the probable introduction of sequence alterations during PCR, several subclones were sequenced. Alternatively,  $\lambda$  DNA was digested with *EcoRI* to release the cDNA insert which was then subcloned into *EcoRI*-digested pBluescript SK(-) (Stratagene).

### DNA sequencing

In order to sequence the derived 11 PCR product, it was excised from a 0.8% low melting point agarose gel, dialysed in distilled water for 30 min, melted at 65°C and an appropriate quantity of melted gel slice added to the sequencing reaction.

Direct cosmid sequencing utilized 0.5–1 µg of cosmid DNA, with 60 ng of primer and ABI PRISM BigDye terminator cycle sequencing ready reaction kits (PE Applied Biosystems, Warrington, UK). Plasmid DNA sequencing reactions were performed using ABI PRISM dye terminator or dRhodamine terminator cycle sequencing ready reaction kits and 20 ng of primer. Products were separated on an ABI 373 or 377 DNA sequencer (PE Applied Biosystems), according to the manufacturer's instructions. Resulting sequence was analysed using the GCG package of sequence analysis software (Wisconsin package version 9.1, Genetics Computer Group, Madison, WI). Sequence contigs were constructed using the Phred, Phrap and Consed software, version 6.0 (24–26).

BLASTN (27) searches using chromosome 1 breakpoint sequence identified only sequence matches to two bacterial artificial chromosome (BAC) end clones (accession nos AQ748746 and AQ105798). Searches using chromosome 1 breakpoint sequence identified sequence matches to a triplet repeat *DIS1621* (accession nos G09671, G09453 and G07779), three BAC end clones (accession nos AQ112950, AQ078498 and B40542) as well as the mRNA and EST matches.

### Sequencing of cosmid ICRFc112I0142Q6

This cosmid spans the chromosome 1 breakpoint (J.K. Millar, S. Christie, D. Lawson, D. Hsiao-Wei Loh, B. Arveiler and D.J. Porteus, manuscript in preparation). It was obtained by screening a chromosome 1 cosmid library provided by the Resource Centre/Primary Database (RZPD) of the German Human Genome Project at the Max-Planck-Institute for Molecular Genetics (<http://www.rzpd.de>). The probe used was the 2.7 kb derived 1 fragment. Cosmid DNA was treated with Plasmid-Safe ATP-dependent DNase (Epicentre Technologies, CAMBIO, Cambridge, UK) to remove contaminating *E.coli* DNA according to the manufacturer's instructions. Cosmid DNA was partially digested with *Sau*III, and the resulting restriction fragments size-fractionated on a 0.8% agarose gel. Fragments of ~900 bp were excised from the gel and subcloned into pBluescript SK(-) (Stratagene). Subclones were picked randomly and sequenced using vector-based primers flanking the cloning site (caggaacagctatgac and gtaaacgacggccagt). Contig overlaps were established by designing primers from the ends of the contigs and subsequent direct cosmid sequencing. This process was repeated to generate the sequence of ICRFc112I0142Q6 as two ordered contigs (GenBank accession no. AF222987).

### PCRs

Unless otherwise stated, PCR was carried out using *AmpliTaq* DNA polymerase (Perkin Elmer Biosystems, Foster City, CA). Each 50 µl reaction contained 1 U of enzyme, 300 ng of each primer, 200 mM of each dNTP, 1.5 mM MgCl<sub>2</sub>, 50 mM KCl and 10 mM Tris-HCl pH 8.3. All reactions utilized 35 cycles with a denaturation step of 30 s at 94°C, an annealing step of 1 min at a temperature appropriate for the primers used, and a synthesis step at 72°C, based on the assumption that 1 min is required to synthesize 1 kb of DNA.

The 1.4 kb derived 11 breakpoint PCR product was amplified (annealing at 60°C) using one primer specific for chromosome 11 (ggctgattgccccttgagccataatt) and one primer

specific for chromosome 1 (agaacagaggaggacgatgatgac). MIS39 genomic DNA was used as template.

Analysis of *DISC1* expression by RT-PCR was performed using proximal primers ggaaggagcaggaggcagcccaggcgga and gcacgctgcaggtgtaagcaatc (152 bp product) with the Advantage-GC cDNA PCR kit (Clontech) and cycling conditions as recommended by the manufacturer. Distal primers ggaagctt-gtcgattgctatcc and agatcttcatcatgactgtggattgc designed from *DISC1* were also utilized for RT-PCR (270 and 336 bp products, annealing at 64°C). Analysis of *DISC2* expression by RT-PCR utilized proximal primers cccaagccttaccctcagatcaa and atcaggcagaatgccacagcgtg (250 bp product, annealing at 65°C) and distal primers gagacgacaagtacagactggag and gtctcaggcataagacactgtgac (486 bp product, annealing at 68°C). In the latter three cases, an initial hot start step was carried out.

### Hybridization

Standard procedures were used for Southern blotting and hybridization (22). Double-stranded probes were labelled with [ $\alpha$ -<sup>32</sup>P]dCTP by random priming using High Prime (Boehringer Mannheim, Lewes, UK) and purified using Amersham Pharmacia Biotech (Little Chalfont, UK) NICK columns. The oligonucleotide probe was labelled with [ $\gamma$ -<sup>32</sup>P]dATP using T4 polynucleotide kinase and hybridized to filters at 58°C.

### Sequencing of *DISC1*

A 694 nt probe containing the 189 nt exon of AB007926 (contained within the 7.3 kb chromosome 1 breakpoint fragment) was prepared by PCR using the chromosome 1 breakpoint fragment as template and primers ccattctggacggc-ctaaagacc and gcaracacttggctaaggcggc (annealing at 58°C). This probe and subsequent rounds of cDNA library rescreening generated several overlapping clones from the fetal brain and fetal heart cDNA libraries spanning 6913 nt of the transcript. 5' RACE products were obtained from *DISC1* using the Advantage-GC cDNA PCR kit (Clontech) and cycling conditions as recommended by the manufacturer, with human 20–25 week fetal brain Marathon-Ready cDNA (Clontech) as template (gene-specific primer 1, gagtcaaggccacct-gtctggc; gene-specific primer 2, gcacgctgcaggtgtaagcaatc). Genomic sequence (2.4 kb) encompassing the 5' end of *DISC1* was obtained as follows. An oligonucleotide (ggaaggagcag-gaggcagcccaggcgga) designed from the 5' end of *DISC1* was hybridized to a genomic *Pst*I fragment of ~2.4 kb. *Pst*I fragments of this size were isolated from a PAC, 135-G6, containing the 5' end of *DISC1*, (J.K. Millar, S. Christie, D. Lawson, D. Hsiao-Wei Loh, B. Arveiler and D.J. Porteus, manuscript in preparation), subcloned into pBluescript SK(-) (Stratagene) using standard methods (22) and sequenced (GenBank accession no. AF222982).

### Identification of *DISC2* transcribed sequence

Probes corresponding to ESTs AA249072 and W04811 were used to screen fetal heart and fetal brain cDNA libraries. A 555 nt probe for EST AA249072 was prepared using the chromosome 1 breakpoint fragment as template and primers gctgctcaattaagcagtaacagtgc and catctctgaacagggtgtgtcc

(annealing at 58°C). The cDNA corresponding to EST W04811 was excised from modified pT7T3 (Pharmacia) by double-digestion with *NotI* and *EcoRI*. Further cDNA clones were identified using cDNAs corresponding to ESTs from Unigene cluster Hs.96883 to rescreen the fetal heart cDNA library (these ESTs were identified from sequencing of cosmid ICRFc112i0142Q6). *DISC2* was extended proximal and distal from the cDNA clones isolated by cDNA library screening using 5' and 3' RACE with human adult heart Marathon-Ready cDNA (Clontech) as template, and Expand LT (Boehringer Mannheim) with cycling conditions as recommended by the manufacturer. Gene-specific primers 1 and 2 were gcttgtgcttat-tctctttgggta, and accatcgtcactgtttctctgct, gcttggcactttggg-ggctgta and tctttcacctcctcctcctcttctt, gccacccatgccagct-cactttta and gctctggcacatattaagaagatccc, tattttccaggttcttcc-cag and ttccctctctctccacaacgt. Gaps were filled by speculative RT-PCR performed using primers designed along the length of the sequence of cosmid ICRFc112i0142Q6, and 10.0 or 13.6 week (gestational age) human fetal heart cDNA as template. Primer pairs (in order, proximal to distal along the cosmid sequence) were ggggtaagggtattgt and cacagagttca-gagttc, tgaggattggcaggtgaaagga and gctgactttaccactctgt-tcca, ttggaacagagtggttaa and cagtgccttggtaaaac, ctgggga-cattttggcagg and tcaaatctggtcctcatgacc, tctcagaagacgtggt-tcagtg and gtgaagtgacaatgatgagattc, ccctaagtggct-taacagctcag and tgtttccacactctcaacccctag, attgaggtgagttggct-taggtg and tcctttctatcacctgattgtct, aggtgcttgagttgctgagttggat and tgaaaaactgctgctgtaaatctgagg, ggacagccctcagattacgcagcag and acaaaatctgctgcttgattctc, ctaagtgagaataccaagcagcaaga and tcctctctctctttttgctctc, actggtgagagaggagcaaaaagag and tgtcccggtctccatttctcctc, cttttattggcagggagaggaggaa and aacccgatgacatcaatttaccta, cagaagaaatgccaatgcaagtgt and caagcccaattcacctcagacagttt, aaactgctgaggtgaattagggttg and gccacatagcaccgcaacactcatct, ccagatgagtggtgctgctatgtg and aggcaaaaatggctgaaagggaataa, gattattcctctcagccattttgc and ttgtgaaaggatgggtgggtg. RT-PCR products were obtained from *DISC2* using Expand LT (Boehringer Mannheim) with cycling conditions as recommended by the manufacturer. The combination of cDNA library screening, RACE and RT-PCR identified 15 178 nt of contiguous genomic sequence which is transcribed. Within this genomic sequence, nucleotides 8213–12 161 and 14 181–14 780 correspond to cDNA clones, nucleotides 6245–8436 and 14 594–15 178 correspond to RACE products and nucleotides 1–6470 and 12 027–14 189 correspond to RT-PCR products. The 5' end of the transcript, which is located distal to cosmid ICRFc112i0142Q6, has not yet been identified due to the lack of further genomic sequence for the design of primers and speculative RT-PCR.

### Human 3' UTR length analysis

3' UTRs were determined for the KIAA subset of mRNA sequences, based on coding sequence annotations. 3' ends of the determined UTRs were checked for the presence of a consensus poly(A) signal (AATAAA) within 400 nt of the presumed end. UTRs failing to meet this criterion were excluded from further analysis. Sequence coordinates for full-length 3' UTRs were determined and poly(A) signals detected using in-house software.

### Northern blot analysis

A probe corresponding to nucleotides 958–1983 of *DISC1* was obtained by excising one of the *DISC1* fetal brain library cDNA clones using *EcoRI*. The cDNA clone corresponding to EST W04811 was used to identify *DISC2* transcripts. Northern blots were obtained from Clontech.

### ACKNOWLEDGEMENTS

We thank Kathy Evans for useful discussion and critical reading of this manuscript. This work was supported by Organon NV and the UK Medical Research Council.

### REFERENCES

- McGuffin, P., Owen, M.J. and Farmer, A.E. (1995) Genetic basis of schizophrenia. *Lancet*, **346**, 678–682.
- St Clair, D., Blackwood, D., Muir, W., Carothers, A., Walker, M., Spowart, G., Gosden, C. and Evans, H.J. (1990) Association within a family of a balanced autosomal translocation with major mental-illness. *Lancet*, **336**, 13–16.
- Millar, J.K., Brown, J., Maule, J.C., Shibasaki, Y., Christie, S., Lawson, D., Anderson, S., Wilson-Annan, J.C., Devon, R.S., St Clair, D.M. *et al.* (1998) A long-range restriction map across 3 Mb of the chromosome 11 breakpoint region of a translocation linked to schizophrenia: localisation of the breakpoint and the search for neighbouring genes. *Psychiatr. Genet.*, **8**, 175–181.
- Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.*, **12**, 857–872.
- Liu, A.Y., Torchia, B.S., Migeon, B.R. and Siliciano, R.F. (1997) The human NTT gene: identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4+ T cells. *Genomics*, **39**, 171–184.
- Deterra-Wadleigh, S.D., Badner, J.A., Berrettini, W.H., Yoshikawa, T., Goldin, L.R., Turner, G., Rollins, D., Moses, T., Sanders, A.R., Karkera, J.D. *et al.* (1999) A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc. Natl Acad. Sci. USA*, **96**, 5604–5609.
- Hovatta, I., Varilo, T., Suvisaari, J., Terwilliger, J.D., Ollikainen, V., Arajärvi, R., Juvonen, H., Kokko-Sahin, M.L., Vaisanen, L., Mannila, H. *et al.* (1999) A genome-wide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. *Am. J. Hum. Genet.*, **65**, 1114–1124.
- Dillman, J.F., Dabney, L.P., Karki, S., Paschal, B.M., Holzbaur, E.L. and Pfister, K.K. (1996) Functional analysis of dynactin and cytoplasmic dynein slow axonal transport. *J. Neurosci.*, **16**, 6742–6752.
- Abe, T.K., Tanaka, H., Iwanaga, T., Odani, S. and Kuwano R. (1997) The presence of the 50 kDa subunit of dynactin complex in the nerve growth cone. *Biochem. Biophys. Res. Commun.*, **233**, 295–299.
- Waterman-Storer, C.M., Karki, S.B., Kuznetsov, S.A., Tabb, J.S., Weiss, D.G., Langford, G.M. and Holzbaur, E.L. (1997) The interaction between cytoplasmic dynein and dynactin is required for fast axonal transport. *Proc. Natl Acad. Sci. USA*, **94**, 12180–12185.
- Ahmad, F.J., Echeverri, C.J., Vallee, R.B. and Baas, P.W. (1998) Cytoplasmic dynein and dynactin are required for the transport of microtubules into the axon. *J. Cell Biol.*, **140**, 391–401.
- Lantz, V.A. and Miller, K.G. (1998) A class VI unconventional myosin is associated with a homologue of a microtubule-binding protein, cytoplasmic linker protein-170, in neurons and at the posterior pole of drosophila embryos. *J. Cell Biol.*, **140**, 897–910.
- Zhang, W., Vazquez, L., Apperson, M. and Kennedy, M.B. (1999) Citron binds to PSD-95 at glutamatergic synapses on inhibitory neurons in the hippocampus. *J. Neurosci.*, **19**, 96–108.
- Kuroda, S., Nakagawa, N., Tokunaga, C., Tatematsu, K. and Tanizawa K. (1999) Mammalian homologue of the *Caenorhabditis elegans* UNC-76 protein involved in axonal outgrowth is a protein kinase C zeta-interacting protein. *J. Cell Biol.*, **144**, 403–411.
- Nagy, J.I., Price, M.L., Staines, W.A., Lynn, B.D. and Granholm, A.C. (1998) The hyaluronan receptor RHAMM in noradrenergic fibers contrib-



- utes to axon growth capacity of locus coeruleus neurons in an intraocular transplant model. *Neuroscience*, **86**, 241–255.
16. Dolnick, B.J. (1997) Naturally occurring antisense RNA. *Pharmacol. Ther.*, **75**, 179–184.
  17. Knee, R. and Murphy, P.R. (1997) Regulation of gene expression by natural antisense RNA transcripts. *Neurochem. Int.*, **31**, 379–392.
  18. Constancia, M., Pickard, B., Kelsey, G. and Reik, W. (1998) Imprinting mechanisms. *Genome Res.*, **8**, 881–900.
  19. Devon, R.S., Evans, K.L., Maule, J.C., Christie, S., Anderson, S., Brown, J., Shibasaki, Y., Porteous, D.J. and Brookes, A.J. (1997) Novel transcribed sequences neighbouring a translocation breakpoint associated with schizophrenia. *Am. J. Med. Genet.*, **74**, 82–90.
  20. Fletcher, J.M., Evans, K., Baillie, D., Byrd, P., Hanratty, D., Leach, S., Julier, C., Gosden, J.R., Muir, W., Porteous, D.J. *et al.* (1993) Schizophrenia-associated chromosome 11q21 translocation—identification of flanking markers and development of chromosome 11q fragment hybrids as cloning and mapping resources. *Am. J. Hum. Genet.*, **52**, 478–490.
  21. Minoru, K., Motoyuki, S., Hiroyuki, M., Hideto, Y. and Mitsuo, O. (1989) Construction of mouse A9 clones containing a single human chromosome tagged with neomycin-resistance gene via microcell fusion. *Jpn J. Cancer Res.*, **80**, 413–418.
  22. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
  23. Boyd, A.C. (1993) Turbo cloning—a fast, efficient method for cloning PCR products and other blunt-ended DNA fragments into plasmids. *Nucleic Acids Res.*, **21**, 817–821.
  24. Ewing, B., Hillier, L., Wendl, M. and Green, P. (1998) Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, **8**, 175–185.
  25. Ewing, B. and Green, P. (1998) Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186–194.
  26. Gordon, D., Abajian, C. and Green, P. (1998) A graphical tool for sequence finishing. *Genome Res.*, **8**, 195–202.
  27. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
  28. Evans, K.L., Brown, J., Shibasaki, Y., Devon, R.S., He, L., Arveiler, B., Christie, S., Maule, J., Baillie, D., Slorach, E. *et al.* (1995) A contiguous clone map over 3 Mb on the long arm of chromosome 11 across a balanced translocation associated with schizophrenia. *Genomics*, **28**, 420–428.
  29. Seki, N., Ohira, M., Nagase, T., Ishikawa, K., Miyajima, N., Nakajima, D., Nomura, N. and Ohara, O. (1997) Characterisation of cDNA clones in size-fractionated cDNA libraries from human brain. *DNA Res.*, **4**, 345–349.
  30. Sonnhammer, E.L. and Durbin, R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, GC1–GC10.

