

ARTICLE

Mutation screening and association analysis of six candidate genes for autism on chromosome 7q

Elena Bonora¹, Janine A Lamb¹, Gabrielle Barnby¹, Nuala Sykes¹, Thomas Moberly¹, Kim S Beyer^{2,14}, Sabine M Klauck², Firtz Poustka³, Elena Bacchelli⁴, Francesca Blasi⁴, Elena Maestrini⁴, Agatino Battaglia⁵, Demetrios Haracopos⁶, Lennart Pedersen⁶, Torben Isager⁶, Gunna Eriksen⁶, Birgitte Viskum⁷, Ester-Ulsted Sorensen⁷, Karen Brondum-Nielsen⁸, Rodney Cotterill⁹, Herman von Engeland¹⁰, Maretha de Jonge¹⁰, Chantal Kemner¹⁰, Karlijn Steggehuis¹⁰, Margret Scherpenisse¹⁰, Michael Rutter¹¹, Patrick F Bolton¹¹, Jeremy R Parr¹², Annemarie Poustka², Anthony J Bailey^{*,12}, Anthony P Monaco^{*,1}, and the International Molecular Genetic Study of Autism Consortium (IMGSAC)¹³

¹The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; ²Department of Molecular Genome Analysis, Deutsches Krebsforschungszentrum, Heidelberg, Germany; ³Department of Child and Adolescent Psychiatry, University of Frankfurt, Frankfurt, Germany; ⁴University of Bologna, Dipartimento di Biologia Evoluzionistica Sperimentale and Lab Genetica Medica, Policlinico S. Orsola-Malpighi, Bologna, Italy; ⁵Stella Maris Clinical Research Institute for Child and Adolescent Neuropsychiatry, Calambrone, Italy; ⁶Videnscenter Og Centre for Autisme, Virum, Denmark; ⁷Borne-Og Ungdomspsykiatrisk Hospital, Risskov, Denmark; ⁸John Kennedy Institutet, Glostrup, Denmark; ⁹Biophysics Group, Danish Technical University, Lyngby, Denmark; ¹⁰Department of Child and Adolescent Psychiatry, Utrecht, Netherlands; ¹¹Centre for Social, Genetic and Developmental Psychiatry and Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London; ¹²Section of Child and Adolescent Psychiatry, Park Hospital for Children, Oxford, UK; ¹³<http://www.well.ox.ac.uk/~maestrini/iat.html>

Genetic studies have provided evidence for an autism susceptibility locus (AUTS1) on chromosome 7q. Screening for mutations in six genes mapping to 7q, *CUTL1*, *SRPK2*, *SYPL*, *LAMB1*, *NRCAM* and *PTPRZ1* in 48 unrelated individuals with autism led to the identification of several new coding variants in the genes *CUTL1*, *LAMB1* and *PTPRZ1*. Analysis of genetic variants provided evidence for association with autism for one of the new missense changes identified in *LAMB1*; this effect was stronger in a subgroup of affected male sibling pair families, implying a possible specific sex-related effect for this variant. Association was also detected for several polymorphisms in the promoter and untranslated region of *NRCAM*, suggesting that alterations in expression of this gene may be linked to autism susceptibility.

European Journal of Human Genetics (2005) 13, 198–207. doi:10.1038/sj.ejhg.5201315

Published online 3 November 2004

Keywords: autism; candidate genes; brain development; mutation screening; linkage disequilibrium; transmission disequilibrium

*Correspondence: Professor AP Monaco, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. Tel: +44 1865 287502; Fax: +44 1865 287650; E-mail: anthony.monaco@well.ox.ac.uk or AJ Bailey, Section of Child and Adolescent Psychiatry, University of Oxford, Park Hospital for Children, Old Road, Headington, Oxford, OX3 7LQ, UK; Tel: +44 1865 226517;

Fax: +44 1865 762358; E-mail: anthony.bailey@psychiatry.oxford.ac.uk
¹⁴Current address: Molecular Neurogenetics Unit, Massachusetts General Hospital, Charlestown, MA 02129, USA
Received 17 May 2004; revised 31 August 2004; accepted 7 September 2004

Introduction

Autism is a severe neurodevelopmental disorder characterised by impairments in reciprocal communication and social interaction, accompanied by unusually restricted and stereotyped patterns of behaviours and interests, and an onset in the first 3 years of life.¹ The population prevalence of autism is approximately 10–30/10 000,² with a male to female ratio of 4:1.³ When other pervasive developmental disorders (PDD) are also considered, the prevalence may be as high as 20–60 in 10 000 children.^{4,5} In several epidemiological studies of autism, the most consistent anatomical result is macrocephaly.^{6,7} Neuroanatomical findings point to abnormalities in the cerebral cortex, cerebellum, and brainstem.^{8–10}

Twin and family studies have indicated a complex genetic predisposition to autism.^{3,11–13} and statistical models suggest that between two and 10 loci are implicated.¹⁴ Several genome scans for autism susceptibility loci have been completed, providing evidence that the long arm of human chromosome 7 is likely to contain an autism susceptibility locus (AUTS1) (reviewed in Folstein and Rosen-Sheidley¹⁵).

In order to identify autism susceptibility genes on 7q, we have systematically screened functional candidate genes, mapping to the region of linkage, for the presence of etiological mutations/variants. Here, we report the analysis of six candidate genes mapping to the region of linkage and with neuronal function: *CUTL1*, *SRPK2*, *SYPL*, *LAMB1*, *NRCAM*, and *PTPRZ1*.

CUTL1 (Cut-like 1) is the human homologue of *Drosophila melanogaster* gene *Cut*, which has a role in determining and maintaining cell-type specificity.¹⁶ The full-length protein contains a homeodomain and acts as a repressor of transcription.¹⁷ One of the alternative forms, (Cut alternative spliced product (*CASP*)) lacks the DNA-binding domains and is a transmembrane protein of the Golgi system.¹⁸ The gene *SRPK2* encodes for serine arginine protein kinase isoform 2, a member of specific kinases for SR-rich splicing factors¹⁹ with a brain-restricted expression pattern.²⁰ *SYPL* encodes for synaptophysin-like protein, a major integral calcium-binding molecule required for vesicle fusion in synapses.²¹ *LAMB1* encodes for the $\beta 1$ chain of laminin, an extracellular matrix (ECM) glycoprotein complex.²² Laminins promote neuronal migration and neurite outgrowth in the developing nervous system.²³ *NRCAM* encodes for Bravo/NrCAM (NgCAM-related cell adhesion molecule) protein, a member of the immunoglobulin superfamily of cell adhesion proteins.²⁴ NrCAM proteins promote directional signaling during axonal cone growth.²⁵ *PTPRZ1* encodes for protein tyrosine phosphatase receptor type Z, a transmembrane protein expressed primarily in the CNS, during development and in adult brain.²⁶

Materials and methods

IMGSAC multiplex and singleton families

The identification of families, assessment methods, and inclusion criteria used by the IMGSAC have been described previously.²⁷ In families passing an initial screen, parents were administered the ADI-R²⁸ and the Vineland Adaptive Behaviour Scales.²⁹ Potential cases were assessed using the ADOS.³⁰ Physical examination was undertaken to exclude recognizable medical causes of autism, particularly tuberous sclerosis. Karyotyping was performed when possible on all affected individuals and molecular genetic testing for Fragile X performed on one case per family.²⁷ Families have been collected in six successive waves for a total of 207 families comprising 219 nonindependent affected sibling pairs (ASP) (145 male–male ASP, 59 male–female ASP and 15 female–female ASP).

The identification and assessment of IMGSAC singletons was similar to the multiplex families; a total of 98 singleton families from the UK, Netherlands and Denmark were included in the study. A total of 112 German singleton families subdivided into groups A (63 male and 21 female cases) and B (24 male and four female cases) with individuals from group B showing no delay in the development of language³¹ and 42 Italian singleton families were also included. Written informed consent was given by all parents/guardians and, where possible, by affected individuals. The study has been approved by the relevant ethical committees.

Gene characterization

The genomic structure for each gene was obtained by BLAST comparison (<http://www.ncbi.nlm.nih.gov/BLAST>) of the coding mRNAs with the genomic sequence (TCAG website; <http://www.chr7.org>). Exon–intron boundaries were identified and primers designed to cover exons and regulatory splice site regions using the program Primer3 (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Promoter regions were determined using Promoterscan (<http://zeon.well.ox.ac.uk>). Sequences and PCR conditions of all primer pairs are available on request. *CUTL1* covers a genomic region of 470 kb and comprises 33 exons. The full-length *CUTL1* mRNA (Accession no. NM_181552) contains exon 1b – 24. *CASP* mRNA (Accession no. NM_001913) contains exons 1a – 14 and 25–33.³² *SRPK2* (mRNA Accession no. NM_182691) extends over a genomic region of 153 kb, comprising 15 exons; *SYPL* (mRNA Accession no. NM_182715) covers a genomic region of 23 kb and is composed of six exons; *LAMB1* (mRNA Accession no. NM_002291) extends over a region of 95 kb, comprising 34 exons. *NRCAM* covers a genomic region of 380 kb and contains 34 exons (Accession no. NM_005010). Different transcripts of *NRCAM* are produced by alternative splicing of exons 10, 19, and 27–29. *PTPRZ1* covers a genomic region of 189 kb and contains 30 exons (Accession no. NM_002851).

Mutation screening by denaturing high-performance liquid chromatography (DHPLC)

Genomic DNA was extracted from blood as described previously.³³ Genomic DNA extracted from buccal swabs was preamplified using GenomiPhi according to manufacturer's instructions (Amersham Pharmacia Biotech). PCR amplifications and DHPLC analysis were performed as described previously.³⁴ Samples showing a variant DHPLC pattern were reamplified and sequenced on both strands using BigDye v3.0 (Applied Biosystems) according to the manufacturer's instruction to determine the nature of the heterozygous changes. Sequences were loaded on ABI377 sequencing machines (Applied Biosystems) and analyzed using Sequence Navigator v3.1.

Prediction analysis of amino-acid substitutions

PolyPhen (<http://tux.embl-heidelberg.de/ramensky/polyphen.cgi>) was used to predict the possible impact of amino-acid substitutions on the protein. The program is based on sequence comparison with homologous proteins; profile scores, position-specific independent counts (PSIC) are generated for the allelic variants and represent the logarithmic ratio of the likelihood of a given amino-acid occurring at a particular site relative to the likelihood of this amino-acid occurring at any site (background frequency). PSIC score differences above 2 indicate a damaging effect; scores between 1.5 and 2 suggest that the variant is possibly damaging, whereas scores below 0.5 indicate that the variant is benign.³⁵

Single-nucleotide polymorphism (SNP) genotyping

The insertion in exon 12 of *PTPRZ1* was fluorescently genotyped on ABI377 sequencing machines, as described previously.²⁷ The SNPs in intron 3 of *CUTL1*, exon 20 of *LAMB1*, and exon 5 of *PTPRZ1* were genotyped by restriction digestion using the enzymes *TaqI*, *AluI*, and *AcI* (New England Biolabs), respectively, according to standard protocols. In the German and Italian singleton samples, the missense change in exon 30 of *LAMB1* was genotyped by restriction digestion using the enzyme *AflIII*. In the IMGSAC sample, the missense change in exon 30 of *LAMB1* was genotyped using the MassARRAY™ primer extension system (see below). The SNP in exon 1 of *NRCAM* could not be distinguished by a commercial restriction enzyme; therefore, mismatch primers inserting a *BsaI* site were created using Insizer (<http://zeon.well.ox.ac.uk>).

MassARRAY™ primer extension

In total 23 SNPs selected from the SNP consortium (<http://snp.cshl.org>) and 17 SNPs identified in our mutation screening were genotyped using the MassARRAY™ system. Genotyping assays were designed using Sequenom's SpectroDESIGNER™ software (Version 1.3.4) and genotypes obtained using the MassARRAY™ system. Multiplex PCR

amplifications were performed in 384-well plates in a final volume of 10 μ l using 24 ng of genomic DNA as described previously.³⁶ Primers and conditions are available on request. Genotyping was performed using the matrix-assisted laser desorption time of flight (MALDI-TOF) technology with the Bruker Biflex III Mass Spectrometer system, as described previously.³⁶ Genotypes were assigned using the SpectroTYPER™ software.

Error checking

The LIMS Integrated Genotyping System database was used to store all genotypic and phenotypic data and to produce files for statistical analysis (<http://bioinformatics.well.ox.ac.uk/project-lims.shtml>). Genotypes were checked for Mendelian consistency using PedCheck.³⁷ Haplotypes were constructed using Genehunter v2.0 and, in cases where apparent excess recombination was observed, genotypes were rechecked and corrected where necessary.

Prior to statistical analysis, SIBMED was run on data from the multiplex families to identify any remaining possible genotyping errors³⁸ using a false-positive rate of 0.001 and a prior genotyping error rate of 0.01. All the SNPs were tested for Hardy–Weinberg equilibrium.

Association analysis

Association was studied using the transmission disequilibrium test (TDT)³⁹ with the *sib_tdt* option from ASPEX v2.3.⁴⁰ This program calculates probabilities for χ^2 statistics by permuting parental alleles while fixing the IBD status of siblings within a family, thereby allowing the use of multiple siblings within a nuclear family. Further analysis at the AUTS1 has suggested that linkage derives mainly from the male ASP, and parent-of-origin linkage modelling indicates two distinct regions of paternal and maternal linkage on chromosome 7 (IMGSAC, unpublished data); therefore, transmissions to male–male ASP and parental transmissions were also examined.

Linkage disequilibrium

The extent of linkage disequilibrium (LD) between intra-genic SNPs was studied using the Haploxt program⁴¹ and characterised with Lewontin's standardised measure of disequilibrium D' .⁴²

Haplotype analysis

Haplotypes were reconstructed for all 219 ASP using MERLIN.⁴³ Haplotypes were recoded as single markers, with each haplotype combination considered as a different 'allele'. Transmission was studied using the *sib_tdt* option from ASPEX v2.3. Haplotype transmission was analysed for SNPs showing nominally significant association at the single-locus level and flanking markers.

Results

Mutation screening

A total of 48 (46 males, two females) unrelated individuals with autism from the multiplex IMGSAC families were screened for sequence variants by DHPLC. Individuals were selected from families showing increased identical by descent (IBD) sharing in the region surrounding the candidate genes, where ASP were IBD1 or IBD2 across a ~15 Mb region containing 15 microsatellite markers. In total, 38 individuals had a clinical diagnosis of autism, met ADI-R and ADOS criteria for autism, and had a history of language delay and a performance IQ ≥ 35 ; the other individuals met the criteria for PPDD or Asperger syndrome. We identified a total of 112 sequence variants: 26 in *CUTL1*, six in *SRPK2*, two in *SYPL*, 32 in *LAMB1*, 25 in *NRCAM*, and 21 in *PTPRZ1*. Comparison with dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) identified 39 changes as known SNPs. Nine changes led to amino-acid substitutions, insertions or deletions. The frequency and positions of the changes identified through our screening are shown in Table 1 and Supplementary Tables 1 and 2 (see Supplementary Tables online). The presence of all missense variants and insertion–deletions was tested in a control group of 192 random Caucasian individuals from the European Collection of Cell Cultures (ECACC). Differences in frequencies of heterozygous individuals in cases (48 individuals) and controls (192 individuals) were calculated using Fisher's exact test. The deletion of amino-acid K1256 in *CUTL1* was not identified in 192 controls. It is located in the homeodomain and maps to a conserved position in Cut proteins. Analysis of the crystal structure of homologous proteins in complexes with DNA suggests that K1256 may interact with the deoxyribose-phosphate backbone (R Esnouf, personal communication). This change is transmitted from the father to all three sons with autism, and not to the unaffected brother, but also to the unaffected sister. Phenotypic investigation of all family members showed that the parents, the non-autistic son,

and daughter present some difficulties in socio-emotional interactions and/or circumscribed interests. However, since both the father and son appear to have the broader autism phenotype,⁴⁴ this variant does not always segregate with the phenotype in the family (see Supplementary Figure 1 online). The deletion was investigated in 342 individuals with autism from 169 multiplex families, and was not identified in other subjects.

LAMB1 and PTPRZ1 coding changes

Four missense changes were identified in *LAMB1* in exons 19, 20, 22, and 30. The G3400A (R1022Q) in exon 22 (rs20556) and G2913A (G860S) in exon 20 were recently described;⁴⁵ C2718G (R795G) in exon 19, and T4975C (I1547T) in exon 30 are new variants. Comparison with homologous proteins in different species (Figure 1) shows that I1547 is conserved, whereas for R795 a positively charged residue is always present. *In silico* methods have been developed to predict the potential of amino-acid substitutions to impact protein structure and activity, based on sequence conservation, physical, and chemical properties of the exchanged residues, and/or protein structural domain information. We used PolyPhen (Polymorphism Phenotyping), one of these algorithms, to study if the new variants might have a functional role. PolyPhen is based on all the previous characteristics, but also values the location of the substitution within identified functional domains and known structural features available in the annotated database (SwissProt). Testing PolyPhen using known variants confirmed its ability of discriminating between benign and deleterious variants and its high concordance with other algorithms.⁴⁶

The I1547T change was predicted to be damaging (PSIC score 2.166). The frequency of this variant was higher in the autistic sample compared to the control group (10.4% versus 3.66%, respectively, $P=0.0466$; Table 1) and it cosegregated with the autism phenotype in the families included in the mutation screening.

Table 1 Nonconservative coding changes identified in the mutation screening of *CUTL1*, *LAMB1*, *NRCAM*, and *PTPRZ1*

| Variant | Position in cDNA | Type of aa change | dbSNP | Het autism (N = 48) (%) ^a | Het ECACC controls (N = 192) (%) ^b | χ^2 P-value ^c |
|--------------------|----------------------|-------------------|-----------|--------------------------------------|---|-------------------------------|
| <i>CUTL1</i> × 23 | Δ(AAG) bp3808–3810 | ΔK1256 | — | 2.1 | 0 | — |
| <i>LAMB1</i> × 19 | C2718G | R795G | — | 2.1 | 3.1 | NS |
| <i>LAMB1</i> × 20 | G2913A | G860S | — | 6.3 | 11.1 | NS |
| <i>LAMB1</i> × 22 | G3400A | R1022Q | rs20556 | 20.8 | 27.9 | NS |
| <i>LAMB1</i> × 30 | T4975C | I1547T | — | 10.4 | 3.7 | 0.0466 |
| <i>NRCAM</i> × 16 | C2045G | P539A | rs6958498 | 27.3 | 21.0 | NS |
| <i>PTPRZ1</i> × 1 | T155G | I3S | — | 29.2 | 13.8 | 0.0201 |
| <i>PTPRZ1</i> × 12 | G4297A | D1384N | — | 4.2 | 10.0 | NS |
| <i>PTPRZ1</i> × 12 | ins(TGA) bp4408–4411 | ins(D1422) | — | 48.8 | 53.0 | NS |

The dbSNP accession number is given where available. Abbreviations: na: not available; NS: not significant.

^aFrequency of autistic individuals heterozygous for the change.

^bFrequency of control individuals heterozygous for the change.

^cP-values for differences in frequency of heterozygous cases and controls.

Table 2 SNP ID, position, sequences and TDT *P*-values of all the markers genotyped across the candidate genes

| SNP ID | Gene | Position on 7q (Mb) | SNP sequences and alleles | TDT combined P 219 ASP | TDT combined P 145 male–male ASP |
|-------------|--------|---------------------|---------------------------|---------------------------|-------------------------------------|
| CUTL1int3 | CUTL1 | 100704361 | TTTGTC[G/A]AGAGGT | 0.3651 | 0.7083 |
| CUTL1int8 | CUTL1 | 100749242 | TGTGTA[T/C]GGATAA | 0.4310 | 0.7401 |
| rs436287 | CUTL1 | 100788259 | AAATAG[T/G]TCAGGT | 0.2579 | 0.7401 |
| rs1107463 | CUTL1 | 100799805 | CTCCTC[A/G]TCTCAA | 0.8065 | 0.7642 |
| CUTL1 × 13 | CUTL1 | 100827596 | TGCACC[G/A]TCCGAG | 0.3994 | 0.1797 |
| rs1297632 | CUTL1 | 100833432 | CTCCCC[A/G]CTTAGG | 0.3833 | 0.0949 |
| CASP × 25 | CUTL1 | 100907195 | CTGTGC[A/G]GAGCTG | 0.8251 | 1.0000 |
| rs1019040 | SRPK2 | 103766255 | CCGCCT[C/T]AGGCAG | 0.1269 | 0.0701 |
| SRPK2int7 | SRPK2 | 103794359 | AAATAC[A/C]GATTTT | 0.4706 | 0.2003 |
| rs1010340 | SRPK2 | 103824126 | TATGAC[C/T]GCCCAT | 0.4976 | 0.3222 |
| rs917114 | SRPK2 | 103900918 | ACTGTG[C/T]ATGAAG | 0.1277 | 0.3078 |
| SYPLint2 | SYPL | 104733287 | CTTTGT[G/A]TAGATT | 0.0212 | 0.2334 |
| rs289178 | SYPL | 104742585 | TAAAAC[A/G]CATGAA | 0.6801 | 0.8065 |
| rs176501 | SYPL | 104746429 | GACGAC[A/G]GGCCGG | 0.9203 | 0.6892 |
| LAMB1 × 30 | LAMB1 | 106563727 | GAAGATA[T/C]ACGTG | 0.0112 | 0.0016 |
| LAMB1int24 | LAMB1 | 106574700 | CTACTT[A/G]ATCATG | 0.1253 | 0.2488 |
| LAMB1 × 22 | LAMB1 | 106587764 | CCCTCC[G/A]GCAGGA | 0.4386 | 0.6242 |
| LAMB1 × 20 | LAMB1 | 106593604 | GAAAGC[C/T]CCAGTG | 0.0942 | 0.7773 |
| LAMB1 × 19 | LAMB1 | 106593999 | TGGGCC[G/C]GCACTG | 0.5270 | 0.1022 |
| LAMB1 × 15 | LAMB1 | 106597087 | CCCTTC[A/C]GGCACT | 0.8624 | 0.8414 |
| rs737955 | LAMB1 | 106608875 | CCCAGA[A/G]GTGGGG | 0.1221 | 0.5776 |
| LAMB1 × 7 | LAMB1 | 106620623 | AAAGTC[G/A]GACGAT | 0.6985 | 0.4386 |
| rs2528658 | LAMB1 | 106634310 | AACAGT[C/G]TGATTC | 0.7083 | 0.3741 |
| NRCAMint31 | NRCAM | 106794000 | TTCCAT[A/G]CACAAG | 0.1532 | 0.3455 |
| NRCAMint30 | NRCAM | 106795500 | TTCCAT[A/G]CACAAG | 0.2488 | 0.3802 |
| NRCAM × 14 | NRCAM | 106838376 | CTGATA[G/T]ACTGCA | 0.8414 | 1.0000 |
| rs1269622 | NRCAM | 106853249 | AAGCAC[A/G]AAGAAC | 0.0135 | 0.0489 |
| NRCAMint8 | NRCAM | 106871290 | ATCTCT[A/G]AAACAT | 0.0578 | 0.1522 |
| NRCAM × 4 | NRCAM | 106880618 | TGCGGG[A/G]AACTGA | 0.3349 | 0.4166 |
| rs2111203 | NRCAM | 106893605 | AGCTGT[A/G]TGAGAA | 0.1659 | 0.3833 |
| rs1269688 | NRCAM | 106925283 | ACAGCT[A/C]CTGGCA | 0.8625 | 0.8624 |
| rs2052314 | NRCAM | 106988688 | CAGTGA[A/T]AAATTT | 0.2059 | 0.1049 |
| rs2395961 | NRCAM | 107010162 | TATAAT[C/T]AAGCAT | 0.8875 | 0.8624 |
| rs2267891 | NRCAM | 107025336 | GTGCTT[C/T]ATGAAG | 0.3833 | 0.2694 |
| rs722519 | NRCAM | 107026484 | GATGCA[A/G]GAAC TA | 0.0078 | 0.0793 |
| rs2160555 | NRCAM | 107027874 | TGAATA[T/C]TTTGCT | 0.4543 | 0.4348 |
| rs1990162 | NRCAM | 107055185 | ACAGAG[A/T]AGGGGG | 0.7082 | 0.9203 |
| rs917251 | NRCAM | 107087507 | CAGAGT[A/G]TCTACT | 0.5541 | 1.0000 |
| rs3763463 | NRCAM | 107098150 | ACCGCG[C/G]ATAGAT | 0.0668 | 0.0853 |
| PTPRZ1 × 1 | PTPRZ1 | 120538409 | TGCGAA[T/G]CCTAAA | 0.1750 | 0.4237 |
| rs740960 | PTPRZ1 | 120551052 | TAAAGA[A/C]CCTGTG | 0.3833 | 0.4166 |
| rs727766 | PTPRZ1 | 120582727 | AATTAT[C/G]AATTTA | 0.6892 | 0.6892 |
| PTPRZ1 × 5 | PTPRZ1 | 120641157 | TGATGC[G/A]GACCGA | 0.2222 | 0.3009 |
| PTPRZ1 × 12 | PTPRZ1 | 120677666 | CAC[TGA]TGATGATGG | 0.2794 | 0.6629 |
| PTPRZ1int27 | PTPRZ1 | 120719926 | TAAC TA[A/C]CATTAC | 0.7642 | 0.8625 |

Position of the SNPs is according to the TCAG consortium release December 19, 2003 (www.chr7.org).

The phenotypic characteristics of autistic individuals carrying the missense change in exon 30 were compared with the other individuals included in *LAMB1* screening who did not carry the change, to identify a possible distinguishing phenotype, but no significant differences were identified.

The T155G change in exon 1 of *PTPRZ1* was found with increased frequency in the autistic cases compared to controls (29.2 versus 13.8%, respectively, $P=0.0201$), but when the analysis was extended to the relatives, the change was found to segregate with the autism phenotype only in 46% of the families. In addition, the change (I3S) resides in

the N-terminal signal domain of the protein and no negative effect on the protein is predicted for this variant.

Association studies

A total of 45 SNPs were genotyped across the candidate genes and tested for association (Table 2). Association analysis was performed using the TDT implemented in ASPEX. Analysis was performed on the whole sample of 219 ASP and in the subset of 145 male–male pairs, considering the sex of the individuals as a possible index of heterogeneity (IMGSAC, unpublished data). Combined *P*-values for all SNPs are reported in Table 2. Transmissions

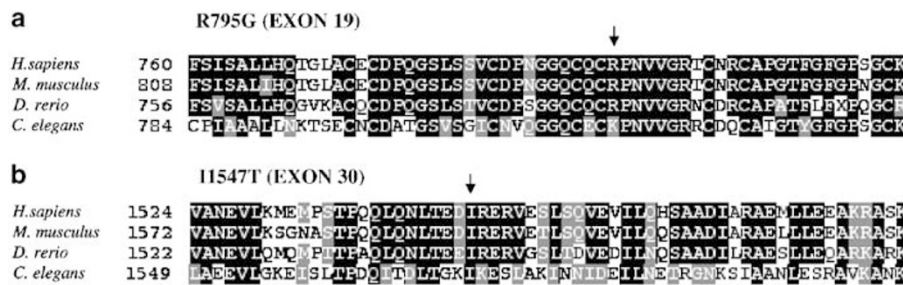


Figure 1 Multiple sequence alignment of laminin $\beta 1$ from different species (Accession No: NP002282 (*H. sapiens*), AAA39407 (*Mus musculus*), NP775382 (*D. rerio*), NP500734 (*C. elegans*)) across the regions with new missense changes. Completely conserved residues are shaded in black, partially conserved residues in grey. Arrows indicate the substituted amino-acids identified through the mutation screening.

for SNPs showing a nominally significant P -value are reported in Table 3. The I1547T variant in *LAMB1* (*LAMB1* $\times 30$, C allele) showed a preferential paternal transmission (15 transmissions *versus* four not transmitted, $P=0.0116$; combined parental, $P=0.0112$) in all families. This association is more significant in the male ASP (13 paternal transmissions *versus* 0 not transmitted, $P=0.0003$, combined parental $P=0.0016$). In *NRCAM*, three SNPs showed transmission disequilibrium (Table 3). Association analysis in the 219 ASP showed a combined preferential transmission of allele A for SNP rs1269622 (combined $P=0.0135$); preferential transmission of allele G for SNP rs722519 (paternal $P=0.0047$; combined $P=0.0078$) – maternal transmission showed the same trend, although not reaching statistical significance – and preferential paternal transmission of allele G ($P=0.0045$; combined $P=0.0668$) for SNP rs3763463. Nominally significant association was detected for the SNP in intron 2 of *SYPL* in all families (combined parental, $P=0.0212$, maternal $P=0.0278$); however, given the number of tests performed, this result may have arisen by chance.

LD patterns

The profile of LD across each gene was studied using the Haploxt program and the Lewontin D' standardised measure of LD. Extensive LD is present between markers rs436287 and rs1297632 in *CUTL1* and across *SRPK2*. In *SYPL*, high LD was detected between the intron 2 SNP and marker rs176501 ($D'=0.839$). No significant LD was found across *PTPRZ1* (data not shown). The profile of LD across *LAMB1* and *NRCAM* is shown in Figure 2.

Haplotype transmission study

Haplotype analysis was performed to further characterise the SNPs showing evidence for association. The determination of haplotypes of SNPs in LD offers more power to detect association than testing of SNPs individually.⁴⁷

Haplotype A–A for markers *SYPL*int2-rs2891878 and C–A for markers *LAMB1* $\times 30$ and *LAMB1*int24 were

overtransmitted, but statistical significance was not increased compared to single markers (data not shown).

In *NRCAM*, analysis was first performed for the two markers showing evidence of association, SNPs rs722519, and rs3763463. rs3763463 and rs722519 are located in the promoter and in the second intron of *NRCAM*, respectively. Subsequent analysis was extended to two, three, and four SNP haplotypes including markers rs1990162 and rs917251 between the two SNPs (SNP rs216055 was not included, since it was not in LD with marker rs722519). Significant preferential transmission was detected for the G–G haplotype of markers rs722519 and rs3763463 ($P=0.0005$; Table 4) and for haplotype G–C–G of rs722519–rs917251–rs3763463 ($P=0.0001$). Analysis of haplotypes containing the A allele of SNP rs1269622 did not increase statistical significance compared to the single-marker test (data not shown).

Replication in singleton families

The five SNPs showing evidence for association were further studied in an independent sample of 98 IMGSAC singleton families. The intron 2 SNP in *SYPL* and SNPs rs1269622, rs722519, and rs3763463 in *NRCAM* did not provide evidence for association ($P>0.05$; data not shown). A nonsignificant trend for a preferential transmission of the C allele was observed for the exon 30 variant in *LAMB1* (six transmissions *versus* one nontransmission, $P=0.0588$). This missense change was further studied in 154 individuals with autism (42 singletons from Italy, 112 singletons from Germany) and was identified at a frequency similar to the control group (data not shown).

Discussion

A defining feature of complex phenotypes is that no single locus contains alleles that are necessary or sufficient for disease susceptibility, but little is known about the nature of genetic variation underlying human complex diseases. One problem is identifying whether genetic variance is due to a small number of loci where susceptibility alleles are

Table 3 TDT results for the SNPs showing association in *SYPL*, *LAMB1*, and *NRCAM*

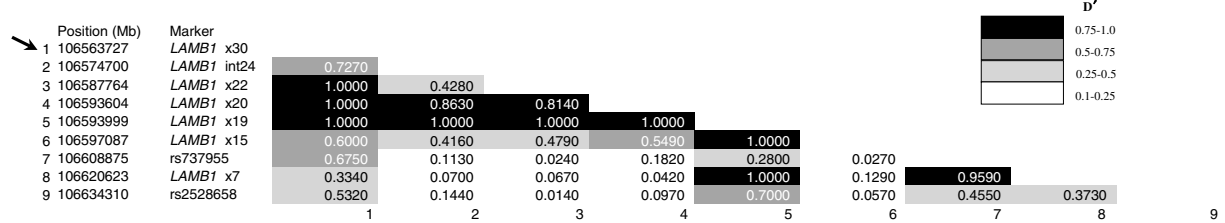
| SNP ID | ASP group | Allele | % | TR | Paternal | | Maternal | | | Combined | | |
|-------------------|--------------|----------|------|----|------------|----------|------------|----|----------|------------|-----|----------|
| | | | | | NT | χ^2 | TR | NT | χ^2 | TR | NT | χ^2 |
| <i>SYPL</i> int2 | 219 ASP | Allele A | 23.9 | 65 | 51 | 1.69 | 61 | 39 | 4.84 | 140 | 104 | 5.31 |
| | | Allele G | 76.1 | 51 | 65 | 1.69 | 39 | 61 | 4.84 | 104 | 140 | 5.31 |
| | | | | | $P=0.1936$ | | $P=0.0278$ | | | $P=0.0212$ | | |
| | 145 male ASP | Allele A | 23.4 | 47 | 39 | 0.74 | 32 | 25 | 0.86 | 87 | 72 | 1.42 |
| | | Allele G | 76.6 | 39 | 47 | 0.74 | 25 | 32 | 0.86 | 72 | 87 | 1.42 |
| | | | | | $P=0.3897$ | | $P=0.3537$ | | | $P=0.2334$ | | |
| <i>LAMB1</i> × 30 | 219 ASP | Allele C | 2.5 | 15 | 4 | 6.37 | 10 | 6 | 1.00 | 25 | 10 | 6.43 |
| | | Allele T | 97.5 | 4 | 15 | 6.37 | 6 | 10 | 1.00 | 10 | 25 | 6.43 |
| | | | | | $P=0.0116$ | | $P=0.3173$ | | | $P=0.0112$ | | |
| | 145 male ASP | Allele C | 2.0 | 13 | 0 | 13.0 | 2 | 2 | 0.00 | 15 | 2 | 9.94 |
| | | Allele T | 98.0 | 0 | 13 | 13.0 | 2 | 2 | 0.00 | 2 | 15 | 9.94 |
| | | | | | $P=0.0003$ | | $P=1.0000$ | | | $P=0.0016$ | | |
| NRCAM rs1269622 | 219 ASP | Allele A | 56.1 | 56 | 35 | 4.85 | 49 | 32 | 3.57 | 138 | 100 | 6.07 |
| | | Allele G | 43.9 | 35 | 56 | 4.85 | 32 | 49 | 3.57 | 100 | 138 | 6.07 |
| | | | | | $P=0.0276$ | | $P=0.0588$ | | | $P=0.0135$ | | |
| | 145 male ASP | Allele A | 57.1 | 36 | 25 | 1.98 | 36 | 22 | 3.38 | 93 | 68 | 3.88 |
| | | Allele G | 42.9 | 25 | 36 | 1.98 | 22 | 36 | 3.38 | 68 | 93 | 3.88 |
| | | | | | $P=0.1593$ | | $P=0.0660$ | | | $P=0.0489$ | | |
| NRCAM rs722519 | 219 ASP | Allele A | 14.7 | 15 | 35 | 8.00 | 22 | 31 | 1.53 | 45 | 74 | 7.07 |
| | | Allele G | 85.3 | 35 | 15 | 8.00 | 31 | 22 | 1.53 | 74 | 45 | 7.07 |
| | | | | | $P=0.0047$ | | $P=0.2161$ | | | $P=0.0078$ | | |
| | 145 male ASP | Allele A | 13.7 | 8 | 17 | 3.24 | 15 | 21 | 1.00 | 29 | 44 | 3.08 |
| | | Allele G | 86.3 | 17 | 8 | 3.24 | 21 | 15 | 1.00 | 44 | 29 | 3.08 |
| | | | | | $P=0.0718$ | | $P=0.3173$ | | | $P=0.0793$ | | |
| NRCAM rs3763463 | 219 ASP | Allele C | 9.2 | 4 | 17 | 8.05 | 21 | 25 | 0.35 | 31 | 48 | 3.66 |
| | | Allele G | 90.8 | 17 | 4 | 8.05 | 25 | 21 | 0.35 | 48 | 31 | 3.66 |
| | | | | | $P=0.0045$ | | $P=0.4583$ | | | $P=0.0668$ | | |
| | 145 male ASP | Allele C | 9.7 | 3 | 11 | 4.57 | 14 | 19 | 0.76 | 22 | 35 | 2.96 |
| | | Allele G | 90.3 | 11 | 3 | 4.57 | 19 | 14 | 0.76 | 35 | 22 | 2.96 |
| | | | | | $P=0.0325$ | | $P=0.3833$ | | | $P=0.0853$ | | |

common, or due to a much larger number of loci where susceptibility alleles are quite rare.⁴⁸ If allelic variation at a complex disease locus is extensive, with multiple susceptibility alleles of independent origin present, the detection of association between marker genotypes and disease phenotype might be negatively affected.⁴⁹ Several independent linkage studies point to the presence of an autism locus on chromosome 7q (reviewed in Folstein and Rosen-Sheidley¹⁵). Nonetheless, the region of interest is very broad and contains at least 190 known genes.⁵⁰ Further characterization of the AUTS1 in 219 ASP suggests that linkage may derive mainly from the male ASP, and parent-of-origin linkage modelling indicates two distinct regions of paternal and maternal linkage (IMGSAC, unpublished data). A combination of approaches has been undertaken to identify the chromosome 7q autism susceptibility gene(s): a candidate gene-screening approach, focused on the direct detection of susceptibility variants even with low

frequency, and family-based association studies on the genes tested for mutation, in order to detect association even if the susceptibility alleles were not directly tested.

A novel missense variant (I1547T) in exon 30 of *LAMB1* showed a preferential paternal transmission of the rare new allele. A possible involvement of this variant in autism susceptibility is intriguing, as amino-acid I1547 is conserved from *Caenorhabditis elegans* to *Homo sapiens*, and this change is predicted to have a damaging effect on protein structure. In addition, this effect is more marked when considering only the male ASP, in concordance with the linkage findings. However, the frequency of this SNP in the multiplex families does not explain all the linkage at the AUTS1 locus and suggests that it might play a role only in a subset of individuals with autism. Interestingly, in an independent sample of 98 singleton families the same trend for preferential transmission of the change was observed, although no difference in frequency was

a LD pattern for *LAMB1*



b LD pattern for *NRCAM*

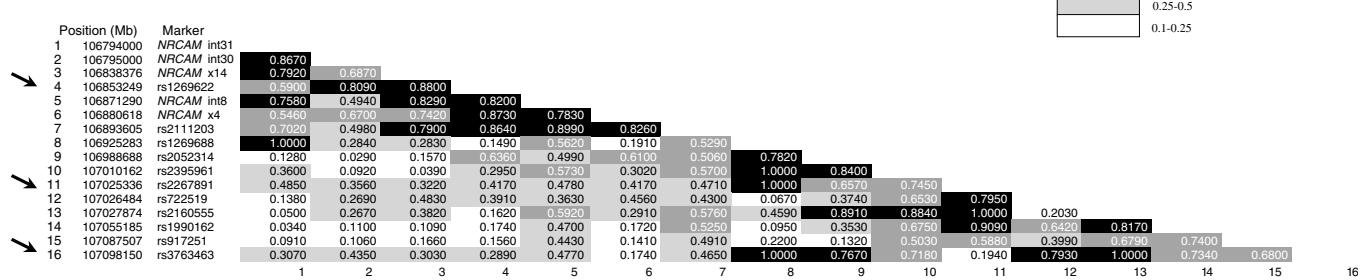


Figure 2 Patterns of LD for *LAMB1* (a) and *NRCAM* (b) genes. Pairwise estimation of D' is shown. Arrows indicate SNPs showing association.

Table 4 Haplotype transmission disequilibrium results for *NRCAM*

| | Haplotypes | % | TR | Combined transmission | | χ^2_{max} |
|--|------------|------|-----|-----------------------|------------|----------------|
| | | | | NT | | |
| 2 SNP haplotype rs722519–rs376463 | G.G | 76.1 | 135 | 78 | 15.25 | |
| | A.G | 13.9 | 51 | 87 | 9.39 | |
| | G.C | 9.6 | 38 | 55 | 3.11 | |
| | | | | P = 0.0005 | | |
| 3 SNP haplotype rs722519–rs1990162–rs3763463 | G.A.G | 53.5 | 118 | 109 | 0.36 | |
| | G.T.G | 23.2 | 103 | 62 | 10.19 | |
| | A.T.G | 10.1 | 33 | 46 | 2.14 | |
| | G.A.C | 9.1 | 31 | 45 | 2.58 | |
| | A.A.G | 3.0 | 4 | 19 | 9.78 | |
| | | | | P = 0.0059 | | |
| 3 SNP haplotype rs722519–rs917251–rs3763463 | G.A.G | 47.7 | 84 | 84 | 0.00 | |
| | G.C.G | 27.4 | 98 | 45 | 19.64 | |
| | A.C.G | 9.5 | 24 | 39 | 3.57 | |
| | G.A.C | 8.3 | 24 | 33 | 1.42 | |
| | A.A.G | 4.8 | 6 | 32 | 17.79 | |
| | G.C.C | 1.7 | 5 | 3 | 0.50 | |
| | | | | P = 0.0001 | | |
| 4 SNP haplotype rs722519–rs1990162–rs917251–rs3763463 | G.A.A.G | 45.7 | 68 | 74 | 0.25 | |
| | G.T.C.G | 18.4 | 52 | 31 | 5.31 | |
| | A.T.C.G | 8.5 | 18 | 26 | 1.45 | |
| | G.A.C.G | 8.2 | 34 | 12 | 10.52 | |
| | G.A.A.C | 8.2 | 17 | 26 | 1.88 | |
| | G.T.A.G | 3.6 | 13 | 10 | 0.39 | |
| | A.A.A.G | 3.4 | 3 | 14 | 7.12 | |
| | G.A.C.C | 1.4 | 6 | 2 | 2.00 | |
| | | | | | P = 0.0046 | |

P_{max} values are reported from the sib_tdt test. Only haplotypes with frequency > 1% are reported.

detected in a second independent sample of 154 singleton cases compared to controls.

Association was identified for SNPs in the promoter (rs3763463) and noncoding regions of *NRCAM* (rs722519; rs1269622) and association was more significant when considering the haplotype transmissions. Analyses of SNPs using MatInspector (<http://genomatix.de>) revealed several potential binding sites for transcription factor, depending on the allele present (data not shown). These SNPs might affect regulatory regions, suggesting a potential alteration in *NRCAM* expression. Loss or abnormal expression of neuronal cell adhesion molecules leads to several neuronal faults.⁵¹ Defects in neuronal organisation have been reported in several *post-mortem* studies of brains of individuals with autism⁹ and there is preliminary evidence of abnormal axonal connectivity (P Luthert, A Dean and AJ Bailey, unpublished data).

The possible role of *LAMB1* and *NRCAM*, both with important roles in brain function, in autism susceptibility is intriguing but not clear. In *LAMB1* we detected association for a paternally transmitted rare allele; in *NRCAM* association derives from common alleles and haplotypes. Association between a complex disorder and common alleles has been shown for type II diabetes and the calpain-10 gene.⁵² Recently, Hutchenson *et al*⁴⁵ reported the analysis of several genes mapping to chromosome 7 in autism, including *LAMB1* and *NRCAM*, and presented evidence for positive association in *LAMB1* in their sample. Although they did not report the presence of the missense change in exon 30, this might have been hampered by the relatively small number of individuals with autism screened for mutations. Nonetheless, the presence of association in the same gene reported by an independent group corroborates our results in *LAMB1* and suggests that further analysis of this gene in autism is warranted. The lack of association for *NRCAM* in the work of Hutchenson *et al*, compared to ours, might be due to differences in the sample collection such as the inclusion–exclusion criteria, or the use of different analytical approaches and marker tested and the different size of family collections. The TDT approach we used is robust to confounder effects, which may hamper the success of classical case–control studies, and is an effective method to refine the localization of a susceptibility locus in a region of linkage, in the presence of LD.⁵³ When testing multiple marker loci, a correction such as the Bonferroni test should be considered to compute *P*-values. However, applying any correction for tightly linked markers would be too conservative;⁵⁴ thus, in the present study, no correction was applied.

In conclusion, the most interesting results we found are for *LAMB1* and *NRCAM*, whereas no evidence for a role of *CUTL1*, *SRPK2*, *SYPL* and *PTPRZ1* in autism susceptibility was detected. Replication of a finding in independent samples of affected individuals is always warranted to provide evidence for a true association. The recent report of association between autism and variants in *LAMB1* offers

the first support to our findings and for a role of *LAMB1* within autism. However, since it is recognised that contributions of multiple genes are probably required for autism, further studies are necessary to better clarify the role of *LAMB1* and *NRCAM* in neuronal development in relation to autism susceptibility.

Acknowledgements

We thank all the families who have participated in the study and the professionals who continue to make this study possible. We would like to thank John Broxholme for help in bioinformatic analysis, Lorne Lonie, Aaron Abbott and Rebecca Redhead for technical support, Clyde Francks, Silvia Paracchini, Angela Marlow, Yvonne Jones and Robert Esnouf for helpful discussions and critical comments. This study is funded in part by support from the UK Medical Research Council, The Wellcome Trust, BIOMED 2 (CT-97-2759), EC Fifth Framework (QLG2-CT-1999-0094), Telethon – Italy (GGP030227), the Janus Korczak Foundation, Deutsche Forschungsgemeinschaft, Fondation France Télécom, Conseil Régional Midi-Pyrénées, Danish Medical Research Council, Sofiefonden, the Beatrice Surovell Haskells Fond for Child Mental Health Research of Copenhagen, the Danish Natural Science Research Council (9802210), the National Institute of Child Health and Development (5-P01-HD-35482-02) and the National Institutes of Health (MO1 RR06022 GCRC NIH, NIH K05 MH01196, K02 MH01389). EB is funded by a University of Oxford Graduate Prize Studentship and APM is a Wellcome Trust Principal Research Fellow. AJB is the Cheryl and Reece Scott Professor of Psychiatry.

References

- 1 Volkmar FR, Lord C, Bailey A, Schultz RT, Klin A: Autism and pervasive developmental disorders. *J Child Psychol Psychiatry* 2004; **45**: 135–170.
- 2 Fombonne E: The prevalence of autism. *JAMA* 2003; **289**: 87–89.
- 3 Smalley SL, Asarnow RE, Spence MA: Autism and genetics. A decade of research. *Arch Gen Psychiatry* 1988; **45**: 953–961.
- 4 Fombonne E: The epidemiology of autism: a review. *Psychol Med* 1999; **29**: 769–786.
- 5 Chakrabarti S, Fombonne E: Pervasive developmental disorders in preschool children. *JAMA* 2001; **285**: 3093–3099.
- 6 Bailey A, Luthert P, Bolton P, Le Couteur A, Rutter M, Harding B: Autism and megalencephaly. *Lancet* 1993; **341**: 1225–1226.
- 7 Bolton PF, Roobol M, Allsopp L, Pickles A: Association between idiopathic infantile macrocephaly and autism spectrum disorders. *Lancet* 2001; **358**: 726–727.
- 8 Bauman ML: Brief report: neuroanatomic observations of the brain in pervasive developmental disorders. *J Autism Dev Disord* 1996; **26**: 199–203.
- 9 Bailey A, Luthert P, Dean A *et al*: A clinicopathological study of autism. *Brain* 1998; **121** (Part 5): 889–905.
- 10 Ritvo ER, Freeman BJ, Scheibel AB *et al*: Lower Purkinje cell counts in the cerebella of four autistic subjects: initial findings of the UCLA-NSAC Autopsy Research Report. *Am J Psychiatry* 1986; **143**: 862–866.
- 11 Folstein S, Rutter M: Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* 1977; **18**: 297–321.
- 12 Bailey A, Le Couteur A, Gottesman I *et al*: Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995; **25**: 63–77.
- 13 Bolton P, Macdonald H, Pickles A *et al*: A case–control family history study of autism. *J Child Psychol Psychiatry* 1994; **35**: 877–900.
- 14 Pickles A, Bolton P, Macdonald H *et al*: Latent-class analysis of recurrence risks for complex phenotypes with selection and

- measurement error: a twin and family history study of autism. *Am J Hum Genet* 1995; **57**: 717–726.
- 15 Folstein SE, Rosen-Sheidley B: Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet* 2001; **2**: 943–955.
- 16 Jack J, Dorsett D, Delotto Y, Liu S: Expression of the cut locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. *Development* 1991; **113**: 735–747.
- 17 Superti-Furga G, Barberis A, Schreiber E, Busslinger M: The protein CDP, but not CP1, footprints on the CCAAT region of the γ -globulin gene in unfractionated B-cell extracts. *Biochim Biophys Acta* 1989; **1007**: 237–242.
- 18 Gillingham AK, Pfeifer AC, Munro S: CASP, the alternatively spliced product of the gene encoding the CCAAT-displacement protein transcription factor, is a Golgi membrane protein related to giantin. *Mol Biol Cell* 2002; **13**: 3761–3774.
- 19 Wang HY, Lin W, Dyck JA *et al*: SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol* 1998; **140**: 737–750.
- 20 Carter AS, Volkmar FR, Sparrow SS *et al*: The Vineland Adaptive Behavior Scales: supplementary norms for individuals with autism. *J Autism Dev Disord* 1998; **28**: 287–302.
- 21 Gincel D, Shoshan-Barmatz V: The synaptic vesicle protein synaptophysin: purification and characterization of its channel activity. *Biophys J* 2002; **83**: 3223–3229.
- 22 Pikkarainen T, Eddy R, Fukushima Y *et al*: Human laminin B1 chain. A multidomain protein with gene (LAMB1) locus in the q22 region of chromosome 7. *J Biol Chem* 1987; **262**: 10454–10462.
- 23 Powell SK, Kleinman HK: Neuronal laminins and their cellular receptors. *Int J Biochem Cell Biol* 1997; **29**: 401–414.
- 24 Grumet M: Nr-CAM: a cell adhesion molecule with ligand and receptor functions. *Cell Tissue Res* 1997; **290**: 423–428.
- 25 Hoffman KB: The relationship between adhesion molecules and neuronal plasticity. *Cell Mol Neurobiol* 1998; **18**: 461–475.
- 26 Levy JB, Canoll PD, Silvennoinen O *et al*: The cloning of a receptor-type protein tyrosine phosphatase expressed in the central nervous system. *J Biol Chem* 1993; **268**: 10573–10581.
- 27 IMGSACa: A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* 2001; **69**: 570–581.
- 28 Lord C, Rutter M, Le Couteur A: Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994; **24**: 659–685.
- 29 Sparrow S, Balla D, Cicchetti D: *Vineland Adaptive Behaviour Scales*. Minn: Pines C, 1984.
- 30 Lord C, Risi S, Lambrecht L *et al*: The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 2000; **30**: 205–223.
- 31 Klauck SM, Poustka F, Benner A, Lesch KP, Poustka A: Serotonin transporter (5-HTT) gene variants associated with autism? *Hum Mol Genet* 1997; **6**: 2233–2238.
- 32 Rong Zeng W, Soucie E, Sung Moon N *et al*: Exon/intron structure and alternative transcripts of the CUTL1 gene. *Gene* 2000; **241**: 75–85.
- 33 IMGSAC: A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Hum Mol Genet* 1998; **7**: 571–578.
- 34 Bonora E, Bacchelli E, Levy ER *et al*: Mutation screening and imprinting analysis of four candidate genes for autism in the 7q32 region. *Mol Psychiatry* 2002; **7**: 289–301.
- 35 Ramensky V, Bork P, Sunyaev S: Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002; **30**: 3894–3900.
- 36 Bonora E, Beyer KS, Lamb JA *et al*: Analysis of reelin as a candidate gene for autism. *Mol Psychiatry* 2003; **8**: 885–892.
- 37 O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
- 38 Douglas JA, Boehnke M, Lange K: A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *Am J Hum Genet* 2000; **66**: 1287–1297.
- 39 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
- 40 Hauser ER, Boehnke M, Guo SW, Risch N: Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet Epidemiol* 1996; **13**: 117–137.
- 41 Moffatt MF, Traherne JA, Abecasis GR, Cookson WO: Single nucleotide polymorphism and linkage disequilibrium within the TCR alpha/delta locus. *Hum Mol Genet* 2000; **9**: 1011–1019.
- 42 Lewontin RC: On measures of gametic disequilibrium. *Genetics* 1988; **120**: 849–852.
- 43 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- 44 Bailey A, Palferman S, Heavey L, Le Couteur A: Autism: the phenotype in relatives. *J Autism Dev Disord* 1998; **28**: 369–392.
- 45 Hutcheson HB, Olson LM, Bradford Y *et al*: Examination of NRCAM, LRRN3, KIAA0716, and LAMB1 as autism candidate genes. *BMC Med Genet* 2004; **5**: 12.
- 46 Xi T, Jones IM, Mohrenweiser HW: Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. *Genomics* 2004; **83**: 970–979.
- 47 Tabor HK, Risch NJ, Myers RM: OPINION: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002; **3**: 391–397.
- 48 Pritchard JK: Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001; **69**: 124–137.
- 49 Reich DE, Schaffner SF, Daly MJ *et al*: Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet* 2002; **32**: 135–142.
- 50 Roberts SB, Mclean CJ, Neale MC, Eaves LJ, Kendler KS: Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Gene* 1999; **65**: 876–884.
- 51 Faivre-Sarrailh C, Falk J, Pollerberg E, Schachner M, Rougon G: NrCAM, cerebellar granule cell receptor for the neuronal adhesion molecule F3, displays an actin-dependent mobility in growth cones. *J Cell Sci* 1999; **112** (Part 18): 3015–3027.
- 52 Horikawa Y, Oda N, Cox NJ *et al*: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000; **26**: 163–175.
- 53 Schaid DJ: Transmission disequilibrium, family controls, and great expectations. *Am J Hum Genet* 1998; **63**: 935–941.
- 54 Clayton D, Jones H: Transmission/disequilibrium tests for extended marker haplotypes. *Am J Hum Genet* 1999; **65**: 1161–1169.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)