

IMMEDIATE COMMUNICATION

Association of the homeobox transcription factor, *ENGRAILED 2, 3*, with autism spectrum disorderN Gharani^{1,5}, R Benayed^{2,5}, V Mancuso², LM Brzustowicz^{1,4} and JH Millonig^{1,2,3}

¹Department of Genetics, Rutgers University, Piscataway, NJ, USA; ²Center for Advanced Biotechnology and Medicine, Piscataway, NJ, USA; ³Department of Neuroscience and Cell Biology, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ, USA; ⁴Department of Psychiatry, UMDNJ-New Jersey Medical School, Piscataway, NJ, USA

Mouse mutants of the homeobox transcription factor *Engrailed2* (*En2*) and autistic individuals display similar cerebellar morphological abnormalities, which include hypoplasia and a decrease in the number of Purkinje cells.^{1–19} Human *EN2* maps to 7q36, a chromosomal region that has demonstrated suggestive linkage to autism spectrum disorder (ASD).^{20–22} To investigate *EN2* for evidence of association with ASD, four single-nucleotide polymorphisms (SNPs) (*rs3735653*, *rs1861972*, *rs1861973*, *rs2361689*) that span the majority of the 8.0 kb gene were assessed by the transmission/disequilibrium test^{23–26}. Initially, 138 triads of autistic individuals and their parents were tested. Two intronic SNPs (*rs1861972* and *rs1861973*) demonstrated significant association with autism (*rs1861972*, $P=0.0018$; *rs1861973*, $P=0.0003$; haplotype, $P=0.000005$). Flanking exonic SNPs (*rs3735653* and *rs2361689*) did not display association. This analysis was then extended to include 167 small nuclear ASD pedigrees and significant association was again only observed for *rs1861972* and *rs1861973* under both the narrow and broad diagnostic criteria (narrow: *rs1861972* $P=0.0290$, *rs1861973* $P=0.0073$, haplotype $P=0.0009$; broad: *rs1861972* $P=0.0175$, *rs1861973* $P=0.0107$, haplotype $P=0.0024$). These data demonstrate association between a cerebellar patterning gene and ASD, suggesting a role for *EN2* as a susceptibility locus and supporting a neurodevelopmental defect hypothesis in the etiology of autism.

Molecular Psychiatry (2004) 9, 474–484. doi:10.1038/sj.mp.4001498

Published online 16 March 2004

Keywords: *ENGRAILED2*; cerebellum; autism; chromosome 7

Autism spectrum disorder (ASD) is composed of three separate diagnoses that include autism and two other milder but qualitatively similar disorders, Asperger's syndrome and Pervasive Developmental Delay-Not Otherwise Specified (PDD-NOS). ASDs are neurodevelopmental disorders characterized by deficits in communication, abnormal social interactions and rigid or repetitive interests and behaviors. Twin, family and disease modeling studies have indicated that ASDs are complex genetic disorders and that an estimated five to 15 interacting genes may be involved in the disease etiology.^{1–5}

The CNS structure most commonly affected in autistic individuals is the cerebellum. Of the 22 autopsy studies, 21 display cerebellar abnormalities including a reduced number of Purkinje cells. These defects occur in the absence of any obvious sign of degeneration, suggesting that autism is caused by developmental defects.^{6–12} Cerebellar hypoplasia has

also been observed in autistic individuals.^{11,13–17} Recently, the growth pattern of the cerebellum during childhood has also been shown to be abnormal. Initially, cerebellar growth is accelerated in autistic individuals compared to unaffected controls but then declines after the age of 6 years.^{18,19} Moreover, functional MRI studies have demonstrated that the cerebellum is active during activities that are deficient in ASD including language generation, attention and problem solving.^{11,20–29} Together, these experiments demonstrate that cerebellar development is perturbed in autistic individuals and that these defects might contribute to the behavioral abnormalities observed in ASD.

Mouse genetics have identified a number of genes that function during cerebellar development.^{30,31} One such gene is *Engrailed 2* (*En2*), a homeobox transcription factor that is orthologous to *Drosophila melanogaster engrailed*. Both loss of function and transgenic misexpression mutants have been generated for the mouse *En2* gene. Interestingly, both types of mutations display a phenotype that is reminiscent of the cerebellar anatomical abnormalities reported for autistic individuals. Adult mice for both mutants are nonataxic, but their cerebella are hypoplastic with a reduction in the number of Purkinje cells and

Correspondence: JH Millonig, PhD, Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, NJ 08854-5638, USA. E-mail: Millonig@CABM.rutgers.edu

⁵These authors contributed equally to this work

Received 20 November 2003; revised 08 January 2004; accepted 04 February 2004

other cell types.^{32–36} Both mouse mutants also display cerebellar foliation defects that have not been reported in autistic individuals. Closer anatomical examination of these mice has revealed that these phenotypes are due to abnormal postnatal development.^{32–37}

Human *EN2* maps to distal chromosome 7 (7q36.3), a region that has provided suggestive evidence for linkage in three studies from two independent genome scans.^{38–40} For these reasons, *EN2* was examined as a susceptibility locus for ASD by performing family-based association analysis.

Materials and methods

Subjects

Families recruited to the Autism Genetic Resource Exchange (AGRE) were used for these studies. AGRE is a central repository of family DNA samples created by The Cure Autism Now (CAN) Foundation and the Human Biological Data Interchange. The selection criteria require that at least two family members have a diagnosis of autism, Asperger's syndrome or PDD-NOS. The diagnosis and characteristics of these families are described in detail elsewhere.^{38,41} For our analysis, a narrow diagnosis is defined as only autism, while a broad diagnosis includes individuals affected with either autism, Asperger's syndrome or PDD-NOS. Although unaffected siblings have not undergone an ADI-R evaluation, extensive medical histories that include neurological, psychological and medical evaluations are available for 17 of the 169 unaffected siblings used in this study. None of the 17 unaffected siblings display characteristics of a broad phenotype. Karyotypic data are now available for 73 families on the AGRE website (www.agre.org). In all, 60 of these families are used in our study. Only two are karyotypically abnormal (AU0065 and AU0106), with a duplication of SNRPN on chromosome 15q12 (a marker for cytogenetic abnormality at the chromosome 15 autism critical region), while the others are either normal or in the process of being analyzed.

In this study, DNA was initially obtained from 138 autistic individuals and their parents (parent-offspring triads). These triads were derived from independent nuclear families by randomly selecting with a diagnosis of autism and both parents. In the second stage of our analysis, the number of our samples was increased to include other affected and unaffected siblings from the 138 original families and an additional 29 nuclear families to give a total of 167 pedigrees. These families comprised 316 triads and 169 phenotypic discordant sib pairs (DSP) under the broad diagnosis (753 total subjects). For the narrow diagnostic classification, 166 pedigrees with 256 triads and 135 DSP were analyzed ($n = 689$).

The 167 nuclear pedigrees were selected based on the initial Columbia University genome scan study³⁸ and subsequent QTL analysis.³⁹ Although these studies report the results of analysis in 110 and 152 families respectively, linkage data on 183 families

were available at the time our samples were selected. Of this total, a number of families were not ordered based either on substantial missing genotyping data or in the case of three families evidence for nonidiopathic autism due to fragile X syndrome. Of our 167 families, 34 include multiple births. There are 20 families with DZ multiple births (18 are DZ twins and two are DZ triplets) and 14 families with MZ twins (of which eight families have DNA available for both co-twins). For the DZ twin families, all siblings are genotyped and included in the data analysis. However, for the MZ twins only one of the co-twins is selected for analysis (five of the eight families show phenotypic concordance between the MZ twins (both are autism : autism) and three MZ twin pairs are autism : PDD discordant. For the latter three families, the twin with the narrow diagnosis of autism was selected for further analysis). The DNA availability of eight MZ twin pairs served as internal controls for the estimation of genotyping errors. Genotypic concordance between MZ co-twins was consistently obtained for all four single-nucleotide polymorphisms (SNP) assays.

DNA analysis

The dbSNP database (www.ncbi.nlm.nih.gov/SNP) was used to identify *EN2* SNPs. The frequency and validity of each SNP was determined by direct sequencing of approximately 200 bp of DNA encompassing each SNP in 24 unrelated individuals (23 Caucasian and one of Hispanic/Latino descent). No additional SNPs were identified by this analysis. Once the SNPs were confirmed, a tetraprimer Amplification Refractory Mutation System (ARMS)-PCR strategy⁴² or primer extension strategy (Pyrosequencing™ system)^{43,44} was used to genotype individuals. Primers were designed using publicly available software (tetraprimer ARMS-PCR-http://cedar.genetics.soton.ac.uk/public_html/primer1.html; Pyrosequencing™-http://www.pyrosequencing.com/pages/technical_supp.html). Sequenced individuals were used as controls for optimization of the genotyping assays. SNPs, *rs3735653* and *rs2361689*, were typed as simplex Pyrosequencing™ assays using the automated PSQ HS 96A platform as described previously.^{43,44} Primer sequences can be obtained from the authors on request. For *rs2361689*, the following PCR amplification conditions were used: 0.25 μM of each primer, 0.2 mM dNTP, 1.875 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100 using standard cycling conditions 94°C, 1 min, 1 ×; 94°C, 30 s, 59°C, 30 s, 74°C, 30 s, 45 ×; 74°C, 10 min, 1 ×. For *rs3735653*, identical conditions were used except 0.1 mM dATP, dCTP, dTTP, 25 μM dGTP and 75 μM 7-deaza-2'-deoxyguanosine triphosphate was included.

For *rs1861972* and *rs1861973*, a tetra-ARMS PCR genotyping assay was employed as described.⁴² Cycling conditions were as described above for *rs2361689* except for the following changes: *rs1861972*, 72°C annealing temperature in a 10 mM Tris-HCl (pH 8.3), 1.5 mM MgCl₂, 25 mM KCl buffer

for 30 cycles; *rs1861973*, 65°C annealing temperature in a 10 mM Tris-HCl (pH 8.8), 1.5 mM MgCl₂, 25 mM KCl 10 × buffer for 30 cycles. As an additional control for the *rs1861972* and *rs1861973* tetra-ARMS PCR assays, pyrosequencing assays have recently been developed for both SNPs and identical genotypes have been generated for the 25 individuals tested.

Error checking of genotype data revealed only two Mendelian inconsistencies (one for SNP *rs1861972* and one for SNP *rs3735653*), both of which were resolved with repeat genotyping. Full four SNP haplotype analysis identified a further 15 likely genotyping errors (two for SNP *rs3735653*; five for SNP *rs1861972*; three for SNP *rs1861973*; and six for SNP *rs2361689*), which again were resolved by repeat genotyping analysis. A number of sample genotypes could not be assigned due to repeated assay failure or unclear/poor-quality genotype results. These represented a total of 39 missing genotypes and consist of 18 genotypes for SNP *rs3735653*; three for SNP *rs1861972*; 10 for SNP *rs1861973*; and eight for SNP *rs2361689*. Therefore of the total 3012 genotypes assayed, we identified two Mendelian inconsistencies, a further 15 genotype errors were revealed by haplotype analysis, and 39 genotypes could not be assigned and represented missing data.

Statistical analysis

All genotype data were checked for Mendelian inconsistencies prior to transmission/disequilibrium test (TDT) analysis using the PEDCHECK program version 1.1.⁴⁵ Each SNP was assessed for deviations from Hardy–Weinberg equilibrium using genotype data from all parents and standard formulae. Error checking for haplotype inconsistencies and all haplotype constructions in the extended pedigrees were carried out using the SIMWALK program version 2.83.⁴⁶ The linkage disequilibrium (LD) coefficient (delta) was calculated for pairs of SNPs according to published methods,⁴⁷ using parental genotypes from the original 138 parent–offspring triads. The sign in front of the coefficient indicates whether LD is positive (allelic association between common alleles of the two loci) or negative (association between the common allele at one locus and the rare allele at the other). Delta = 0 indicates no LD between the two loci. The significance of the delta can be tested by taking $N(\text{delta})^2$ (where N is the total number of haplotypes observed), this is distributed as χ^2 with one degree of freedom.

Single marker and haplotype TDT analyses for the initial 138 triads were performed using the TRANSMIT program (version 2.5.4).⁴⁸ Compared to other programs, TRANSMIT is advantageous because it can handle both missing parental data and the transmission of multilocus haplotypes, even if phase is unknown. It uses a score test, which is based on the ‘conditional on parental genotype’ likelihood. When transmissions are fully observed, the score test reduces to the familiar Pearson χ^2 test. Minimum haplotype frequencies were set at 0.05 for the TRANSMIT analysis (using the *-c#* flag), ie rare

haplotypes with frequency of less than 5% were pooled. For multilocus analysis, both the global P -value (which assess the significance of transmission distortion for all test haplotypes) and P -values, which assess the significance of transmission distortion for specific haplotypes are calculated. The bootstrap simulation procedure implemented in TRANSMIT carries out multiple samplings of the data to control for haplotype ambiguities and derive exact P -values (if sufficient bootstrap samples are drawn).

The TRANSMIT output gives the total number of observed and expected transmissions for each allele/haplotype, which includes uninformative transmissions from homozygous parents. The number of informative transmissions from the heterozygous parents was calculated as stated in the TRANSMIT documentation (www-gene.cimr.cam.ac.uk/clayton/software/transmit.txt) as explained below:

- (i) Multiply the tabulated value for Var ($O-E$) (Variance of Observed–Expected) by 4 to give N (where N is equivalent to the number of fully informative transmissions, that is, total number of heterozygous parents).
- (ii) $N/2 =$ expected informative transmissions under null hypothesis.
- (iii) The difference between the expected counts from TRANSMIT output table and the informative expected counts ($N/2$) = number of expected transmissions coming from uninformative homozygous parents. This number is subtracted from the observed transmissions from the TRANSMIT output to give the observed transmissions (T) from the informative heterozygous parents. We then replaced the expected column with the ‘Untransmitted’ counts (UT) and this was derived as $N-T=UT$.

To test for sex difference in susceptibility at this locus, the haplotype transmissions of the two intronic markers for the 138 triads were examined by TRANSMIT using the *-s#* flag (considers each sex separately by recoding the other sex as unknown prior to analysis). In this series, there are 105 male and 33 female autistic probands. A 2×2 contingency table and the Fisher’s exact test was then used to compare the transmission ratios obtained for the male triad and the female triad families. A significant difference would be indicative of a sex bias effect in susceptibility at this locus.

Haplotype constructions were conducted using SIMWALK version 2.83,⁴⁶ assuming no recombination events between markers. In the absence of parental phase ambiguities, each haplotype was recoded as a single allele (creating a pseudosingle marker). This allowed haplotype transmissions to be tested by programs such as pedigree-based transmission/disequilibrium test (PDT)⁴⁹ and ETDT version 2.4,⁵⁰ which are designed for single locus analysis. The ETDT program⁵⁰ was used to assess evidence for a parent-of-origin effect (using the intronic haplotype data from the original 138 triads) since the TRANS-

MIT program lacks this feature. Haplotype data for the triads were extracted from the nuclear family data after recoding of haplotypes as a pseudosingle marker (see above), prior to ETDT analysis.

The PDT version 4.0⁴⁹ was used for TDT analysis in the extended pedigrees. PDT was designed to allow the use of data from related triads and disease discordant sibships from extended pedigrees when testing for transmission disequilibrium. It determines the presence of association by testing for unequal transmission of either allele from parents to affected offspring and/or unequal sharing of either allele between discordant sibships. Informative extended pedigrees contain at least one informative triad (ie an affected child with at least one parent heterozygous at the marker) and/or discordant sibship (ie at least one affected and one unaffected sibling with different marker genotypes). PDT has been shown to provide substantial gains in power over other similar tests that utilize only a subset of the family data. Furthermore, although misclassification of affected individuals as unaffected is expected to lead to a loss of power, PDT has been shown to be more robust than other tests when extended family data are available.⁴⁹ PDT has two global scores: the 'PDT^{sum}', which sums the level of significance from all families, and the 'PDT^{ave}', for which equal weight is given to all families in a data set. For our data, the χ^2 distribution and *P*-values obtained under both test statistics gave very similar results. We report the PDT^{sum} data in our results section.

Results

EN2 is composed of two exons separated by a single 3.3 kb intron. Four SNPs (*rs3735653*, *rs1861972*, *rs1861973* and *rs2361689*) that span the majority of the *EN2* gene were tested for transmission disequilibrium

in parent-offspring triads and small nuclear families obtained from the AGRE data set. SNP *rs3735653* in exon 1 alters the coding sequence from a Leu to a Phe, while *rs2361689* in exon 2 results in a synonymous change for Leu. SNPs *rs1861972* and *rs1861973* are located 152 bp apart in the intron, approximately 1.3 kb 5' of *rs2361689* (exon 2) and 2.5 kb 3' of *rs3735653* (exon 1). The allele frequencies for each SNP are given in Table 1a. Varying strengths of LD is observed between all four SNPs ranging from $\delta = 0.89$ ($P = 6.4 \times 10^{-49}$) for the intronic SNPs *rs1861972* and *rs1861973*, to $\delta = -0.22$ ($P = 0.0004$) for *rs3735653* and *rs1861972* (Table 1b).⁴⁷

To test initially whether certain variants of *EN2* were transmitted more frequently than expected by chance to autistic individuals, the genotype of each SNP was determined in 138 parent-offspring triads that fit the narrow diagnostic criteria. Once all genotypes were verified to be in Hardy-Weinberg equilibrium, allelic transmissions were assessed for each SNP by the TDT.⁵¹ This analysis was conducted using the TRANSMIT program.⁴⁸ A significant over-transmission of the *rs1861972* A-allele to affected offspring was observed, with 73 of 113 heterozygous parents (65%) transmitting the A-allele and only 40 transmitting the G-allele ($P = 0.0018$) (Table 2a). A similar distortion of transmissions was observed for the *rs1861973* SNP with 71 of 104 heterozygous parents (68%) preferentially transmitting the C-allele to their autistic offspring ($P = 0.0003$) (Table 2a). However, no evidence for association was obtained for either of the flanking exonic SNPs (Table 2a).

Haplotype transmissions for these SNPs were then examined using the TRANSMIT software.⁴⁸ These tests were reiterated 1 000 000 times by a bootstrap simulation procedure in order to control for ambig-

Table 1 (a) *EN2* SNP characteristics and (b) LD between SNPs

dbSNP no.	Position	Polymorphism ^a	Amino acid	Frequency ^b
<i>(a) EN2 SNP characteristics</i>				
<i>rs3735653</i>	Exon 1	C/T	Leu > Phe	0.489
<i>rs1861972</i>	Intron	A/G		0.281
<i>rs1861973</i>	Intron	C/T		0.287
<i>rs2361689</i>	Exon 2	T/C	Leu > Leu	0.320
SNP	<i>Delta</i> ^c	<i>Delta-χ^2</i> ^d	<i>P-value</i> ^d	
<i>(b) LD between SNPs</i>				
<i>rs3735653 + rs1861972</i>	-0.21	12.71	0.00036	
<i>rs3735653 + rs1861973</i>	-0.23	15.12	0.0001	
<i>rs3735653 + rs2361689</i>	-0.34	29.97	4.39×10^{-8}	
<i>rs1861972 + rs1861973</i>	0.89	216.11	6.39×10^{-49}	
<i>rs1861972 + rs2361689</i>	-0.26	18.90	0.000014	
<i>rs1861973 + rs2361689</i>	-0.25	16.40	0.000051	

^aSecond allele is rare allele.

^bFrequency of rare allele.

^cLD correlation coefficient *delta*.⁵⁰

^d*Delta- χ^2* (1 df) and associated one-sided *P*-value.

Table 2 (a) TDT results for *rs3735653*, *rs1867972*, *rs1867973* and *rs2361689*; (b) *rs1861972*-*rs1861973* haplotype; and (c) *rs3735653*-*rs1867972*-*rs1867973*-*rs2361689* haplotypes in 138 autism parent-offspring triads

Parental transmissions ^a					
<i>dbSNP no.</i>	T	U	χ^{2b}	<i>P-value</i> ^c	
<i>(a) rs3735653, rs1867972, rs1867973 and rs2361689</i>					
<i>rs3735653</i>	57	75	2.46	0.1168	
<i>rs1867972</i>	73	40	9.80	0.0018	
<i>rs1867973</i>	71	34	13.23	0.0003	
<i>rs2361689</i>	60	57	0.07	0.7853	
Parental transmissions ^a					
<i>Haplotype</i>	<i>Frequency</i> ^d	T	U	χ^{2b}	<i>P-value</i> ^e
<i>(b) rs1861972 and rs1861973 haplotype</i>					
A-C	0.688	76	34	16.26	0.00011
A-T	0.022	2	8	—	—
G-C	0.031	2	11	—	—
G-T	0.259	39	66	7.08	0.0096
Global ^f				22.42	0.000005
Parental transmissions ^a					
<i>Haplotype</i>	<i>Frequency</i> ^d	T	U	χ^{2b}	<i>P-value</i> ^e
<i>(c) rs3735653-rs1867972-rs1867973-rs2361689 haplotype</i>					
C-A-C-T	0.0924	26	16	2.27	0.2467
T-A-C-T	0.3191	74	46	6.42	0.1048
T-G-C-T	0.0139	3	4	—	—
C-A-T-T	0.0110	1	5	—	—
T-A-T-T	0.0094	0	3	—	—
C-G-T-T	0.1669	30	49	4.26	0.1362
T-G-T-T	0.0690	12	19	1.35	0.3150
C-A-C-C	0.2051	43	41	0.04	0.8519
T-A-C-C	0.0713	17	14	0.26	0.6188
C-G-C-C	0.0059	0	3	—	—
C-A-T-C	0.0105	0	3	—	—
C-G-T-C	0.0217	5	6	—	—
T-G-T-C	0.0019	0	1	—	—
Global ^g				22.22	0.0559

^aT = number of times common allele transmitted; U = number of times common allele not transmitted from heterozygous parents.

^b χ^2 values calculated by TRANSMIT.²⁵

^cTwo-sided *P*-value (1 df).

^dFrequency of haplotype.

^eExact *P*-values based on 1 000 000 bootstrap samples.

^fGlobal test restricted to common haplotypes with frequency greater than 5% (2 df).

^gGlobal test restricted to common haplotypes with frequency greater than 5% (6 df).

uous haplotypes and to derive exact *P*-values. Two types of haplotype analysis were performed. First, the intronic markers were analyzed alone since association was only observed for these SNPs. Next, the entire four SNP haplotype was examined for transmission distortions. In our population, the *rs1861972* and *rs1861973* A-C haplotype has an observed frequency of 68.8%. This haplotype is specifically overtransmitted to affected individuals (bootstrap

P = 0.00011) (Table 2b). All other haplotypes are undertransmitted (Table 2b). Global χ^2 tests based on the common haplotypes A-C and G-T (frequency > 5%) yielded a bootstrap *P*-value of 0.000005, demonstrating significant evidence and association between these markers and the autism phenotype (Table 2b).

Extension of the haplotypes to include the exonic markers *rs3735653* and *rs2361689* resulted in an overall reduction in the significance of the haplotype

transmission distortions (Table 2c). Four of the 13 haplotypes observed in this population were composed of the core intronic A–C haplotype and although each demonstrated some degree of excess transmission, inclusion of the exonic markers resulted in a dilution of this effect and statistical significance was no longer achieved (Table 2c).

The haplotype transmissions of the two intronic markers were also investigated for parent-of-origin and sex differences using ETDT⁵⁰ and TRANSMIT⁴⁸ software, respectively. When maternal and paternal transmissions were investigated separately, no parental differences in transmission ratios were observed (data not shown). Analysis of the haplotype transmission data separately by sex of proband also demonstrated no significant difference in transmission ratios (data not shown). Therefore, male and female affected individuals appear equally likely to inherit the A–C haplotype. Thus, these data do not support significant parent-of-origin or sex bias effects at this locus.

To determine whether these association findings could be extended to a larger sample, the entire pedigree for the initial 138 triads (ie including other affected and unaffected siblings) plus a further 29 nuclear families (167 in total) were genotyped for the four SNPs. Since these pedigrees include siblings affected with the broad ASD phenotype of Asperger's syndrome or PDD-NOS, association could be tested under both a narrow (autism) and broad (autism, Asperger's syndrome, PDD-NOS) diagnostic schemes. The PDT (PDT version 4.0)⁴⁹ was chosen for this analysis because the program was designed specifically to test for association using data from multiple related triads as well as information with regard to allele sharing among discordant sibships within the extended pedigrees.

The result of the PDT analysis for each SNP is presented in Table 3a. Once again significant evidence for association was observed for the intronic markers *rs1861972* and *rs1861973* under both the narrow and broad diagnostic schemes (*rs1861972*, narrow: $P=0.0290$, broad: $P=0.0175$; *rs1861973*, narrow: $P=0.0073$, broad: $P=0.0107$). The A-allele of *rs1861972* and the C-allele of *rs1861973* were significantly overtransmitted from heterozygous parents and were over-represented in affected sibs compared to unaffected sibs of DSPs (Table 3a). Analysis of the exonic SNPs (*rs3735653* and *rs2361689*) once again demonstrated a lack of association when tested under both the narrow and broad diagnostic classifications (Table 3a).

Next, we performed haplotype TDT analysis in the extended pedigrees. Since these markers are in tight LD and are only separated by a distance of less than 4 kb of DNA, recombination events between these SNPs are highly unlikely in a sample of this size. Assuming the absence of recombination, unambiguous haplotypes could be assigned to all individuals in 158 of the 167 pedigrees (for the two intronic marker analysis) and 146 of 167 pedigrees for the four marker haplotypes. The pedigrees in which parental

haplotype phase could not be distinguished were omitted from further analysis. Each haplotype was recoded as a single allele. By recoding the haplotypes as a single locus, the transmissions could be analyzed by PDT. The A–C haplotype is again overtransmitted under both diagnostic classifications and over-represented in affected siblings of DSPs at statistically significant levels (narrow, $P=0.0018$; broad, $P=0.0035$) (Table 3b). Conversely, the A–T, G–C and G–T haplotypes are all undertransmitted and are not observed more frequently in affected siblings of DSPs (Table 3b). Global χ^2 tests for all haplotypes also yielded significant P -values (narrow, $P=0.0009$; broad, $P=0.0024$) (Table 3b).

Inclusion of the flanking *rs3735653* and *rs2361689* SNPs in the full four marker haplotype analysis once again failed to provide significant evidence for association of any specific haplotype with ASD under both diagnostic schemes (Table 4). Three SNP haplotype analyses consisting of the two intronic markers with each of the exonic markers in turn (*rs3735653*–*rs1861972*–*rs1861973* and *rs1861972*–*rs1861973*–*rs2361689*) were also carried out in an attempt to further elucidate the disease association in this region. However, these analyses also failed to identify a single significantly predisposing haplotype (data not shown). Overall, these results are consistent with the weaker LD observed between the rare alleles of the exonic markers (*rs3735653* and *rs2361689*) and the common predisposing alleles of the intronic markers *rs1861972* and *rs1861973* (Table 1a).

Discussion

In this study, we report the investigation of a candidate susceptibility gene *ENGRAILED2*, for ASD. This gene was selected for analysis based on functional data from mouse mutant studies and the genomic localization of the human *EN2* gene to a region on chromosome 7q that has previously displayed suggestive linkage to ASD.^{6–10,36,38–40,37} Four SNPs were examined for evidence of excess transmission from parents to affected offspring. Our data have demonstrated significant association between the intronic A–C *rs1861972*–*rs1861973* haplotype and ASD in both the 138 triads and 167 extended pedigrees. These data represent one of the more significant association between a candidate gene and ASD in the publicly available AGRE data set (www.agre.org).^{52–61} The triad data also define a minimal population for future family-based association studies that is both a powerful and cost-efficient alternative to using extended nuclear families.

No evidence for association was observed for the two exonic markers. Furthermore, four SNP haplotype analysis in the extended pedigrees has revealed slight undertransmission for one of the four core intronic A–C haplotypes (C–A–C–C) (Table 4), indicating that only a subpopulation of the common A–C *rs1861972*–*rs1861973* haplotype is in association with ASD. Three SNP haplotype analysis of the two

Table 3 (a) *rs3735653*, *rs1867972*, *rs1867973* and *rs2361689* PDT results in 167 extended pedigrees and (b) *rs1861972* and *rs1861973* haplotype PDT results for 158 extended pedigrees

<i>dbSNP no.</i>	<i>Diagnosis</i>	<i>Parental transmissions^a</i>		<i>DSP^b</i>		χ^{2c}	<i>P-value^d</i>
		<i>T</i>	<i>U</i>	<i>A</i>	<i>UA</i>		
<i>(a) rs3735653, rs1867972, rs1867973 and rs2361689</i>							
<i>rs3735653</i>	Narrow	118	137	111	97	1.23	0.2685
	Broad	150	165	144	97	0.89	0.3444
<i>rs1867972</i>	Narrow	113	87	162	122	4.77	0.0290
	Broad	141	110	204	123	5.65	0.0175
<i>rs1867973</i>	Narrow	110	74	159	121	7.19	0.0073
	Broad	133	95	195	122	6.51	0.0107
<i>rs2361689</i>	Narrow	119	114	143	115	0.02	0.8877
	Broad	146	153	172	117	0.39	0.5323
<i>Haplotype^e</i>	<i>Diagnosis</i>	<i>T</i>	<i>U</i>	<i>A</i>	<i>UA</i>	χ^{2c}	<i>P-value^d</i>
<i>(b) rs1861972 and rs1861973</i>							
A-C	Narrow	115	67	146	108	—	0.0018
A-T		3	19	5	2	—	0.0112
G-C		3	12	4	2	—	0.1985
G-T		75	98	43	44	—	0.0391
Global							16.53
A-C	Broad	133	86	177	109	—	0.0035
A-T		6	22	5	2	—	0.0222
G-C		4	13	4	2	—	0.2076
G-T		97	119	56	45	—	0.0509
Global							14.37

^aParental transmissions: *T* = number of times common allele transmitted; *U* = number of times common allele not transmitted from heterozygous parents. Total number of triads included in the analyses in: (a) 256 (narrow diagnosis) 316 (broad diagnosis) and (b) 245 (narrow diagnosis) 300 (broad diagnosis).

^bDiscordant sibpair counts: *A* = total number of the common allele in affected siblings; *UA* = total number of the common allele in unaffected siblings of DSPs. Total number of DSPs included in the analyses in: (a) 135 (narrow diagnosis) 169 (broad diagnosis) and (b) 122 (narrow diagnosis) 149 (broad diagnosis).

^cGlobal χ^2 values calculated by PDT^{sum, 26}

^d*P*-value generated by PDT (in (a) 1 df and (b) 1 df for single haplotype *P*-values and 3 df for the global tests).

^eHaplotypes recoded as alleles prior to PDT analysis.

intronic markers with each of the exonic markers (*rs3735653*–*rs1861972*–*rs1861973* and *rs1861972*–*rs1861973*–*rs2361689*) also failed to identify a single significantly predisposing haplotype (data not shown). These results argue against a functional role for the A- and C-alleles of *rs1861972* and *rs1861973* and suggest that they are likely to be nonfunctional polymorphisms in LD with other variant(s) that contribute to autism.

In this study, we have used a family-based association method to demonstrate evidence for association at the *EN2* locus on chromosome 7q36.3 with ASD. Previously, three separate studies of two independent genome scans yielded only suggestive evidence for linkage between ASD and 7q36. However, only one of these studies used markers that spanned the *EN2*

locus. In this Finnish report, suggestive linkage to a combined phenotype of ASD and dysphasia was obtained at marker *D7S550* (LOD = 2.02), which is located about 170 kb distal of *EN2*.⁴⁰ The other two studies were carried out using a subset of the AGRE families that largely overlap with our data set.^{38,39} Liu *et al*³⁸ carried out fine mapping analysis of the region proximal to *EN2* in the original 110 AGRE families and reported a LOD score of 2.13 for *D7S483* located approximately 5.5 Mb proximal of *EN2*.³⁸ In a further study, Alarcon and co-workers used the same set of microsatellite markers and 152 AGRE families to map quantitative trait loci implicated in language deficits to distal chromosome 7 at a distance of less than 1 Mb from the *EN2* locus (*P* = 0.001).³⁹ Recently, Yonan *et al* performed a genome scan on the complete set of 345

Table 4 *rs3735653*, *rs1867972*, *rs1867973* and *rs2361689* haplotype PDT results in 146 extended pedigrees analyzed under the narrow and broad diagnoses

Haplotype ^a	Diagnosis	Parental transmissions ^b		DSP ^c		χ^2 ^d	P-value ^e
		T	U	A	UA		
C-A-C-T	Narrow	45	34	23	21	—	0.3800
T-A-C-T		119	86	65	45	—	0.0402
T-G-C-T		3	6	2	2	—	0.1573
C-A-T-T		1	5	1	0	—	0.0896
T-A-T-T		1	5	2	1	—	0.2059
C-G-T-T		53	63	28	27	—	0.2004
T-G-T-T		19	29	9	10	—	0.2050
C-A-C-C		72	78	47	36	—	0.8658
T-A-C-C		32	23	13	10	—	0.3402
C-G-C-C		1	4	1	0	—	0.4142
C-A-T-C		1	6	1	1	—	0.2253
C-G-T-C		11	12	3	6	—	0.5791
T-G-T-C		1	1	1	1	—	0.3173
Global							17.63
C-A-C-T	Broad	57	42	33	21	—	0.2924
T-A-C-T		137	112	70	46	—	0.1532
T-G-C-T		3	9	2	2	—	0.1573
C-A-T-T		3	10	1	0	—	0.1573
T-A-T-T		2	5	2	1	—	0.4054
C-G-T-T		67	77	34	27	—	0.2322
T-G-T-T		25	37	12	11	—	0.1228
C-A-C-C		88	95	56	36	—	0.8383
T-A-C-C		43	29	21	10	—	0.1093
C-G-C-C		1	5	1	0	—	0.3173
C-A-T-C		1	7	1	1	—	0.1797
C-G-T-C		15	14	6	6	—	0.7456
T-G-T-C		1	1	1	1	—	0.3173
Global							16.76

^aHaplotypes recoded as alleles prior to PDT analysis.

^bParental Transmissions: T=number of times common allele transmitted; U=number of times common allele not transmitted from heterozygous parents. A total of 271 triads were included in the analysis.

^cDiscordant sibpair counts: A = total number of the common allele in affected siblings; UA = total number of the common allele in unaffected siblings of genotypic discordant DSPs. A total of 151 DSPs were included in the analysis.

^dGlobal χ^2 values calculated by PDT^{sum,26}

^eP-value generated by PDT (1 df for single haplotype P-values and 12 df for the global test).

AGRE families and observed only minimal linkage at distal chromosome 7 (LOD < 1.3).⁶² However, once again only markers proximal to the *EN2* locus were utilized. Hence, to further investigate linkage of this genomic region to ASD, additional markers spanning the *EN2* locus will be analyzed for linkage and association with both qualitative and quantitative variables of the ASD phenotype in the complete set of AGRE samples.

Two previous studies have investigated *EN2* as an autism susceptibility locus. In a case-control study performed on a Northern French population, significant association ($P < 0.01$) was observed for a *PvuII* polymorphism located 5' of the *EN2* promoter,⁶³ supporting a role for this gene in autism development. Future analysis will investigate whether this *PvuII* polymorphism is associated with ASD in the AGRE population and whether it is in LD with the

rs1861972-rs1861973 A-C haplotype. More recently, Zhong *et al*⁶⁴ examined *rs3735653* from exon 1 using a subset of families from the AGRE data set. They demonstrated a lack of association between *rs3735653* and autism ($P = 0.58$). We have also observed no association between *rs3735653* and ASD in the AGRE data set, and these results are consistent with the weaker LD observed between alleles of this marker and the ASD-associated intronic markers.

In summary, these data demonstrate the significant association of a cerebellar patterning gene with ASD, suggesting a role for *EN2* as a susceptibility locus and supporting the hypothesis that genetic alterations that affect cerebellar development could predispose individuals to both autism and related ASD. Further LD mapping of the region spanning 5'-3' of the *EN2* gene is required to elucidate the role of the ASD-associated intronic markers and to reveal the identity of the

putative functional variant(s). *Cis*-regulatory sequences are often located within the first intron of genes,^{65,66} so it is possible that the functional variant associated with ASD maps to this region and causes the misregulation of *EN2* during cerebellar development. As expected, when the intron is scanned for *cis*-regulatory sequences using computer prediction programs (www.genomatix.de:80), multiple potential binding sites for developmentally regulated transcription factors are observed. Consistent with this idea of functional variants affecting the expression of associated genes, causative alleles responsible for bipolar disorder and rheumatoid arthritis have recently been identified in the promoter and intron of XBP1 and SLC22A4, respectively.^{67,68}

During mouse cerebellar development, the spatial and temporal regulation of *En2* is tightly regulated. For example, from E17.5 to postnatal day 4 (P4), *En2* is expressed in spatially restricted 'stripes'. Within these 'stripes', *En2* is expressed in all primary cerebellar cell types (granule, deep nuclei and Purkinje cells). However, by P4, a developmental switch occurs so that *En2* is no longer expressed in Purkinje or deep nuclei cells, but is now restricted to differentiating granule cells. In an *En2* transgenic that causes the prolonged expression of *En2* in Purkinje cells past P4, adult mice exhibit an 'autistic-like' cerebellar phenotype, a decreased number of Purkinje cells and hypoplasia. Interestingly, the *En2* mouse knockout also displays hypoplasia and reduced number of Purkinje cells, indicating that decreased levels of *En2* could result in a similar phenotype. Both mutants disrupt the topographic mapping of spinocerebellar mossy fibers,^{36,37} which could in turn affect the electrophysiological function of the cerebellum. Together, these data suggest that functional variants in human *EN2* that affect either the level or the spatial/temporal expression of the gene during human cerebellar development might contribute to the anatomical cerebellar phenotypes observed in autism which could play a role in the underlying etiology of autism. These data make *EN2* an excellent candidate for an autism susceptibility locus. Future work will include both further LD mapping to determine whether other variants within *EN2* are associated with ASD and analysis of additional data sets to investigate whether association of *EN2* to ASD is observed in other populations.

Acknowledgements

We thank Cure Autism Now (CAN) and the Autism Genetic Resource Exchange (AGRE) for supplying the resources necessary for this study, Christopher Bartlett for scientific discussions and Emanuel DiCicco-Bloom for critical reading of the paper. We also thank Jay Tischfield and Rutgers University Cell Repository under contract to HBDI for generously providing the AGRE DNA samples. This work was supported in part by research grants from The March of Dimes Birth Defects Foundation (No. 12-FY01-110), National

Alliance for Autism Research and The New Jersey Governor's Council of Autism to LMB and March of Dimes Basil O'Connor Starter Research Award (5-FY00-582), NJ Governor's Council for Autism, Whitehall Foundation (2001-12-54-APL), National Ataxia Foundation and National Alliance for Autism Research to JHM. Most importantly, we thank the families who have participated in and contributed to these studies.

Appendix

Members of the AGRE Consortium

Daniel H Geschwind, University of California at Los Angeles, Los Angeles, USA; Maya Bucan, University of Pennsylvania, Philadelphia, USA; W Ted Brown, New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY, USA; Joseph Buxbaum, Mt Sinai School of Medicine, New York, USA; Edwin H Cook Jr, University of Chicago, Chicago, USA; T Conrad Gilliam, Columbia Genome Center, New York, USA; David A Greenberg, Mt Sinai Medical Center, New York, USA; David H Ledbetter, University of Chicago, Chicago, USA; Bruce Miller, University of California at San Francisco, San Francisco, USA; Stanley F Nelson, University of California at Los Angeles School of Medicine, Los Angeles, USA; Jonathon Pevsner, Kennedy Krieger Institute, Baltimore, USA; Jerome I Rotter, Cedar-Sinai Medical Center, Los Angeles, USA; Gerald D Schellenberg, University of Washington, Seattle, USA; Carol A Sprouse, Children's National Medical Center, Baltimore, USA; Rudolph E Tanzi, Massachusetts General Hospital, Boston, USA; Kirk C Wilhelmsen, University of California at San Francisco, San Francisco, USA; and Jeremy M Silverman, Mt Sinai Medical School, New York, USA.

References

- 1 Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* 1977; **18**: 297-321.
- 2 Ritvo ER, Freeman BJ, Mason-Brothers AM, Mo A, Ritvo AM. Concordance for the syndrome of autism in 40 pairs of affected twins. *Am J Psychiatry* 1985; **142**: 74-77.
- 3 Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J et al. A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* 1999; **65**: 493-507.
- 4 Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet* 2001; **2**: 943-955.
- 5 Lamb JA, Moore J, Bailey A, Monaco AP. Autism: recent molecular genetic advances. *Hum Mol Genet* 2000; **9**: 861-868.
- 6 Bauman ML, Kemper TL. Histoanatomic observations of the brain in early infantile autism. *Neurology* 1985; **35**: 866-874.
- 7 Bauman ML, Kemper TL. Developmental cerebellar abnormalities: a consistent finding in early infantile autism. *Neurology* 1986; **36**(Suppl. 1): 190.
- 8 Kemper TL, Bauman ML. The contribution of neuropathologic studies to the understanding of autism. *Behav Neuro* 1993; **11**: 175-187.

- 9 Bauman ML, Kemper TL. Neuroanatomic observations of the brain in autism. In: Bauman M, Kemper T (eds). *The Neurobiology of Autism*. John Hopkins University Press: Baltimore, 1994 pp 119–145.
- 10 Ritvo ER, Freeman BJ, Scheibel AB, Duong T, Robinson H, Guthrie D et al. Lower Purkinje cell count in the cerebella of four autistic subjects: initial findings of the UCLA-NSAC Autopsy Research report. *Am J Psychiatry* 1986; **143**: 862–866.
- 11 Courchesne E. Brainstem, cerebellar and limbic neuroanatomical abnormalities in autism. *Curr Opin Neurobiol* 1997; **7**: 269–278.
- 12 Bailey A, Luthert P, Dean A, Harding B, Janota I, Montgomery M et al. A clinicopathological study of autism. *Brain* 1998; **121**: 889–905.
- 13 Gaffney GR, Kuperman S, Tsai LY, Minchin S, Hassanein KM. Diffusional magnetic resonance imaging of autism. *Br J Psychiatry* 1987; **151**: 831–833.
- 14 Courchesne E, Yeung-Courchesne R, Press GA, Hesselink JR, Jernigan TL. Hypoplasia of cerebellar vermal lobules VI and VII in autism. *N Engl J Med* 1988; **318**: 1349–1354.
- 15 Murakami JW, Courchesne E, Press GA, Yeung-Courchesne R, Hesselink JR. Reduced cerebellar hemisphere size and its relationship to vermal hypoplasia in autism. *Arch Neurol* 1989; **46**: 689–694.
- 16 Kleiman MD, Neff S, Rosman NP. The brain in infantile autism: are posterior fossa structures abnormal? *Neurology* 1992; **42**: 753–760.
- 17 Hashimoto T, Tayama M, Murakawa K, Yoshimoto T, Miyazaki M, Harada M et al. Development of the brainstem and cerebellum in autistic patients. *J Autism Dev Discord* 1995; **25**: 1–18.
- 18 Courchesne E, Karns CM, Davis HR, Ziccardi R, Carper RA, Tigue ZD et al. Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study. *Neurology* 2001; **57**: 245–254.
- 19 Courchesne E, Carper R, Akshoomoff N. Evidence of brain overgrowth in the first year of life in autism. *JAMA* 2003; **290**: 337–344.
- 20 Allen G, Buxton RB, Wong EC, Courchesne E. Attentional activation of the cerebellum independent of motor involvement. *Science* 1997; **275**: 1940–1943.
- 21 Allen G, Courchesne E. Differential effects of developmental cerebellar abnormality on cognitive and motor functions in the cerebellum: an fMRI study of autism. *Am J Psychiatry* 2003; **160**: 262–273.
- 22 Courchesne E, Townsend J, Akshoomoff NA, Saitoh O, Yeung-Courchesne R, Lincoln AJ et al. Impairment in shifting attention in autistic and cerebellar patients. *Behav Neurosci* 1994; **108**: 848–865.
- 23 Courchesne E, Allen G. Prediction and preparation, fundamental functions of the cerebellum. *Learn Mem* 1997; **4**: 1–35.
- 24 Gao JH, Parsons LM, Bower JM, Xiong J, Li J, Fox PT. Cerebellum implicated in sensory acquisition and discrimination rather than motor control. *Science* 1996; **272**: 545–547.
- 25 McDermott KB, Petersen SE, Watson JM, Ojemann JG. A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 2003; **41**: 293–303.
- 26 Corina DP, San Jose-Robertson L, Guillemin A, High J, Braun AR. Language lateralization in a bimanual language. *J Cogn Neurosci* 2003; **15**: 718–730.
- 27 Akshoomoff NA, Courchesne E, Townsend J. Attention coordination and anticipatory control. *Int Rev Neurobiol* 1997; **41**: 575–598.
- 28 Raichle ME, Fiez JA, Videen TO, MacLeod AM, Pardo JV, Fox PT et al. Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb Cortex* 1994; **4**: 8–26.
- 29 Kim SG, Ugurbil K, Strick PL. Activation of a cerebellar output nucleus during cognitive processing. *Science* 1994; **265**: 949–951.
- 30 Hatten ME, Heintz N. Mechanisms of neural patterning and specification in the developing cerebellum. *Annu Rev Neurosci* 1995; **18**: 385–408.
- 31 Hatten ME, Alder J, Zimmerman K, Heintz N. Genes involved in cerebellar cell specification and differentiation. *Curr Opin Neurobiol* 1997; **7**: 40–47.
- 32 Millen KJ, Wurst W, Herrup K, Joyner AL. Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse Engrailed-2 mutants. *Development* 1994; **120**: 695–706.
- 33 Millen KJ, Hui CC, Joyner AL. A role for En-2 and other murine homologues of *Drosophila* segment polarity genes in regulating positional information in the developing cerebellum. *Development* 1995; **121**: 3935–3945.
- 34 Kuemerle B, Zanjani H, Joyner A, Herrup K. Pattern deformities and cell loss in Engrailed-2 mutant mice suggest two separate patterning events during cerebellar development. *J Neurosci* 1997; **17**: 7881–7889.
- 35 Baader SL, Sanlioglu S, Berrebi AS, Parker-Thornburg J, Oberdick J. Ectopic overexpression of Engrailed-2 in cerebellar Purkinje cells causes restricted cell loss and retarded external germinal layer development at lobule junctions. *J Neurosci* 1998; **18**: 1763–1773.
- 36 Baader SL, Vogel MW, Sanlioglu S, Zhang X, Oberdick J. Selective disruption of 'late onset' sagittal banding patterns by ectopic expression of Engrailed-2 in cerebellar Purkinje cells. *J Neurosci* 1999; **19**: 5370–5379.
- 37 Vogel MW, Ji Z, Millen K, Joyner AL. The Engrailed-2 homeobox gene and patterning of spinocerebellar mossy fiber afferents. *Brain Res Dev Brain Res* 1996; **96**: 210–218.
- 38 Liu J, Nyholt DR, Magnussen P, Parano E, Pavone P, Geschwind D et al. A genomewide screen for autism susceptibility loci. *Am J Hum Genet* 2001; **69**: 327–340.
- 39 Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH. Autism Genetic Research Exchange Consortium. Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am J Hum Genet* 2002; **70**: 60–71.
- 40 Auranen M, Vanhala R, Varilo T, Ayers K, Kempas E, Ylisaukko-Oja T et al. A genomewide screen for autism-spectrum disorders: evidence for a major susceptibility locus on chromosome 3q25–27. *Am J Hum Genet* 2002; **71**: 777–790.
- 41 Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P et al. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet* 2001; **69**: 463–466.
- 42 Ye S, Dhillon S, Ke X, Collins AR, Day IN. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* 2001; **29**: E88–E98.
- 43 Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science* 1998; **281**: 363–365.
- 44 Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M et al. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* 2000; **280**: 103–110.
- 45 O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
- 46 Weeks DE, Lathrop M. Polygenic disease: methods for mapping complex disease traits. *Trends Genet* 1995; **11**: 513–519.
- 47 Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet* 1968; **38**: 226–231.
- 48 Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 1999; **65**: 1170–1177.
- 49 Martin ER, Monks SA, Warren LL, Kaplan NL. A test for linkage and association in general Pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000; **67**: 146–154.
- 50 Sham PC, Curtis D. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet* 1995; **59**: 323–336.
- 51 Spielman RS, McGinnis RE, Ewing WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
- 52 Lauritsen MB, Nyegaard M, Betancur C, Colinaux C, Josiassen TL, Kruse TA et al. Analysis of transmission of novel polymorphisms in the somatostatin receptor 5 (SSTR5) gene in patients with autism. *Am J Med Genet* 2003; **121**: 100–104.
- 53 Nabi R, Zhong H, Serajee FJ, Huq AH. No association between single nucleotide polymorphisms in DLX6 and Piccolo genes at 7q21–q22 and autism. *Am J Med Genet* 2003; **119**: 98–101.
- 54 Zhang H, Liu X, Zhang C, Mundo E, Macciardi F, Grayson DR et al. Reelin gene alleles and susceptibility to autism spectrum disorders. *Mol Psychiatry* 2002; **7**: 1012–1017.
- 55 Lauritsen MB, Borglum AD, Betancur C, Philippe A, Kruse TA, Leboyer M et al. Investigation of two variants in the DOPA

- decarboxylase gene in patients with autism. *Am J Med Genet* 2002; **114**: 466–470.
- 56 Veenstra-VanderWeele J, Kim SJ, Lord C, Courchesne R, Akshoomoff N, Leventhal BL *et al*. Transmission disequilibrium studies of the serotonin 5-HT_{2A} receptor gene (HTR2A) in autism. *Am J Med Genet* 2002; **114**: 277–283.
- 57 Bonora E, Bacchelli E, Levy ER, Blasi F, Marlow A, Monaco AP *et al*. Mutation screening and imprinting analysis of four candidate genes for autism in the 7q32 region. *Mol Psychiatry* 2002; **7**: 289–301.
- 58 Jamain S, Betancur C, Quach H, Philippe A, Fellous M, Giros B *et al*. Linkage and association of the glutamate receptor 6 gene with autism. *Mol Psychiatry* 2002; **7**: 302–310.
- 59 Cook Jr EH, Courchesne RY, Cox NJ, Lord C, Gonen D, Guter SJ *et al*. Linkage-disequilibrium mapping of autistic disorder with 15q11–13 markers. *Am J Hum Genet* 1998; **62**: 1077–1083.
- 60 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.
- 61 Cook Jr EH, Courchesne R, Lord C, Cox NJ, Yan S, Lincoln A *et al*. Evidence of linkage between the serotonin transporter and autistic disorder. *Mol Psychiatry* 1997; **2**: 247–250.
- 62 Yonan AL, Alarcon M, Cheng R, Magnossun PKE, Spence SJ, Palmer AA *et al*. A genomewide screen of 345 families for autism-susceptibility loci. *Am J Hum Genet* 2003; **73**: 886–897.
- 63 Petit E, Hérault J, Martineau J, Perrot A, Barthelemy C, Hameury L *et al*. Association study with two markers of a human homeogene in infantile autism. *J Med Genet* 1995; **32**: 269–274.
- 64 Zhong H, Serajee FJ, Nabi R, Huq AH. No association between the EN2 gene and autistic disorder. *J Med Genet* 2003; **40**: e4.
- 65 Le Hir H, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* 2003; **28**: 215–220.
- 66 Fedorova L, Fedorov A. Introns in gene evolution. *Genetica* 2003; **118**: 123–131.
- 67 Kakiuchi C, Iwamoto K, Ishiwata M, Bundo M, Kasahara T, Kusumi I *et al*. Impaired feedback regulation of XBP1 as a genetic risk factor for bipolar disorder. *Nat Genet* 2003; **35**: 171–175.
- 68 Tokuhiro S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T *et al*. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat Genet* 2003; **35**: 341–348.