Linkage and linkage disequilibrium scan for autism loci in an extended pedigree from Finland

Helena Kilpinen1, Tero Ylisaukko-oja1,2, Karola Rehnström1,2, Emilia Gaál1, Joni A. Turunen1, Elli Kempas1, Lennart von Wendt3, Teppo Varilo1,2 and Leena Peltonen1,2,4,5,*

1Department of Molecular Medicine, Institute for Molecular Medicine, Finland (FIMM), National Public Health Institute, Biomedicum, Haartmaninkatu 8 00251, Helsinki, Finland, 2Department of Medical Genetics, University of Helsinki, Biomedicum, Haartmaninkatu 8 00251, Helsinki, Finland, 3Unit of Child Neurology, Hospital for Children and Adolescents, Lastenlinnantie 2 00029 HUS, Helsinki, Finland, 4The Broad Institute, MIT and Harvard University, 7 Cambridge Center, Cambridge, MA 02141-2023, USA and 5Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Received February 17, 2009; Revised and Accepted May 11, 2009

Population isolates, such as Finland, have proved beneficial in mapping rare causative genetic variants due to a limited number of founders resulting in reduced genetic heterogeneity and extensive linkage disequilibrium (LD). We have here used this special opportunity to identify rare alleles in autism by genealogically tracing 20 autism families into one extended pedigree with verified genealogical links reaching back to the 17th century. In this unique pedigree, we performed a dense microsatellite marker genome-wide scan of linkage and LD and followed initial findings with extensive fine-mapping. We identified a putative autism susceptibility locus at 19p13.3 and obtained further evidence for previously identified loci at 1q23 and 15q11–q13. Most promising candidate genes were TLE2 and TLE6 clustered at 19p13 and ATP1A2 at 1q23.

INTRODUCTION

Autism [AD (MIM 209850)] is a severe childhood-onset disorder characterized by impaired development in social interaction and communication as well as presence of stereotyped patterns of interests and behavior before the age of 3 years (1,2). It is the most recognizable syndrome in a group of autism spectrum disorders (ASDs; or PDDs, pervasive developmental disorders), which include also Asperger syndrome (AS) and atypical forms of autism. The currently reported population prevalence for autism is 4–10 of 10,000, whereas the total prevalence for ASDs is reported as 10–60 of 10,000 (3–5). Awareness of ASDs has substantially increased during the past few years, and it is now recognized that when the early onset of the disorder is taken into account, the public health burden of autism in patient-years is substantial.

On the basis of twin and family studies, predisposition to autism is strongly heritable. The estimated sibling risk of 2–4% indicates significant familial clustering of the disorder, whereas twin studies have suggested heritability estimates of over 90% (6–8). Nevertheless, only a few confirmed genetic disorders that may lead to an autistic phenotype are currently known. These include syndromes, such as Fragile X, and chromosomal aberrations, especially on chromosomes X and 15q11–q13 (9–12). These and some more recent findings, such as identification of mutations in the neurexin genes (12,13), SHANK3 (14,15) and CNTNAP2 (16–19) point out that rare high-penetrance mutations may be causative for the autistic phenotype. In addition, de novo DNA copy number variation seems to influence autism predisposition (20–22). Yet, the predominant hypothesis is that a combination of multiple predisposing genetic and environmental factors is required in most cases. In fact, assuming multifactorial inheritance, it is estimated that potentially more than 15 individual loci each having a minor effect might be involved in the etiology of autism (23). Assuming such small effect variants, and thus extensive genetic heterogeneity, it is not surprising that the numerous genome-wide linkage scans performed to date have yielded poorly replicated findings and modest levels of significance. However, on the basis of a recent meta-analysis of six separate genome-wide linkage scans comprising a total of 771 affected sib-pairs, especially the locus at 7q22–q32 appears as a promising
candidate for autism (24), as well as 11p12–p13, which was the single major locus identified recently by the Autism Genome Project Consortium (25). Genetic heterogeneity and contribution of numerous rare alleles most probably complicate also ongoing genome-wide association studies, which rely on association to common variants. This heterogeneity was recently highlighted in a study by Morrow et al. (26), where homozygosity mapping was carried out in autistic children of consanguineous marriages. All of the mutations identified in this study were associated to autism only within single families.

Due to the limited success in linkage-based genome-wide scans in autism, novel approaches are needed to reveal the molecular mechanism underlying ASDs. If the genetic background of autism reflects that of other complex diseases, massive study samples are required for genome-wide association studies which can be expected to expose common, most probably low-impact alleles (27). On the other hand, linkage studies in large pedigrees are likely to contribute to our understanding of rare mutations with a high impact. Yet an alternative approach is to focus on population isolates, or preferably on further restricted sub-isolates with a well-established genealogy, in which a single (or few) causative variant(s) can be expected to be enriched (28). Such approaches focusing on the rare forms of common diseases have proved to be of high importance and revealed genes, which have led to a significant improvement in our understanding of molecular pathways underlying the disease. Further, there is an increasing number of examples of susceptibility genes initially mapped in rare families that have proved to be important also at the population level (29).

Population isolates, such as Finland, offer several well-recognized advantages for disease gene mapping. The results have been most striking in the mapping of rare Mendelian traits, but many of them hold also for mapping of common complex diseases. The benefits include for e.g. (i) shared common environment and culture of the study families, (ii) well-standardized diagnostic criteria and common training of clinicians, (iii) centralized healthcare records, (iv) opportunity to identify small sub-isolates or reconstruct large pedigrees based on population registers, and (v) reduced genetic heterogeneity at least in the cases where rare alleles confer significant increase in susceptibility to complex diseases (28,30). Moreover, recent evidence verifies that population isolates exhibit substantially higher linkage disequilibrium (LD) around common alleles than outbred samples suggesting that relatively sparse marker maps might be sufficient for initial disease gene localization (31). Here, we describe an extended ASD pedigree constructed from 20 nuclear families (Fig. 1) scattered all over the country by systematically examining the genealogy of all of the families in our nationwide ASD study sample. We carried out a dense microsatellite-based genome-wide scan in this unique pedigree and report a putative ASD susceptibility locus at 19p13.3 and further evidence for previously identified loci at 1q23 and 15q11–q13.

RESULTS

We performed a dense genome-wide scan in the 20 nuclear families of the extended pedigree (n_{affected} = 33) with a set of 1109 microsatellites, which yielded an average intermarker distance of 3.43 cM. Individual families were analyzed separately, except in the three cases where two nuclear families can be connected as one family on the level of first or second cousins (Fig. 1). Two primary strategies were selected to analyze the data. Since the fundamental hypothesis behind the current study was that the observed genealogical links reflect identical-by-decent sharing of the same ancestral susceptibility variant(s) in the current set of families, we wanted to maximally extract information of allele sharing both within and across the families. Our primary approach was the LD + Linkage method of Pseudomarker analysis program, as described by Goring and Terwiliger (32). Both dominant and recessive Pseudomarker analyses were conducted. However, the structure of the pedigree and the obtained results support recessive inheritance due to which
we mainly focused on the results of the ‘recessive Pseudomarker analysis’. The ‘dominant Pseudomarker analysis’ is analogous to affected relative pair methods in large families, and it weights the sharing between parents affected with ASDs and their affected children more strongly than that between unaffected parents and affected children. In the ‘recessive Pseudomarker analysis’ contributions of both parents are weighted equally. The Pseudomarker approach enables the combination of linkage and association evidence in various types of samples (family-based and case-control samples) in the same analysis (32). We also employed non-parametric multi-point linkage (NPL) analysis of Simwalk2 v. 2.91 software, which is especially suitable for complex pedigrees, to monitor allele sharing within the families (linkage) (33). Of the five NPL statistics produced by Simwalk2, we chose to report the two that have been shown to be most powerful and best suitable for dominant and recessive traits (‘BLOCKS’, referred here as ‘NPL_recessive’ and ‘MAX-TREE’, referred here as ‘NPL_dominant’) (34). These analyses were additionally run using all known genealogical connections (Fig. 1) to the common ancestors, but this had no marked effect on the results (data not shown). In the initial scan, genotypes from 22 regionally matched controls were included in the analyses to improve power in LD analyses as well as to better estimate allele frequencies in the linkage analysis. In the extended pedigree, Family 2 includes a twin pair with severe infantile autism as well as 11 individuals with AS, of which four have a relatively mild AS phenotype. To avoid the dominance of this single family in the initial LD + Linkage analysis, all AS individuals in the family were excluded from the primary Pseudomarker analysis (designated as Set 1). The strategy of assigning these individuals as both unaffected (Set 1) and affected (Set 2) was employed also in the multipoint linkage analyses by Simwalk2 as well as in the follow-up and fine-map of the best loci.

Initial genome-wide scan

Summary of the genome-wide LD + Linkage results from the recessive Pseudomarker analysis in the extended pedigree (Set 1) is given in Figure 2. Altogether nine loci exceeded the $-\log(\rho)$ value of 2.5 which was chosen based on the distribution of the results as an arbitrary cut-off for the selection of a reasonable number of follow-up loci (Table 1). For these nine loci, we additionally monitored the LD|Linkage value in Pseudomarker. This test of LD allowing, but not assuming, for linkage is used to demonstrate that some of the signals are obtained from haplotype sharing among individuals instead of linkage only. The best LD + Linkage $P$-values were observed with markers D1S2707 on 1q23.2 ($P = 0.00082$, Set 1) and D15S156 on 15q12 ($P = 0.00081$, Set 1). The evidence for D1S2707 was almost entirely attributable to LD ($P = 0.00079$, LD|Linkage, Set 1). The flanking markers of D1S2707 (D1S1653 and D1S484, 4.5 cM proximally and 1.4 cM distally, respectively) yielded no evidence of either linkage or LD in the Pseudomarker analysis. On 15q12, in contrast, also D15S975 (2.1 cM proximally to D15S156) yielded a suggestive $P$-value in the Linkage + LD analysis ($P = 0.0065$, Set 1), with most evidence for both markers again resulting from LD ($P = 0.02$ to $P = 0.0009$, LD|Linkage). In the dominant Pseudomarker analysis (Set...
1), only one locus exceeded the $-\log(p) = 2.5$ cut-off. Results of the dominant analysis are given in Supplementary Material, Figure S1.

In addition to the nine loci identified in the recessive Pseudomarker analysis (Table 1), we identified one significant locus in the Simwalk2 analysis with NPL of $2\log(p) = 3.57$ at D19S591 (Set 2, ‘NPL_dominant’; see Fig. 4) located at 19p13.3. These ten loci were selected for the follow-up stage, as discussed below (see Fig. 3, for a flowchart of the study). The second most significant linkage in the Simwalk2 analysis was observed on chromosome 6 with $2\log(p) = 2.15$ (D6S958, Set 2, ‘NPL_dominant’) while the results for all other chromosomes were below $2\log(p) = 1.5$ (data not shown).

### Follow-up stage and candidate gene analyses

Altogether 44 additional microsatellites from the ten genomic regions were analyzed in the extended pedigree in the follow-up stage. Detailed information of the follow-up markers in each locus is included in Supplementary Material, Table S1, whereas all follow-up results are reported in Supplementary Material, Table S2 and Figure S2. Two of the loci selected for follow-up (1q23 and 15q12) have been indicated in earlier autism studies. Two independent genome-wide screens for ASDs performed in Finnish families found evidence for 1q23, reporting highest LOD scores of 2.63 (D1S1653; ASD scan) and 3.58 (D1S484; AS scan) within 4.5 cM of D1S2707 (35,36). The extended pedigree of the current study overlaps with the previous ASD scan with four families (eight affected individuals); with the AS scan there is no overlap. The exclusion of the overlapping samples from the analyses was not considered because this would have decreased the already small sample size of the pedigree and broken up some of the genealogical links, resulting in a significant loss of information. The prior evidence for 15q11–q13 locus in autism arises from cytogenetic studies demonstrating that some 1–3% of autism cases are caused by inverted maternal duplications of this region (10,11,37). 15q11–q13 is also a well-known imprinted locus playing a key role in Angelman and Prader–Willi syndromes.

| Ranking from screen | Marker   | Chromosome | Genetic distance (cM) | LD + Linkage [−log(ρ)] | LD|Linkage [−log(ρ)] |
|---------------------|----------|------------|----------------------|------------------------|------------------------|
| 1                   | D15S156  | 15q12      | 15.1                 | 3.09                   | 3.05                   |
| 2                   | D1S2707  | 1q23       | 156.1                | 3.09                   | 3.10                   |
| 3                   | D13S232  | 13q12      | 8.7                  | 3.01                   | 2.85                   |
| 4                   | D14S283  | 14q11      | 14.7                 | 2.96                   | 2.52                   |
| 5                   | D8S1132  | 8q23       | ~113.1               | 2.68                   | 2.84                   |
| 6                   | D6S1279  | 6p24       | ~30                  | 2.66                   | 1.59                   |
| 7                   | D5S2090  | 5q32       | 150.0                | 2.62                   | 2.81                   |
| 8                   | D5S2066  | 5q35       | 205.7                | 2.56                   | 2.74                   |
| 9                   | D6S422   | 6p22       | 42.8                 | 2.51                   | 2.58                   |

All markers with LD + Linkage $-\log(ρ) > 2.5$ are displayed, together with the LD|Linkage values for these loci. Results produced by recessive Pseudomarker analysis (32) in the extended pedigree (Set 1).

*UCSC Human Genome Browser, May 2004 Assembly (http://genome.ucsc.edu).

*From ptel based on deCODE Genetic map (70).
From the cluster, was included due to a previous report of association to ASDs (43), and its role in phenotypically related Angelman syndrome. With the 41 SNPs analyzed, LD + Linkage was detected in the extended pedigree with six SNPs from the GABA A cluster (from $P = 0.02$ to $P = 0.0023$, Set 2, recessive Pseudomarker), with the most significant results again originating from four consecutive SNPs within the GABRB3 gene, yielding LD|Linkage from $P = 0.03$ to $P = 0.00084$ (best $P$-value rs71773713). Again, when analyzed in the nationwide study sample (autism and AS families), only modest evidence of LD + Linkage or LD|Linkage was seen at the region (from $P = 0.04$ to $P = 0.002$).

On 19p, we selected 13 biologically relevant candidate genes in the best multipoint linkage region for further analysis: PALM, GRIN3B, EFNA2, MBD3, GNG7, TLE6, TLE2, AES, GNA15, SH3GL1, SEMA6B, NRTN and PSPN (Table 2). Altogether 80 SNPs for these candidates were analyzed in the extended pedigree. Of these genes, TLE6, TLE2 and AES (also known as TLE5) are situated as a cluster 428 kb from D19S565 and 12.8 kb from D19S591, the two best markers of the initial scan (Fig. 4 and Supplementary Material, Fig. S3). These three genes belong to a TLE family of proteins homologous to the Drosophila Groucho (Gro) protein which is involved in neurogenesis during embryonic development. Since the most significant linkage signal from the initial Simwalk2 analysis was detected in a Set 2 analysis, where the multiple AS individuals of Family 2 markedly contribute to the linkage signal, we here focused on Set 1 results to identify Family 2-independent association signals. Interestingly, the most significant results at 19p13, and in the whole study, were seen within the TLE6–TLE2–AES gene cluster with altogether eight consecutive SNPs yielding LD + Linkage $P$-values $< 0.04$ in the same analysis (Set 1) with dominant Pseudomarker, consistently with the original multipoint linkage. Of these, rs4806893 and rs216283 yielded the best results (both $P = 0.000078$, LD + Linkage, Set 1, dominant Pseudomarker) together with rs216276 ($P = 0.00063$). For five of eight SNPs, there was also evidence of sharing across families ($P$-values from 0.0019 to 0.05, LD|Linkage, Set 1, dominant Pseudomarker). When comparing allele frequencies of these SNPs between affected individuals and controls, the minor allele of rs216276 was notably overrepresented in cases compared with controls (0.15 versus 0.06). With rs4806893 and rs216283, the major allele was instead overrepresented (0.74 in cases versus 0.54 in controls, both SNPs). The eight SNPs cover a region of 16.5 kb and are located mainly within TLE2 and the 3′-UTR/intergenic region of both TLE2 and TLE6 (transcribed in reversed directions; see Supplementary Material, Fig. S3 for details). Some evidence of LD + Linkage ($P < 0.01$) in the extended pedigree was additionally seen with the GNA15 gene (best $P = 0.004$, Set 1, dominant Pseudomarker). Owing to the encouraging results with the TLE-cluster, we additionally genotyped the SNPs ($n = 26$) for these genes in the nationwide autism and AS study samples. However, the analyses disclosed no comparable evidence of association in these study samples outside the extended pedigree (best LD + Linkage $P = 0.02$; majority $> 0.05$). In order to further investigate, the variation at this specific locus we constructed haplotypes of the eight associat-

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome</th>
<th>Position (bp)</th>
<th>Size (bp)</th>
<th>Genotyped SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>KCNJ10</td>
<td>1q23</td>
<td>156821108–156822207</td>
<td>11199</td>
<td>6</td>
</tr>
<tr>
<td>KCNJ9</td>
<td>1q23</td>
<td>15688433–15682285</td>
<td>7852</td>
<td>2</td>
</tr>
<tr>
<td>ATP1A2</td>
<td>1q23</td>
<td>156898621–156924645</td>
<td>27842</td>
<td>5</td>
</tr>
<tr>
<td>ATP1A4</td>
<td>1q23</td>
<td>156934346–156969836</td>
<td>35400</td>
<td>7</td>
</tr>
<tr>
<td>NOS1AP</td>
<td>1q23</td>
<td>158771239–159069886</td>
<td>298674</td>
<td>6</td>
</tr>
<tr>
<td>RG5</td>
<td>1q23</td>
<td>159770809–159770836</td>
<td>7227</td>
<td>5</td>
</tr>
<tr>
<td>UBE3A</td>
<td>15q12</td>
<td>23133489–23201746</td>
<td>68257</td>
<td>10</td>
</tr>
<tr>
<td>GABRB3</td>
<td>15q12</td>
<td>24342556–24470020</td>
<td>227464</td>
<td>19</td>
</tr>
<tr>
<td>GABRA5</td>
<td>15q12</td>
<td>24663151–24776749</td>
<td>113598</td>
<td>6</td>
</tr>
<tr>
<td>GABRG3</td>
<td>15q12</td>
<td>24799263–25451729</td>
<td>652466</td>
<td>6</td>
</tr>
<tr>
<td>PALM</td>
<td>19p13</td>
<td>660002–699328</td>
<td>39326</td>
<td>9</td>
</tr>
<tr>
<td>GRIN3B</td>
<td>19p13</td>
<td>951437–960723</td>
<td>9286</td>
<td>2</td>
</tr>
<tr>
<td>EFNA2</td>
<td>19p13</td>
<td>1237168–1250944</td>
<td>13776</td>
<td>2</td>
</tr>
<tr>
<td>MBD3</td>
<td>19p13</td>
<td>1527678–1543652</td>
<td>15974</td>
<td>4</td>
</tr>
<tr>
<td>GNG7</td>
<td>19p13</td>
<td>2462218–2653663</td>
<td>191445</td>
<td>21</td>
</tr>
<tr>
<td>TLE6</td>
<td>19p13</td>
<td>2984637–2980022</td>
<td>31385</td>
<td>11</td>
</tr>
<tr>
<td>AES</td>
<td>19p13</td>
<td>3003908–3013571</td>
<td>9463</td>
<td>8b</td>
</tr>
<tr>
<td>GNA15</td>
<td>19p13</td>
<td>3078230–3114747</td>
<td>32752</td>
<td>12</td>
</tr>
<tr>
<td>SH3GL1</td>
<td>19p13</td>
<td>4313236–4351471</td>
<td>40104</td>
<td>4</td>
</tr>
<tr>
<td>SEMA6B</td>
<td>19p13</td>
<td>4493600–4509503</td>
<td>15904</td>
<td>2</td>
</tr>
<tr>
<td>NRTN</td>
<td>19p13</td>
<td>5774818–5779335</td>
<td>4518</td>
<td>1</td>
</tr>
<tr>
<td>PSPN</td>
<td>19p13</td>
<td>6326305–6326860</td>
<td>556</td>
<td>1</td>
</tr>
</tbody>
</table>

*UCSC Human Genome Browser, May 2004 Assembly (http://genome.ucsc.edu).

*Includes one intergenic SNP.
ing SNPs using Phase v2.1.1 program, separately for the cases in the extended pedigree and the regional controls. The distribution of the different haplotypes between cases and controls revealed three common haplotypes (Supplementary Material, Table S3), of which one was notably more frequent in the cases (59%) compared with controls (38%) suggesting the presence of a susceptibility variant on this haplotype. However, on the basis of the distribution of the haplotype frequencies, no single haplotype could be expected to account for the entire association signal.

DISCUSSION

Linkage disequilibrium has been efficiently used for disease gene identification in monogenic disorders in the Finnish population. As few as four affected individuals are sufficient for disease gene localization due to extensive haplotype sharing around the disease-causing mutation (44,45). The starting point for this study is in many ways analogous to these early examples: the Finnish small founder population as a whole shows extensive degree of LD, not to mention young sub-isolates, such as the region from which this extended ASD pedigree originates (31). Additionally, this pedigree contains strikingly many individuals with an ASD, indicating possible enrichment of causative variants and thus providing an ideal setting for genetic mapping based on linkage and haplotype sharing.

We obtained evidence of linkage at three loci in the extended pedigree. Our results provide additional support to previously reported ASD susceptibility loci at 1q23 and 15q11–q13, and reveal an additional interesting locus at 19p13. Results at 19p13 were obtained with the Set 2 analysis in which the signal mostly originates from the multiple individuals with AS in Family 2. Also, the results at 19p13 were obtained with a model best suitable for dominant inheritance in both Simwalk2 and Pseudomarker. In contrast, the most significant results at 1q23 and 15q12 were obtained with a Set 1 recessive Pseudomarker analysis. Due to the unique structure of the extended pedigree, it is important to be able to separate out the effects of linkage from the results, because linkage regions generally are much larger than LD regions. Therefore, using the test of LD allowing for linkage in Pseudomarker, it is possible to model only for the sharing between families and not within families. The test does not assume linkage, but instead subtracts out the information about linkage from the joint analysis of LD + Linkage, making it applicable in situations where the linkage signal might not be formally significant. Also, since the power of SNPs to detect linkage is smaller than with microsatellites, significant linkage cannot necessarily be expected. At all three loci, 1q23, 15q12 and 19p13, substantial evidence of sharing across families, or LD, was detected as expected, pointing to greater genetic homogeneity due to isolation. In the association study of regional candidates, the most plausible and interesting findings were ATP1A2 at 1q and the TLE gene cluster at 19p, both of which provided considerable evidence of association.

At 1q23, the most significant results were obtained with five consecutive SNPs spanning 25.8 kb of the ATP1A2 gene in the extended pedigree (best $P = 0.00055$, LD|Linkage, Set 1). Linkage to this locus has been detected in at least two previous ASD studies (35,36) as well as in multiple schizophrenia studies (40,46). ATP1A2 is one of the four genes that make up the syntenic seizure susceptibility locus (Szs1) in mouse, originally identified by quantitative trait locus mapping (38,39). Association between seizure susceptibility and idiopathic generalized epilepsy and KCNJ10 has since been detected in both in mice and humans (38,39,47). Association of this locus with ASDs is of interest since up to 30% of individuals with autism suffer from epilepsy (48).
Suggestive linkage at 19p13 has been reported in previous genome-wide scans for ASDs, but reported values have been modest (49–54). Therefore, to observe a multipoint $-\log(p)$ value of 3.57 ($P = 0.000029$) at this locus with only 20 nuclear families, seems encouraging. In fact, at 19p13, all of the families in the extended pedigree show complete segregation with the trait (linkage), implying that this locus might contribute to the disease risk in this pedigree. In particular Family 2 displays significant sharing across affected individuals (Supplementary Material, Fig. S4), which explains in part the observed linkage signal.

In the fine-mapping stage, the most significant results in the whole study were observed within a cluster of three genes located just 12.8 kb away from D19S591, the best marker in the initial scan. Eight consecutive SNPs, located in the borderline of TLE6 and TLE2 genes (Supplementary Material, Fig. S3), displayed significant LD + Linkage (P-values from 0.04 to 0.000078, Set 1). With five of these SNPs, the signal was mostly attributable to sharing across families (P-values from 0.05 to 0.00019, LD)Linkage, Set 1). Due to the lack of established methods to correct for multiple testing in multilevel gene-mapping studies, we have not attempted to correct for multiple testing or LD structure in our study. Taking the small sample size into account, the results would not remain significant if corrected for all markers and tests. However, the most significant association signal with TLE2 and TLE6 ($P = 0.000078$) does remain significant if it is corrected for the total number of SNPs ($n = 152$) and tests (dominant and recessive Pseudomarker analysis) performed (Bonferroni, $P = 0.024$). The different tests performed by Pseudomarker (LD + Linkage and LD)Linkage cannot be included in this correction since the tests are not independent and the Bonferroni correction assumes independence of tests. The genes, TLE6, TLE2 and AES (also known as AES), belong to the human TLE (transducin-like enhancer of split) protein family that is extensively homologous with the Drosophila Gro (http://flybase.org) protein which, together with proteins of the hairy/enhancer of split (HES) family, is involved in neurogenesis during embryonic development as components of the Notch signaling pathway (55,56). All of the members in the human Gro/TLE family share a conserved TLE_N domain and act as transcriptional corepressors. They have been suggested to perform functions analogous to their Drosophila counterparts, that is, negatively regulating neuronal development and differentiation (57). Loss of function of Gro and HES proteins, and other components of the Notch signalling pathway, results in the overproduction of central and peripheral neurons (58), which is of interest in regard of macrocephaly reported in ~20% of individuals with autism (59) and increased brain volume frequently observed in autistic cases (60). In Drosophila, Gro is also known to interact with the conserved Engrailed protein (61), whose human homologue EN2 was recently associated with autism (62).

As a conclusion, on the basis of this study, we suggest that loci at 1q23, 15q12 and 19p13 are likely to contain genes that increase susceptibility to ASDs. In particular, the results obtained with SNPs at 1p are promising with considerable linkage supporting the association evidence. In the next step, extensive sequencing of the candidates is required to reveal the complete allelic variability of the genes. However, the fact that none of the association results could be replicated outside the extended pedigree in the nationwide study samples indicates that the putative predisposing loci represent an enrichment of these loci in a small founder population. Since comparable information from single consanguineous autism families has provided useful also in other studies, as highlighted recently by Morrow et al. (26), the results obtained with this pedigree should not be ignored simply because all of them do not reach genome-wide significance. We are currently working with detailed phenotypic analyses and a 317k SNP array to further characterize this unique ASD pedigree with the hope to eventually identify the specific genetic factors responsible for the disease phenotype in the pedigree. Such apparently rare factors could provide new insight to the molecular mechanism of the autistic phenotype, as has been the case with some of the rare high penetrance mutations identified in individuals with autism to date.

MATERIALS AND METHODS

Study samples and genealogy

We first identified a total of 10 families with autism whose ancestors originated from a single small farm in a village of the late-settlement region of Central Finland some 5–10 generations ago. When we followed up all the ancestral trees back up to 12 generations, we were able to distinguish nine ancestors connecting the 10 autism families, which, most interestingly, were born on the same small farm 215–350 years ago. Local church and civil registers were utilized for information after year 1850 and the Finnish National Archives for the earlier periods in accordance with published criteria (63). It is thus probable that these families share one common ancestor, although the archives did not reach back enough to allow the identification of the founder couple. We were also able to link 10 additional nuclear families with ASDs to this core pedigree as well as to reveal additional genealogical links among the 20 nuclear families, as shown in Figure 1. The consanguinity of these autism families leading to the identified pedigree structure here is strikingly similar to the pedigrees we have uncovered in numerous rare recessive Mendelian disorders of the Finnish disease heritage, such as variant form of late infantile neuronal ceroid lipofuscinosis (vLINCL [MIM 256731]) (64) or infantile onset spinocerebellar ataxia (IOSCA [MIM 271245]) (65), thus providing an optimal setting for genetic mapping studies and identification of the disease gene(s) (66,67). In total, our extended pedigree consists of 20 Finnish nuclear families with altogether 34 individuals affected with an ASD, of which 25 are males and 9 are females. Of these, 17 are diagnosed with infantile autism, 14 with AS, and three with a PDD not otherwise specified (PDD-NOS). Since two of the males affected with infantile autism are monozygotic twins, we have used only one of them in the analyses, making the total number of affected individuals 33. In all Pseudomarker analyses performed in the extended pedigree, we utilized genotype data from regionally matched controls to properly control for the diversity of alleles in the general population. The control samples were gathered from the same village.
where the families in the extended pedigree originate. In the initial scan we analysed 22 controls and for the follow-up and fine-map stages the number of controls was increased to 93.

We followed the most promising loci emerging from the initial scan by genotyping regional SNP markers in the complete pedigree as well as in the nationwide sample collection of 238 familial autism and AS cases and their family members (Table 3), for which the diagnostic procedures have been described earlier (35,68,69). The carefully phenotyped autism study sample consists of 97 Finnish families with 119 affected individuals diagnosed with infantile autism according to the ICD-10 (1) and DSM-IV (2) criteria. Only families with at least one child with infantile autism were included, whilst families with associated medical conditions such as Fragile X syndrome or profound mental retardation were excluded. The AS sample contains 28 large Finnish pedigrees with 119 affected individuals. The pedigrees contain only AS cases fulfilling the ICD-10 criteria in multiple subsequent generations (Table 3). (It should be noted that in the extended pedigree four individuals in Family 2 display notable AS-like features [here assigned as AS] but do not completely meet all of the ICD-10 criteria for AS.) Only individuals with normal overall cognitive development before the age of 3 were included in the AS sample. Of the 20 nuclear families in the extended pedigree, 16 are included in the current nationwide autism study sample, one in both autism and AS samples, and three in neither. This study has been approved by relevant ethical committees and informed written consent was received from all the participating families.

### Laboratory methods

The microsatellite markers of the initial scan \((n = 1109)\) were genotyped by standard procedures at deCODE Genetics Inc. (Reykjavik, Iceland). Follow-up microsatellites were genotyped with the ABI 3730 DNA sequencer, analysed with GeneMapper v.3.0 software (Applera Corporation, Norwalk, CT, USA) and verified by two individuals independently. In all multipoint analyses, we used the deCODE high-resolution genetic map (70). SNP markers in the fine-map stage were genotyped either with Sequenom’s homogenous MassEXTEND (hME) and iPLEX technology using the Mass ARRAY Platform, as specified by manufacturer’s instructions (Sequenom, San Diego, CA, USA), or by fluorogenic 5’ nuclease allelic discrimination chemistry (TaqMan) with an ABI Prism 7900 Sequence Detection System (Applied Biosystems, Foster City, CA, USA). All genotypes were checked for correct Mendelian transmission with PEDCHECK v.1.1 (71) and monitored for Hardy–Weinberg equilibrium. Markers accepted for analysis displayed a minimum genotyping success rate of 90%, with the majority of markers having a success rate of >95%. The borderline for the MAF of SNPs was 5%, with most of the SNPs having MAF of >10%. Individuals were treated either as affected or unknown.

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

### ACKNOWLEDGEMENTS

We wish to thank Drs Irma Järvelä, Reija Alen, Raija Vanhala, Raili Riihonnen, Taina Nieminen-von Wendt, Ismo Makkonen and Mari Auranen for their contribution in collecting and characterizing the sample. Professor Joseph D. Terwilliger is thanked for advice in statistical issues concerning Pseudomarker.

Conflict of Interest statement. None declared.

### FUNDING

This work was supported by the Center of Excellence in Complex Disease Genetics of the Academy of Finland; Biocentrum Helsinki; Päiviikki and Sakari Sohlberg Foundation; and Autism Speaks/Cure Autism Now [to T.Y.].

### REFERENCES


32. Gurling, H.M., Kalsi, G., Brynjolfsson, J., Sigmundsson, T., Sherrington, R., Mankoo, B.S., Read, T., Murphy, P., Blaveri, E., McQuillen, A. et al. (2001) Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q22-23, 2q33.2-3, and 8p21–22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.11-12.3. Am. J. Hum. Genet., 68, 661–673.


