HAL
open science

Novel method for combined linkage and genome-wide association analysis finds evidence of distinct genetic architecture for two subtypes of autism

Veronica Vieland, Joachim Hallmayer, Yungui Huang, Alistair Pagnamenta, Dalila Pinto, Hameed Khan, Anthony Monaco, Andrew Paterson, Stephen Scherer, James Sutcliffe, et al.

To cite this version:

Veronica Vieland, Joachim Hallmayer, Yungui Huang, Alistair Pagnamenta, Dalila Pinto, et al.. Novel method for combined linkage and genome-wide association analysis finds evidence of distinct genetic architecture for two subtypes of autism. Journal of Neurodevelopmental Disorders, 2011, 3 (2), pp.113-123. 10.1007/s11689-011-9072-9 . inserm-03135308

HAL Id: inserm-03135308

https://inserm.hal.science/inserm-03135308

Submitted on 8 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

J Neurodevelop Disord (2011) 3:113–123
DOI 10.1007/s11689-011-9072-9

Novel method for combined linkage and genome-wide association analysis finds evidence of distinct genetic architecture for two subtypes of autism

Veronica J. Vieland · Joachim Hallmayer · Yungui Huang · Alistair T. Pagnamenta · Dalila Pinto · Hameed Khan · Anthony P. Monaco · Andrew D. Paterson · Stephen W. Scherer · James S. Sutcliffe · Peter Szatmari · The Autism Genome Project (AGP)

Received: 20 October 2010 / Accepted: 4 January 2011 / Published online: 19 January 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The Autism Genome Project has assembled two large datasets originally designed for linkage analysis and genome-wide association analysis, respectively: 1,069 multiplex families genotyped on the Affymetrix 10 K platform, and 1,129 autism trios genotyped on the Illumina 1 M platform. We set out to exploit this unique pair of resources by analyzing the combined data with a novel statistical method, based on the PPL statistical framework, simultaneously searching for linkage and association to loci involved in autism spectrum disorders (ASD). Our analysis also allowed for potential differences in genetic architecture for ASD in the presence or absence of lower IQ, an important clinical indicator of ASD subtypes. We found strong evidence of multiple linked loci; however, association evidence implicating specific genes was low even under the linkage peaks. Distinct loci were found in the lower IQ families, and these families showed stronger and more numerous linkage peaks, while the normal IQ group yielded the strongest association evidence. It appears that presence/absence of lower IQ (LIQ) demarcates more genetically homogeneous subgroups of ASD patients, with not just different sets of loci acting in the two groups, but possibly distinct genetic architecture between them, such that the LIQ group involves more major gene effects (amenable to linkage mapping), while the normal IQ group potentially involves more common alleles with lower penetrances. The possibility of distinct genetic architecture across subtypes of ASD has implications for further research and perhaps for research approaches to other complex disorders as well.

Keywords Autism · Linkage analysis · Genome-wide association · PPL · PPLD · IQ

V. J. Vieland (✉) · Y. Huang
Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH 43205, USA
e-mail: Veronica.Vieland@nationwidechildrens.org

J. Hallmayer
Department of Psychiatry, Division of Child and Adolescent Psychiatry and Child Development, Stanford University School of Medicine,
Stanford, CA 94304, USA

A. T. Pagnamenta · A. P. Monaco
Wellcome Trust Centre for Human Genetics, University of Oxford,
Oxford OX3 7BN, UK

D. Pinto · H. Khan · A. D. Paterson · S. W. Scherer
The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children and Department of Molecular Genetics, University of Toronto,
Toronto, ON M5G 1L7, Canada

J. S. Sutcliffe
Department of Molecular Physiology and Biophysics, Vanderbilt Kennedy Center, and Centers for Human Genetics Research and Molecular Neuroscience, Vanderbilt University,
Nashville, TN 37232, USA

P. Szatmari
Department of Psychiatry and Behavioural Neurosciences, McMaster University,
Hamilton, ON L8N 3Z5, Canada

Introduction

Autism spectrum disorders (ASDs) are heritable, genetically complex neurodevelopmental conditions. In this paper, we

search for ASD genes through combined analysis of two datasets collected by the Autism Genome Project (AGP). The AGP previously reported linkage to 11p12–p13 along with notable copy number variations in the largest collection of multiplex ASD families analyzed to date (Szatmari et al. 2007; see also Liu et al. 2008), followed by copy-number variation (CNV) and association analysis in a large cohort of trios (Anney et al. 2010; Pinto et al. 2010). Here, we use a distinct statistical method, based on the PPL framework (Vieland 1998, 2006; Yang et al. 2005; Vieland et al. 2008; Wratten et al. 2009; Huang and Vieland 2010), to reconsider the combined multiplex and trio data sets.

The PPL statistical framework has three principle advantages in this context. (1) It handles genetic heterogeneity via "sequential updating" across data subsets (Vieland et al. 2001; Huang and Vieland 2001; Govil and Vieland 2008). The posterior evidence from previously analyzed data is carried forward as prior evidence as new data subsets are analyzed, with underlying genetic parameters (allele frequencies, penetrances, and levels of heterogeneity) allowed to vary between subsets. This can be a powerful method for discovering genetic signals arising from even very small subsets of the data, provided only that we have classification variables allowing the division of families (or cases) into relatively more homogeneous groups (Govil and Vieland 2008). (2) The PPL accumulates evidence against genetic hypotheses as well as in favor of them. Thus inspection of subset-specific contributions to the omnibus signal can distinguish among subsets that are supporting the hypothesis and subsets that are actually contributing evidence against it. (3) The PPL permits analysis of multiplex families and trios in a unified manner. The multiplex families provide linkage information, while the trios provide information on allelic associations. Here, we introduce a novel method for genome-wide analysis based on simultaneous use of linkage and association information from two different sets of data.

It is widely accepted that ASD is genetically heterogeneous, but less clear whether clinical features can be used to demarcate more homogeneous subclasses. Familial concordances for specific ASD symptoms are not strong, and there is generally high intrafamilial variability. However, familiality for nonverbal IQ has been reported in several studies (Le Couteur et al. 1996; Silverman et al. 2002; Szatmari et al. 2008; MacLean et al. 1999). This is in line with more recent family and twin studies suggesting that IQ is the most heritable component of the ASD phenotype (Szatmari et al. 2008). Furthermore, subgrouping ASD patients on the basis of IQ has provided the most consistent method for distinguishing patients on a number of dimensions. At the lower end of the IQ range, there is considerable overlap between autistic features and chromosomal syndromes (Xu et al. 2004); epilepsy is more prevalent, and the ratio of females to males approaches unity in contrast to the preponderance of males among higher IQ cases (Amiet et al. 2008). Moreover, there is compelling emerging evidence of considerable etiologic overlap between the clinical classification of intellectual disability, various mental retardation syndromes, and ASD in terms of rare de novo and inherited CNVs (Guilmatre et al. 2009; Bijlsma et al. 2009; Marshall et al. 2008). Indeed, the distinction of "high" and "low" functioning autism is often based on IQ, and indicates groups that differ with respect to associated brain dysfunction, outcome, and response to treatment (Lotspeich et al. 2004; Allen et al. 2001; Stevens et al. 2000).

Here, we accumulate the total, or "omnibus," evidence across subsets of families characterized by the presence or absence of lower IQ autistic individuals while allowing for the fact that the subsets may differ substantially from one another. We find compelling evidence that, indeed, the lower IQ group appears to be genetically distinct.

Methods

Participants

Multiplex families (N=1,069), each containing at least 2 individuals diagnosed with autism by the Autism Diagnostic Interview (ADI; Le Couteur et al. 1989) and clinical best estimate, were contributed by 10 sites. (See Szatmari et al. (2007) for additional details. Note that some "sites", or research groups, covered multiple data collection locations; however, sample sizes precluded further subdivision of the data.) IQ was recorded as a dichotomous trait. Families were classified as lower IQ (LIQ, N=255) if at least one ASD individual had performance IQ≤50 or was coded as "missing due to low functioning"; as normal IQ (NIQ, N=580) if all ASD individuals had IQ>50; and as missing IQ (MIQ, N=234) if there were no lower IQ individuals and at least one affected individual missing IQ information.

Trios (N=1,129) were contributed by 8 sites. (See Anney et al. (2010) for additional details. Again, some sites covered multiple data collection locations). Children met criteria for either autism or ASD based on ADI and ADOS criteria. A trio was classified as LIQ (N=285) if the child had performance IQ≤70 or was coded as "missing due to low functioning", as NIQ (N=394) if the child had IQ>70, and as MIQ (N=450) if the IQ information was missing. Changes in IQ classification by the AGP over time have led to the slight difference in IQ classification compared to the multiplex families. However, IQ is used only to subdivide the sample into relatively more homogeneous subsets, not as an outcome variable in its own right, and this is therefore unlikely to appreciably affect the results. Note

too that the proportion of LIQ families is similar (25% vs. 24%) in the trios and multiplex families, respectively, suggesting that the change in cutoff might actually appropriately compensate for differences between the two datasets. Of the trios, 283 overlapped with the multiplex families, but only a single case from each overlapping family was used in the LD analyses; thus there is no overlap in the information extracted from the overlapping samples. All trios were of European ancestry (Anney et al. 2010). All sites had Institutional Review Board approval for this study, and the research was conducted in accordance with the World Medical Association Declaration of Helsinki (2000). Written informed consent was obtained from all subjects after the study had been fully explained.

Genotyping and data cleaning

Details of genotyping methods are given in Szatmari et al. (2007) and Anney et al. (2010). In preparation for linkage analysis, marker data were cleaned for family structure problems and Mendelian inconsistencies. Merlin (Abecasis et al. 2002) was run to detect and remove unlikely double recombinants, and to cluster any SNPs in LD groups. (Most parents were genotyped and LD in the marker map proved not to affect the results.) In preparation for LD analyses, marker data were additionally cleaned for marker missingness (>5%), sample missingness (>5%), and excess Mendel errors both by SNP and individual. Markers with minor allele frequency <1% were dropped, as were SNPs with a Hardy–Weinberg (HW) p value$<1\times10^{-10}$ in at least one data subset or HW p value<0.05 in at least three subsets. After cleaning, 749,933 SNPs remained in the analyses.

Statistical methods

All analyses were conducted using the software package KELVIN (Huang et al. 2006), which implements the PPL class of models for measuring the strength of genetic evidence (Vieland 1998, 2006). The two specific statistics employed were the PPL itself (posterior probability of linkage) and the posterior probability of trait-marker linkage disequilibrium (PPLD). Linkage analyses utilized LOD scores computed in Merlin (Abecasis et al. 2002; Lander and Green 1987) as input to PPL calculations (Vieland 1998). The genetic map is based on http://compgen.rutgers.edu/mapopmat (Matise et al. 2007; release 10/09/06).

The PPL as applied here is parameterized as a dichotomous trait model with parameters α (the admixture parameter of Smith (1963), representing the proportion of 'linked' pedigrees), p (the disease allele frequency), and the penetrance vector f_i, representing the probability that an individual with genotype i develops disease, for $i-1..3$. All trait parameters are integrated out of the final statistic, using uniform prior distributions, implicitly allowing for dominant, recessive, and additive models along with intra-subset heterogeneity. This provides a robust approximation for mapping complex traits in terms of the marginal model at each locus, and because the parameters are integrated out, no specific assumptions regarding their values are required. The likelihood also contains two location parameters: the recombination fraction θ and the standardized LD parameter D′, representing trait–marker association due to physical proximity.

The PPL framework accumulates evidence across data subsets by integrating the trait parameters out of the likelihood separately for each subset, using Bayesian sequential updating to combine the marginal information regarding θ and D′ across subsets. This procedure allows for genetic differences among data subsets, and is far more robust in retaining true signals originating from individual subsamples than analyses that simply combine subsets for a single analysis (Vieland et al. 2001; Huang and Vieland 2001; Govil and Vieland 2008). Here, we have subdivided the data and sequentially updated across IQ groups. Because the AGP families have been contributed by multiple research groups, we also sequentially update over "site." Sites can vary with respect to the populations from which they recruit, ascertainment strategies and criteria, and subtle differences in clinical evaluations; simple sampling variability can also lead to inter-site differences. While not usually considered as a separate source of variation in genetic studies, the importance of allowing for site effects has been long appreciated in other settings, such as clinical trials. After dividing by IQ and site, subset sizes ranged from N=20–148 (mean=62) for the multiplex families and N=20–169 (mean=71) for the trios.

The PPL is on the probability scale, and its interpretation is therefore straightforward: e.g., PPL=40%, means that there is a 40% probability of a trait gene at the given location based on these data. Based on earlier calculations (Elston and Lange 1975), the prior probability at each location is set to 2%, so that PPLs>2% indicate (some degree) of evidence in favor of a trait gene at that locus, while PPLs<2% represent evidence against the location. The prior probability of LD given linkage (L) is also set to 0.02, so that in the absence of prior linkage information P(L&LD)=0.0004 (see also Welcome Trust Case Control Consortium 2007 for justification of a comparable figure).

Novel here is a mathematically rigorous method for using linkage information from the multiplex families to inform the association analyses, based on the fact that PPLD=PP(LD|L)×PPL (see Huang and Vieland 2010 for additional details). We interpolated the PPL results onto the physical map, and inserted the measured PPL into this equation. Thus, PPLs<2% will depress PPLDs, and PPLs>

2% will increase PPLDs, by increasing the prior probability of LD under a linkage peak, up to a maximum of 2% prior probability of LD when PPL=1 (see Roeder et al. (2007) for a related approach). This assumes that at least some ASD genes are etiologically relevant to both the multiplex and trio sets.

The PPL and PPLD are measures of statistical evidence, not decision-making procedures; therefore, there are no "significance levels" associated with them and they are not interpreted in terms of associated error probabilities (Royall 1997; Vieland and Hodge 1998). By the same token, no multiple testing corrections are applied to the PPL or PPLD, just as one would not "correct" a measure of the temperature made in one location for readings taken at different locations (Vieland 2006). Nevertheless, it may assist readers to have some sense of scale relative to more familiar frequentist test statistics. In simulations of 10,000 replicates of sets of 1,000 affected sib-pairs under the null hypothesis (no linkage), PPLs of 5%, 25%, and 80% were associated type 1 error probabilities of 0.00128, 0.00002, and <0.00001, respectively. In 10,000 null (no linkage, no LD) replicates of sets of 1,000 trios, no PPLDs>1% were observed, while PPLD>0.1% occurred in just 0.04% of replicates. At a locus with PPL<2%, this represents a PPLD<0.1%; while at a locus with PPL=80% this would still only correspond to a PPLD of 3.9%.

It is also of interest to consider "power" in the trio sample in particular. For relative risk (RR) of 1.3–1.7, our ability to detect association in regions lacking evidence of linkage is low; e.g., for RR=1.7, PPLDs>10% occur just 10% of the time. However, LD under linkage peaks is expected to be considerably stronger. For RR=2.0, with PPL=80%, 91% of PPLDs are>29% and 59% of PPLDs are >82%; with RR=2.5 99.6% of PPLDs are >82%. (Here we generated data with disease and marker minor allele frequencies of 0.1, varying D′ and the penetrances to achieve different RRs; actual power can obviously deviate from these results.) Thus, we are unable to draw definitive conclusions regarding absence of LD in unlinked regions of the genome based on the current sample size. However, the sample size appears adequate for detection of moderate allelic effects under linkage peaks.

Results

Omnibus linkage analysis

Figure 1a shows genome-wide PPL results for the omnibus (all groups) analysis. 92.6% of the genome showed evidence against linkage, 97.4% of the genome had PPLs <5%, and 98.7% of PPLs were <10% (99.6% ignoring chromosome 11, which shows several broad peaks). Against this backdrop, several peaks stand out. Two peaks on chromosome 11 coincide with locations reported in the two previous AGP analyses of this dataset (PPL=60%@11p13; PPL=93%@11p15.2). Also noteworthy is the very high PPL=87% on 16q21, as well as several other peaks including: 2p25 (PPL=12%), 4q31 (PPL=33%), 6q14 (PPL=11%), 18q22 (PPL=18%), and possibly two additional peaks on 11p15 and 11q14, which are more moderate in size although still salient relative to the background. We note that the detection of multiple loci in this dataset is attributable largely to the PPL's use of sequential updating. For instance, if we simply "pool" all sites and IQ groups together for a single analysis, on 16q21, the PPL at the peak is just 4%, compared to 87% based on sequential updating.

Linkage analysis by IQ group

Plotting the IQ groups separately (Fig. 1b–d), we see that the linkage plots suggest substantially different genetic profiles, with peaks occurring at different positions and more peaks in the LIQ group than the NIQ group. Notably, in several cases in which one IQ group gives evidence in favor of linkage, the other IQ group is actually giving evidence against linkage across the region. For example, the NIQ group gives PPL<2% across the entire region surrounding the peak on 16q21 in which the LIQ group is giving evidence for linkage.

In this context, the MIQ group serves as a kind of control. Combining data from two genetically distinct groups in a single analysis tends to attenuate linkage signals (Govil and Vieland 2008). Thus, if the LIQ and MIQ groups differ in their underlying genetic etiology, then the MIQ group, presumably comprised of a mixture of LIQ and NIQ families, should produce smaller linkage signals overall. On the other hand, if the appearance of two distinct genomic patterns comparing the LIQ and NIQ groups were the result of random variations rather than true genetic differences, the larger MIQ group would be expected to yield larger linkage peaks, perhaps in separate locations. The observed pattern in the MIQ group corroborates the interpretation of these graphs as indicating that IQ is indeed demarcating genetically different subsets of the data.

The linkage signals on 1q31.3, 13q22.1, 14q24.2, and 16q21 are clearly driven by one IQ group in particular (with the other giving evidence against linkage), and in three of the four cases it is the LIQ group that is driving the signal. The peaks on 11p13, 11p15.2 are more difficult to parse: on the one hand, the omnibus PPLs are higher than the PPLs from either the LIQ or the MIQ subset; on the other hand, Fig. 1 strongly suggests that there are multiple loci on this chromosome, and possibly distinct genes operating in the two IQ groups (see below), which is consistent with the

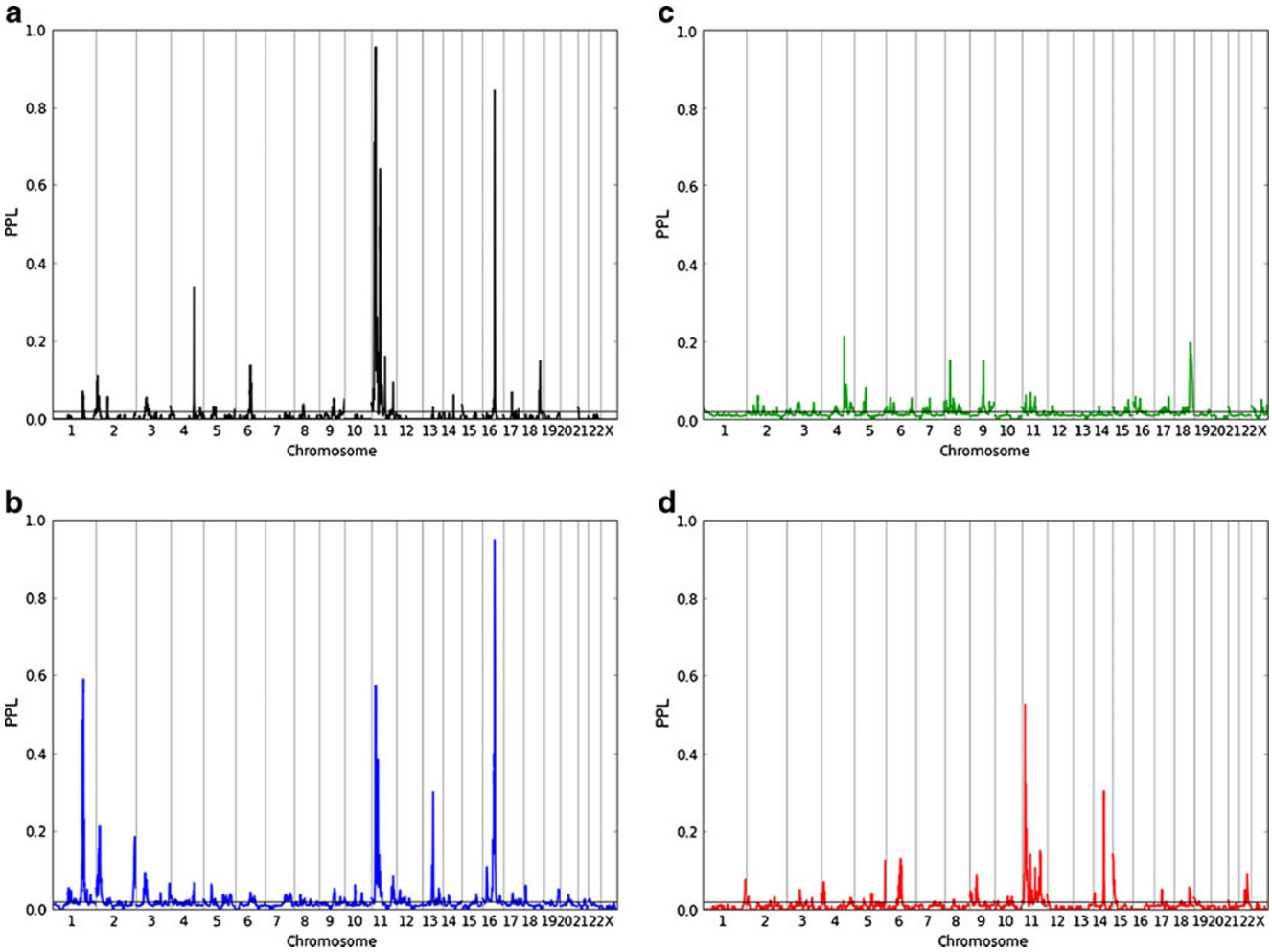

Fig. 1 Genome-wide linkage analyses in **a** omnibus, **b** LIQ, **c** MIQ, and **d** NIQ groups. The PPL (posterior probability of linkage) represents the probability of an ASD gene at each position. The *x*-axis represents chromosomes 1–23 (*X*) on the Kosambi cM scale; the *y*-axis is on the probability scale. The horizontal line at PPL=0.02 corresponds to the prior probability of linkage. Values below this line represent evidence against linkage, while values above the line represent evidence for linkage, at the given position

absence of appreciable signals from the MIQ group. Note too the small but visible omnibus signal on 15q11.2 (PPL=4%), which rises to PPL=14% in the NIQ group. This signal is directly over the known Prader–Willi ASD locus (van der Zwaag et al. 2010; Vorstman et al. 2006).

Omnibus combined linkage and association analysis

Figure 2a shows omnibus PPLD results. Against a very clean background, two modest peaks stand out. These occur at rs11603469 (11p15.2, PPLD=26%) and rs10221112 (16q21, PPLD=15%). In both cases, surrounding SNPs are also giving PPLDs elevated above the baseline (prior) probability of LD. On 11p15.2, rs11603469 is one of a small cluster of SNPs showing some LD evidence and overlapping the gene *FAR1* (rs11603469 itself is 10 kb from the *FAR1* start site); on 16q21 the SNP falls 351 kb from the nearest annotated gene (*GOT2*). A third, smaller, LD signal stands out on 4q31.23 (PPLD=6% at rs7668351, which falls in BC031092). In each case, these SNPs fall directly under corresponding linkage peaks (Fig. 3). It is noteworthy that in each case, multiple data subsets (sites) support LD, but also, multiple sites give evidence against LD, and some are merely neutral. In situations where allelic effects may vary across strata, pooling data across strata will tend to wash out these types of signals.

Combined linkage and association analysis by IQ group

Because the linkage results strongly suggest distinct etiology in the LIQ and NIQ groups, it is also of interest to consider the two groups separately in the association analyses. As expected, different SNPs are salient in the two groups (Fig. 2b–c). In general, the LIQ plot is slightly noisier than the NIQ plot, with smaller maximum peak and more "chatter" at the bottom of the plot. In part, this is

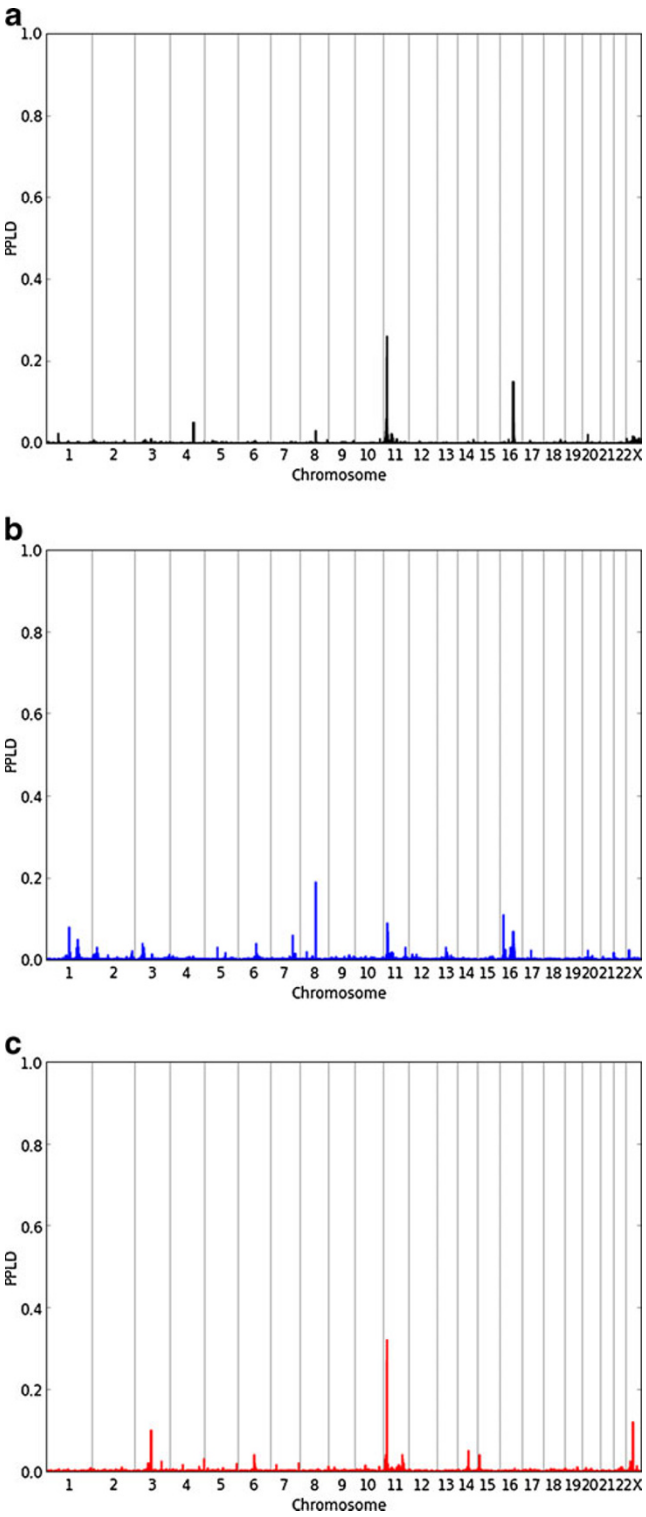

Fig. 2 Genome-wide combined linkage and association results from **a** omnibus, **b** LIQ, and **c** NIQ analyses. The PPLD (posterior probability of LD) represents the probability of allelic association with ASD due to LD for each SNP in turn, and utilizes both linkage information from the multiplex families and association information from the trios. The *x*-axis represents the physical map for chromosomes 1–23 (*X*); the *y*-axis is on the probability scale. An additional 151 markers from the pseudoautosomal region of *X* are not shown on the graph; none had PPLD exceeding the prior probability of LD

consistent with smaller sample size. However, "power" is not merely a function of sample size, but also of the underlying genetic model. The LIQ multiplex family dataset is also smaller than the multiplex NIQ dataset, yet the linkage signals are more numerous and higher in the LIQ group. The different pattern observed for the PPLD analyses may therefore be revealing real differences in the underlying genetic architecture, and not just reflecting relative sample sizes. We return to this point below.

Table 1 shows all PPLDs≥10% from the separate LIQ and NIQ analyses. Compared with the omnibus results, on 11p15.2, the omnibus signal in *FAR1* is driven by the NIQ group (maximum PPLD=32%). On 16q21, the omnibus signal is driven by the LIQ group, which on its own gives a PPLD=7%, bolstered by a small signal from the MIQ group (not shown); none of these SNPs falls in an annotated gene. Some additional signals also appear in the subgroup analyses that were not salient in the omnibus results (see Fig. 4; this figure also shows the distinct genetic linkage patterns on chromosome 11). On 8q21.12 (LIQ, not in an annotated gene), a pair of SNPs is showing evidence of LD in a region not showing evidence of linkage (the second SNP, rs7007634, has PPLD=8%). Additional association signals from the separate analyses are found on 3p12.1 (NIQ) and Xq13.1 (NIQ, with no clear difference between males and females) and 16p13.2 (LIQ).

Discussion

These analyses represent an examination of the AGP data from a unique statistical perspective. In contrast to the original analyses of the multiplex families (Szatmari et al. 2007), we have found multiple strong linkage signals. Disappointingly, however, PPLD analysis failed to find strong evidence of allelic effects under the linkage peaks, which would point us to the individual genes driving the linkage results. (We note, however, that follow up molecular work focused on one of the linkage peaks has established a strong prima facie case for involvement of the gene *CDH8* (Pagnamenta et al. 2011).) The apparent absence of allelic effects could reflect a genuine absence of LD under the peaks, or limitations in 1 M coverage of the peaks for LD mapping purposes. Another possibility is that there is sufficient heterogeneity that the trios are simply too dissimilar to the multiplex families to be informative at the same genes. The absence of dense SNP array data in the majority of the multiplex families makes direct evaluation of this possibility difficult.

It is also important to keep in mind that the trio sample is still relatively small, and in particular, the LIQ and NIQ groups individually may be too small to provide strong evidence on their own. The AGP is currently completing a

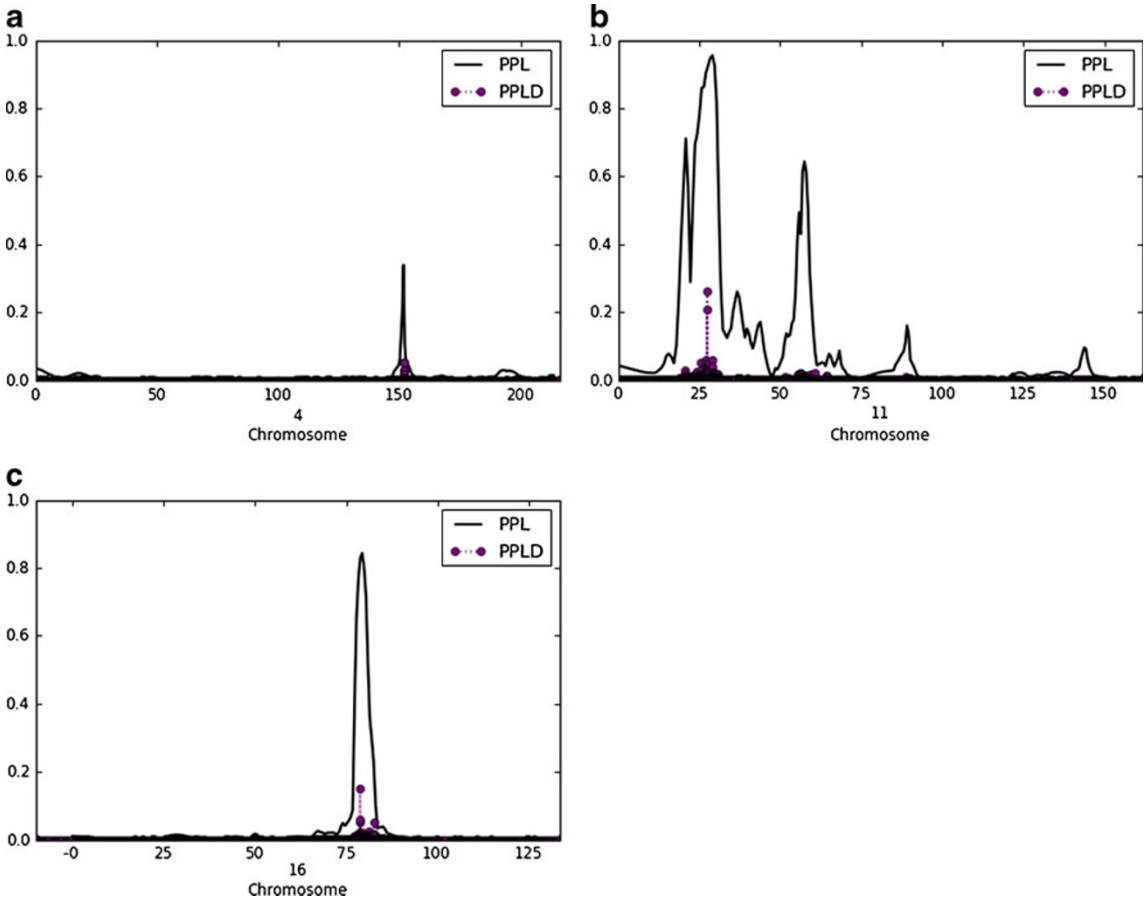

Fig. 3 Omnibus PPL and PPLD for chromosomes **a** 4, **b** 11, **c** 16. Units on the *x*-axis are in cM

second phase of trio data collection and genotyping, which will effectively double the sample size, and sequentially updating with the new dataset will provide better differentiation between SNPs truly supporting LD and SNPs with evidence against LD.

However, the overall pattern of results might reflect heterogeneity between the IQ groups rather than sample size. The linkage analysis finds more loci in the LIQ analyses than the NIQ analyses, despite the fact that the multiplex NIQ sample is 2.3 times the size of the multiplex LIQ sample; while in the LD analyses, where the sample sizes are better matched (the NIQ trio sample is just 1.4 times as large as the LIQ trio sample), the strongest signals are found in the NIQ group. Linkage analysis is powerful for identifying relatively major effects, that is, those in which mutations at a single locus greatly increase disease risk, even if only in a small subset of cases or against specific genetic and environmental backgrounds. Association analysis is particularly powerful for detecting alleles that individually confer small effects on disease risk, but do so in a relatively homogeneous manner across the study population.

Table 1 All PPLDs≥10% from the LIQ, NIQ analyses

IQ group	Chromosome	SNP name	cM position	Physical position[a]	PPLD	Gene
NIQ	11p15.2	rs11603469	27.41	13636952	0.32	8 kb from *FAR1*
NIQ	11p15.2	rs10500796	27.41	13655409	0.27	*FAR1*
LIQ	8q21.12	rs3885022	92.28	79890601	0.19	17.6 kb from *IL7*
NIQ	Xq13.1	rs12556351	59.34	71509736	0.12	*HDAC8*
LIQ	16p13.2	rs17722735	17.10	6660421	0.11	*A2BP1*
NIQ	3p12.1	rs9812103	117.58	86069410	0.1	*CADM2*

[a] Physical positions are based on NCBI build 36.1.

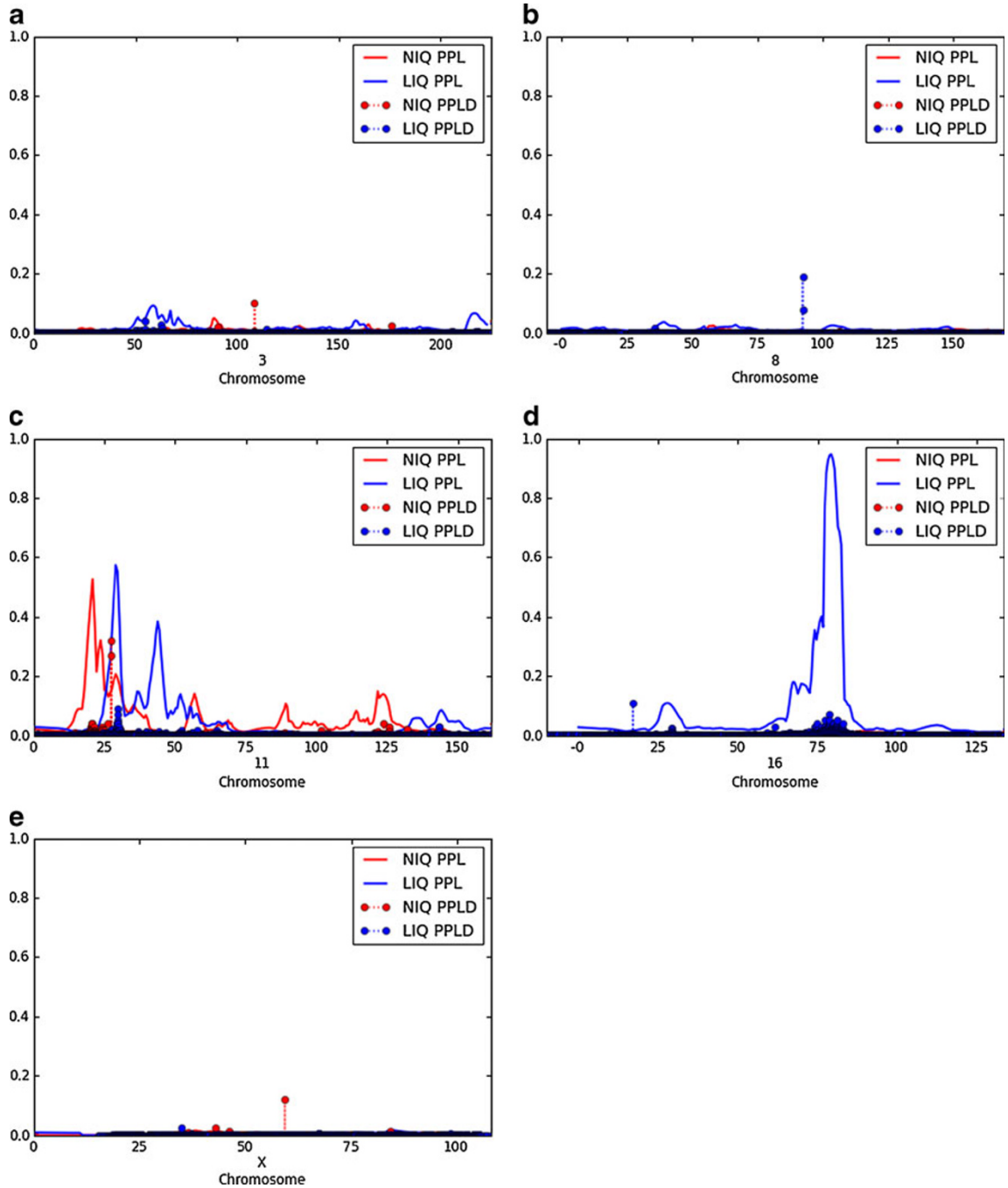

Fig. 4 PPL and PPLD for LIQ, NIQ groups respectively, for chromosomes **a** 3, **b** 8, **c** 11, **d** 16, **e** X containing SNPs shown in Table 1. Units on the *x*-axis are in cM

Thus, the two sets of results can be interpreted as telling a complementary story. The LIQ families may represent more strongly "genetic" forms of disease, in which a single gene or a small number of genes cause the disorder in any given individual, with sufficient overlap in causal genes across families to permit linkage mapping. The NIQ families, on the other hand, could involve more of a spectrum of conditions, possibly more highly influenced by the accumulation of variants in multiple genes each of smaller effect, or perhaps simply involving even higher levels of heterogeneity and/or many private mutations.

Of course until more data are available, this remains highly speculative. Further work to fully characterize the distinction between the LIQ and NIQ groups, combined with additional genetic analyses, will be needed to refine and test this hypothesis. But the results obtained thus far require us to at least consider the possibility that subtypes of autism have

distinct genetic architectures. This means that no single study design or experimental approach is likely to be optimal for all subtypes, and that we must be prepared for disparate results across different types of studies, or across data sets comprising different mixtures of subtypes. This point almost certainly applies to other complex disorders as well.

Finally, it is interesting to note that the association signal on 16p13.2, which does not fall under a PPL linkage peak, does fall within a linkage interval previously reported in a subset of AGP families, using a very different approach to untangling clinical heterogeneity based on latent class modeling (Bureau et al. 2008). Thus, the PPLD may be indicating a true association, but at a locus that our linkage analysis lacked power to detect, given the particular phenotypic classifications used here.

Acknowledgments This work was funded in part by NIH grant MH086117 (VJV) and NS042165 to the Autism Genetics Collaborative. The authors gratefully acknowledge the families participating in the study and the main funders of the AGP: Autism Speaks (USA), the Health Research Board (HRB, Ireland), The Medical Research Council (MRC, UK), Genome Canada/Ontario Genomics Institute, and the Hilibrand Foundation (USA). Additional support for individual groups was provided by the US National Institutes of Health (NIH grants: HD055751, HD055782, HD055784, HD35465, MH52708, MH55284, MH061009, MH06359, MH066673, MH080647, MH081754, MH66766, NS026630, NS049261), the Canadian Institutes for Health Research (CIHR), AP-HP Autism Speaks UK, Canada Foundation for Innovation/Ontario Innovation Trust, Deutsche Forschungsgemeinschaft (grant: Po 255/17-4) (Germany), EC Sixth FP AUTISM MOLGEN, Fundação Calouste Gulbenkian (Portugal), Fondation de France, Fondation FondaMental (France), Fondation Orange (France), Fondation pour la Recherche Médicale (France), Fundação para a Ciência e Tecnologia (Portugal), GlaxoSmithKline-CIHR Pathfinder Chair (Canada), the Hospital for Sick Children Foundation and University of Toronto (Canada), INSERM (France), Institut Pasteur (France), the Italian Ministry of Health convention 181 of 19.10.2001, the John P Hussman Foundation (USA), McLaughlin Centre (Canada), Netherlands Organization for Scientific Research (Rubicon 825.06.031), Ontario Ministry of Research and Innovation (Canada), Royal Netherlands Academy of Arts and Sciences (TMF/DA/5801), the Seaver Foundation (USA), the Swedish Science Council, The Centre for Applied Genomics (Canada) and the Utah Autism Foundation (USA).

Conflict of interest The authors have no conflicts of interest to report.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002;30(1):97–101.

Allen DA, Steinberg M, Dunn M, Fein D, Feinstein C, Waterhouse L, et al. Autistic disorder versus other pervasive developmental disorders in young children: same or different? Eur Child Adolesc Psychiatry. 2001;10(1):67–78.

Amiet C, Gourfinkel-An I, Bouzamondo A, Tordjman S, Baulac M, Lechat P, et al. Epilepsy in autism is associated with intellectual disability and gender: evidence from a meta-analysis. Biol Psychiatry. 2008;64(7):577–82.

Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, et al. A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet. 2010;19(20):4072–82.

Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, van de Putte DE Fransen, Anderlid BM, et al. Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. Eur J Med Genet. 2009;52(2–3):77–87.

Bureau A, Labbe A, Croteau J, Merette C. Using disease symptoms to improve detection of linkage under genetic heterogeneity. Genet Epidemiol. 2008;32(5):476–86.

Elston RC, Lange K. The prior probability of autosomal linkage. Ann Hum Genet. 1975;38(3):341–50.

Govil M, Vieland VJ. Practical considerations for dividing data into subsets prior to PPL analysis. Hum Hered. 2008;66:223–37. PMID: 18612207.

Guilmatre A, Dubourg C, Mosca AL, Legallic S, Goldenberg A, Drouin-Garraud V, et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. Arch Gen Psychiatry. 2009;66(9):947–56.

Huang J, Vieland VJ. Comparison of 'model-free' and 'model-based' linkage statistics in the presence of locus heterogeneity: single data set and multiple data set applications. Hum Hered. 2001;51 (4):217–25. PMID: 11287743.

Huang Y, Vieland VJ. Association Statistics under the PPL Framework. Genetic Epidem. 2010;34:835–45.

Huang Y, Segre A, O'Connell J, Wang H, Vieland VJ. KELVIN: a 2nd generation distributed multiprocessor linkage and linkage disequilibrium analysis program. American Society of Human Genetics; 2006; ASHG 56th annual meeting.

Lander ES, Green P. Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA. 1987;84(8): 2363–7.

Le Couteur A, Rutter M, Lord C. Autism diagnostic interview: a semi-structure interview for parents and caregivers of autistic persons. J Autism Dev Disord. 1989;19:363–87.

Le Couteur A, Bailey A, Goode S, Pickles A, Robertson S, Gottesman I, et al. A broader phenotype of autism: the clinical spectrum in twins. J Child Psychol Psychiatry Allied Discipl. 1996;37 (7):785–801.

Liu XQ, Paterson AD, Szatmari P, et al. Genome-wide linkage analyses of quantitative and categorical autism subphenotypes. Biol Psychiatry. 2008;64(7):561–70.

Lotspeich LJ, Kwon H, Schumann CM, Fryer SL, Goodlin-Jones BL, Buonocore MH, et al. Investigation of neuroanatomical differences between autism and Asperger syndrome. Arch Gen Psychiatry. 2004;61(3):291–8.

MacLean JE, Szatmari P, Jones MB, Bryson SE, Mahoney WJ, Bartolucci G, et al. Familial factors influence level of functioning in pervasive developmental disorder. J Am Acad Child Adolesc Psychiatry. 1999;38(6):746–53.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural variation of chromosomes in autism spectrum disorder. J Hum Genet. 2008;82:477–88.

Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, et al. A second-generation combined linkage physical map of the human genome. Genome Res. 2007;17(12):1783–6.

Pagnamenta A, Khan H, Walker S, Gerrelli D, Wing K, Bonaglia MC, et al. Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability. J Med Genet. 2011;48:48–54.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010;466(7304):368–72.

Roeder K, Devlin B, Wasserman L. Improving power in genome-wide association studies: weights tip the scale. Genet Epidemiol. 2007;31(7):741–7.

Royall R. Statistical evidence: a likelihood paradigm. London: Chapman & Hall; 1997.

Silverman JM, Smith CJ, Schmeidler J, Hollander E, Lawlor BA, Fitzgerald M, et al. Symptom domains in autism and related conditions: evidence for familiality. Am J Med Genet. 2002;114 (1):64–73.

Smith CAB. Testing for heterogeneity of recombination fraction values in human genetics. Ann Hum Genet. 1963;27:175–82.

Stevens MC, Fein DA, Dunn M, Allen D, Waterhouse LH, Feinstein C, et al. Subgroups of children with autism by cluster analysis: a longitudinal examination. J Am Acad Child Adolesc Psychiatry. 2000;39(3):346–52.

Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet. 2007;39(3):319–28.

Szatmari P, Merette C, Emond C, Zwaigenbaum L, Jones MB, Maziade M, et al. Decomposing the autism phenotype into familial dimensions. Am J Med Genet B Neuropsychiatr Genet. 2008;147B(1):3–9.

van der Zwaag B, Staal WG, Hochstenbach R, Poot M, Spierenburg HA, de Jonge MV, et al. A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder. Am J Med Genet B Neuropsychiatr Genet. 2010;153B(4):960–6.

Vieland VJ. Bayesian linkage analysis, or: how I learned to stop worrying and love the posterior probability of linkage. Am J Hum Genet. 1998;63(4):947–54. PMID: 9758634.

Vieland VJ. Thermometers: something for statistical geneticists to think about. Hum Hered. 2006;61(3):144–56. PMID: 16770079.

Vieland VJ, Hodge SE. Review of statistical evidence: a likelihood paradigm. Am J Hum Genet. 1998;63:283–9.

Vieland VJ, Wang K, Huang J. Power to detect linkage based on multiple sets of data in the presence of locus heterogeneity: comparative evaluation of model-based linkage methods for affected sib pair data. Hum Hered. 2001;51(4):199–208. PMID: 11287741.

Vieland VJ, Huang Y, Bartlett C, Davies TF, Tomer Y. A multilocus model of the genetic architecture of autoimmune thyroid disorder, with clinical implications. Am J Hum Genet. 2008;82 (6):1349–56. PMID: 18485327.

Vorstman JA, Staal WG, van Daalen E, van Engeland H, Hochstenbach PF, Franke L. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. Mol Psychiatry. 2006;11 (1):18–28.

Welcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661–78. PMID: 17554300.

Wratten NS, Memoli H, Huang Y, Dulencin AM, Matteson PG, Cornacchia MA, et al. Identification of a schizophrenia-associated functional noncoding variant in NOS1AP. Am J Psychiatry. 2009;166(4):434–41.

Xu J, Zwaigenbaum L, Szatmari P, Scherer SW. Molecular cytogenetics of autism. Curr Genomics. 2004;5(4):347–64.

Yang X, Huang J, Logue MW, Vieland VJ. The posterior probability of linkage allowing for linkage disequilibrium and a new estimate of disequilibrium between a trait and a marker. Hum Hered. 2005;59:210–9. PMID: 16015031.

Members of the Autism Genome Project Consortium

Lambertus Klei[7], Richard Anney[8], Daniele Merico[9], Regina Regan[10], Judith Conroy[10], Tiago Magalhaes[11], Catarina Correia[11], Brett S. Abrahams[12], Joana Almeida[13], Elena Bacchelli[14], Gary D. Bader[9], Anthony J. Bailey[15], Gillian Baird[16], Agatino Battaglia[17], Tom Berney[18], Nadia Bolshakova[8], Sven Bölte[19], Patrick F. Bolton[20], Thomas Bourgeron[21], Sean Brennan[8], Jessica Brian[22], Susan E. Bryson[23], Andrew R. Carson[4], Guillermo Casallo[4], Jillian Casey[10], Lynne Cochrane[8], Christina Corsello[24], Emily L. Crawford[5], Andrew Crossett[25], Geraldine Dawson[26,27], Maretha de Jonge[25], Richard Delorme[29], Irene Drmic[22], Eftichia Duketis[19], Frederico Duque[13], Annette Estes[30], Penny Farrar[3], Bridget A. Fernandez[31], Tiziana Filippi[17], Eric Fombonne[32], Christine M. Freitag[19], John Gilbert[33], Christopher Gillberg[34], Joseph T. Glessner[35], Jeremy Goldberg[6], Andrew Green[10], Jonathan Green[36], Stephen J. Guter[37], Hakon Hakonarson[35, 38], Elizabeth A. Heron[8], Matthew Hill[8], Richard Holt[3], Jennifer L. Howe[4], Gillian Hughes[8], Vanessa Hus[24], Roberta Igliozzi[17], Cecilia Kim[35], Sabine M. Klauck[39], Alexander Kolevzon[40], Olena Korvatska[41], Vlad Kustanovich[42], Clara M. Lajonchere[42], Janine A. Lamb[43], Magdalena Laskawiec[15], Marion Leboyer[44], Ann Le Couteur[18], Bennett L. Leventhal[37], Anath C. Lionel[4], Xiao-Qing Liu[4], Catherine Lord[24], Linda Lotspeich[2], Sabata C. Lund[5], Elena Maestrini[14], William Mahoney[45], Carine Mantoulan[46], Christian R. Marshall[4], Helen McConachie[18], Christopher J. McDougle[47], Jane McGrath[8], William M. McMahon[48], Alison Merikangas[8], Ohsuke Migita[4], Nancy J. Minshew[49], Ghazala K. Mirza[3], Jeff Munson[28], Stanley F. Nelson[50], Carolyn Noakes[22], Abdul Noor[51], Gudrun Nygren[34], Guiomar Oliveira[13], Katerina Papanikolaou[52], Jeremy R. Parr[15], Barbara Parrini[17], Tara Paton[4], Andrew Pickles[53], Marion Pilorge[54], Joseph Piven[55], Chris P. Ponting[56], David J Posey[47], Annemarie Poustka[39X], Fritz Poustka[19], Aparna Prasad[4], Jiannis Ragoussis[3], Katy Renshaw[15], Jessica Rickaby[4], Wendy Roberts[22], Kathryn Roeder[25], Bernadette Roge[46], Michael L. Rutter[57], Laura J. Bierut[58], John P. Rice[58], SAGE Consortium, Jeff Salt[37], Katherine Sansom[4], Daisuke Sato[4], Ricardo Segurado[8], Lili Senman[22], Naisha Shah[10], Val C. Sheffield[59], Latha Soorya[40], Inês Sousa[3], Olaf Stein[1], Vera Stoppioni[60], Christina Strawbridge[6], Raffaella Tancredi[17], Katherine Tansey[8], Bhooma Thiruvahindrapduram[4], Ann P. Thompson[6], Susanne Thomson[5], Ana Tryfon[40], John Tsiantis[52], Herman Van Engeland[28], John B. Vincent[51], Fred Volkmar[61], Simon Wallace[15], Kai Wang[35], Zhouzhi Wang[4], Thomas H. Wassink[62], Caleb Webber[56], Kirsty Wing[3], Kerstin Wittemeyer[46], Shawn Wood[7], Jing Wu[25], Brian L. Yaspan[5], Danielle Zurawiecki[40], Lonnie Zwaigenbaum[63], Joseph D. Buxbaum[40], Rita M. Cantor[50], Edwin H. Cook[37], Hilary Coon[48], Michael L. Cuccaro[33], Bernie Devlin[7], Sean Ennis[10], Louise Gallagher[8], Daniel H. Geschwind[12], Michael Gill[8], Jonathan L. Haines[64], Judith Miller[48], John I. Nurnberger Jr.[47], Margaret A. Pericak-Vance[33], Gerard D. Schellenberg[65], Astrid M. Vicente[11], Ellen M. Wijsman[66], Catalina Betancur[54]

[1]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH, 43205, USA. [2]Department of Psychiatry, Division of Child and Adolescent Psychiatry and Child Development, Stanford University School of Medicine, Stanford, CA, 94304, USA. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, OX3 7BN, UK. [4]The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children and

Department of Molecular Genetics, University of Toronto, Ontario, M5G 1 L7, Canada. [5]Department of Molecular Physiology and Biophysics, Vanderbilt Kennedy Center, and Centers for Human Genetics Research and Molecular Neuroscience, Vanderbilt University, Nashville, TN, 37232, USA. [6]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada. [7]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, 19104-6100, PA, USA. [8]Autism Genetics Group, Department of Psychiatry, School of Medicine, Trinity College Dublin 8, Ireland. [9]Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto. [10]School of Medicine and Medical Science University College, Dublin 4, Ireland. [11]Instituto Nacional de Saude Dr Ricardo Jorge and Instituto Gulbenkian de Cîencia Lisbon, Portugal. [12]Department of Neurology, University of California–Los Angeles School of Medicine, Los Angeles, CA, 90095, USA. [13]Hospital Pediatrico de Coimbra, Coimbra, Portugal. [14]Department of Biology, University of Bologna, 40126 Bologna, Italy. [15]Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford, OX3 7JX, UK. [16]Newcomen Centre, Guy's Hospital, London, SE1 9RT, UK. [17]Stella Maris Institute for Child and Adolescent Neuropsychiatry, 56128 Calambrone (Pisa), Italy. [18]Child and Adolescent Mental Health, University of Newcastle, Sir James Spence Institute, Newcastle upon Tyne, NE1 4LP, UK. [19]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, J.W. Goethe University Frankfurt, 60528 Frankfurt, Germany. [20]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London, SE5 8AF, UK. [21]Human Genetics and Cognitive Functions, Institut Pasteur; University Paris Diderot-Paris 7, Fondation FondaMental, 75015 Paris, France. [22]Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehabilitation, University of Toronto, Toronto, Ontario, M5G 1Z8, Canada. [23]Department of Pediatrics and Psychology, Dalhousie University, Halifax, Nova Scotia, B3K 6R8, Canada. [24]Autism and Communicative Disorders Centre, University of Michigan, Ann Arbor, MI, USA. [25]Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, USA. [26]Autism Speaks, USA. [27]Department of Psychiatry, University of North Carolina, Chapel Hill, NC, 27599-3366, USA. [28]Department of Child Psychiatry, University Medical Center, Utrecht, The Netherlands. [29]APHP, Hôpital Robert Debré, Child and Adolescent Psychiatry, 75019 Paris, France. [30]Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, 98195, USA. [31]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St. John's Newfoundland, Canada. [32]Division of Psychiatry, McGill University, Montreal, Quebec, Canada. [33]The John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, 33101, USA. [34]Department of Child and Adolescent Psychiatry, Goteborg University, Goteborg, S41345, Sweden. [35]The Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA. [36]Academic Department of Child Psychiatry, Booth Hall of Children's Hospital, Blackley, Manchester, M9 7AA, UK. [37]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA. [38]Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA. [39]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany. [40]The Seaver Autism Center for Research and Treatment, Department of Psychiatry, Mount Sinai School of Medicine, NY, 10029, USA. [41]Department of Medicine, University of Washington, Seattle, WA, 98195, USA. [42]Autism Genetic Resource Exchange, Autism Speaks, Los Angeles, CA, 90036-4234, USA. [43]Centre for Integrated Genomic Medical Research, University of Manchester, Manchester, M13 9PT, UK. [44]INSERM U995, Department of Psychiatry, Groupe hospitalier Henri Mondor-Albert Chenevier, AP-HP; University Paris 12, Fondation FondaMental, Créteil, France. [45]Department of Pediatrics, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada. [46]Centre d'Eudes et de Recherches en Psychopathologie, University de Toulouse Le Mirail, Toulouse 31200, France. [47]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, 46202, USA. [48]Psychiatry Department, University of Utah Medical School, Salt Lake City, UT, 84108, USA. [49]Departments of Psychiatry and Neurology, University of Pittsburgh School of Medicine, 15213. [50]Department of Human Genetics, University of California - Los Angeles School of Medicine, Los Angeles, CA, 90095, USA. [51]Centre for Addiction and Mental Health, Clarke Institute and Department of Psychiatry, University of Toronto, Toronto, Ontario M5G 1X8, Canada. [52]University Department of Child Psychiatry, Athens University, Medical School, Agia Sophia Children's Hospital, 115 27 Athens, Greece. [53]Department of Medicine, School of Epidemiology and Health Science, University of Manchester, Manchester, M13 9PT, UK. [54]INSERM U952 and CNRS UMR 7224 and UPMC Univ Paris 06, UMR-S 952, Paris 75005, France. [55]Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, NC, 27599-3366, USA. [56]MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom. [57]Social, Genetic and Developmental Psychiatry Centre, Institute Of Psychiatry, London, SE5 8AF, UK. [58]Department of Psychiatry, Washington University in St. Louis, School of Medicine, St. Louis, MO, 63130, USA. [59]Department of Pediatrics and Howard Hughes Medical Institute Carver College of Medicine, University of Iowa, Iowa City, IA, 52242, USA. [60]Neuropsichiatria Infantile, Ospedale Santa Croce,61032 Fano, Italy. [61]Child Study Centre, Yale University, New Haven, CT, 06520, USA. [62]Department of Psychiatry, Carver College of Medicine, Iowa City, IA, 52242, USA. [63]Department of Pediatrics, University of Alberta, Edmonton, Alberta T6G 2J3, Canada. [64]Center for Human Genetics Research, Vanderbilt University Medical Centre, Nashville, TN, 37232, USA.[65]Pathology and Laboratory Medicine, University of Pennsylvania, PA, 19104, USA. [66]Departments of Biostatistics and Medicine, University of Washington, Seattle, WA, 98195, USA.