Glutathione pathway gene variation and risk of autism spectrum disorders

Katherine Bowers · Qing Li · Joseph Bressler ·
Dimitrios Avramopoulos · Craig Newschaffer ·
M. Daniele Fallin

Abstract Despite evidence that autism is highly heritable with estimates of 15 or more genes involved, few studies have directly examined associations of multiple gene interactions. Since inability to effectively combat oxidative stress has been suggested as a mechanism of autism, we examined genetic variation 42 genes (308 single-nucleotide polymorphisms (SNPs)) related to glutathione, the most important antioxidant in the brain, for both marginal association and multi-gene interaction among 318 case–parent trios from The Autism Genetic Resource Exchange. Models of multi-SNP interactions were estimated using the trio Logic Regression method. A three-SNP joint effect was observed for genotype combinations of SNPs in glutaredoxin, glutaredoxin 3 (*GLRX3*), and cystathione gamma lyase (*CTH*); OR=3.78, 95% CI: 2.36, 6.04. Marginal associations were observed for four genes including two involved in the three-way interaction: *CTH*, alcohol dehydrogenase 5, gamma-glutamylcysteine synthetase, catalytic subunit and *GLRX3*. These results suggest that variation in genes involved in counterbalancing oxidative stress may contribute to autism, though replication is necessary.

Keywords Autism · Oxidative stress · Gene–gene interaction

The Autism spectrum disorders (ASD) are a collection of pervasive developmental disorders characterized by delayed or absent language development, lack of interest in other people, stereotyped or repetitive behaviors, and in some cases regression of early speech and sociability (Rapin and Autism 1997) . While several lines of evidence suggest a genetic etiology (Veenstra-VanderWeele and Cook 2004), only a few genes have been established as risk factors for ASDs (Bill and Geschwind 2009). Although interactive models involving multiple genes are commonly proposed as likely models to describe autism risk (Risch et al. 1999; Lamb et al. 2000), few studies have empirically explored gene–gene interactions in ASDs to date (Anderson et al. 2008, 2009; Campbell et al. 2008; Coutinho et al. 2007;

K. Bowers · M. D. Fallin
Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health,
615 N. Wolfe St. room W6509,
Baltimore, MD 21205, USA

Q. Li · M. D. Fallin (✉)
Department of Biostatistics,
Johns Hopkins Bloomberg School of Public Health,
615 N. Wolfe St,
Baltimore, MD 21205, USA
e-mail: dfallin@jhsph.edu

J. Bressler
Department of Environmental Health Sciences,
Johns Hopkins Bloomberg School of Public Health,
615 N. Wolfe St,
Baltimore, MD 21205, USA

J. Bressler
Kennedy-Krieger Institute,
707 North Broadway,
Baltimore, MD 21205, USA

Ashley-Koch et al. 2007; Ma et al. 2005; Kim et al. 2008). While genome-wide linkage and association studies have provided initial discoveries of potential genetic risk factors and have implicated particular molecular pathways in autism, most of these studies have neglected gene–gene interaction, primarily due to methodologic limitations, as examination of interaction effects on the genomic scale is computationally difficult. In addition, genome-wide markers may not fully cover implicated regions. Candidate genes studies are therefore an appropriate design to evaluate gene–gene interaction. One approach is to focus on genes known to code for proteins that play a role in a particular biological pathway implicated in ASDs, as these are likely to interact on risk. Of the handful of candidate gene studies that have undertaken gene–gene interaction in ASDs, none have explored high-order interaction in a single pathway, with adequate gene coverage. Previous studies did not have the advantage of very high density marker information on each gene, and were limited to only two or three-way interaction based on traditional regression methods (Campbell et al. 2008; Kim et al. 2008). The few family-based three-way interaction approaches that have been reported for ASD could not take full advantage of the family data because they were limited to multiple dimensionality reduction (MDR) methods developed for case–control designs (Anderson et al. 2008, 2009; Coutinho et al. 2007; Ashley-Koch et al. 2007; Ma et al. 2005), although newer family-based MDR methods are now available (Ritchie et al. 2001). We attempted to overcome these limitations by focusing on a single pathway, using an information-based gene-selection algorithm (based on location in autism linkage regions and expression patterns), selecting tag-single-nucleotide polymorphisms (SNPs) to achieve very high coverage of genes and flanking regions, and by evaluating high order gene–gene interactions using a newly developed method adapted for family data. Logic regression is an adaptive regression methodology that uses and/or combinations of SNPs to predict disease risk.

D. Avramopoulos
McKusick Nathans Institute of Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA

D. Avramopoulos
Department of Psychiatry, School of Medicine, Johns Hopkins University, Broadway Research Building Room 509, 733 N. Broadway, Baltimore, MD 21205, USA

C. Newschaffer
Department of Epidemiology and Biostatistics, Drexel University School of Public Health, 1505 Race Street Mailstop 1033, Philadelphia, PA 19102, USA

Oxidative stress during prenatal and early postnatal development, resulting from polymorphisms in key antioxidant genes in the fetus/infant is hypothesized to contribute to the susceptibility of autism (James et al. 2006). Oxidative stress is believed to contribute to neuronal damage for a wide range of neurological diseases including Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, and HIV-associated dementia (Perry et al. 2004) with a suggested contribution to developmental disorders, such as fetal alcohol syndrome (Cohen-Kerem and Koren 2003). Further, susceptibility to oxidative stress is not limited to the CNS. Deficits in this system would help explain the full range of medical symptoms observed in autism, such as gastrointestinal (Valicenti-McDermott et al. 2006) and immune system (Pardo et al. 2005) irregularities.

Decreased total glutathione and a decreased ratio of reduced (active) to oxidized (inactive) glutathione has been observed in plasma of autistic children (James et al., 2004, 2006). Plasma (Sogut et al. 2003; Yorbik et al. 2002) and erythrocyte (Yorbik et al. 2002) glutathione peroxidase were also shown to differ between autistic children and controls; however, this was increased in one study (Sogut et al. 2003) and decreased in the other (Yorbik et al. 2002). One genetic association study has shown an association between autism risk and the null allele of glutathione-S-transferase Mu 1 implying reduced antioxidant capacity due to alterations in glutathione associated genes (James et al. 2006; Buyske et al. 2006). Additionally, a haplotype of another glutathione gene, glutathione-S-transferase Pi 1, was observed to be overtransmitted in mothers of autistic children (Williams et al. 2007) .

Glutathione and its metabolic cofactors provide the primary defense against oxidative stress (Maher 2006) including: directly scavenging free radicals, reducing peroxides and conjugations with toxic electrophilic compounds (Maher 2006). Therefore, alternations in enzymes related to the metabolism of glutathione or any of these three processes may increase the potential for oxidative damage that could contribute to the development of autism.

Our objective is to more fully characterize the contribution of a child's genetic variation in glutathione pathway genes to autism risk and to specifically explore potential interactions between genes in this pathway. We genotyped SNPs in 42 glutathione-related genes to determine their marginal and potential interacting effects on autism risk in a collection of 318 families from the Autism Genetic Resource Exchange (AGRE) repository. We examined 308 SNPs in these 42 genes for marginal effects and an uncorrelated set of 138 of these SNPs from these genes for potential interaction. To our knowledge, this analyses the most extensive search for gene–gene interaction along a suspected etiologic pathway.

Methods

Study population

The AGRE was created by Cure Autism Now and the Human Biological Data Exchange to advance genetic research in autism spectrum disorders by consolidating large numbers of families into one collection (www.agre. org; Geschwind et al. 2001). Genetic biomaterials and clinical data were obtained from families with a child diagnosed with an ASD based on evaluation by the Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994 Oct). The 1,149 individuals from 318 families included in the present analyses are a subset of the initial 331 families recruited by AGRE. For a subset of participants, an additional diagnosis is available based on the Autism Diagnostic Observational Schedule (ADOS) (Lord et al. 2000). Cases in the studies presented here met criteria for Autistic Disorder based on the ADI-R and confirmed by ADOS if available. Participant characteristics are described in Table 1.

Gene selection

Gene selection began with an initial list of 64 genes identified via searches of the University of California at Santa Cruz genome database (http://genome.ucsc.edu/) and the GO database (geneontology.org), using the keyword 'glutathione'. Each gene was further entered in the Entrez database to collect additional information, such as function, and to review previous studies of the gene (http://www. ncbi.nlm.nih.gov/). Of these 64 genes, 59 had sufficient data available and were then ranked using a weighted sum of non-standardized values from three domains: (a) location in an autism linkage region; (b) expressed in the brain; and (c) having a pattern of expression in the brain that is correlated with expression of other glutathione genes. A final list of the 59 genes is available in Supplemental Table 1.

Genes located in the same chromosomal band as an autism linkage signal, based on a 2007 review of genetic studies of autism (Freitag 2007) were given a linkage domain weight of 10 for our gene-ranking algorithm, while those that were not in a linkage region were given a weight of 0. Gene expression weights were based on expression scores for 17 brain regions from two replicates using Affymetrix microarrays in the GNF2 Atlas track of the University of California, Santa Cruz (UCSC) genome browser. For each gene, if the gene's standardized, relative expression score was >0 for a particular brain region, it was given a brain expression weight of 1, 0 otherwise, for that region. Therefore, across 17 brain regions, expression weights for this domain ranged from 0 to 17 per gene.

Table 1 Participant characteristics of 457 Autistic Children form Autism Genetic Resource Exchange (n=318 families)

	Count (%)
Sibships	321 (mean 1.67)
Male	352 (77.62)
Race	
White	419 (91.68)
Black	1 (0.22)
Unknown	37 (8.10)
Ethnicity	
Non-Hispanic	380 (83.15)
Hispanic	40 (8.75)
Unknown	37 (8.10)
Family type	
Multiplex	446 (98.02)
Simplex	9 (1.98)
Zygosity	
DZ Twin	33 (7.22)
Triplet	6 (1.31)

These first two criteria facilitated the selection of genes that are plausibly associated with ASD. Because we were focused on genes that may interact together biologically, we included an additional domain favoring genes that have similar co-expression in the brain. For this, we used the same UCSC brain expression data to calculate a quantitative co-expression weight for each gene determined by the number of other glutathione genes for which each gene's brain expression is correlated greater than 0.4, resulting in co-expression weights between 0 and 63. The correlation was calculated using the correlation function in excel 2007.

We then ranked the 64 genes based on this non-standardized weighted sum. The highest scoring gene, GSTM2, had a score of 42 and the lowest scoring gene, GSR, had a score of 2. Starting with the highest scoring gene, we included as many genes as possible using tagSNPs (described below) to fill a 384-SNP Illumina Oligo Assay Pool (www.illumina.com). A total of 42 genes were included.

SNP selection and genotyping

We selected SNPs within each gene based on information available in The International HapMap Project (HapMap; www.hapmap.org) for the Centre de'Etude du Polymorphism Humain (CEPH) population using build 35, release 21a (Thorisson et al. 2005). We implemented the program Tagger (http://www.broad.mit.edu/mpg/tagger/) to select tagSNPs to cover all known CEPH HapMap SNPs in that gene and 10 kb of flanking sequence with an $r^2 > 0.8$ (de Bakker et al. 2005). In addition, we included potentially functional SNPs as well as SNPs in conserved regions,

based on synonymous and non-synonymous SNPs in HapMap and conservation information in UCSC. A conservation score >0.5 for any position within five base pairs of a particular SNP, using the Mammal Cons (phastCons44wayPlacental) track, was considered conserved. We achieved at least 85% coverage of HapMap CEPH SNPs (minor allele frequency (MAF)>1%) with an $r^2 > 0.8$ for the majority of genes, with nearly half covered at 100%. Because tagSNP selection was based on CEPH samples, there may be reduced coverage of these candidate genes among non-CEPH families in our sample, such as those who report Hispanic ancestry.

Blood was taken from each participant by researchers in the AGRE program and then sent to the Rutgers University Cell and DNA Repository in New Jersey for cell line creation, DNA isolation and storage. DNA samples for 1,518 participants in 20 μg aliquots at 0.1 μg/μl concentration were shipped to The Johns Hopkins Bloomberg School of Public Health (JHSPH). Sample plating was carried out at JHSPH and samples shipping for genotyping at the JHU SNP Center using Illumina GoldenGate Genotyping Assay on a Beadlab system (www.illumina.com).

Data cleaning

SNPs and DNA samples were examined for MAF, Hardy–Weinberg Equilibrium (HWE), Mendelian inconsistencies (MI), and missing data. MAF were calculated using SAGE (S.A.G.E. [2008] Release 5.4.2: http://darwin.cwru.edu/) and SNPs were dropped if the MAF was <1%. HWE was estimated using Stata (StataCorp. 2001. Release 10.0) and SNPs were dropped if the p value <0.001. MI was estimated using SAGE. SNPs were dropped if greater than 5% of the individuals genotyped for that SNP contained errors; individuals were dropped if greater than 5% of their SNPs contained errors. Inconsistencies within families were verified by estimating pairwise identity-by-decent (IBD) using Plink (http://pngu.mgh.harvard.edu/purcell/plink/) and comparing the estimated values to expected IBD sharing for a given relative. Genotypes were defined as missing if the Illumina Gene Call score was <0.6. Individuals and SNPs were dropped if they were missing greater than 5% of their data. There were 308 SNPs remaining after data cleaning.

Statistical analysis

Interaction models

Models of possible joint effects of SNPs from genes related to glutathione were built and evaluated using the logic regression method adapted for trio data (Li et al. 2009a, b), denoted as "trioLR". The objective of trioLR is to search

Boolean (and/or/not) combinations of covariates (in this case, SNPs) for the best model to predict transmitted genotype status, compared to other possible child genotypes, given parent genotypes. We assume two risk groups in the population defined by some genotype pattern G and assume that the probability of disease p is given via logistic regression as: $\ln\left(\frac{p}{1-p}\right) = \alpha + \beta I_G$ where I_G indicates whether a particular Boolean statement of interaction is met, and α and β are effect parameters to be estimated. The genotype interaction patterns G are based on Boolean combinations of SNPs in dominant and recessive coding, such as (SNP_{R1} ^SNP_{D15}), indicating that subjects with the combination of two variant alleles at SNP 1 (recessive coding) and at least one variant allele at SNP 15 (dominant coding) define the high-risk group. This combination can also be displayed as a "tree" defining the particular interaction model.

Given our trio data, trioLR is embedded in a conditional logistic regression framework, conditioning on a case–parent trio with one observed case genotype and three matched pseudo-control genotypes. It uses the deviance function to evaluate the fit of each model. A search algorithm evaluates all possible trees (interaction models) of a specified size (number of interacting SNPs) to identify specific combinations that minimize the deviance. We evaluated models containing between one and six SNPs, and report the best models for trees of size 1–4, as higher models did not add predictive information in our data set.

The conditional logistic model for a single SNP compares one case genotype to three possible pseudo-controls based on parent genotypes. However, when more than one SNP is included in a logic tree (as in the two-SNP example above), >3 psuedo-controls are possible if the SNPs are uncorrelated. trioLR handles this by randomly selecting three multi-SNP combinations consistent with Mendelian possibilities from among the full set of multi-SNP combinations possible given the parent genotypes. Since pseudo-controls are randomly selected, different data sets can be constructed for the same observed data resulting in different trioLR results. To address this issue, we constructed 10 separate data sets, with unique pseudo-control draws per Boolean combination, and ran each in trioLR. The best logic tree models for each of the 10 data sets were recorded and the most common logic tree result among the 10 data sets was used for final analysis presentations and permutation testing described below.

To determine which model size is most appropriate for our data, we employed permutation testing. For trioLR, the response variable, or residual conditioned on a smaller trioLR model, is permuted within each matched set after conditioning on a particular fitted logic model. Beginning with a null model (no predictors), we permuted response (case genotype versus pseudo-controls) 500 times, per-

formed trioLR, and plotted the final deviance scores for each of the iterations in a histogram (top row of Fig. 1). Lower scores suggest better performance and it is expected that larger tree models will perform better; therefore the majority of the permuted data set scores, which fit non-null models, should fall to the left of the null model. We next fitted the best one-SNP model from our data, then permuted residuals from this model to create 500 data sets, conditional on a one-leaf model. trioLR was carried out for each of these data sets and the deviance scores plotted. Again, all permuted data set results, allowing higher-order models, performed better than the one-leaf model. We continued this by conditioning on the best two-leaf model, three-leaf model, etc. up to six leaves. As the histogram

ceases to drift left, less residual risk can be explained by larger models and the fitted model size is selected as the last model size for which an improvement can be seen. Since logic regression does not involve hypothesis testing, formal power calculations are not relevant and were not computed. However, the recent methods paper by Li et al. (2009a) discusses the "hit rate" for trio logic models under simulations of two- and three-way interaction and shows that sample sizes in the range we present here do have high probabilities of detecting at two-way interactions and for detecting at least one of the causal SNPs in three-way interactions.

An important aspect of trioLR, similar to case-only studies of interaction, is that SNPs included in model searches be uncorrelated in the general population. There-

Fig. 1 Results of the logic regression permutation tests. Histograms of score results per permutation given model size (leaf) 0 through 4. Scores for the null model through 4 leaf model are 0.347, 0.343, 0.34, 0.337 and 0.334, respectively
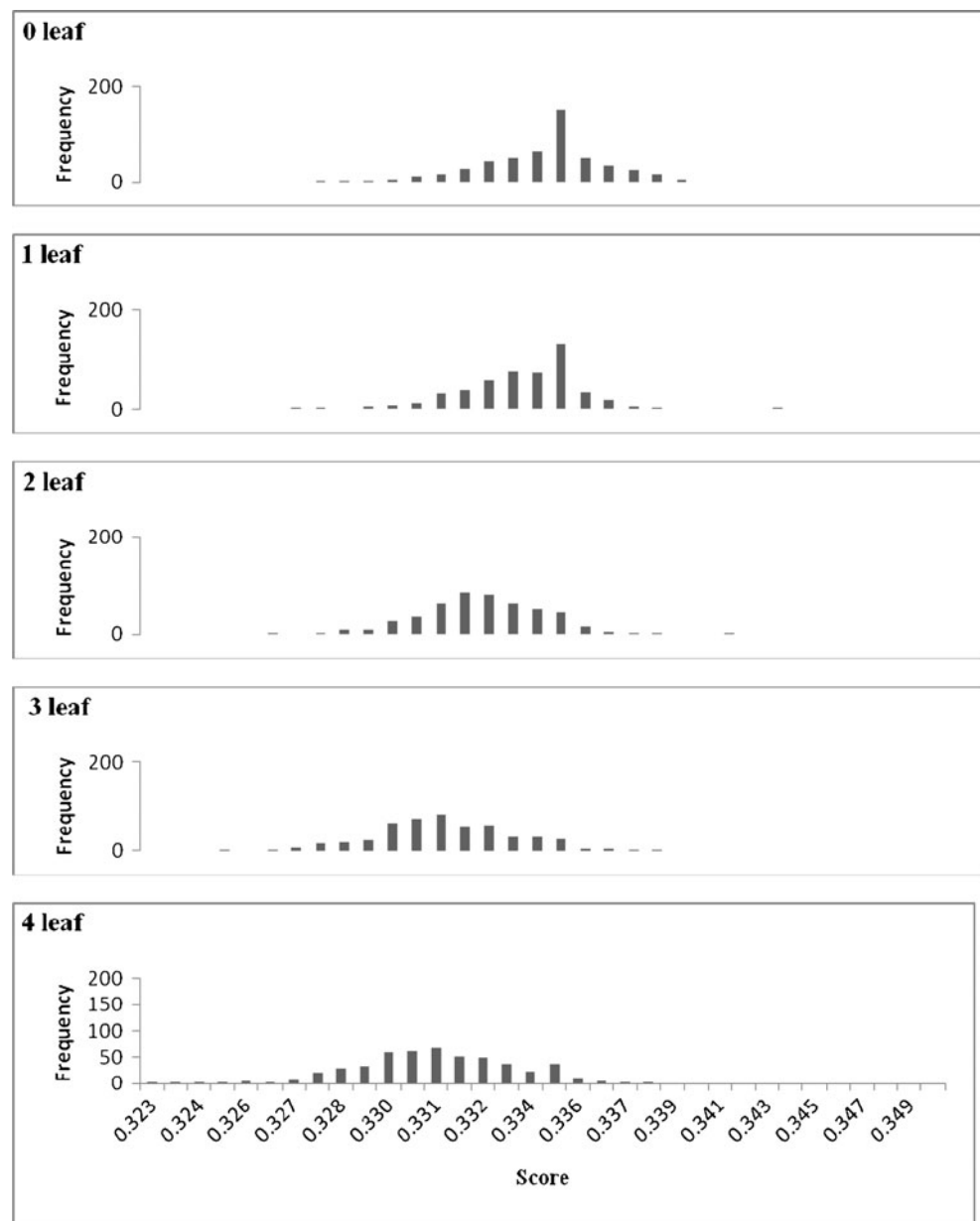
fore, we trimmed our set of 308 SNPs to only 138 SNPs with an $r^2 \leq 0.2$ between them for LR analyses. To ensure that all glutathione-related genes are represented in the interaction models, and that marginal signals are investigated for potential interactions, the most significant SNP from each gene was force included as well as any SNPs with a p value <0.10. LR analyses were conducted in R (R Development Core Team. R: A Language and Environment for Statistical Computing 2008).

Single SNP analyses

Odds ratios and 95% confidence intervals were estimated using conditional logistic regression with a robust variance estimator to account for multiple affected children per family. Matched sets consisted of an affected child and three pseudo-controls based on the other Mendelian-possible child genotypes given the parent genotypes. P values were generated using likelihood ratio tests of conditional logistic models with and without terms for genotype effects, with heterozygotes and minor allele homozygotes modeled separately compared to major allele homozygotes as the reference. These analyses were conducted in Stata (StataCorp. 2001. Release 10.0).

Additional AGRE families

The present study evaluated SNPs in 318 families, reflecting the earliest recruitment of AGRE samples. AGRE now has a collection of over 1,000 families available, and two genome-wide association studies (GWAS) have been conducted on the majority of these families including an Affymetrix 5.0 (http://www.affymetrix.com) panel of 500,000 SNPs on 777 families (Weiss et al. 2008) and an Illumina Hap550 (www.Illumina.com) panel of 550,000 SNPs on 943 families (Wang et al. 2009). For each of the associated SNPs in our analysis, we

attempted to identify exact SNP matches or SNPs that were correlated with an $r^2>0.9$, as evaluated using HapMap, in either GWAS panel to allow additional families to be analyzed as a replication of our initial findings. This was only possible for six of the none SNPs with $p<0.05$. For these, we employed identical statistical methods to estimate conditional logistic regression odds ratios in the subset of AGRE families not previously included in our analyses. SNPs from the two GWAS data sets were analyzed separately.

Results

Summary of quality assessment

A total of 1,158 individuals from 318 families were genotyped and 336 SNPs were released by SNP Center Quality Control criteria (genotype quality score >0.25). Of these, eight had atypical clustering and were not included in the analyses, 13 were dropped due to missing data, five were dropped due Mendelian inconsistencies, and two were dropped due to a low MAF. After these filters, no additional SNPs were out of HWE among parents. Nine individuals were dropped as a result of Mendelian inconsistencies that could not be resolved. The data cleaning resulted in 308 SNPs and 1,149 individuals from 318 families with a mean sibship size of 1.67. Participant characteristics are listed in Table 1.

Interaction models

Table 2 displays trioLR results for the best models of size 1–4 interacting SNPs. We have included odds ratios generated through trioLR as well as in separate conditional logistic regression models with 95% confidence intervals, estimated via STATA. The one-SNP model suggests carrying either one or two common alleles at cystathione

Table 2 Logic trees, odds ratios, and 95% confidence intervals for trees including one through four SNPs

Logic tree model size	Logic tree model	Logic regression OR	Conditional logistic regression OR (95% CI)
1	*not*CTH -r (rs12737233) [i.e., *0 or 1 copies of minor allele]*	0.19	0.19 (0.08, 0.49)
	Complement interpretation: CTH-r	–	5.14 (2.02, 13.06)
2	*not*CTH -r (rs12737233) AND *not*GLRX3-r (rs3750834)	0.39	0.39 (0.26, 0.58)
	Complement interpretation: [CTH-r OR GLRX3-r]	–	2.58 (1.72, 3.87)
3	[CTH-r (rs12737233) OR GLRX3-r (rs3750834)] AND GLRX⁻d (rs11135434)	3.78	3.78 (2.36, 6.04)
4	{[*not*CTH-r (rs12737233) AND *not*GLRX3-r (rs3750834)] OR GSTA2-d (rs2608615)} OR GLRX-d (rs11135434)	0.17	0.17 (0.10, 0.30)
	Complement interpretation: {[CTH-r (rs12737233) OR GLRX3-r (rs3750834)] AND notGSTA2-d (rs2608615)}AND not GLRX-d (rs11135434)	–	5.79 (3.31, 10.15)

gamma lyase (CTH) SNP rs12737233 is protective for autism (OR=0.19, 95% CI: 0.08–0.49), consistent with our single-SNP marginal effect results. This is referred to as "notCTH-r" for the complement of the recessive (−r) coding. One could also interpret this as a strong risk effect for carrying two copies of the minor allele at this SNP (OR=5.14, 95% CI: 2.02, 13.06). The two-SNP model suggests carrying either one or two common alleles of this same CTH SNP and carrying one or two common alleles the GLRX3 SNP rs3750834 also reduces risk of autism. This may be more easily understood via the alternative interpretation that being homozygote for the minor allele at either SNP (CTH-r or GLRX3-r) confers increased risk (OR=2.58, 95% CI: 1.72, 3.87). The three-SNP model reflects the same combination of these two SNPs with the additional inclusion of carriers of the minor allele at rs17216887 in GLRX (OR=3.78, 95% CI: 2.36–6.04). This result is consistent with the one- and two-leaf models. The four-leaf model includes a fourth gene, GSTA2, and results in a combined risk set with an OR=5.79.

Based on the permutation tests, the optimal model size for these data appears to be three SNPs (see Fig. 1). Figure 2 shows the risk and protective group memberships for this LR model graphically.
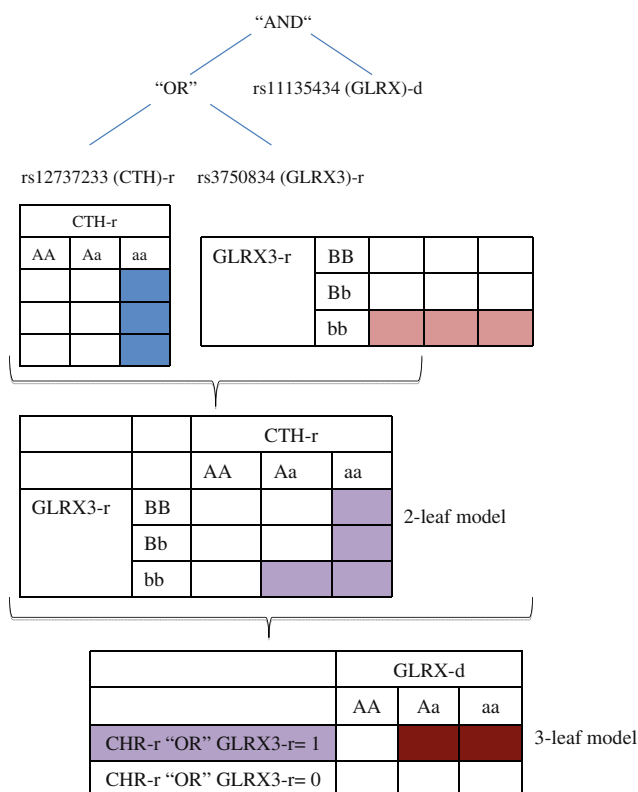
Fig. 2 Final three-leaf model (rs12737233 (CTH)-r "OR" rs3750834 (GLRX3)-r). Two-leaf model shown in *purple* and final, three-leaf model shown in *red*

Single-SNP results

Results of the single SNP analyses are presented in Fig. 3. The majority of SNPs with a p value <0.05 are located in four genes including CTH, alcohol dehydrogenase 5 (ADH5), gamma-glutamylcysteine synthetase, catalytic subunit (GCLC), and GLRX3. Figures 4 shows each of the four genes enlarged and superimposed with the linkage disequilibrium structure (as indicated by the recombination hot spots) of that gene. The strongest association was observed for SNP rs12737233 in CTH, with a heterozygote (OR=0.91, 95% CI: 0.65, 1.28) and a homozygote (OR=4.83, 95% CI: 1.85, 12.59). Two other SNPs in this gene showed associations for homozygote risk as well (Fig. 4a). Using the program String (http://string-db.org/), we created Fig. 5 to present the known interactions between all 42 glutathione-related genes included in the present study. Associations identified in the analyses are denoted with a star for the marginally significant single SNP results and an 'I' for the members of our three-SNP model.

Replication in independent AGRE samples

Of the nine SNPs with p values <0.05 in our primary single-SNP analyses, six were represented either directly or through an LD proxy on one of the GWAS panels available on additional AGRE samples (Table 3). Two SNPs, rs524553 and rs761141, both located in GCLC, approached nominal statistical significance in independent AGRE samples (Table 3), with similar odds ratio profiles.

Discussion

We evaluated gene–gene interactions between SNPs located in 42 genes encoding enzymes involved in glutathione (GSH) metabolism or that utilize GSH as a co-factor in enzymatic reactions. The most compelling evidence from interaction models point to the genes CTH and GLRX. Our three-SNP model showed that having both minor allele variants of either rs12737233 (CTH) or rs3750834 (GLRX3) combined with at least one variant allele of rs11135434 in GLRX increases the risk for autism almost fourfold. Additional stepwise conditional logistic regression models were run to evaluate whether this three SNP result is primarily due to the marginal effects of a single SNP. While we could not fit a fully saturated model due to lack of observations in some three-way interaction groups, stepwise logistic regression models did show that no main effect or two-way interaction, adjusting for all three SNPs, was as strong as this three-way combination.

The permutation results suggest that the three-SNP model is the best fit for our data, though several SNPs may be good candidates for future studies. For example, we

Fig. 3 Negative log *p* value for 308 SNPs located in glutathione-related genes in 318 families from Autism Genetic Resource Exchange. Genes are in positional order, delineated by *color*
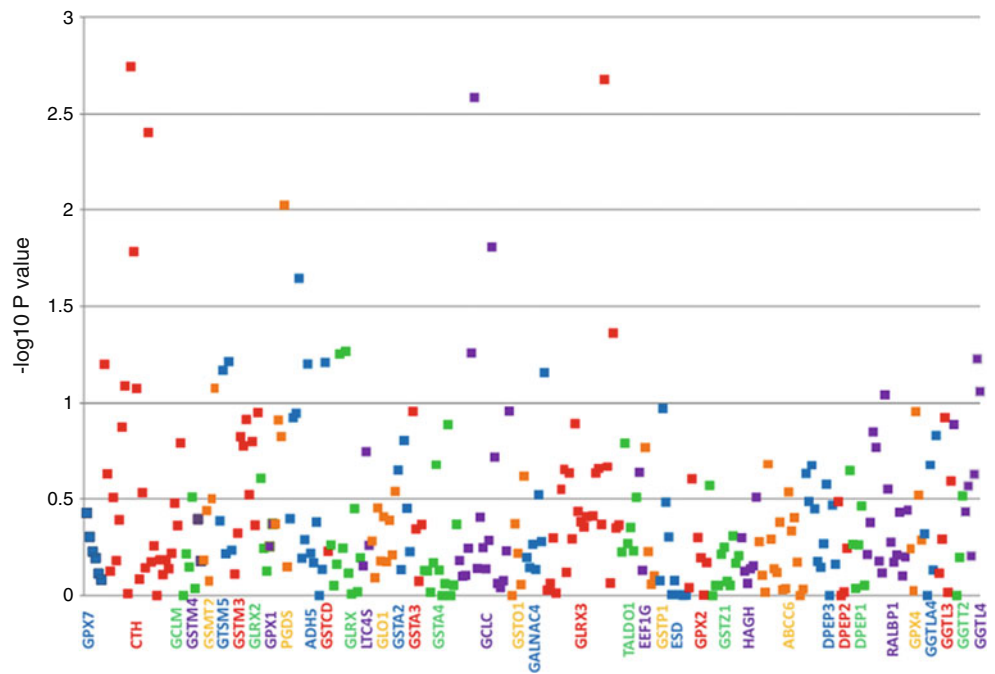
ran 10 iterations of pseudo-control data sets and while the two-SNP model with *CTH* and *GLRX3* had the highest posterior frequency (50% of the 10 replicates), a model with the recessive rs10439143 of *GPX4* or the recessive rs11017128 (*GLRX3*) was the best two-SNP model in four of the 10 iterations, or 40% posterior frequency (not shown). Additionally, rs3743853 of hydroxyacyl glutathione hydrolase, also known as *GLO2*, was in several three-
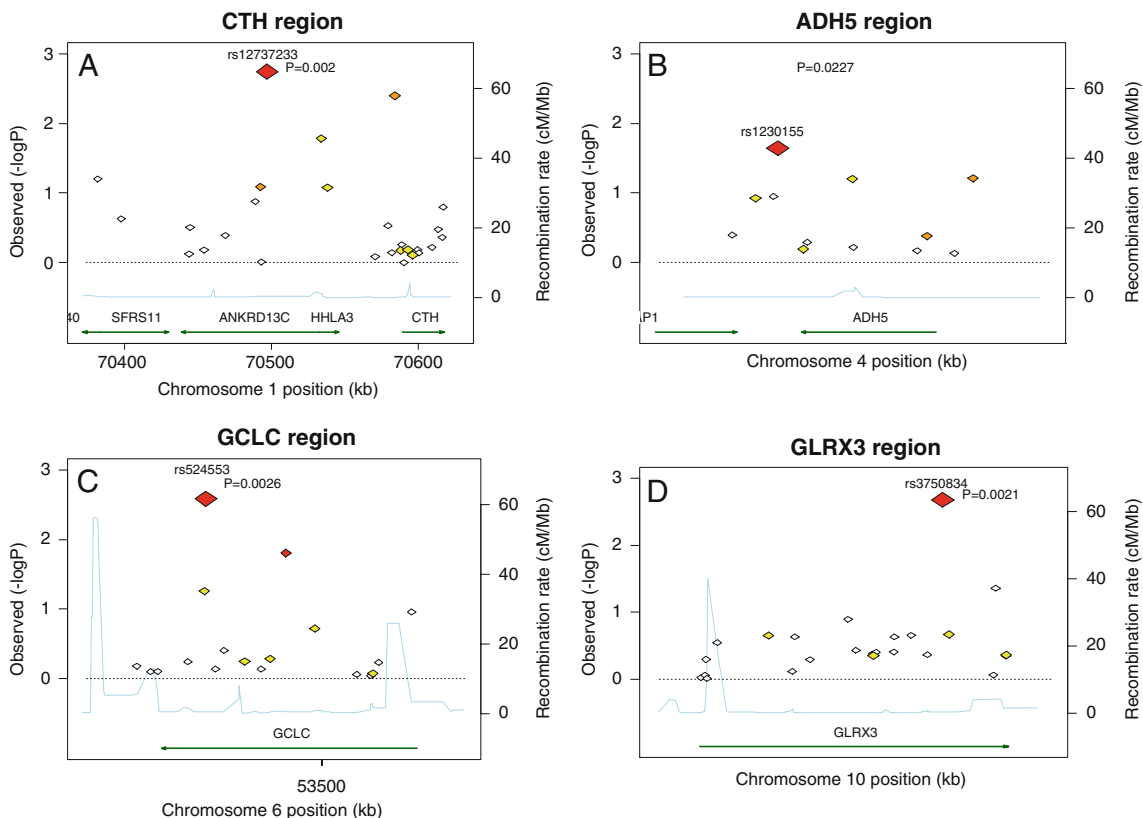
Fig. 4 Negative log (*p* values), linkage disequilibrium structure, and recombination rates for SNPs located in **a** CTH, **b** ADH5, **c** GCLC, **d** GLRX3

Table 3 Replication results for 6 SNPs with significant marginal effects

SNP (gene)	Primary analysis (n=1,149)			Independent AGRE families			Combined
	Genotype	OR (95% CI)	P value	OR (95% CI)	Sample size (n)	P value	P value[b]
rs12737233 (CTH)	C/TT/T	0.91 (0.65, 1.28) 4.83 (1.85, 12.59)	0.002	1.02 (0.81,1.29) 0.62 (0.31, 1.23)	3,327	0.30	0.84
rs7658602 (PGDS)	G/TT/T	0.27 (0.10, 0.78) 0.20 (0.07, 0.60)	0.01	0.71 (0.20, 2.49) 0.77 (0.22, 2.76)	3,327	0.73	0.06
rs761141 (GCLC)	C/TT/T	2.10 (1.20, 3.69) 1.67 (0.91, 3.09)	0.02	1.10 (0.70, 1.71) 1.34 (0.85, 2.13)	3,327	0.10	0.11
rs524553 (GCLC)[a]	C/TT/T	2.70 (1.47, 4.98) 2.23 (1.16, 4.28)	0.003	1.22(0.81, 1.83) 1.46 (0.94, 2.28)	3,327	0.08	n/a[c]
rs1230155 (ADH5)	A/GG/G	0.70 (0.54, 0.89) 0.69 (0.44, 1.07)	0.02	0.92 (0.76, 1.11) 0.74 (0.52, 1.03)	3,327	0.20	0.08
rs1017128 (GLRX3)	A/GG/G	0.69 (0.51, 0.93) 0.78 (0.36, 1.70)	0.04	0.90 (0.74, 1.10) 0.91 (0.49, 1.69)	2,858	0.59	0.09

[a] Replication of SNP rs524553 assessed through a proxy (rs2066511) with r^2 =0.9

[b] Identical methods were used to calculate p value with combined data set including the primary analysis and independent families.

[c] Did not combine p values since replication analysis used proxy SNP

ously associated with oxidative stress. Furthermore, comparisons of AD versus control brain samples revealed a reduction in *GLRX1* expression levels (Lillig et al. 2008). Similarly, a protective role of *GLRXs* in PD has also been observed (Lillig et al. 2008). Glutaredoxin 1 is also induced during pregnancy and expression levels have been associated with preeclampsia and growth restriction (Lillig et al. 2008). The role of GLRX3 has been primarily identified for its role in the life cycle of viruses and may have a role in cell activation associated signaling pathways or in response to stress signals (Lillig et al. 2008).

One advantage of logic regression is the ability to detect gene–gene interactions even in the absence of marginal effects, therefore enhancing gene discovery. While several of the SNPs in our final interaction models did show significant marginal effects (suggestive marginal associations ($p < 0.05$) were observed between several SNPs located in four genes including: *CTH*, *ADH5*, *GCLC*, and *GLRX3*), others did not including: rs11135434 (*GLRX*), with a marginal effect p value=0.89, rs2608615 (*GSTA2*), p value=0.74, as well as the suggestive SNP rs10439143 (*GPX4*) which had a marginal effect p value=0.11. A second advantage is an ability to discover interactions that may not have been found using traditional interaction models in logistic regression. For example, our two-SNP model revealed an interaction between SNPs in *CTH* and *GLRX3*. However, using traditional stepwise conditional logistic regression methods, we would not have been able to model this as a multiplicative interaction since there are no individuals who are recessive for both the *CTH* and *GLRX3* SNPs simultaneously.

The potential implications of variation in the remaining genes with borderline marginal significance, including *GCLC* and *ADH5*, are worth mentioning. As with CTH, GCLC is vital in the production of glutathione. In addition to the significant association in the primary data set, two SNPs of *GCLC*, rs524553 and rs761141, approached nominal statistical significance in independent AGRE samples, with similar odds ratio profiles. There are two enzymes responsible for the synthesis of glutathione and include gamma-glutamylcysteine lygase (GCL; formally called synthetase) and glutathione synthase (Townsend et al. 2003). GCL catalyzes the first and rate-limiting step in GSH synthesis. Gamma-glutamyl cysteine ligase is composed of two subunits, a regulatory (GCLM) and catalytic subunit (GCLC; Maher 2006). Gamma-glutamyl cysteine ligase catalyzes the integration of L-glutamate and cysteine forming gamma-glutamyl-cysteine. This is the final product prior to the addition of glycine to form the full tripeptide of glutathione. In mice, a GCLC deficiency leads to embryonic death (Maher 2006). In studies of rat neurons, the absence of either the regulatory or catalytic subunit of GCL led to cell death in the presence of nitric oxide or glutamate (two neurotransmitters), suggesting particular importance in

neuronal cells (Maher 2006). Small changes in GCLC resulting from polymorphisms would be expected to have more subtle effects.

Despite only one nominally significant SNP, *ADH5* contained two other SNPs with borderline significance including one SNP (rs1154415) with a significant odds ratio (OR=1.54, 95% CI: 1.06, 2.24) comparing the homozygous GG genotype to the homozygous AA (Figs. 3 and 4). Alcohol dehydrogenase 5 oxidizes glutathione conjugates, most importantly formaldehyde–glutathione (Koivusalo et al. 1989) thereby eliminating the formaldehyde and allowing GSH to be available for further reduction reactions. ADH5 protects from nitrositive stress and is specific for nitro-soglutathione (Liu et al. 2001). In addition, *ADH5* and other alcohol dehydrogenases have been studied in association with alcohol metabolism and dependence (Edenberg 2007). Oxidative stress has been implicated as the mechanism by which alcohol causes detrimental affects to the fetus (Henderson et al. 1999) and altered alcohol metabolism due to variations in alcohol metabolizing genes, could increase oxidative stress. However, it should be noted that the ability of ADH5 to oxidize ethanol is limited.

Each of these results should be interpreted with caution, as we did not adjust for multiple comparisons. Adjusting for multiple testing is an ongoing debate in genetics research. Traditional adjustments include those that control the family wise error rate (FWER), such as Bonferroni and Sidak. However, for our analyses, these adjustments are overly conservative given the correlation between SNPs (Fallin and Matteini 2009) and the assumption of a global null hypothesis. While methods exist to take the correlation into consideration, for example estimating and adjusting for the number of independent tests rather than the overall number of tests, even these aim to protect the FWER, which controls for the probability of at least 1 significant result assuming all tests are truly null (Fallin and Matteini 2009 Jan). However, given the a priori gene selection and prioritizing with consideration of multiply associated genes, as well as SNP selection focusing on coverage as well as potentially functional SNPs, this global null hypothesis may not be appropriate. Therefore, we have displayed unadjusted *p* values and encourage both replication and biological plausibility as validation techniques for these initial findings. The interaction modeling should be considered as hypothesis generating in its origin, and is thus not directly amenable to the multiple testing frameworks. In the trioLR analyses, we relied on permutation testing to gage whether signals occurred by chance, and again encourage both replication and biological plausibility as additional milestones for validation.

It is important to note that we did not observe associations with glutathione genes previously implicated in ASDs (James et al. 2006; Buyske et al. 2006; Williams et al. 2007; Ming et

al. 2010). We believe our pathway approach has advantages over looking gene-by-gene and may not have replicated these particular findings due to differences in samples. Further, we did not replicate some of our single-gene findings in the independent set of AGRE families. This is however difficult to interpret since many of our SNPs/genes were not well represented on the GWAS panels available. Nevertheless, our interaction approach should be considered exploratory and provoke more attention to glutaredoxins via replication and functional analyses.

In conclusion, our results confirm that variation in glutathione pathway genes, perhaps through multiple gene effects, may contribute to autism risk. Using novel methods, we were able to detect potential gene–gene interactions, in some cases in the absence of marginal effects. This demonstrates the importance of evaluating gene–gene interaction in future genetic studies of autism and further characterizing the role of oxidative stress in the etiology of autism.

References

Rapin I, Autism N. Engl J Med. 1997;337(2):97–104.

Veenstra-VanderWeele J, Cook Jr EH. Molecular genetics of autism spectrum disorder. Mol Psychiatry. 2004;9(9):819–32.

Bill BR, Geschwind DH. Genetic advances in autism: heterogeneity and convergence on shared pathways. Curr Opin Genet Dev. 2009;19(3):271–8.

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(2):493–507.

Lamb JA, Moore J, Bailey A, Monaco AP. Autism: recent molecular genetic advances. Hum Mol Genet. 2000;9(6):861–8.

Anderson BM, Schnetz-Boutaud N, Bartlett J, Wright HH, Abramson RK, Cuccaro ML, et al. Examination of association to autism of common genetic variationin genes related to dopamine. Autism Res. 2008;1(6):364–9.

Anderson BM, Schnetz-Boutaud NC, Bartlett J, Wotawa AM, Wright HH, Abramson RK, et al. Examination of association of genes in the serotonin system to autism. Neurogenetics. 2009;10(3):209.

Campbell DB, Li C, Sutcliffe JS, Persico AM, Levitt P. Genetic evidence implicating multiple genes in the MET receptor tyrosine kinase pathway in autism spectrum disorder. Autism Res. 2008;1(3):159–68.

Coutinho AM, Sousa I, Martins M, Correia C, Morgadinho T, Bento C, et al. Evidence for epistasis between SLC6A4 and ITGB3 in autism etiology and in the determination of platelet serotonin levels. Hum Genet. 2007;121(2):243–56.

Ashley-Koch AE, Jaworski J, de Ma Q, Mei H, Ritchie MD, Skaar DA, et al. Investigation of potential gene-gene interactions between APOE and RELN contributing to autism risk. Psychiatr Genet. 2007;17(4):221–6.

Ma DQ, Whitehead PL, Menold MM, Martin ER, Ashley-Koch AE, Mei H, et al. Identification of significant association and gene–gene interaction of GABA receptor subunit genes in autism. Am J Hum Genet. 2005;77(3):377–88.

Kim SJ, Brune CW, Kistner EO, Christian SL, Courchesne EH, Cox NJ, et al. Transmission disequilibrium testing of the chromosome 15q11–q13 region in autism. Am J Med Genet B Neuropsychiatr Genet. 2008;147B(7):1116–25.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen–metabolism genes in sporadic breast cancer. Am J Hum Genet. 2001;69(1):138–47.

James SJ, Melnyk S, Jernigan S, Cleves MA, Halsted CH, Wong DH, et al. Metabolic endophenotype and related genotypes are associated with oxidative stress in children with autism. Am J Med Genet B Neuropsychiatr Genet. 2006;141(8):947–56.

Perry SW, Norman JP, Litzburg A, Gelbard HA. Antioxidants are required during the early critical period, but not later, for neuronal survival. J Neurosci Res. 2004;78(4):485–92.

Cohen-Kerem R, Koren G. Antioxidants and fetal protection against ethanol teratogenicity. I. Review of the experimental data and implications to humans. Neurotoxicol Teratol. 2003;25(1):1–9.

Valicenti-McDermott M, McVicar K, Rapin I, Wershil BK, Cohen H, Shinnar S. Frequency of gastrointestinal symptoms in children with autistic spectrum disorders and association with family history of autoimmune disease. J Dev Behav Pediatr. 2006;27(2 Suppl):S128–36.

Pardo CA, Vargas DL, Zimmerman AW. Immunity, neuroglia and neuroinflammation in autism. Int Rev Psychiatry. 2005;17(6):485–95.

James SJ, Cutler P, Melnyk S, Jernigan S, Janak L, Gaylor DW, et al. Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. Am J Clin Nutr. 2004;80(6):1611–7.

Sogut S, Zoroglu SS, Ozyurt H, Yilmaz HR, Ozugurlu F, Sivasli E, et al. Changes in nitric oxide levels and antioxidant enzyme activities may have a role in the pathophysiological mechanisms involved in autism. Clin Chim Acta. 2003;331(1–2):111–7.

Yorbik O, Sayal A, Akay C, Akbiyik DI, Sohmen T. Investigation of antioxidant enzymes in children with autistic disorder. Prostaglandins Leukot Essent Fatty Acids. 2002;67(5):341–3.

Buyske S, Williams TA, Mars AE, Stenroos ES, Ming SX, Wang R, et al. Analysis of case-parent trios at a locus with a deletion allele: association of GSTM1 with autism. BMC Genet. 2006;7(1):8.

Williams TA, Mars AE, Buyske SG, Stenroos ES, Wang R, Factura-Santiago MF, et al. Risk of autistic disorder in affected offspring of mothers with a glutathione S-transferase P1 haplotype. Arch Pediatr Adolesc Med. 2007;161(4):356–61.

Maher P. Redox control of neural function: background, mechanisms, and significance. Antioxid Redox Signal. 2006;8(11-12):1941–70.

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(2):463–6.

Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994;24(5):659–85.

Lord C, Risi S, Lambrecht L, Cook Jr EH, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000;30(3):205–23.

Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. Mol Psychiatry. 2007;12(1):2–22.

Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. Genome Res. 2005;15(11):1592–3.

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet. 2005;37(11):1217–23.

Li Q, Louis T, Fallin MD, Ruczinski I. Trio logic regression—detection of SNP–SNP interactions in case–parent trios. Genetic epidemiology 2009; Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 194.

Li Q, Fallin D, Louis T, McGrath J, Avramapoulos D, Wolyniec P, et al. Detection of SNP–SNP interactions in trios of parents with schizophrenic children. Genetic epidemiology. 2010;34(5):396–406.

R Development Core Team. R: A language and environment for statistical computing. 2008; 2.7.2.

Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. 2008;358(7):667–75.

Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature. 2009;459(7246):528–33.

Townsend DM, Tew KD, Tapiero H. The importance of glutathione in human disease. Biomed Pharmacother. 2003;57(3-4):145–55.

Wang J, Huff AM, Spence JD, Hegele RA. Single nucleotide polymorphism in CTH associated with variation in plasma homocysteine concentration. Clin Genet. 2004;65(6):483–6.

Lillig CH, Berndt C, Holmgren A. Glutaredoxin systems. Biochim Biophys Acta. 2008;1780(11):1304–17.

Koivusalo M, Baumann M, Uotila L. Evidence for the identity of glutathione-dependent formaldehyde dehydrogenase and class III alcohol dehydrogenase. FEBS Lett. 1989;257(1):105–9.

Liu L, Hausladen A, Zeng M, Que L, Heitman J, Stamler JS. A metabolic enzyme for S-nitrosothiol conserved from bacteria to humans. Nature. 2001;410(6827):490–4.

Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. Alcohol Res Health. 2007;30(1):5–13.

Henderson GI, Chen JJ, Schenker S. Ethanol, oxidative stress, reactive aldehydes, and the fetus. Front Biosci. 1999;4:D541–50.

Fallin MD, Matteini A. Genetic epidemiology in aging research. J Gerontol A Biol Sci Med Sci. 2009;64(1):47–60.

Ming X, Johnson WG, Stenroos ES, Mars A, Lambert GH, Buyske S. Genetic variant of glutathione peroxidase 1 in autism. Brain Dev. 2010;32(2):105–9.