An Analysis Paradigm for Investigating Multi-locus Effects in Complex Disease: Examination of Three GABA$_A$ Receptor Subunit Genes on 15q11-q13 as Risk Factors for Autistic Disorder

A. E. Ashley-Koch[*], H. Mei, J. Jaworski, D. Q. Ma, M. D. Ritchie[1], M. M. Menold, G. R. Delong[2], R. K. Abramson[3], H. H. Wright[4], J. P. Hussman[4], M. L. Cuccaro, J. R. Gilbert, E. R. Martin and M. A. Pericak-Vance

Center for Human Genetics, Department of Medicine, Duke University Medical Center, Durham, NC
[1]*Vanderbilt Center for Human Genetics Research, Vanderbilt University, Nashville, TN*
[2]*Department of Pediatrics, Duke University Medical Center, Durham, NC*
[3]*W.S. Hall Psychiatric Institute, Columbia, SC*
[4]*John P Hussman Foundation, Ellicott City, MD*

Summary

Gene-gene interactions are likely involved in many complex genetic disorders and new statistical approaches for detecting such interactions are needed. We propose a multi-analytic paradigm, relying on convergence of evidence across multiple analysis tools. Our paradigm tests for main and interactive effects, through allele, genotype and haplotype association. We applied our paradigm to genotype data from three GABA$_A$ receptor subunit genes (*GABRB3, GABRA5,* and *GABRG3*) on chromosome 15 in 470 Caucasian autism families. Previously implicated in autism, we hypothesized these genes interact to contribute to risk. We detected no evidence of main effects by allelic (PDT, FBAT) or genotypic (genotype-PDT) association at individual markers. However, three two-marker haplotypes in *GABRG3* were significant (HBAT). We detected no significant multi-locus associations using genotype-PDT analysis or the EMDR data reduction program. However, consistent with the haplotype findings, the best single locus EMDR model selected a *GABRG3* marker. Further, the best pairwise genotype-PDT result involved *GABRB3* and *GABRG3,* and all multi-locus EMDR models also selected *GABRB3* and *GABRG3* markers. GABA receptor subunit genes do not significantly interact to contribute to autism risk in our overall data set. However, the consistency of results across analyses suggests that we have defined a useful framework for evaluating gene-gene interactions.

Introduction

Gene-gene interactions are hypothesized to play an important role in the etiology of many complex genetic disorders. In spite of this, most candidate gene association studies typically assess effects of candidate genes independently of each other. Studies of the joint effect of other candidate genes are rare. It is our working hypothesis that for some genes these epistatic, or gene-gene, interactions may be more important than the independent effects of the single genes. The idea that gene-gene interactions play an important role in human biology is not new. Wright (1932) emphasized that the relationship between genotype and phenotype is dependent on dynamic interactive networks of genes and environmental factors. This idea holds true today. Gibson (1996) stressed that gene-gene and gene-environment

*Correspondence author: Allison Ashley-Koch, Ph.D., Center for Human Genetics, Duke University Medical Center, Box 3400, 2007 Snyderman Genomic Sciences Building, 595 LaSalle Street, Durham, NC 27710, Phone: (919) 684-1805; Fax: (919) 684-0912; E-mail: allison.ashleykoch@duke.edu.

interactions must be ubiquitous, given the complexities of intermolecular interactions that are necessary to regulate gene expression and the hierarchical complexity of metabolic networks. The identification and characterization of common complex disease susceptibility genes remains one of the great challenges facing human geneticists.

Traditional parametric statistical methods, such as logistic regression, can be applied to the detection of gene-gene interactions. But for a few reasons, these traditional methods may not be sufficiently powered or flexible to detect genetic effects that are only observed in the presence of other genes or environmental factors (Templeton, 2000; Schlichting & Pigliucci, 1998). For example, with many parametric statistical methods modeling high-order interactions may result in several contingency table cells that have no observations, and therefore lead to very large standard errors of coefficient estimates (Hosmer & Lemeshow, 2000). This situation is exacerbated as the number of polymorphisms increases and the number of combinations of variations that needs to be evaluated increases exponentially. Some methods, such as stepwise logistic regression, deal with this difficulty by only considering factors that have a statistically significant marginal or main effect for the model. The limitation with this approach is that factors with purely interactive effects will be missed.

To address these issues, the multifactor dimensionality reduction (MDR) method, a data reduction approach, has been developed (Ritchie *et al.* 2000). We have subsequently modified this approach into the 'extended' MDR or EMDR which allows for more analysis options, including additional test statistics and permutation tests (Mei *et al.*, in press). Recently, Martin and colleagues have also expanded the original pedigree disequilibrium test (Martin, 2000b) to the genotype-PDT in order to examine genotypic association, allowing for analysis of not only a single locus, but also multi-locus effects (Martin *et al.* 2003). These are just a few examples of recently developed methods that can be applied to detect gene-gene interactions. However, this methodology is still in its infancy, and no single method has thus far proven to be superior to others. Furthermore, in order to sort out true interactive effects from joint effects that are driven by a strong main effect, one must also concurrently assess for single locus effects. Thus, given

all these issues, a multi-analytic approach to analysis of gene-gene interactions in complex disease, searching for consistency of results and preponderance of evidence to make conclusions, should prove most useful.

One of the prototypical complex diseases hypothesized to include epistatic genetic effects is autistic disorder (AutD; MIM209850), a neurodevelopmental disorder characterized by disturbances in social, communicative and behavioural functioning. Significant evidence from twin and family studies supports a strong genetic component (Bailey *et al.* 1995; Lotspeich & Ciaranello, 1993; Folstein & Piven, 1991; Steffenburg *et al.* 1989; Ritvo *et al.* 1985; Folstein & Rutter, 1977), but the inheritance of AutD appears to be quite complex. The concordance rates among monozygotic and dizygotic twins are not consistent with Mendelian modes of inheritance (Bailey *et al.* 1995), and the pattern of familial recurrence risks in AutD is most consistent with an oligogenic model including epistasis (Jorde *et al.* 1991).

Examination of functional candidate genes is one approach that has been utilized by many investigators in the genetic dissection of the complex AutD phenotype. The GABAergic system, in particular, has received much attention. GABA is the primary inhibitory neurotransmitter in the adult brain, but during development GABA acts as an excitatory neurotransmitter, due to high intracellular chloride concentration in immature neurons (Jentsch *et al.* 2002). In brain, GABA acts on the GABA receptor complex, a heteropentameric structure forming a central chloride channel. Eighteen different receptor subunit genes have been characterized in mammals. Classes of subunits include α, β, δ, ϵ, γ, π, θ, and ρ. In addition to providing binding sites for GABA, the GABA receptor contains sites for several therapeutic agents and drugs, including benzodiazepines, barbiturates, anesthetics and alcohols. Binding studies using labeled ligands in human children and nonhuman primates indicate that GABA receptor density is greatest early in life, and then dramatically decreases to adult levels (Chugani *et al.* 2001).

Several lines of genetic evidence specifically implicate the involvement of three $GABA_A$ receptor subunit genes located on 15q11-q13 ($\alpha 5$, $\beta 3$, $\gamma 3$) in AutD susceptibility. First and foremost, these three $GABA_A$ receptor subunit genes are physically positioned in the region

on chromosome 15q which is the most common site of chromosomal abnormalities observed in autistic patients (Wolpert *et al*. 2000; Gillberg, 1998; Schroer *et al*. 1998). But there is also evidence connecting the individual genes to Aut D. Genetic markers within the *GABRB3* gene have been implicated in AutD susceptibility through both genetic linkage (Shao *et al*. 2002, 2003; Philippe *et al*. 1999; Liu *et al*. 2001) and linkage disequilibrium (Cook *et al*. 1997; Martin *et al*. 2000a; Buxbaum *et al*. 2002) analyses, making *GABRB3* an excellent candidate gene for AutD. However, this association with *GABRB3* has not been universally confirmed (Maestrini *et al*. 1999; Salmon *et al*. 1999). Phenotypic subsets of AutD have provided linkage evidence for even stronger association with the *GABRB3* region. Shao and colleagues (Shao *et al*. 2003) found a significant increase in the linkage evidence for the *GABRB3* region in families in which affected individuals had a high degree of insistence on sameness. Similarly, recent examination of this region in Aut D multiplex families (2 or more affected individuals) revealed that the evidence for linkage to the *GABRB3* region was increased in the subset of families with probands who had greater savant skills (Nurmi *et al*. 2003). This association between savant skills and the GABRB3 region, however, was not replicated in our own data set (Ma *et al*. 2005a). In addition to evidence implicating the *GABRB3* region, we have also observed evidence for association between AutD and the *GABRG3* region (Menold *et al*. 2001). Further evidence implicating a role for *GABRG3* is the observation that *GABRG2* mutations have been detected in families with epilepsy phenotypes (Wallace *et al*. 2001). This may be important with regard to *GABRG3* because seizures are observed in subjects with autism more than in the general population (Pavone *et al*. 2004; Tuchman & Rapin, 2002).

There are functional data to support a role of the GABA$_A$ receptors as well. PET imaging *in vivo* using [^{11}C]flumazenil (FMZ) shows decreased binding in autistic children compared with controls (Chugani, 2001). Further, there is a significant decrease in GABA$_A$ receptors in the hippocampus from autistic brains compared with control brains (Blatt *et al*. 2001). The hippocampus composes part of the limbic system, controlling both emotion and long-term memory. Disruptions of the limbic system could lead to some of the hallmark characteristics of autism, namely difficulties in social interactions.

Due to these functional and genetic data, we have examined the possibility that this cluster of GABA genes on chromosome 15q11-q13 may be acting epistatically to contribute to autism susceptibility. We have chosen to address this complex question with a multi-analytic paradigm that emphasizes reproducibility of results across different analytic tools. As a prototype for analysis of gene-gene interactions, we have utilized in our analysis approach both methods to assess main effects (PDT and genotype-PDT, Martin *et al*. 2000b, 2003) and the Haplotype Based Association Test (Horvath, 2004), as well as methods to detect multi-locus effects (EMDR, Mei *et al*. in press), and the multi-locus genotype-PDT (Martin *et al*. 2003).

Materials and Methods

Data Set

Statistical analyses were performed on a total of 470 Caucasian AutD families (265 multiplex (two or more affected individuals) and 205 trios (parents and affected offspring)). The Collaborative Autism Team (CAT) from the Duke Center for Human Genetics and the WS Hall Psychiatric Institute contributed 246 families. Two-hundred-and-twenty-four families were ascertained by the Autism Genetic Resource Exchange (AGRE). All affected individuals were ascertained on the basis of a clinical diagnosis of AutD and were between 3 and 21 years of age. The Autism Diagnostic Interview-Revised (ADI-R) was used to confirm the clinical diagnosis of autism. The classification of an individual with autism required that their ADI-R scores exceed cutoffs in each of the three critical areas: social behaviour, communication (nonverbal or verbal) and restricted, repetitive behaviours. Families were excluded in cases where the AutD diagnosis was not idiopathic (e.g. Fragile X Syndrome, Tuberous Sclerosis Complex) or was associated with a cytogenetic abnormality. Blood was obtained from patients and other family members under IRB-approved procedures. DNA was extracted from whole blood using standard protocols (Vance, 1998).

GENE	SNP Ref ID NCBI or Celera	SNP location within gene	SNP chromosomal location (bp) (NCBI build 35)
GABRB3	rs2081648	Intron 8	24349292
	rs1426217	Intron 6	24372218
	rs754185	Intron 3	24438972
	rs890317/hcv8865209	Intron 3	24473294
	rs2059574	Intron 3	24548136
GABRA5	hcv42974	Intron 7	24743281
	rs7173260	Intron 7	24754690
	rs140681	Intron 7	24764958
	rs140683	Intron 9	24771081
GABRG3	rs7172534/hcv2078506	Intron 3	24855745
	rs208129	Intron 3	25007653
	rs897173	Intron 3	25052647
	hcv428306	Intron 6	25406295
	rs140679	Exon 8	25446271

Table 1 SNP Locations within the GABA$_A$ receptor subunit genes

SNP Genotyping

SNPs within each candidate region were identified using NCBI's Single Nucleotide Polymorphism database (dbSNP) (http://www.ncbi.nlm.nih.gov/SNP). Table 1 describes the locations of the SNPs that were genotyped within *GABRB3*, *GABRA5* and *GABRG3*. We genotyped 5 SNPs within *GABRB3* and *GABRG3* and 4 SNPs within the *GABRA5* gene. SNPs were selected for analysis based on availability of assay from the manufacturer (Applied Biosystems, Foster City, CA), coverage across the gene, and rare allele frequencies. Most markers had a heterozygosity of over 35%, with the exception of markers rs2081648 in *GABRB3* (23%), rs140681 in *GABRA5* (16%) and hcv428306 in *GABRG3* (21%). Since we examined multiple SNPs per gene, we used Haploview (Barrett *et al*. 2004) to assist us in identifying 'haplotype tagging' SNPs, such that per gene we reduced the repetitive information coming from SNPs within the same haplotype block. However, as demonstrated by the linkage disequilbrium table (Table 2), there were only two markers in significant LD as defined by r^2 > 0.20 (hcv42974 and rs7173260 in *GABRA5*). Thus, we used all markers in all analyses, with the exception of the EMDR analysis where we omitted rs7173260.

SNP genotyping was performed by TaqMan, using 'Assays-on-Demand' or 'Assays-by-Design' SNP genotyping products (Applied Biosystems, Foster City, CA). For all genotype assays, quality control measures were applied, including genotyping a series of blinded duplicate samples and CEPH controls. The genotypes of all duplicate samples had to match in order for the assay to pass quality control. Further, we required that each assay achieve 95% efficiency (i.e. the genotypes of at least 95% of the samples could be called with certainty) to be considered for statistical analysis. Assays were performed on ABI 9700 dual 384-well Geneamp PCR systems according to the manufacturer's instructions (Applied Biosystems, Foster City, CA). Genotypes were analyzed using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Each reaction contained 2.7ng of total genomic DNA.

Statistical Methods

The first step of the multi-analytic paradigm described the characteristics of each marker with respect to deviations from Hardy-Weinberg equilibrium, and linkage disequilibrium with the other markers under examination. Hardy-Weinberg equilibrium was assessed using exact tests implemented in the Genetic Data Analysis program (Zaykin *et al*. 1995). Pairwise linkage disequilibria (D$'$ and r^2) between markers within each gene were calculated in the affected individuals and unaffected individuals separately, using one affected and one unaffected from each family by applying the software package GOLD (Abecasis & Cookson, 2000). The patterns of linkage disequilibrium amongst the markers are shown in Table 2.

The second step of the paradigm defines single locus allelic and genotypic associations with the phenotype.

Table 2 Linkage disequilibrium between all SNPs typed, calculated by R^2 and D′

R^2 — Affected Individuals (upper-right triangle) / Normal Individuals (lower-left triangle)

	rs2081648	rs1426217	rs754185	hcv8865209	rs2059574	hcv42974	rs7173260	rs140681	rs140683	hcv2078506	rs208129	rs897173	hcv428306	rs140679
rs2081648		0.113	0.025	0.002	0.005	0.003	0.003	0.005	0.001	0.002	0.000	0.001	0.001	0.001
rs1426217	0.111		0.000	0.006	0.000	0.004	0.009	0.000	0.001	0.000	0.002	0.000	0.000	0.000
rs754185	0.020	0.000		0.002	0.006	0.001	0.001	0.000	0.001	0.003	0.011	0.000	0.000	0.000
hcv8865209	0.018	0.021	0.007		0.012	0.001	0.001	0.000	0.001	0.002	0.002	0.000	0.000	0.001
rs2059574	0.017	0.011	0.000	0.016		0.001	0.000	0.008	0.000	0.001	0.004	0.002	0.000	0.000
hcv42974	0.027	0.005	0.001	0.000	0.012		0.362	0.002	0.126	0.003	0.000	0.000	0.009	0.001
rs7173260	0.004	0.001	0.024	0.019	0.021	0.273		0.002	0.103	0.000	0.003	0.005	0.002	0.002
rs140681	0.000	0.001	0.041	0.025	0.000	0.001	0.001		0.061	0.010	0.001	0.004	0.007	0.001
rs140683	0.017	0.001	0.001	0.000	0.013	0.126	0.006	0.062		0.000	0.007	0.006	0.002	0.006
hcv2078506	0.008	0.031	0.003	0.004	0.004	0.000	0.001	0.000	0.000		0.004	0.001	0.004	0.011
rs208129	0.000	0.007	0.007	0.003	0.000	0.003	0.019	0.004	0.001	0.006		0.001	0.004	0.001
rs897173	0.011	0.000	0.000	0.001	0.000	0.000	0.007	0.014	0.000	0.000	0.012		0.000	0.000
hcv428306	0.033	0.013	0.000	0.000	0.005	0.012	0.002	0.003	0.001	0.000	0.000	0.011		0.000
rs140679	0.004	0.015	0.001	0.017	0.001	0.000	0.000	0.004	0.004	0.001	0.016	0.013	0.006	

D′ — Affected Individuals (upper-right triangle) / Normal Individuals (lower-left triangle)

	rs2081648	rs1426217	rs754185	hcv8865209	rs2059574	hcv42974	rs7173260	rs140681	rs140683	hcv2078506	rs208129	rs897173	hcv428306	rs140679
rs2081648		1.000	0.280	0.171	0.180	0.095	0.140	0.568	0.059	0.086	0.032	0.184	0.242	0.063
rs1426217	1.000		0.006	0.137	0.011	0.102	0.104	0.003	0.038	0.004	0.048	0.026	0.035	0.026
rs754185	0.307	0.001		0.090	0.109	0.032	0.047	0.080	0.037	0.064	0.111	0.004	0.035	0.020
hcv8865209	0.644	0.257	0.184		0.169	0.026	0.054	0.001	0.040	0.049	0.052	0.016	0.035	0.047
rs2059574	0.366	0.115	0.027	0.211		0.053	0.020	0.298	0.011	0.042	0.084	0.076	0.012	0.001
hcv42974	0.334	0.106	0.027	0.045	0.157		0.926	0.234	0.428	0.090	0.002	0.018	0.177	0.041
rs7173260	0.226	0.031	0.277	0.184	0.198	1.000		0.127	0.423	0.002	0.081	0.124	0.089	0.053
rs140681	0.187	0.148	0.994	0.999	0.013	0.065	0.064		1.000	0.267	0.127	0.106	0.103	0.092
rs140683	0.353	0.035	0.035	0.016	0.120	0.466	0.121	1.000		0.030	0.096	0.172	0.145	0.097
hcv2078506	0.198	0.253	0.060	0.080	0.076	0.010	0.056	0.033	0.006		0.097	0.051	0.135	0.122
rs208129	0.025	0.103	0.083	0.124	0.027	0.095	0.142	0.341	0.047	0.129		0.071	0.216	0.048
rs897173	0.585	0.023	0.002	0.036	0.018	0.018	0.121	0.237	0.031	0.048	0.284		0.007	0.019
hcv428306	0.186	0.374	0.038	0.013	0.208	0.232	0.087	0.073	0.087	0.027	0.046	0.595		0.006
rs140679	0.158	0.123	0.041	0.250	0.026	0.026	0.022	0.208	0.065	0.045	0.148	0.193	0.207	

To accomplish this, we used the pedigree disequilibrium test (Martin, 2001; Martin, 2000, the genotype-PDT (Martin *et al.* 2003), and the family based association test (Horvath, 2004). The PDT and FBAT are similar allele-based tests of association. However, PDT is a valid test of both linkage and association in extended pedigrees. The FBAT treats nuclear families within an extended pedigree as independent. The genotype-PDT is a genotype-based association test and an extension of the original PDT. All approaches provide both allele- or genotype-specific P-values, as well as global P-values adjusting for all alleles/genotypes. We report here only the global P-values.

The third step of the paradigm defines haplotype associations. We used the haplotype based association test (Horvath, 2004) to accomplish this. For the HBAT we obtained global scores for haplotypes for each pair of SNPs within each gene. That is, the HBAT calculates both haplotype-specific and global P-values for significance testing. Here we report only the global P-values which adjust for all possible haplotypes. For each SNP pair observed haplotypes with fewer than 10 families were not considered, because these haplotypes were observed in only 2% of our data set. We did not examine haplotypes across genes as this violates the assumption of tight linkage between markers.

The fourth step of the paradigm defines possible multi-locus or gene-gene interactions that may occur with the phenotype. The multi-locus genotype-PDT, an extension of the genotype-PDT (Martin *et al.* 2003), was one method used to examine the joint effects of genes. We also used the EMDR (Mei *et al.* in press), an extension of the MDR data reduction program (Ritchie *et al.* 2000), to examine possible multi-locus effects. The EMDR differs from the original MDR program in that it allows missing data for individuals who were incompletely genotyped, includes a chi-square statistic and allows for a non-fixed permutation test (described below); please also see Ma *et al.* 2005b for further description.

The current version of the EMDR is restricted to either unrelated case-control data or matched pairs (e.g. discordant sibpairs or constructed from family triad data). In our analysis, we selected the proband (or most completely genotyped affected child) from each multi-plex and triad family (n = 470 total) as the affected, and then generated an inferred 'unaffected' sibling (e.g. the untransmitted alleles) based on parental genotypes. The use of an "inferred" unaffected which is constructed from untransmitted alleles in triad data (AFBAC population) has been previously examined in detail by Thomson (1995) and shown to be a reasonable approach. We chose to use untransmitted alleles as the control, because in complex diseases individuals who do not meet established diagnostic criteria may still carry the at-risk genotype(s), making the untransmitted alleles of the cases a better control. And in fact it has been shown that for diseases with low prevalence, unaffected siblings provide little information in family-based association tests (Kaplan & Martin, 2001). For each of the EMDR runs we allowed the program to find the best possible combination of loci that predicted affection status. We did this for single, double, triple and quadruple combinations of loci in our data set. While it is possible to examine even higher order interactions with this methodology, the interpretation of results becomes quite difficult, and the power is also reduced because of the inherent correction for multiple testing in the EMDR permutation test.

To evaluate which combination of loci was the best at predicting affection status, two test statistics were used for each analysis run: chi-square statistic, and classification error (CE). CE is the misclassification error in the entire data set, and is calculated in a marginal table as the percentage of controls classified as high-risk and the percentage of cases classified as low-risk. Therefore, the best model is the locus model that yields the largest chi-square or smallest CE. This approach differs from the original MDR by Ritchie *et al.* (2001) in that it does not use the cross-validation approach, which divides the data set into n-1 subsets for training and uses the remaining subset for validation. For example, in the original MDR 10-fold cross-validation our data set of 470 families would have been subdivided into 10 subsets of 47 families each, where 9 subsets would be used for training to identify a model and the final subset for validation of the model. We previously found that cross-validation produced a large variation in test statistics, leading to inconsistent conclusions (Mei *et al*, in press). Because cross-validation subdivides the data set into several smaller data sets, rather than utilizing the

entire sample as a whole, the inconsistencies that we observed could be due to sample size and/or genetic heterogeneity. Thus, for the purposes of our analysis, we concluded that non-cross-validation performs more reliably than cross-validation.

To determine statistical significance for the best over-all predicted combinations of loci, the EMDR provides three kinds of permutation tests (fixed, non-fixed and omnibus), each of which attempts to adjust for the data reduction technique across many locus combinations to a different extent. All permutation tests hypothesize that a specific n-locus model is independent of case or randomly associated with risk of case. Data are simulated under the null hypothesis by permuting case and control status within each family triad (e.g., for a family triad transmitted and non-transmitted genotypes are permuted randomly). For a review of these permutation tests, please see Mei *et al*. (in press). Briefly, the fixed permutation test only permutes a specific n-locus model (i.e. if marker A and marker C were selected as the best model for the 2-locus model, only that combination would be permuted), the non-fixed permutation test permutes over all possible models within a n-locus test (i.e. all possible 2 locus combinations), and the omnibus test permutes over the entire set of models (i.e. all possible 1-locus, 2-locus...k-locus models). For the purpose of this analysis we used the non-fixed permutation test, which accounts for the multiple testing of all possible models for each n-locus combination.

Results

We began first with the analyses of main effects with single loci. None of the SNPs examined (Table 1) deviated from Hardy-Weinberg equilibrium expectations in the subset of independent unaffected individuals observed in our sample. However, we did observe a deviation from equilibrium in the affected individuals at marker rs1426217 in *GABRB3* ($p = 0.02$). This could represent a potentially interesting disease association, since the deviation was observed only in the affected individuals. For the single locus PDT and FBAT, as well as the genotype-PDT, none of the markers provided a statistically significant association when considering allelic or genotypic association at each marker independently (Table 3). However, when we examined haplotype associations within each gene, we did observe an

Table 3 Single locus PDT, FBAT and Genotype-PDT results

Gene	SNP	P-value for PDT	P-value for FBAT	Global P-value for Genotype-PDT
GABRB3	rs2081648	0.60	0.63	0.84
	rs1426217	0.19	0.17	0.31
	rs754185	0.67	0.73	0.85
	hcv8865209	0.34	0.35	0.52
	rs2059574	0.30	0.25	0.41
GABRA5	hcv42974	0.65	0.72	0.07
	rs7173260	0.94	0.90	0.85
	rs140681	0.89	0.81	0.76
	rs140683	0.83	0.83	0.98
GABRG3	hcv2078506	0.08	0.08	0.24
	rs208129	0.28	0.25	0.27
	rs897173	0.24	0.24	0.45
	hcv428306	1.00	0.95	0.61
	rs140679	0.41	0.39	0.59

association at the *GABRG3* locus. As shown in Table 4, for pairwise analyses of markers within *GABRG3*, three pairwise combinations of markers provided global p-values of 0.05 or less. Furthermore, when we simultaneously examined all 5 markers that were genotyped in *GABRG3*, we observed a single haplotype that was over-transmitted to the affected individuals ($p = 0.01$). However, this haplotype was only present at a frequency of 2.5%. When we examined all markers for *GABRB3* and *GABRA5* simultaneously within each gene, there was no significant haplotype association detected for either gene (data not shown).

For the multi-locus analyses, we began by performing pairwise analyses with the multi-locus genotype PDT. None of the global p-values were less than 0.05 (data not shown). The smallest global p-value obtained was 0.12 for hcv8865209 in *GABRB3* with rs140679 in *GABRG3*.

Table 5 describes the results of the best locus models identified by the EMDR analysis. As shown in the Table, no single or multi-locus model provided evidence for association with the AutD phenotype when considering either the chi-square or prediction error statistic. The best single locus model selected marker rs208129 in *GABRG3*. The best two-locus model selected markers rs2059574 in *GABRB3* and rs208129 in *GABRG3*. The best three-locus model selected marker hcv8865209 in *GABRB3* and markers rs208129 and

Table 4 Pairwise HBAT analysis within each gene

Gene	SNP A	SNP B	Global P-value
GABRB3	rs2081648	rs1426217	0.18
	rs2081648	rs754185	0.93
	rs2081648	hcv8865209	0.17
	rs2081648	rs2059574	0.59
	rs1426217	rs754185	0.69
	rs1426217	hcv8865209	0.36
	rs1426217	rs2059574	0.47
	rs754185	hcv8865209	0.35
	rs754185	rs2059574	0.79
	hcv8865209	rs2059574	0.39
GABRA5	hcv42974	rs7173260	0.69
	hcv42974	rs140681	0.89
	hcv42974	rs140683	0.62
	rs7173260	rs140681	0.93
	rs7173260	rs140683	0.92
	rs140681	rs140683	0.69
GABRG3	hcv2078506	rs208129	**0.03**
	hcv2078506	rs897173	**0.04**
	hcv2078506	hcv428306	**0.05**
	hcv2078506	rs140679	0.13
	rs208129	rs897173	0.36
	rs208129	hcv428306	0.15
	rs208129	rs140679	0.64
	rs897173	hcv428306	0.46
	rs897173	rs140679	0.44
	hcv428306	rs140679	0.66

rs140679 in *GABRG3*. And finally, the best four-locus model selected markers hcv8865209 and rs2059574 in *GABRB3* and markers rs208129 and rs140679 in *GABRG3*.

Discussion

Whilst it is hypothesized that gene-gene interactions play an important role in the etiology of many complex disorders, the methodology for detecting gene-gene interactions is still in its infancy. In this manuscript we present an analysis paradigm for examining multi-locus effects in complex diseases, and test this approach on real data from families with autistic disorder. We chose to take a multi-analytic approach and looked for convergence of evidence among the various methods, rather than relying solely on the results from a single analytic tool. We believe that it will be necessary to use multiple analysis tools in order to interpret findings of higher-order interactions among such data. In particular, one must not consider only interactive effects, but simultaneously assess main effects of the genes as well. The results from the main effects analyses will be necessary to sift through true interactive effects versus effects that are solely driven by a strong main effect of a single gene. We do acknowledge that one of the difficulties with this multi-analytic approach is the issue of multiple testing. We did not apply a correction for multiple testing across

Table 5 EMDR results on chromosome 15

Location (cM)	Marker	Marker number	Best-model*	P-value for chi-square	P-value for classification error
11.07	rs2081648	**1**	**10**	0.22	0.71
11.08	rs1426217	**2**	**5 10**	0.49	0.56
11.23	rs754185	**3**	**4 10 13**	0.84	0.85
11.33	hcv8865209	**4**	**4 5 10 13**	0.62	0.88
11.54	rs2059574	**5**			
12.06	hcv42974	**6**			
12.12	rs140681	**7**			
12.14	rs140683	**8**			
12.38	hcv2078506	**9**			
12.81	rs208129	**10**			
12.94	rs897173	**11**			*GABRB3*
14.46	hcv428306	**12**			*GABRA5*
14.66	rs140679	**13**			*GABRG3*

*Best-model numbers refer to the "marker number" in the previous column. i.e. Single locus best-model is marker number "10" which is rs208129 in *GABRG3*.

the various methods that we used. As described by the linkage disequilibrium amongst the markers in Table 2, the markers that we examined were not completely independent. Thus, applying a Bonferroni correction for multiple testing would be conservative. However, it is unclear what the correction should be. This issue of multiple testing is undoubtedly an area for future development. In conjunction with a multi-analytic approach, it will be necessary for replication in other independent data sets, and for molecular biological experiments to conclusively elucidate complex genetic interactions.

The results of the application of our analysis paradigm to the real genotype data in autistic disorder do not support the presence of multi-locus effects amongst the GABA receptor subunit genes on chromosome 15 as a major contributor to autism etiology. Nonetheless, we were still pleased with the consistency of results that were obtained from the analysis paradigm. For example, in the main effects analysis the marker with the smallest p-value for the allele-based tests (PDT p = 0.08 and FBAT p = 0.08) was hcv2078506 in *GABRG3*. When we performed the HBAT analysis, it was again markers in *GABRG3* that provided the smallest p-values. These results were also supported by the results of the EMDR, which selected marker rs208129 in *GABRG3* as the best single locus model. Thus, although the results of the single locus EMDR analysis were not statistically significant, they were consistent with the more traditional allele- and haplotype-based analyses. That is to say that if there is a main effect in our data set of one of the GABAR subunit genes in Aut D susceptibility, the consistency across the various analyses would suggest that it may lie within the *GABRG3* gene. Furthermore, in the EMDR analysis, the two-, three- and four-locus models consistently selected markers in both *GABRB3* and *GABRG3*, even though they were not statistically significant. And although we observed no significant global p-values for the multi-locus genotype-PDT analysis, the best result was obtained with marker hcv8865209 in *GABRB3* and marker rs140679 in *GABRG3*, both markers also having been selected in EMDR multi-locus analyses. Thus the results, although not statistically significant, were consistent across analyses suggesting a possible interactive effect between *GABRB3* and *GABRG3*. It is interesting to note that these results are

also consistent with the previous main effects analyses of the GABA receptor subunit genes on chromosome 15 in autism. That is, both the *GABRB3* and *GABRG3* regions have been implicated by linkage and/or association analysis, but *GABRA5* has not. Thus, we conclude that our multi-analytic paradigm is a useful approach for evaluating the presence of multi-locus effects in complex disease.

Since we did observe consistency of results, albeit not statistically significant, it is possible that our sample size was insufficient to detect multi-locus interactions of small effect size. The power of the original MDR has been described previously in both simulated and real data from breast cancer patients (Ritchie *et al.* 2003). This analysis showed that the approach had reasonable power for sample sizes of 200 cases and 200 controls. This is less than half the sample size used in the current analysis (n = 470 cases and 470 controls). The EMDR, using the non-fixed permutation test, no cross-validation and a sample of 440 triads, has demonstrated sufficient power to identify the underlying multi-locus effects generated in the GAW14 simulated data (Mei *et al.* in press). While it is possible that the effect sizes that we were trying to detect our analysis were so small that the present data set was under-powered, our previous experience with both the MDR and the EMDR would suggest that this is not the case.

Another possibility, given our previous findings, is that there is a specific phenotypic subset of patients with autistic disorder that harbours susceptibility in these GABA receptor subunit genes. In particular, we previously found that multiplex families with a high-degree of insistence on sameness provided the most evidence for linkage at *GABRB3* in our data set (Shao *et al.* 2003). Thus, we also analyzed these 23 multiplex families for association with the markers in this current analysis. The most significant result was obtained at marker hcv8865209 in *GABRB3* (PDT p = 0.02 and genotype-PDT p = 0.04), consistent with our previous linkage findings. Furthermore, we identified an additional 69 singleton families whose probands also fit our definition of the 'insistence on sameness' subphenotype. When we combined the singleton and multiplex families, for a total of 92 families in the association analysis, we again found that marker hcv8865209 provided the most significant evidence of association (PDT

p = 0.004, genotype-PDT p = 0.008). Thus, there does appear to be a specific subgroup of patients with autistic disorder that is providing the most evidence for autism susceptibility on chromosome 15. Because the combined subset of singleton and multiplex families who fit the criteria for insistence on sameness was less than 100, we did not attempt to perform the more complicated multi-locus analyses with these families. However, as our data set grows, we will be very interested in looking more closely at this subset. Thus, these data suggest that the GABA receptor subunit genes remain excellent candidates for autism.

In conclusion, we have presented a multi-analytic framework for the examination of multi-locus effects in complex diseases. This analysis paradigm has combined multiple analysis tools that test for both main and interactive effects, at the level of allele, genotype and haplotype. The consistency of results that we obtained with this approach in our application to autistic disorder data suggests that this multi-analytic paradigm performs well, and is a reasonable framework for approaching the analysis of gene-gene interactions in all complex disorders. While other investigators may decide to use different software in this paradigm approach, we believe the keys to success are to use multiple analysis tools to identify both the main and interactive effects of the markers, as well as looking for consistency of results across those various analysis tools.

Acknowledgements

References

Abecasis G. R. & Cookson W. O. (2000) GOLD–graphical overview of linkage disequilibrium. *BioInformatics* **16**, 182–183.

Bailey A., Le Couteur A., Gottesman I., Bolton P., Simonoff E., Yuzda E. & Rutter M. (1995) Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* **25**, 63–77.

Barrett J. C., Fry B., Maller J. & Daly M. J. (2004) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*.

Blatt G. J., Fitzgerald C. M., Guptill J. T., Booker A. B., Kemper T. L. & Bauman M. L. (2001) Density and distribution of hippocampal neurotransmitter receptors in autism: an autoradiographic study. *J Autism Dev Disord* **31**, 537–543.

Buxbaum J. D., Silverman J. M., Smith C. J., Greenberg D. A., Kilifarski M., Reichert J., Cook E. H., Jr., Fang Y., Song C. Y. & Vitale R. (2002) Association between a GABRB3 polymorphism and autism. *Mol Psychiatry* **7**, 311–316.

Chugani, D. C. Abnormal Serotonin Synthesis and GABAA Receptors Measure with Pet in Autistic Children. 2001. NIH/ACC Autism Meeting, Bethesda.

Chugani D. C., Muzik O., Juhasz C., Janisse J. J., Ager J. & Chugani H. T. (2001) Postnatal maturation of human GABAA receptors measured with positron emission tomography. *Ann Neurol* **49**, 618–626.

Cook E. H., Jr., Lindgren V., Leventhal B. L., Courchesne R., Lincoln A., Shulman C., Lord C. & Courchesne E. (1997) Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am J Hum Genet* **60**, 928–934.

Folstein S. & Rutter M. (1977) Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* **18**, 297–321.

Folstein S. E. & Piven J. (1991) Etiology of autism: genetic influences. *Pediatrics* **87**, 767–773.

Gibson G. (1996) Epistasis and pleiotropy as natural properties of transcriptional regulation. *Theor Popul Biol* **49**, 58–59.

Gillberg C. (1998) Chromosomal disorders and autism. *J Autism Dev Disord* **28**, 415–425.

Horvath S., Xu X., Lake S. L., Silverman E. K., Weiss S. T. & Laird N. M. (2004) Family-based tests for associating haplotypes with general phenotype data: Application to asthma genetics. *Genetic Epidemiology* **26**, 61–69.

Hosmer D. & Lemeshow S. (2000) *Applied Logistic Regression.* New York: John Wiley & Sons, Inc.

Jentsch T. J., Stein V., Weinreich F. & Zdebik A. A. (2002) Molecular structure and physiological function of chloride channels. *Physiol Rev* **82**, 503–568.

Jorde L. B., Hasstedt S. J., Ritvo E. R., Mason-Brothers A., Freeman B. J., Pingree C., McMahon W. M., Petersen B., Jenon W. R. & Mo A. (1991) Complex segregation analysis of autism. *Am J Hum Genet* **49**, 932–938.

Kaplan N. L. & Martin E. R. (2001) Power calculations for a general class of tests of linkage and association that use nuclear families with affected and unaffected sibs. *Theor Popul Biol* **60**, 193–201.

Liu J., Nyholt D. R., Magnussen P., Parano E., Pavone P., Geschwind D., Lord C., Iversen P., Hoh J., Ott J. & Gilliam

T. C. (2001) A genomewide screen for autism susceptibility loci. *Am J Hum Genet* **69**, 327–340.

Lotspeich L. J. & Ciaranello R. D. (1993) The neurobiology and genetics of infantile autism. *Int Rev Neurobiol* **35**, 87–129.

Ma D. Q., Jaworski J., Menold M., Donnelly S., Abramson R., Wright H. H., Gilbert J. R., Pericak-Vance M. A. & Cuccaro M. L. Ordered subset analysis of savant skills in autistic disorder in 15q11-q13. *Neuropsych Genet* (in press).

Ma, D. Q., Whitehead, P. L., Menold, M. M., Martin, E. R., Ashley-Koch, A. E., Mei, H., Ritchie, M. D., DeLong, G. R., Abramson, R. K., Wright, H. H., Cuccaro, M. L., Hussman, J. P., Gilbert, J. R., & Pericak-Vance, M. A. Identification of Significant Association and Gene-Gene Interaction on GABA Receptor Subunit Genes in Autism. AJHG. 3-11-2005b.

Maestrini E., Lai C., Marlow A., Matthews N., Wallace S., Bailey A., Cook E. H., Weeks D. E. & Monaco A. P. (1999) Serotonin transporter (5-HTT) and gamma-aminobutyric acid receptor subunit beta3 (GABRB3) gene polymorphisms are not associated with autism in the IMGSA families. The International Molecular Genetic Study of Autism Consortium. *Am J Med Genet* **88**, 492–496.

Martin E. R., Bass M. P., Gilbert J. R., Pericak-Vance M. A. & Hauser E. R. (2003) Genotype-based association test for general pedigrees: the genotype-PDT. *Genet Epidemiol* **25**, 203–213.

Martin E. R., Menold M. M., Wolpert C. M., Bass M. P., Donnelly S. L., Ravan S. A., Zimmerman A., Gilbert J. R., Vance J. M., Maddox L. O., Wright H. H., Abramson R. K., DeLong G. R., Cuccaro M. L. & Pericak-Vance M. A. (2000a) Analysis of linkage disequilibrium in gamma-aminobutyric acid receptor subunit genes in autistic disorder. *Am J Med Genet* **96**, 43–48.

Martin E. R., Monks S. A., Warren L. L. & Kaplan N. L. (2000b) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* **67**, 146–154.

Menold M. M., Shao Y., Wolpert C. M., Donnelly S. L., Raiford K. L., Martin E. R., Ravan S. A., Abramson R. K., Wright H. H., DeLong G. R., Cuccaro M. L., Pericak-Vance M. A. & Gilbert J. R. (2001) Association analysis of chromosome 15 gabaa receptor subunit genes in autistic disorder. *J Neurogenet* **15**, 245–259.

Nurmi E. L., Dowd M., Tadevosyan-Leyfer O., Haines J. L., Folstein S. E. & Sutcliffe J. S. (2003) Exploratory subsetting of autism families based on savant skills improves evidence of genetic linkage to 15q11-q13. *J Am Acad Child Adolesc Psychiatry* **42**, 856–863.

Pavone P., Incorpora G., Fiumara A., Parano E., Trifiletti R. R. & Ruggieri M. (2004) Epilepsy is not a prominent feature of primary autism. *Neuroped* **35**, 207–210.

Philippe A., Martinez M., Guilloud-Bataille M., Gillberg C.,

Rastam M., Sponheim E., Coleman M., Zappella M., Aschauer H., Van Maldergem L., Penet C., Feingold J., Brice A., Leboyer M. & van Malldergerme L. (1999) Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study. *Hum Mol Genet* **8**, 805–812.

Ritchie M. D., Hahn L. W. & Moore J. H. (2003) Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genet Epidemiol* **24**, 150–157.

Ritchie M. D., Roodi N., Bailey L. R., Dupont W. D., Plummer W. D., Hahn L. W., Parl F. F. & Moore J. H. (2000) Detection of high-order interactions among estrogen metabolism genes in sporadic breast cancer. *Am J Hum Genet* **67**, 97.

Ritvo E. R., Freeman B. J., Mason-Brothers A., Mo A. & Ritvo A. M. (1985) Concordance for the syndrome of autism in 40 pairs of afflicted twins. *Am J Psychiatry* **142**, 74–77.

Salmon B., Hallmayer J., Rogers T., Kalaydjieva L., Petersen P. B., Nicholas P., Pingree C., McMahon W., Spiker D., Lotspeich L., Kraemer H., McCague P., Dimiceli S., Nouri N., Pitts T., Yang J., Hinds D., Myers R. M. & Risch N. (1999) Absence of linkage and linkage disequilibrium to chromosome 15q11-q13 markers in 139 multiplex families with autism. *Am J Med Genet* **88**, 551–556.

Schlichting C. D. & Pigliucci M. (1998) *Phenotypic evolution: A reaction norm perspective*. Sinauer Associates, Inc.

Schroer R. J., Phelan M. C., Michaelis R. C., Crawford E. C., Skinner S. A., Cuccaro M., Simensen R. J., Bishop J., Skinner C., Fender D. & Stevenson R. E. (1998) Autism and maternally derived aberrations of chromosome 15q. *Am J Med Genet* **76**, 327–336.

Shao Y., Cuccaro M. L., Hauser E. R., Raiford K. L., Menold M. M., Wolpert C. M., Ravan S. A., Elston L., Decena K., Donnelly S. L., Abramson R. K., Wright H. H., DeLong G. R., Gilbert J. R. & Pericak-Vance M. A. (2003) Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. *Am J Hum Genet* **72**, 539–548.

Shao Y., Wolpert C. M., Raiford K. L., Menold M. M., Donnelly S. L., Ravan S. A., Bass M. P., McClain C., von Wendt L., Vance J. M., Abramson R. H., Wright H. H., Ashley-Koch A., Gilbert J. R., DeLong R. G., Cuccaro M. L. & Pericak-Vance M. A. (2002) Genomic screen and follow-up analysis for autistic disorder. *Am J Med Genet* **114**, 99–105.

Steffenburg S., Gillberg C., Hellgren L., Andersson L., Gillberg I. C., Jakobsson G. & Bohman M. (1989) A twin study of autism in Denmark, Finland, Iceland, Norway, and Sweden. *J Child Psychol Psychiatry* **30**, 405–416.

Templeton A. R. (2000) Epistasis and complex traits. In: *Epistasis and the Evolutionary Process* (eds. M. Wade, N. Brodie & J. Wolf), pp. 41–57, Oxford University Press.

Thomson G. (1995) Mapping disease genes: family–based association studies. *Am J Hum Genet* **57**, 487–498.

Tuchman R. & Rapin I. (2002) Epilepsy in autism. *Lancet Neurol* **1**, 352–358.

Vance J. M. (1998) The Collection of Biological Samples for DNA Analysis. In: *Approaches to Gene Mapping in Complex Human Diseases* (eds.J. L. Haines & M. A. Pericak-Vance), pp. 201–211 New York: Wiley-Liss.

Wallace R. H., Marini C., Petrou S., Harkin L. A., Bowser D. N., Panchal R. G., Williams D. A., Sutherland G. R., Mulley J. C., Scheffer I. E. & Berkovic S. F. (2001) Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures. *Nat Genet* **28**, 49–52.

Wolpert C. M., Menold M. M., Bass M. P., Qumsiyeh M. B., Donnelly S. L., Ravan S. A., Vance J. M., Gilbert J. R., Abramson R. K., Wright H. H., Cuccaro M. L. & Pericak-Vance M. A. (2000) Three probands with autistic disorder and isodicentric chromosome 15. *Am J Med Genet* **96**, 365–372.

Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. 1, 356–366. 1932. Proceedings of the 6th International Congress of Genetics.

Zaykin D., Zhivotovsky L. & Weir B. S. (1995) Exact tests for association between alleles at arbitrary numbers of loci. *Genetica* **96**, 169–178.