Investigation of Maternal Genotype Effects in Autism by Genome-Wide Association

Han Yuan and Joseph D. Dougherty

Like most psychiatric disorders, autism spectrum disorders have both a genetic and an environmental component. While previous studies have clearly demonstrated the contribution of in utero (prenatal) environment on autism risk, most of them focused on transient environmental factors. Based on a recent sibling study, we hypothesized that environmental factors could also come from the maternal genome, which would result in persistent effects across siblings.

In this study, the possibility of maternal genotype effects was examined by looking for common variants (single-nucleotide polymorphisms or SNPs) in the maternal genome associated with increased risk of autism in children. A case/control genome-wide association study was performed using mothers of probands as cases, and either fathers of probands or normal females as controls. Autism Genetic Resource Exchange and Illumina Genotype Control Database were used as our discovery cohort ($n = 1616$). The same analysis was then replicated on Simon Simplex Collection and Study of Addiction: Genetics and Environment datasets ($n = 2732$).

We did not identify any SNP that reached genome-wide significance ($P < 10^{-8}$), and thus a common variant of large effect is unlikely. However, there was evidence for the possibility of a large number of alleles of effective size marginally below our power to detect. ***Autism Res*** *2014, 7: 245–253.* © 2014 International Society for Autism Research, Wiley Periodicals, Inc.

Keywords: maternal genotype effect; autism; GWAS; AGRE; SSC

Introduction

Epidemiological family and twin studies have determined that many psychiatric disorders have both a genetic and an environmental component [Shih, Belmonte, & Zandi, 2004]. Autism spectrum disorders (ASDs) are pervasive developmental disorders, affecting one out of every 88 children aged 8 years in the United States [Autism and Developmental Disabilities Monitoring Network Surveillance Year Principal Investigators and Centers for Disease Control and Prevention, 2012]. They are characterized by deficits in social interaction, restricted interests, and resistance to change [Moldin & Rubenstein, 2006]. Previous studies on recurrence rates of monozygotic twins versus dizygotic twins indicated very high heritability for ASDs [Ronald & Hoekstra, 2011; Schaaf & Zoghbi, 2011]. However, in 70% of ASDs cases, the underlying genetic cause cannot be identified [Schaaf & Zoghbi, 2011]. This suggests a more complex relationship between genetic factors and environmental factors in contributing to the risk of the disorder. Environmental factors in autism include prenatal, perinatal, and postnatal environments. They have been investigated in a number of studies [Folstein & Rosen-Sheidley, 2001; Gardener, Spiegelman, & Buka, 2009], and recent work suggested that they might have a more important role than previously considered [Hallmayer et al., 2011].

The effects of prenatal environment on autism risk have been suggested by many studies. One line of evidence comes from family and twin studies: even though both dizygotic twins and full siblings share half of their genome, the recurrence rate in dizygotic twins is around 30% [Hallmayer et al., 2011; Rosenberg et al., 2009], whereas in nontwin siblings it is 3–14% [Bolton et al., 1994; Constantino, Zhang, Frazier, Abbacchi, & Law, 2010; Sumi, Taniai, Miyachi, & Tanemura, 2006]. This increased risk in dizygotic twins suggests the possibility of the contribution of transient factors in the prenatal environment to the risk of this disorder. Studies also indicated an association between autism with maternal hypothyroidism [Gillberg, Gillberg, & Kopp, 1992; Haddow et al., 1999], maternal thalidomide use [Miller & Strömland, 1993; Stromland, Nordin, Miller, Akerstrom, & Gillberg, 1994], maternal valproic acid use [Christianson, Chesler, & Kromberg, 1994; Williams & Hersh, 1997], maternal alcohol use [Aronson, Hagberg, & Gillberg, 1997; Nanson, 1992], and maternal viral infection [Stubbs, 1978; Stubbs, Ash, & Williams, 1984]. Association between maternal viral infection and autism was further supported by recent studies using a mouse model, which showed that

From the Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA (H.Y.); Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA (J.D.D.)

Received May 17, 2013; accepted for publication January 18, 2014

Address for correspondence and reprints: Joseph Dougherty, Department of Genetics, Washington University School of Medicine, 4566 Scott Avenue, Campus Box 8232, St. Louis, MO 63110. E-mail: jdougherty@genetics.wustl.edu

Published online 25 February 2014 in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.1363

maternal viral infection can cause behavioral and neuropathological symptoms in offspring [Hsiao, McBride, Chow, Mazmanian, & Patterson, 2012]. Thus, it is clear that prenatal environment contributes to autism risk.

The factors we discussed above have transient influence on the prenatal environment, but it is also possible that some prenatal environmental factors could be a product of the mother's genome (which has permanent influence on prenatal environment): we hypothesize that some genetic variants carried by the mother could increase risk of autism in her children through alteration of the prenatal environment (hypothetical examples could be polymorphisms influencing thyroid levels, uterine nutrient levels, or immune system function). This hypothesis is supported by a recent half-sibling study [Constantino et al., 2012]. The study showed that recurrence rate in maternal half siblings, who would have been exposed to the same maternal genome, was 5.2% ($n = 619$) while the paternal half-sibling recurrence rate was 0% ($n = 55$). Even though the sample size of paternal half sibling was very small, this suggested ($P < .07$, Fisher's exact test) the possibility of some influence on autism risk derived from the mother that persists across multiple siblings.

The mother's genome influences development of the offspring through mechanisms such as maternal genotype effects, maternal–fetal genotype interactions, and parent-of-origin effects (imprinting) [Ainsworth, Unwin, Jamison, & Cordell, 2011]. In fact, a variety of diseases are hypothesized to have contributions from such mechanisms. For example, maternal genotype effects have been implicated in risk of spina bifida in offspring [Jensen, Etheredge, Brown, Mitchell, & Whitehead, 2006]; human leukocyte antigen (HLA) maternal–fetal genotype matching has been associated with low birth weight [Larizza et al., 2002; Ober et al., 1987] and schizophrenia [Palmer et al., 2006]; imprinting effects have been implicated in children affected with congenital toxoplasmosis [Jamieson et al., 2008]. Previous authors have referred to such alleles embedded in the maternal genome and increasing the risk of disease in the children as "teratogenic alleles" [Johnson, 2003]. They identified two possible teratogenic alleles in mothers of autism probands, GSTP*A [Williams et al., 2007] and HLA-DR4 [Johnson et al., 2009]. In this paper, we performed a genome-wide search in mothers for such alleles.

We propose and test here a direct approach to examine the contribution of the maternal genotype effects to autism risk using existing data sources: we reexamined the Autism Genetic Resource Exchange (AGRE) [Geschwind et al., 2001] dataset by conducting a case–control genome-wide association analysis treating mothers of probands as cases and either fathers of probands, or females from an independent dataset, as controls based on a logistic model. Then we repeated this analysis with the Simons Simplex Collection (SSC) [Fischbach & Lord, 2010].

Methods

Datasets

Discovery stage. Case samples in the discovery stage came from the AGRE. Parents of individuals scored as autistic based on the Autism Diagnostic Interview-Revised were selected from AGRE dataset (mother, $n = 572$; father, $n = 547$). For multigenerational pedigrees, one mother–father pair was randomly selected. Self-reported Caucasians were selected from the Illumina Genotype Control Database (iCon) [Illumina iControlDB, 2013] as controls. Caucasian females were selected from the iCon ($n = 508$). Individuals from both datasets were previously genotyped on the Illumina HumanHap 550 platform for 550,000 single-nucleotide polymorphism (SNPs).

Replication stage. Samples in the replication stage were obtained from the SSC and Study of Addiction: Genetics and Environment (SAGE), a control dataset from an addiction study. In the SSC dataset, individuals were genotyped on either Illumina 1Mv1 or Illumina 1Mv3 Duo Bead array [Sanders et al., 2011]; in the SAGE dataset, genotyping was performed using Illumina Human 1Mv1_C Beadchips [Bierut et al., 2010]. Individuals from both iCon and SAGE datasets were selected to match the ethnicity of case datasets (mostly Caucasian). While they were not ascertained for autism, we assumed that the prevalence in these two datasets would be no higher than that in the general population (~1%). Two thousand one hundred seventy-seven individuals were collected from the SSC dataset (mother, $n = 1089$; father, $n = 1088$) and 2024 females were collected from the SAGE dataset. Individuals from both datasets were genotyped on the Illumina Human 1M platform.

Quality Control

In each study stage, case and control samples were merged and quality control was performed using PLINK [Purcell et al., 2007].

Individuals matching the following criteria were excluded:

1. Individuals with genotyping rate <0.1.
2. Individuals whose reported sex did not match with predicted sex (based on chromosome X data).

SNPs matching the following criteria were excluded:

1. SNPs with a maximum per-SNP missing >0.1.
2. SNPs with minor allele frequency <0.05.
3. SNPs that violated Hardy–Weinberg disequilibrium (P-value <0.001).

Sample sizes and number of SNPs in each stage that survived quality control were listed in (Table 1). Even though samples of the replication stage were genotyped on the Illumina Human 1M platform and we expect about

Table 1. Sample Size and Single-Nucleotide Polymorphism (SNP) Number that Survived Quality Control

N	Case mother	Case father	Control female	SNPs
Discovery stage	561	547	508	503,972
Replicate stage	982	1,007	743	583,160

Note. After quality control, there are 561 mothers of probands, 547 fathers of probands, and 508 control females with 503 972 SNPs remaining in the discovery stage, 982 mothers of probands, 1007 fathers of probands, and 743 control females with 583 160 SNPs remaining in the replication stage.

one million SNPs, there were 893,308 SNPs overlapping between SSC dataset and SAGE dataset because of slight differences in chip version. Only about 583,160 SNPs survived the merging and quality control in the replication stage.

Principal Component Analysis

Because population substructure confounds association study, principal component analysis (PCA) was utilized to account for this in our analysis. PCA reduces the genetic variation of samples to a small number of principal components, describing as much variation as possible. In the discovery stage, PCA was done using EIGENSTRAT [Price et al., 2006] with 10 principal components based on about 50,000 independent SNPs (selected to not be in linkage disequilibrium, LD) extracted from the sample. In most cases, each principal component has a geographical or ethnic interpretation. For example, in (Fig. 1A), we plotted the distribution of samples on the first versus second principal components. It is clear that the first principal component separated Caucasians from African-Americans, and the second principal component separated Caucasians from Asians. Thus, PCA successfully accounted for the population substructure in our samples. Because the first five principal components explained most of the variation (Fig. 1B), these were taken as covariates in the association tests to account for population substructure. The same type of PCA was also done for the replication stage samples, and the first five principal components were again used as covariates for the replication association studies.

Genotype Imputation

In order to increase the coverage of the genome, SNP imputations were performed in both discovery and replication stages. Because both our discovery and replication stage samples are composed of various racial groups, imputation of individuals' genotypes in both stages was done with 460 HapMap haplotypes including different racial groups (CEU + YRI + CHB + JPT) from HapMap phase 2 release 22 using MACH [Li, Willer, Ding, Scheet, & Abecasis, 2010]. Using a combination of CEU + YRI + CHB + JPT haplotype panels also have performance advantage at rare SNPs compared with using single haplotype panel [Marchini & Howie, 2010]. Following the criteria described in Wang et al.'s [2009] paper, we removed markers with Rsq <0.3 or posterior probability <0.9. After another round of quality control (same as above), a total of 1,776,310 high-quality SNPs remained in discovery stage, and 1,619,647 SNPs remained in replicate stage.

Association Test Design

A two-step association analysis was conducted on samples of the discovery stage (Fig. 2) and then repeated in the replication stage. Association testing was implemented using PLINK based on a logistic model using the first five principal components from PCA analysis as covariates. In the first association test, AGRE mothers were treated as cases, and AGRE fathers as controls. This test identifies associations from three sources:

1. Alleles more prevalent in mothers of probands (maternal genotype effects).
2. Alleles more prevalent in fathers of probands (paternal genotype effects).
3. Alleles with apparent differences between sexes (e.g., three "SNPs" [e.g., rs12734338] on chromosome 1, in a region homologous to chromosome Y, appear to be 100% "heterozygous" for male and 100% "homozygous" for female as previously reported [Boraska et al., 2012]).

To identify only those alleles due to maternal genotype effects, a second association test was conducted using the same cases but treating females from iCon as controls. This second test detects any maternal effect alleles, but can also be confounded by batch effects: systematic technical differences between samples genotyped at different times or by different groups. Focusing on only those nominally significant SNPs ($P < 0.05$) in both tests should remove both batch and sex effects as SNPs associated with a maternal contribution to autism risk should be significant in both tests. We set the threshold for nominally significant SNPs at the level of 0.05 for the initial data exploration. At the end of the analysis, we looked for hits with a combined *P*-value of 10e-8.

Results

In the discovery stage, 14,454 SNPs were nominally significant ($P < 0.05$) in both AGRE mothers versus fathers and AGRE mothers versus iCon females (Fig. 3, Table 2). The most significant SNP with high concordance across

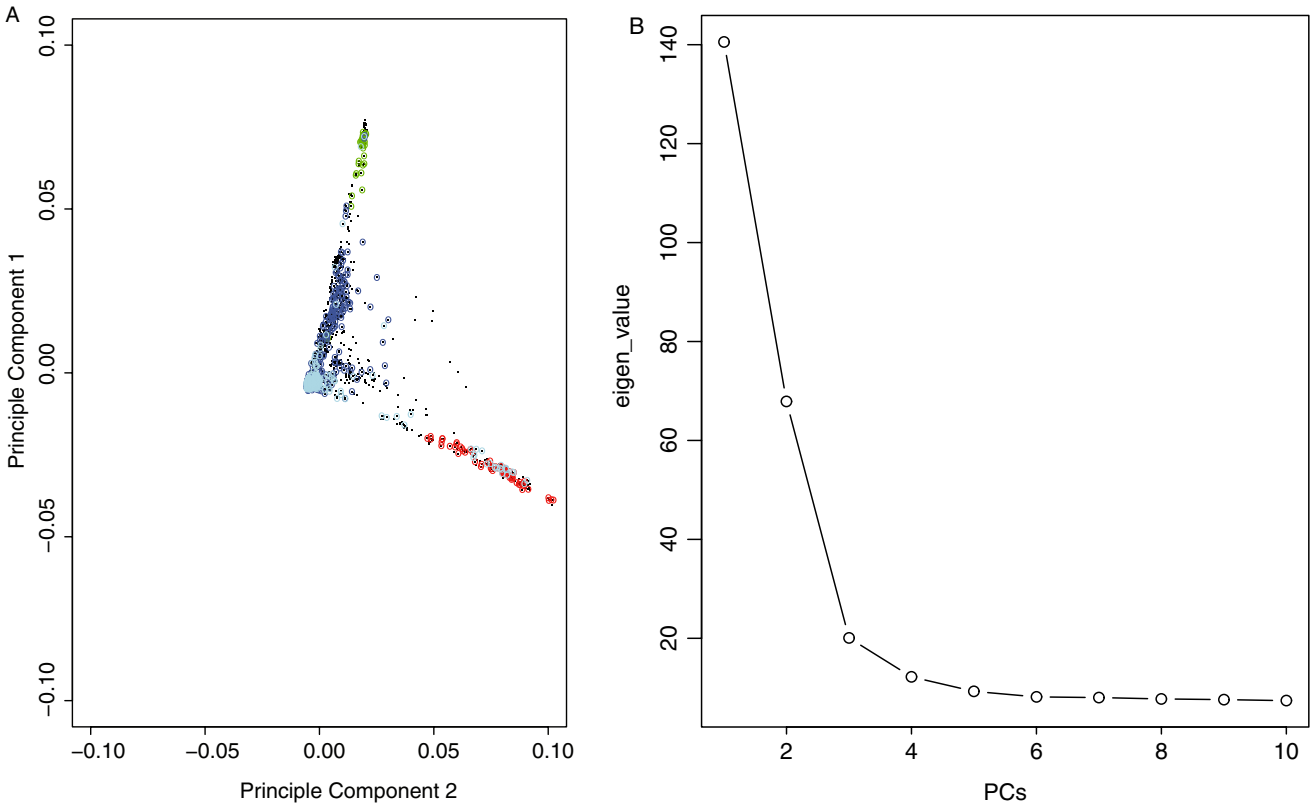

Figure 1. (A) Sample distribution on the first versus second principal components. Figure 1A shows the distribution of samples in discovery stage on the first versus second principal components. Green dots are self-reported Asians in Autism Genetic Resource Exchange (AGRE) dataset; red dots are self-reported African-Americans in AGRE dataset; dark blue dots are self-reported Caucasians in AGRE dataset; and light blue are Caucasians in Illumina Genotype Control Database (iCon) dataset. From the figure, we can see that the first principal component separates African-American from Caucasians and Asians, and the second principal component separates Asians from Caucasians and African-Americans. (B) Screeplot from principal component analysis (PCA) analysis for samples in discovery stage. Figure 1B shows the screeplot for PCA analysis for samples in discovery stage. The *x*-axis is principal component number, and the *y*-axis is the eigenvalue, which is directly proportional to proportion of variation. From this figure, we can see that the first principal component explains the largest proportion of variation and the proportion of variation explained by each subsequent principal component decreases.

both discovery analyses was rs12431425, located in the intergenic region on 14q21.1 ($P = 7.8E-06$, odds ratio [OR] = 0.516 [95% confidence interval: 0.386–0.689] for AGRE mother vs. father, $P = 4.95E-05$, OR = 0.532 [0.393–0.717] for AGRE mother vs. iCon female). The single most significant association was rs2473147 ($P = 0.01474$, OR = 0.741 [0.582–0.942] for AGRE mother vs. father, $P = 1.82E-07$, OR = 0.506 [0.393–0.650] for AGRE mother vs. iCon female), located in the intronic region of utrophin (*UTRN*) on chromosome 6. Figure 3 shows the Manhattan plots for both tests, aligned together by genomic coordinates. As estimated by the CaTS power calculator [Skol, Scott, Abecasis, & Boehnke, 2006], with our sample size, we had 80% power to detect a susceptibility variant with an allele frequency >0.05 and genotype relative risk >2.63. Because we did not find any common variant that reaches genome-wide significance ($P < 10^{-8}$), a common variant with large effect in the maternal genome is unlikely. However, power analysis suggested that we had less than 11% power to detect common variant (disease allele frequency >0.05) of relatively small effect (genotype relative risk <2). Note that even in autism probands, no alleles with a genotype relative risk of <1.2 have been replicated as genome-wide significant by association testing [Anney et al., 2012].

The two best hits described above (rs12431425 and rs2473147) were not replicated as nominally significant in the replication stage. Association testing of all 1,619,647 SNPs did identify 13,440 that were nominally significant both when SSC mothers were compared with SSC fathers and when they were compared with females from the SAGE collection, but none reaching genome-wide significance. Three hundred seventy-three out of the nominally significant 14,454 SNPs from the discovery phase had their significance level raised when combining discovery with replicate stages (by merging the two datasets and then performing genome-wide association study on the combined dataset). The most significant SNP

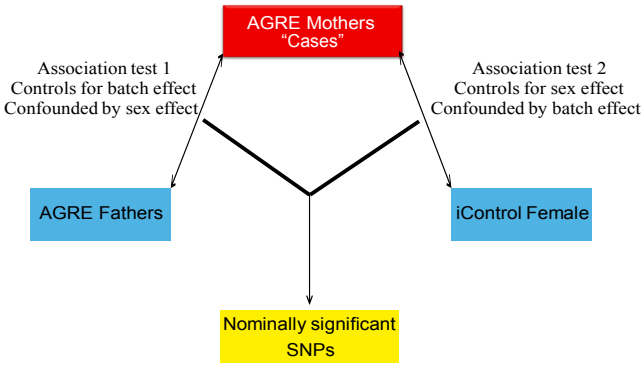

Figure 2. Two-step design for association analysis. The diagram illustrates our two-step association design to permit the identification of maternal genetic effects on risk from existing datasets. Intersecting the results of the two association tests controls for potential confounds from either sex effect or batch effects. The *P*-value thresholds are at 0.05 to discover nominally significant single-nucleotide polymorphisms (SNPs) for the Autism Genetic Resource Exchange (AGRE) mother versus father test and the AGRE mother versus Illumina Genotype Control Database (iCon) female test.

from the joint analysis was rs2670494 (combined *P*-value for proband mother vs. father test is 5.02E-05, combined *P*-value for proband mother vs. control female test is 8.95E-05), mapped to the intronic region of spermatogenesis associated 21 (*SPATA21*). However, SSC dataset was ascertained to contain simplex families, and a substantial fraction of the autism probands in this cohort carry significant de novo mutations [Iossifov et al., 2012; Sanders et al., 2011, 2012]. Thus, it was less likely that we would find significant association because of maternal genotype effects in the replication stage.

Discussion

To our knowledge, this is the first genome-wide search for maternal genotype effects in autism. Because we were using datasets that were not collected primarily for this purpose, we took a very conservative approach to our analysis. To eliminate any confounding sex or batch effects, we required alleles from the mothers to reach a nominally significant *P*-value in two separate comparisons. With these filters, we did not identify any SNP that reached genome-wide significance, and we did not replicate earlier teratogenic alleles [Johnson et al., 2009; Williams et al., 2007], although we had limited power to detect rare polymorphisms or polymorphisms with smaller effect sizes (OR < 2).However, we did identify some interesting candidate genes near our most significant SNPS such as PDE11A. PDE11A is expressed in the adrenal cortex and catalyzes the hydrolysis of cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP) [Carney, Gaillard, Bertherat, & Stratakis, 2010]. Mutation of this gene is suggested to affect the size of adrenal gland and levels of multiple hormones such as cortisol [Carney et al., 2010; Ceyhan, Birsoy, & Hoffman, 2012]. Because adrenal gland is the key for stress response and regulates both hormone level and immune system, mutation in PDE11A in mothers could affect children's risk of autism through mechanisms such as hormone regulation or immune response.

Here, we used a logarithmic model for association testing implemented in PLINK because it readily permitted the inclusion of population stratification as a covariate. Using this approach, we only explored maternal genotype effects while in fact the maternal genome could also influence risk of autism in offspring through maternal–fetal genotype interactions and imprinting. For example, HLA maternal–fetal genotype matching has been associated with schizophrenia in a previous study [Palmer et al., 2006]. Although it can be difficult to distinguish each of these three possibilities statistically [Ainsworth et al., 2011], there are elegant linear models for fitting each [Howey & Cordell, 2012]. However, they currently do not implement the inclusion of covariates for coping with population stratification, which is clearly present in our datasets. It is also worth noting that using PCA to exclude non-Caucasians, rather than to generate covariates, still did not uncover genome-wide significant alleles, although our sample size, and thus our power, was substantially lower with that approach (not shown). For our 10 most significant regions (Table 1), we used Cordell's linear model, EMIM [Howey & Cordell, 2012], to determine if our findings are consistent with a maternal genotype effect (Supporting Information Table S2). This analysis indicates that any risk attributable to these alleles is most consistent with a maternal genotype effect rather than child genotype effect or imprinting effect.

Overall, we did not discover any polymorphisms in the mother's genome that reached genome-wide significance. However, it is worth noting that the three largest common variant studies to date on the children themselves also did not replicate each other for any common variants reaching genome-wide significance [Anney et al., 2010; Wang et al., 2009; Weiss, Arking, Daly, & Chakravarti, 2009]. They did, however, identify an excess of independent regions with $P < 10E\text{-}5$ and $P < 10E\text{-}4$, suggesting that there may be a large number of alleles of effective size marginally below their power to detect [Weiss et al., 2009]. Likewise, we found in our set 14 SNPs with $P < 10E\text{-}6$ (8 by expectation), in AGRE mother versus iCon female test, suggesting that alleles in the mother's genome may collectively make an important contribution. We do not believe that these SNPs represent a batch effect because none of these SNPs reaches 10E-4 in an AGRE father versus iCon male test. It is worth noting that no regions identified from the studies of the children

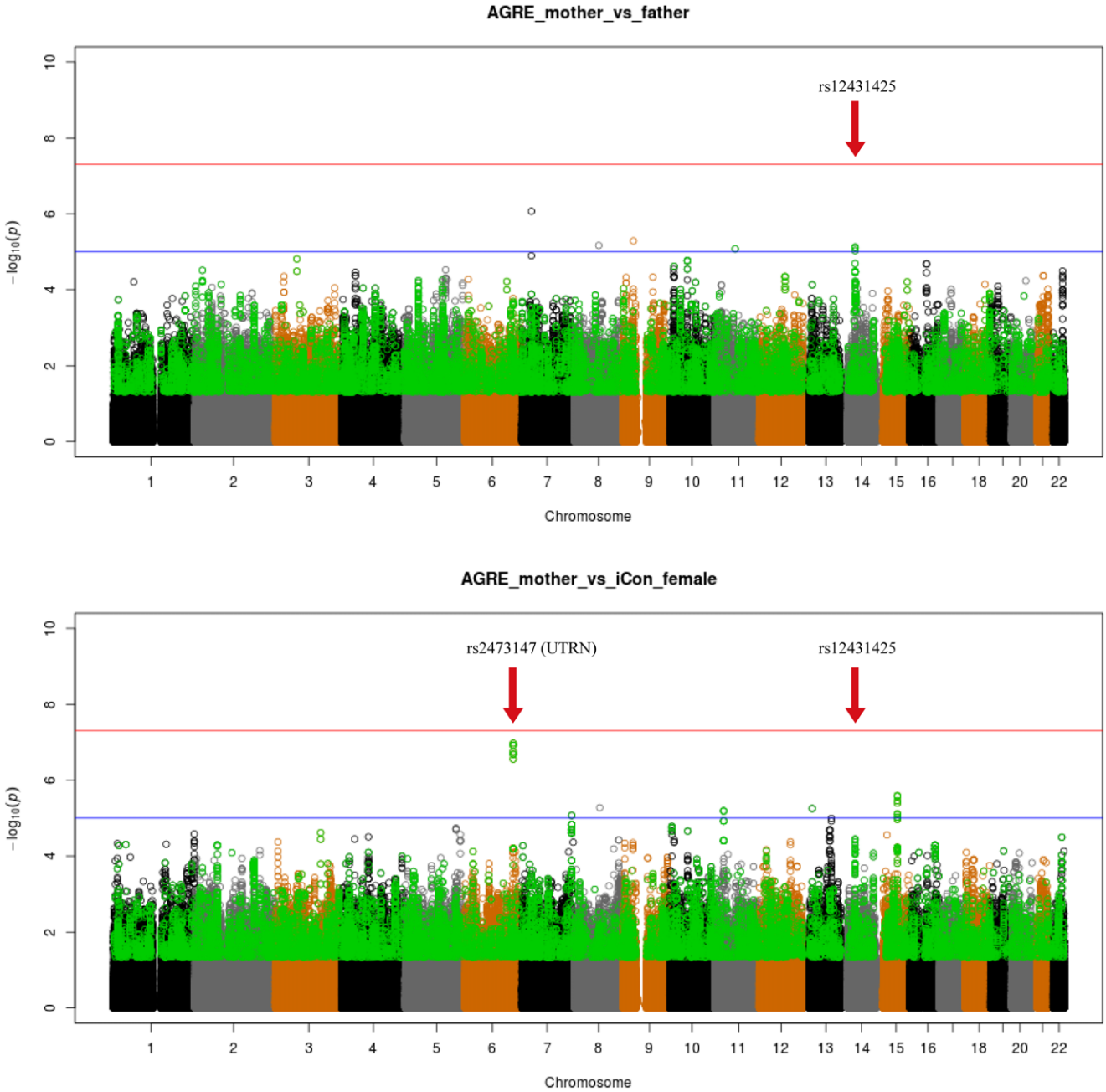

Figure 3. Association tests for samples in discovery stage. The top panel shows the Manhattan plot for Autism Genetic Resource Exchange (AGRE) mother versus father test. The bottom panel shows the Manhattan plot for AGRE mother versus Illumina Genotype Control Database (iCon) female test. The two plots are aligned together by genomic coordinate (*x*-axis). The *y*-axis is the negative logarithm of the association *P*-value for each single-nucleotide polymorphism (SNP). Green dots indicate SNPs that are nominally significant ($P < 0.05$) in both tests. The most significant SNPs, rs2473147 and rs12431425, are pointed out by red arrows.

were among our top candidates, suggesting that alleles highlighted here may have their influence on risk in the children whether or not they are transmitted, consistent with a maternal genetic effect. Thus, based on our analysis, a polygenic model maybe more informative in future investigation because of the overabundance of alleles of small effective size. Likewise, our analysis would not have detected rare alleles that contribute to risk through a maternal genetic mechanism.

Acknowledgments

We gratefully thank John Constantino, Jennifer K. Lowe, Don Conrad, Laura Bierut, Nancy Saccone, and members

Table 2. Association Results for Samples in Discovery Stage

SNP	Region	Gene	*P*-value (discovery)		*P*-value (replication)		R^2 discovery	R^2 replicate
			Test_1	Test_2	Test_3	Test_4		
rs12431425	14q21.1	none	7.81E-06	4.95E-05	0.9228	0.4513	0.992	0.9925
rs475916	13q12.2	POLR1D	7.62E-05	8.25E-06	0.6229	0.6188	Genotyped	Genotyped
rs10174604	2q31.2	PDE11A	0.000269	0.000263	0.4631	0.3713	0.999	0.9982
rs11762727	7q26.1	GIMAP6	0.000357	2.80E-05	0.5532	0.5665	0.976	Genotyped
rs4750249	10p13	CAMK1D	0.000213	0.000365	0.3785	0.374	0.942	Genotyped
rs1456820	4q13.1	none	9.33E-05	0.00051	0.3085	0.9204	0.914	0.9186
rs3850729	5q23.1	none	0.000573	0.000209	0.1618	0.5682	0.952	0.9421
rs1978656	2p14	none	0.000579	0.000121	0.6359	0.9327	Genotyped	Genotyped
rs12774114	10q26.13	none	0.000586	0.000387	0.1535	0.4506	0.981	0.9157
rs2473147	6q24.2	UTRN	0.01474	1.82E-07	0.107	0.9045	0.989	0.9424

Note. The table shows the most significant single-nucleotide polymorphisms (SNPs) in the discovery stage, plus rs2473147, the SNP with single most significant association. Test_1 is Autism Genetic Resource Exchange (AGRE) mother versus father test and Test_2 is AGRE mother versus iCon female test. As expected, we detected large blocks of linkage disequilibrium surrounding our most significant SNPs (Fig. 3). Here, we only reported the most significant SNP from each block. Test_3 and Test_4 are the *P*-values of the replicated tests in SSC. Test_3 is SSC mother versus father test and Test_4 is SSC mother versus Study of Addiction: Genetics and Environment (SAGE) female test.

of Dougherty lab for their suggestions and support. Funding was provided by the Mallinkrodt Foundation (J. D. D.) and the National Institutes of Health (NIH): 4R00NS067239 -03 (J. D. D.) and 9R01MH100027-06. The authors declare no conflicts of interest.

We are also grateful for the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium* and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to Clara M. Lajonchere (PI). We are likewise grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, and E. Wijsman). Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative (GEI) (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C).

References

Ainsworth, H.F., Unwin, J., Jamison, D.L., & Cordell, H.J. (2011). Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. Genetic Epidemiology, 35, 19–45. doi:10.1002/gepi.20547

Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Human Molecular Genetics, 21, 4781–4792. doi:10.1093/hmg/dds301

Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. Human Molecular Genetics, 19, 4072–4082. doi:10.1093/hmg/ddq307

Aronson, M., Hagberg, B., & Gillberg, C. (1997). Attention deficits and autistic spectrum problems in children exposed to alcohol during gestation: A follow-up study. Developmental Medicine and Child Neurology, 39, 583–587.

Autism and Developmental Disabilities Monitoring Network Surveillance Year Principal Investigators and Centers for Disease Control and Prevention (2012). Prevalence of autism spectrum disorders–Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. MMWR. Surveillance Summaries: Morbidity and Mortality Weekly Report. Surveillance Summaries, 61, 1–19.

Bierut, L.J., Agrawal, A., Bucholz, K.K., Doheny, K.F., Laurie, C., et al. (2010). A genome-wide association study of alcohol dependence. Proceedings of the National Academy of Sciences of the United States of America, 107, 5082–5087. doi:10.1073/pnas.0911109107

Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., et al. (1994). A case-control family history study of autism. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 35, 877–900.

Boraska, V., Jeroncic, A., Colonna, V., Southam, L., Nyholt, D.R., et al. (2012). Genome-wide meta-analysis of common variant differences between men and women. Human Molecular Genetics, 21, 4805–4815. doi:10.1093/Hmg/Dds304

Carney, J.A., Gaillard, R.C., Bertherat, J., & Stratakis, C.A. (2010). Familial micronodular adrenocortical disease, Cushing syndrome, and mutations of the gene encoding phosphodiesterase 11A4 (PDE11A). The American Journal of Surgical Pathology, 34, 547–555. doi:10.1097/PAS.0b013e3181d31f49

Ceyhan, O., Birsoy, K., & Hoffman, C.S. (2012). Identification of biologically active PDE11-selective inhibitors using a yeast-based high-throughput screen. Chemistry and Biology, 19, 155–163. doi:10.1016/j.chembiol.2011.12.010

Christianson, A.L., Chesler, N., & Kromberg, J.G. (1994). Fetal valproate syndrome: Clinical and neuro-developmental features in two sibling pairs. Developmental Medicine and Child Neurology, 36, 361–369.

Constantino, J.N., Todorov, A., Hilton, C., Law, P., Zhang, Y., et al. (2012). Autism recurrence in half siblings: Strong support for genetic mechanisms of transmission in ASD. Molecular Psychiatry, 18, 137–138. doi:10.1038/mp.2012.9

Constantino, J.N., Zhang, Y., Frazier, T., Abbacchi, A.M., & Law, P. (2010). Sibling recurrence and the genetic epidemiology of autism. The American Journal of Psychiatry, 167, 1349–1356. doi:10.1176/appi.ajp.2010.09101470

Fischbach, G.D., & Lord, C. (2010). The Simons simplex collection: A resource for identification of autism genetic risk factors. Neuron, 68, 192–195. doi:10.1016/j.neuron.2010.10.006

Folstein, S.E., & Rosen-Sheidley, B. (2001). Genetics of autism: Complex aetiology for a heterogeneous disorder. Nature Reviews. Genetics, 2, 943–955. doi:10.1038/35103559

Gardener, H., Spiegelman, D., & Buka, S.L. (2009). Prenatal risk factors for autism: Comprehensive meta-analysis. The British Journal of Psychiatry: The Journal of Mental Science, 195, 7–14. doi:10.1192/bjp.bp.108.051672

Geschwind, D.H., Sowinski, J., Lord, C., Iversen, P., Shestack, J., et al. (2001). The autism genetic resource exchange: A resource for the study of autism and related neuropsychiatric conditions. American Journal of Human Genetics, 69, 463–466. doi:10.1086/321292

Gillberg, I.C., Gillberg, C., & Kopp, S. (1992). Hypothyroidism and autism spectrum disorders. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 33, 531–542.

Haddow, J.E., Palomaki, G.E., Allan, W.C., Williams, J.R., Knight, G.J., et al. (1999). Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. The New England Journal of Medicine, 341, 549–555. doi:10.1056/NEJM199908193410801

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., et al. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. Archives of General Psychiatry, 68, 1095–1102.

Howey, R., & Cordell, H.J. (2012). PREMIM and EMIM: Tools for estimation of maternal, imprinting and interaction effects using multinomial modelling. BMC Bioinformatics, 13, 149. doi:10.1186/1471-2105-13-149

Hsiao, E.Y., McBride, S.W., Chow, J., Mazmanian, S.K., & Patterson, P.H. (2012). Modeling an autism risk factor in mice leads to permanent immune dysregulation. Proceedings of the National Academy of Sciences of the United States of America, 109, 12776–12781. doi:10.1073/pnas.1202556109

Illumina iControlDB (2013). http://www.illumina.com/science/icontroldb.ilmn

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., et al. (2012). De novo gene disruptions in children on the autistic spectrum. Neuron, 74, 285–299. doi:10.1016/j.neuron.2012.04.009

Jamieson, S.E., de Roubaix, L.-A., Cortina-Borja, M., Tan, H.K., Mui, E.J., et al. (2008). Genetic and epigenetic factors at COL2A1 and ABCA4 influence clinical outcome in congenital toxoplasmosis. PLoS ONE, 3, 6.

Jensen, L.E., Etheredge, A.J., Brown, K.S., Mitchell, L.E., & Whitehead, A.S. (2006). Maternal genotype for the monocyte chemoattractant protein 1 A(-2518)G promoter polymorphism is associated with the risk of spina bifida in offspring. American Journal of Medical Genetics. Part A, 140, 1114–1118. doi:10.1002/ajmg.a.31212

Johnson, W.G. (2003). Teratogenic alleles and neurodevelopmental disorders. Bioessays: News and Reviews in Molecular, Cellular and Developmental Biology, 25, 464–477. doi:10.1002/bies.10268

Johnson, W.G., Buyske, S., Mars, A.E., Sreenath, M., Stenroos, E.S., et al. (2009). HLA-DR4 as a risk allele for autism acting in mothers of probands possibly during pregnancy. Archives of Pediatrics and Adolescent Medicine, 163, 542–546.

Larizza, D., Martinetti, M., Dugoujon, J.M., Tinelli, C., Calcaterra, V., et al. (2002). Parental GM and HLA genotypes and reduced birth weight in patients with Turner's syndrome. Journal of Pediatric Endocrinology and Metabolism, 15, 1183–1190.

Li, Y., Willer, C.J., Ding, J., Scheet, P., & Abecasis, G.R. (2010). MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic Epidemiology, 34, 816–834. doi:10.1002/gepi.20533

Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. Nature Reviews. Genetics, 11, 499–511. doi:10.1038/nrg2796

Miller, M.S., & Strömland, K. (1993). Thalidomide embryopathy: An insight into autism?

Moldin, S.O., & Rubenstein, J.L.R. (2006). Understanding autism from basic neuroscience to treatment. Boca Raton, FL: CRC/Taylor & Frances.

Nanson, J.L. (1992). Autism in fetal alcohol syndrome: A report of six cases. Alcoholism, Clinical and Experimental Research, 16, 558–565.

Ober, C., Simpson, J.L., Ward, M., Radvany, R.M., Andersen, R., et al. (1987). Prenatal effects of maternal-fetal HLA compatibility. American Journal of Reproductive Immunology and Microbiology, 15, 141–149.

Palmer, C.G., Hsieh, H.J., Reed, E.F., Lonnqvist, J., Peltonen, L., et al. (2006). HLA-B maternal-fetal genotype matching increases risk of schizophrenia. American Journal of Human Genetics, 79, 710–715. doi:10.1086/507829

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38, 904–909. doi:10.1038/ng1847

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81, 559–575. doi:10.1086/519795

Ronald, A., & Hoekstra, R.A. (2011). Autism spectrum disorders and autistic traits: A decade of new twin studies. American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics, 156B, 255–274. doi:10.1002/Ajmg.B.31159

Rosenberg, R.E., Law, J.K., Yenokyan, G., McGready, J., Kaufmann, W.E., & Law, P.A. (2009). Characteristics and concordance of autism spectrum disorders among 277 twin pairs. Archives of Pediatrics and Adolescent Medicine, 163, 907–914. doi:10.1001/archpediatrics.2009.98

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron, 70, 863–885. doi:10.1016/j.neuron.2011.05.002

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature, 485, 237–241. doi:10.1038/nature10945

Schaaf, C.P., & Zoghbi, H.Y. (2011). Solving the autism puzzle a few pieces at a time. Neuron, 70, 806–808. doi:10.1016/j.neuron.2011.05.025

Shih, R.A., Belmonte, P.L., & Zandi, P.P. (2004). A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. International Review of Psychiatry, 16, 260–283. doi:10.1080/09540260400014401

Skol, A.D., Scott, L.J., Abecasis, G.R., & Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies (vol 38, pg 209, 2006). Nature Genetics, 38, 390–390. doi:10.1038/Ng0306-390

Stromland, K., Nordin, V., Miller, M., Akerstrom, B., & Gillberg, C. (1994). Autism in thalidomide embryopathy: A population study. Developmental Medicine and Child Neurology, 36, 351–356.

Stubbs, E.G. (1978). Autistic symptoms in a child with congenital cytomegalovirus infection. Journal of Autism and Childhood Schizophrenia, 8, 37–43.

Stubbs, E.G., Ash, E., & Williams, C.P. (1984). Autism and congenital cytomegalovirus. Journal of Autism and Developmental Disorders, 14, 183–189.

Sumi, S., Taniai, H., Miyachi, T., & Tanemura, M. (2006). Sibling risk of pervasive developmental disorder estimated by means of an epidemiologic survey in Nagoya, Japan. Journal of Human Genetics, 51, 518–522. doi:10.1007/s10038-006-0392-7

Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature, 459, 528–533. doi:10.1038/nature07999

Weiss, L.A., Arking, D.E., Daly, M.J., & Chakravarti, A. (2009). A genome-wide linkage and association scan reveals novel loci for autism. Nature, 461, 802–808. doi:10.1038/nature08490

Williams, P.G., & Hersh, J.H. (1997). A male with fetal valproate syndrome and autism. Developmental Medicine and Child Neurology, 39, 632–634.

Williams, T.A., Mars, A.E., Buyske, S.G., Stenroos, E.S., Wang, R., et al. (2007). Risk of autistic disorder in affected offspring of mothers with a glutathione S-transferase P1 haplotype. Archives of Pediatrics and Adolescent Medicine, 161, 356–361. doi:10.1001/archpedi.161.4.356

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. *P*-values of top hits after combining discovery and replicate stages. Only single-nucleotide polymorphisms (SNPs) that had their significance level raised after merging were included in this table. For each SNP, we included *P*-values of case mother versus father test in discovery stage (AGREMvF), case mother versus control female test in discovery stage (AGREFvF), case mother versus father test in replicate stage (simonMvF), case mother versus control female test in replicate stage (simonFvF), case mother versus father test in merged dataset (mergeMvF), and case mother versus control female test in merged dataset (mergeFvF).

Table S2. Comparison of significant levels between our study and EMIM result for top 10 hits. We tested the 10 most significant single-nucleotide polymorphisms (SNPs) in Table 1 and used Cordell's linear model (EMIM) to predict their association with autism under child genotype effect model, maternal genotype effect model, and maternal imprinting model, respectively. For SNPs that are not directly genotyped, we looked for genotyped SNPs in LD ($R^2 > 0.9$), with target SNPs and performed EMIM on them. *P*-values for each SNP under each model is reported in the table. pvaluesCG represents *P*-values under child genotype effect model; pvaluesMG represents *P*-values under maternal genotype model; and pvaluesIM represents *P*-values under maternal imprinting model. NA,no SNP in LD available for testing.