

Runs of Homozygosity Associated with Speech Delay in Autism in a Taiwanese Han Population: Evidence for the Recessive Model

Ping-I Lin^{1,2}, Po-Hsiu Kuo³, Chia-Hsiang Chen^{4,5}, Jer-Yuarn Wu^{6,7}, Susan S-F. Gau^{2,3,4,8*}, Yu-Yu Wu⁹, Shih-Kai Liu¹⁰

1 Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States of America, **2** Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan, **3** Graduate Institute of Epidemiology and Preventive Medicine, National Taiwan University College of Public Health, Taipei, Taiwan, **4** Department of Psychiatry, National Taiwan University College of Medicine, Taipei, Taiwan, **5** Center for Neuropsychiatric Research, National Health Research Institutes, Zhunan, Taiwan, **6** Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, **7** School of Chinese Medicine, China Medical University, Taichung, Taiwan, **8** Graduate Institute of Brain and Mind Sciences, Graduate Institute of Clinical Medicine, Department of Psychology, and School of Occupational Therapy, National Taiwan University, Taipei, Taiwan, **9** Department of Psychiatry, Chang Gung Memorial Hospital- Linkou Medical Center, Chang Gung University College of Medicine, Tao-Yuan, Taiwan, **10** Department of Child and Adolescent Psychiatry, Taoyuan Mental Hospital, Department of Health, Executive Yuan, Tao-Yuan, Taiwan

Abstract

Runs of homozygosity (ROH) may play a role in complex diseases. In the current study, we aimed to test if ROHs are linked to the risk of autism and related language impairment. We analyzed 546,080 SNPs in 315 Han Chinese affected with autism and 1,115 controls. ROH was defined as an extended homozygous haplotype spanning at least 500 kb. Relative extended haplotype homozygosity (REHH) for the trait-associated ROH region was calculated to search for the signature of selection sweeps. Totally, we identified 676 ROH regions. An ROH region on 11q22.3 was significantly associated with speech delay (corrected $p = 1.73 \times 10^{-8}$). This region contains the NPAT and ATM genes associated with ataxia telangiectasia characterized by language impairment; the CUL5 (culin 5) gene in the same region may modulate the neuronal migration process related to language functions. These three genes are highly expressed in the cerebellum. No evidence for recent positive selection was detected on the core haplotypes in this region. The same ROH region was also nominally significantly associated with speech delay in another independent sample ($p = 0.037$; combinatorial analysis Stouffer's z trend = 0.0005). Taken together, our findings suggest that extended recessive loci on 11q22.3 may play a role in language impairment in autism. More research is warranted to investigate if these genes influence speech pathology by perturbing cerebellar functions.

Citation: Lin P-I, Kuo P-H, Chen C-H, Wu J-Y, Gau S-S-F, et al. (2013) Runs of Homozygosity Associated with Speech Delay in Autism in a Taiwanese Han Population: Evidence for the Recessive Model. PLoS ONE 8(8): e72056. doi:10.1371/journal.pone.0072056

Editor: Balraj Mittal, Sanjay Gandhi Medical Institute, India

Received: May 2, 2013; **Accepted:** July 5, 2013; **Published:** August 16, 2013

Copyright: © 2013 Lin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was mainly supported by grants from National Science Council (NSC96-3112-B-002-033, NSC97-3112-B-002-009, NSC98-3112-B-002-004, and NSC 99-3112-B-002-036), National Taiwan University (AIM for Top University Excellent Research Project: 10R81918-03, 101R892103, 102R892103), and National Taiwan University Hospital (NTUH101-N2017) and partially supported by Academia Sinica Genomic Medicine Multicenter Study (Academia Sinica 40-05-GMM) and National Center for Genome Medicine at Academia Sinica (NCGM, NSC-101-2319-B-001-001) of National Core Facility Program for Biotechnology (NCFPB) and Translational Resource Center for Genomic Medicine (TRC, NSC-101-2325-B-001-035) of National Research Program for Biopharmaceuticals (NRPB), National Science Council, Taiwan. The authors gratefully acknowledge the resources provided by the AGRE consortium and the participating AGRE families. AGRE 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). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gaushufe@ntu.edu.tw

Introduction

Autistic disorder (henceforth denoted as autism) is a neurodevelopmental disorder characterized by deficits in communication, social interaction, and behavioral patterns. Family and twin studies have strongly suggested that genetic factors contribute to the development of autism [1]. Most genome-wide association studies (GWAS) have investigated the impact of genetic variants on the risk of autism one at a time [2–4]. However, many of these GWAS-derived findings could not be successfully replicated across different populations [5]. The failure to replicate previous findings may be, at least in part, attributed to the negligence of multi-locus effects [6]. To evaluate all possible multi-locus effects in the context of hypothesis-free GWAS, one has to overcome the computational and statistical burden. Some of prior studies have

focused on genes with relevant biological functions to investigate multi-locus effects on the risk of autism [7–9]. Additionally, whole-genome scans also suggest that a cluster of rare variants across different genes may collectively predict the risk of autism [10,11]. Therefore, systemic approaches to investigating the effect of clusters of multiple loci from the whole genome may lead to discoveries that complement the GWAS-derived findings.

Runs of homozygosity (ROHs) may play a role in neuropsychiatric diseases, such as schizophrenia [12,13] and Alzheimer's disease [14]. A recent study also identified several novel candidate genes characterized by ROHs associated with the risk of autism [15]. Compared to the number of SNPs in the whole genome, the number of ROHs is apparently more tractable, and hence requires a less stringent significance threshold to search for significant findings. Therefore, a ROH-based approach may provide

opportunities of revealing multi-locus effects on phenotypes. The link between common ROHs and diseases may reflect several different non-mutually exclusive mechanisms. First, a haplotype at high frequency with high homozygosity spanning over a large region is a sign of an incomplete selective sweep. Under such circumstances, an individual may carry consecutive homozygous SNPs due to identical-by-descent haplotypes that harbor ancestral alleles with an advantageous effect [16]. In case-control studies, an ROH over-represented in cases may be attributed to a disease-linked variant with an advantageous effect, while an ROH over-represented in controls may stem from a protective effect of recent mutation. On the other hand, selection pressure may not purge all deleterious mutations and hence inbreeding might cause the accumulation of multiple variants of adverse effects, which leads to a multi-locus recessive disease model. Alternatively, a disease-associated ROH may arise when a deleterious mutation is in linkage disequilibrium with another variant that undergoes recent positive selection [17]. Second, an ROH over-represented in cases may simply stem from a multi-locus recessive disease model. Third, a disease-associated ROH may indicate the difference in relatedness between cases and controls [18]. The ROH-based analysis is a novel approach to identifying clustering patterns of variants to unmask ambiguous disease-genotype associations. To explore the relationships between ROHs and autism, we conducted a genome-wide association study in a Taiwanese Han population. Our core hypothesis posits that several novel genes characterized by ROHs are associated with autism and its related language impairment. We selected speech delay as the primary clinical feature as previous evidence suggests that language impairment is the most important predictor for the prognosis and developmental course of autism [19,20].

Results

The descriptive analysis results for demographic and clinical features are summarized in Table 1. Verbal IQ and Performance IQ had the highest percentage of missing data, and hence we compared the association test results with and without Verbal IQ/Performance IQ in the regression model. Since the effect of IQ on the association between ROH markers and traits was limited, the missing data of IQ might not pose a great concern in the current study. The case-control association analysis did not yield genome-wide significant findings after multiple-testing correction (Table 2). There was no statistically significant difference in the ROH length between cases and controls (mean length: 658 kb vs. 645 kb; $z = 0.62$, $p = 0.229$). We also calculated the value of Froh (total length of all their ROHs in the autosome and divided by the total SNP-mappable autosomal distance) for the ROH burden analysis [21], and did not find any significant difference in Froh between cases and controls. The genome-wide significant finding ($p = 0.05/676 = 7.4 \times 10^{-3}$) was obtained from the association analysis for speech delay (surrogated by “age-of-first-phrase or AFP”). The association analysis findings across the whole autosome are illustrated in Figure 1. We also used k-means clustering algorithm to classify the population into two subgroups, and identified the early-AFP group and late-AFP group with the cutoff at 45 months of age. The results suggest that the distributions of ROHs of early-AFP versus late-AFP groups appeared to be similar to each other (Figure 2). One ROH region on chromosome 11q22.3 was significantly associated with risk of speech delay (Bonferroni-corrected $p = 1.73 \times 10^{-8}$). The significant results (corrected with Bonferroni method) for AFP are summarized in Table 3. This ROH marker was found to be positively associated with AFP as a continuous variable. This result

remained statistically genome-wide significant after we adjusted for IQ, gender, and education levels of parents. We also assessed the relationship between this ROH region and AFP as a dichotomous variable using the logistic regression model, and the association remained significant ($p < 0.0001$). This region contains nine genes, none of which has been found to be associated with the risk of autism in previous studies or specific language disorders.

We further examined if the ROH region on 11q22.3 might arise from selection sweeps. The distributions of relative extended haplotype homozygosity (REHH) of the early-AFP and late-AFP groups (classified using the k-means clustering algorithm) appeared to be similar to each other (Figure 3A versus Figure 4A). The distributions of REHH (the factor by which EHH decays on the tested five-SNP core haplotype “rs1074014-rs1072877-rs1564582-rs11212724-rs11211725” in the genes on 11q22.3) of early-AFP and late-AFP groups seemed to differ by the core haplotype with strongest evidence for incomplete selection sweep (Figure 3C versus Figure 4C). Additionally, these two groups might have different ancestral haplotypes (Figure 3E versus 4E), although their frequency distributions of core haplotypes were similar (Figure 3B, 3D versus Figure 4B, 4D). The Nevertheless, none of these core haplotypes appeared to have remarkable signatures of recent positive selection based on the REHH distributions (i.e., REHH exceeding 2 at 200 Kb away from the core haplotype). We also searched for the signature of selection sweeps in genes proximal to 11q22.3, and found that the CWF19L2 (CWF19-like 2, cell cycle control) gene located 1 Mb upstream to this region based on the phase-I Hapmap Asian-descent population (CHB+JPT) has an iHS (Integrated Haplotype Score) score = 1.7 ($p = 0.0237$) based on the query using the webtool Haplotter [22]. We hence calculated the linkage disequilibrium (LD) coefficients D' between the ROH region and CWF19L2 gene, and found that a locus (rs1046094, a 3' UTR variant) within the CWF19L2 gene was correlated with another locus (rs4754276, an intronic variant) within the RAB39 (ras-related protein Rab-39A) gene ($D' = 0.9$) (Figure 5).

The inbreeding coefficient F value was < 0.01 based on the SNP data on chromosome 11 in either the early-AFP or late-AFP groups. Therefore, the ROH markers associated with speech delay might not be caused by the difference in the degree of consanguinity between these two subpopulations. Additionally, we queried the CNV data generated by the same SNP arrays in the discovery sample, and did not find any deletions or duplications in this 11q region. Therefore, the ROHs on 11q22.3 were not likely attributed to hemizygous deletions.

Replication Study

The recruitment of subjects under the auspice of Autism Genetic Resource Exchange (AGRE) has been described elsewhere [23]. Briefly, AGRE is a joint effort of the Cure Autism Now (CAN) Foundation and the Human Biological Data Interchange (HBDI). The diagnosis was made by all of the NIH autism collaborative groups using the Autism Diagnostic Interview-Revised (ADI-R) [24] and the Autism Diagnostic Observational Schedule (ADOS) [25]. We have downloaded the clinical and SNP data (generated by the Affymetrix SNP 5.0 platform) for all probands. We implemented the same data-cleaning algorithm used in the discovery sample. A total of 325,971 valid SNPs for 1,387 subjects diagnosed with autism were obtained. The age of first phrase (AFP) distributions of the AGRE sample and our discovery sample are shown in Supporting Information (Figure S1). We did not find significant difference in the distributions of AFP between the discovery population (Taiwan) and replication sample (AGRE) (Mann-Whitney U test $p > 0.05$). We attempted to

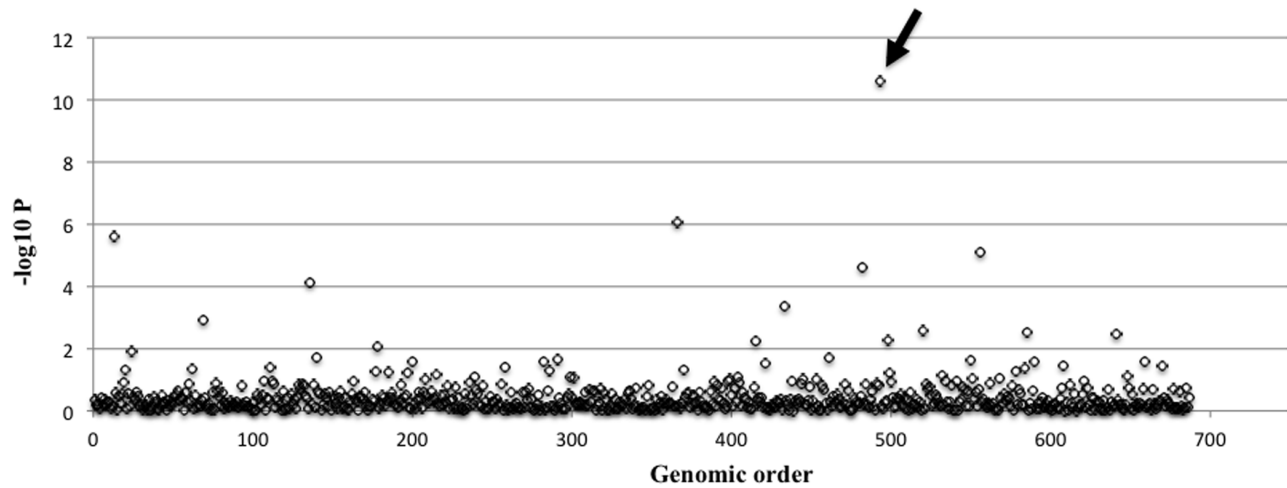


Figure 1. The association findings for age of first phrase (AFP) are presented as $-\log_{10}$ p-values (unadjusted by multiple tests) across the whole autosome. The arrow indicates the ROH region at 11q22.3.
doi:10.1371/journal.pone.0072056.g001

replicate the association between the ROH region on 11q22.3 and AFP in another independent population. The SNP data on chromosome 11 were retrieved from 1,387 individuals affected with autism recruited through multi-site collaborative efforts of Autism Genetic Resource Exchange (AGRE). We performed the same statistical methods as what we used in the discovery sample described in the Methods section and identified 31 ROH regions

on chromosome 11. When AFP was treated as a continuous outcome, no significant association was detected on 11q22.3. However, when we chose 49 months as a cutoff using the k-means clustering algorithm to define the presence of “speech delay,” we found that the ROH region on 11q22.3 (117.5 Mb-113.1 Mb) was nominally significantly associated with speech delay ($P=0.0377$). We then calculated the combined p-values from these two samples

Distribution of ROHs

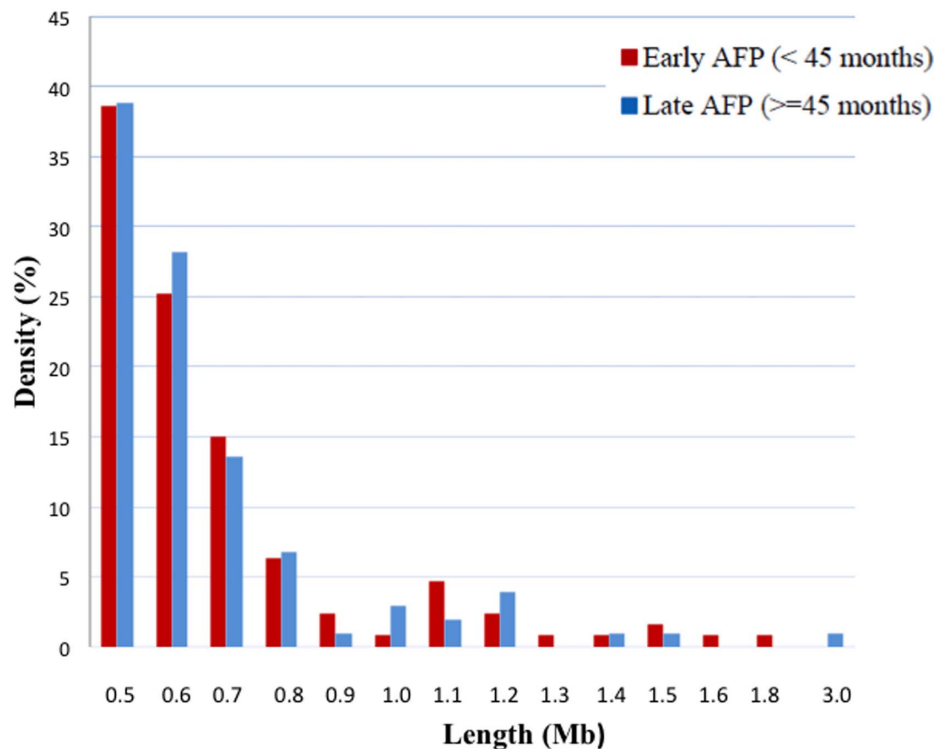


Figure 2. The distribution of runs of homozygosity (ROH) regions (by length of the ROH region) is shown. Age of first phrase (AFP) was classified into early-AFP and late-AFP groups by the k-means clustering algorithm.
doi:10.1371/journal.pone.0072056.g002

Table 1. Demographic features of cases in the discovery population.

Variable	Minimum	Median	Mean	Maximum	Standard Deviation	# Missing	# Non-missing
Maternal Education ¹	1	4	3.7	5	0.76	0	315
Paternal Education ¹	1	4	3.9	5	0.77	0	315
SRST ²	32	97.0	99.4	183	26.63	11	304
SCQ ³	3	18	18.19	36	7.03	59	256
FSTBEH ⁴	3	18	18.66	37	6.76	58	257
Age of First Phrase ⁵	5	36.0	39.6	156	18.29	85	230
VIQ ⁶	44	100.0	95.2	148	24.24	114	201
PIQ ⁶	41	98.0	97.1	145	20.71	114	201
Paternal age ⁷	22.8	33.7	34.1	59.3	4.88	57	258
Maternal age ⁷	19.6	30.5	30.9	44.5	4.53	57	258

¹Maternal and paternal education: 1 ≤ 6 years, 2 = 7–9 years, 3 = 10–12 years, 4 = 13–16 years, 5 = >16 years.

²SRST: Total social responsiveness score assessed by the Autism Diagnostic Interview-revised.

³SCQ: social communication quotient total score.

⁴FSTBEH: Stereotype behavior/interest score assessed by the Autism Diagnostic Interview-revised.

⁵The unit of age of phrase is month.

⁶VIQ = verbal IQ; PIQ = performance IQ.

⁷The unit of paternal/maternal age is year.

doi:10.1371/journal.pone.0072056.t001

based on the Stouffer method, and obtained Stouffer z value and z trend of 0.0007 and 0.0005, respectively. Note that these SNP data were based on Affymetrix SNP 5.0 platform that had lower marker density than Affymetrix SNP 6.0 data. The AGRE sample had a European origin, which might also contribute to different ROH patterns from our sample with an Asian origin.

Discussion

There has been limited research on the role of ROHs in autism in Asian populations. A recent study identified several novel candidate genes in ROH regions associated with the risk of autism in a European-descent population [15]. However, most of these loci reported by this study would not remain to be significantly associated with the disease risk after multi-testing corrections. Implementing stricter correction methods, we failed to detect significant disease-associated ROHs at a genome-wide level in our population. We speculate that the effect size of single ROH region associated with the risk of autism might be too small to be detected in a genome-wide scan. Another recent study reported that the length and number of ROHs in autistic cases were higher than

controls in a southern European-descent population [26]. However, our study shows that either lengths of ROHs or Froh values were similar in cases and controls. The inconsistent findings may stem from the difference in the population history of different samples. Additionally, consanguinity is unlikely to explain the relationship between ROHs and speech delay, as our findings do not reveal a remarkable difference in the degree of inbreeding between subgroups with speech delay and without speech delay. Furthermore, recent positive selection may play a limited role in the ROHs associated with speech delay in autism, as none of the candidate genes were found to have a strong signature of selection sweeps. However, we found that the patterns of extended homozygosity decay from the core haplotype on 11q22.3 might vary by the presence of speech delay. Our results suggest the variant within the RAB39 gene might be associated with the variant within the CWF19L2 gene under recent positive selection.

The current findings reveal a few novel candidate genes on 11q22.3 associated with speech delay in a Taiwanese Han population of autism. Among these genes, NPAT and ATM genes are associated with ataxia telangiectasia, one of the most frequent autosomal recessive cerebellar ataxias. Ataxia telangiect-

Table 2. Case-control association test results for 4 runs of homozygosity (ROH) regions nominally associated with the risk of autism (unadjusted p-value <0.01) in the discovery sample.

Start position	End position	Length	Chromosome	Fisher P ^A	Odds ratio	R _{Case} [*]	R _{Ctrl} [*]	Genes
208,824,998	209,180,938	355,940	2	0.000827	20.32	0.018	0.001	IDH1, PIP5K3, PTH2R
120,293,987	120,445,970	151,983	8	0.000827	20.32	0.018	0.001	MAL2
46,630,743	47,455,038	824,295	20	0.002221	5.10	0.027	0.005	PREX1, ARFGF2, FKSG61, CSE1L, CSE1, STAU1, DDX27, ZNF1, C20orf199, KCNB1
120,782,861	121,320,700	537,839	8	0.006451	6.76	0.018	0.003	TAF2, DSCC1, DEPDC6, COL14A1

*R_{Case} = prevalence rate of the ROH marker in cases; R_{Ctrl} = prevalence rate of the ROH marker in controls;

^AUnadjusted p-values based on the Fisher's exact tests.

doi:10.1371/journal.pone.0072056.t002

Table 3. Case-only association test results for age of first phrase (only unadjusted p-value $< 1 \times 10^{-5}$ were shown).

Start position	End position	Length	Chrom	Regression coefficient	P _{Bonf*}	Genes
107,309,153	108,201,671	892,518	11	116.9 (±10.6)	1.7×10^{-8}	RAB39, CUL5, ACAT1, NPAT, ATM, C11orf65, KDELC2, EXPH, DOX10
89,171,609	90,594,484	1,422,875	7	51.08 (±6.1)	7.3×10^{-7}	DPY19L2P4, STEAP1, STEAP2, C7orf63, GTPBP10, CLDN12, PFTK1, KIAA0834
42,126,282	43,130,565	1,004,283	1	49.05 (±5.6)	0.0035	GUCA2B, FOXJ3, ZMYND12, RIMKLA, PPCS, LOC728621, DKFZp686k01114, PPH, AF086102, YBX1, CLDN19, LEPRE1, CR623026, C1orf50, CDC23, ERMAP, ZNF691
76404435	77605674	1,201,239	13	56.91 (±5.1)	0.0113	CLN5, FBXL3, KIAA0916, MYCBP2, DKFZp586G0322, SEL, SLAIN1, EDNRB, AK095779

*Bonferroni-corrected p-value.
doi:10.1371/journal.pone.0072056.t003

tasia is also characterized by impairment in verbal fluency. Individuals affected with Ataxia telangiectasia often show weak oral motor performance [27]. It is unclear whether ataxia telangiectasia and autism has similar defects in the speech pathologies. Another gene located in the same region, EXPH5 (exophilin 5) is a cerebellum-expressed gene [28]. Cerebellum modulates motor coordination that also regulates the speech function. Additionally, individuals with autism and speech delay and individuals with autism without speech delay have marked difference in metabolic ratio in cerebellar regions [29]. Therefore, the ATM, NPAT, and EXPH5 genes may influence some neural correlates associated with the cerebellum. Variants in these three genes may thus influence the language function linked to the cerebellum in autism.

Additionally, the CUL5 (culin 5) gene in the region has been found to regulate cortical layering by modulating the neuronal migration process [30,31]. The protein culin 5 encoded by the CUL5 gene plays a pivotal role if degradation of an intracellular signaling molecule, Disabled-1, which is activated by reelin encoded by the RELN gene. Previous studies have shown mixed evidence for the association between the RELN gene and the risk of autism [32,33]. It has been shown that subtle dysregulated neuronal migration, such as perisylvian polymicrogyria, is associated with the developmental language disorder [34]. An animal study also showed that homozygous mutants for the CUL5 variant is defective in Notch signaling as indicated by the impaired expression of Notch target genes, which affects the initiation of Notch signaling during neurogenesis [35]. These findings may comprise the lines of indirect evidence for the relationship between the CUL5 gene and speech delay.

The association of chromosome 11q structural variants with language impairment has been documented by several studies. For instance, at least half of the individuals afflicted by 11q terminal deletion syndrome might be affected by mild to moderate impairment in expressive language [36]. A case report documents a girl with a 11q21–22.3 deletion manifested multiple congenital abnormalities, including speech delay [37]. Two case studies also report the association between 11q24 deletion and developmental speech delay in Jacobson syndrome [38]. Mosaic 11q deletions have also been noted in metopic synostosis associated with an increased risk of speech delay [39]. Additionally, the deletion of 11q23.3 might be associated with speech delay [40,41], while the duplication of the 11q23.3 region might also lead to speech delay [42]. Taken together, these findings suggest that the chromosome 11q21–q24 might harbor genes that play a role in language development.

Speech delay has been regarded as an endophenotype of autism. Some prior studies used speech delay as a clinical marker to identify homogeneous subgroups of autism, while others treated speech delay as an independent trait. Several regions have been found to be associated with speech delay in autism. For instance, the chromosome 7q31–q33 is one of the regions that have been found to contain genetic polymorphisms linked to speech delay in autism [43–47]. It has also been suggested that the 7q11–q12 duplication may be linked to speech delay in autism [48,49]. Additionally, the chromosome 2q is another region that might contain genetic variants associated with speech delay in autism [50,51]. Some of these candidate regions associated with speech delay in sporadic case reports. However, the CNTNAP2 (contactin associated protein-like 2) gene on 7q, which has been found to be linked to language impairment in some large-scale studies [52,53], was not included in the ROH regions that were significantly associated with speech delay in our sample. The CNTNAP2 gene, as well as other candidate risk genes for autism,

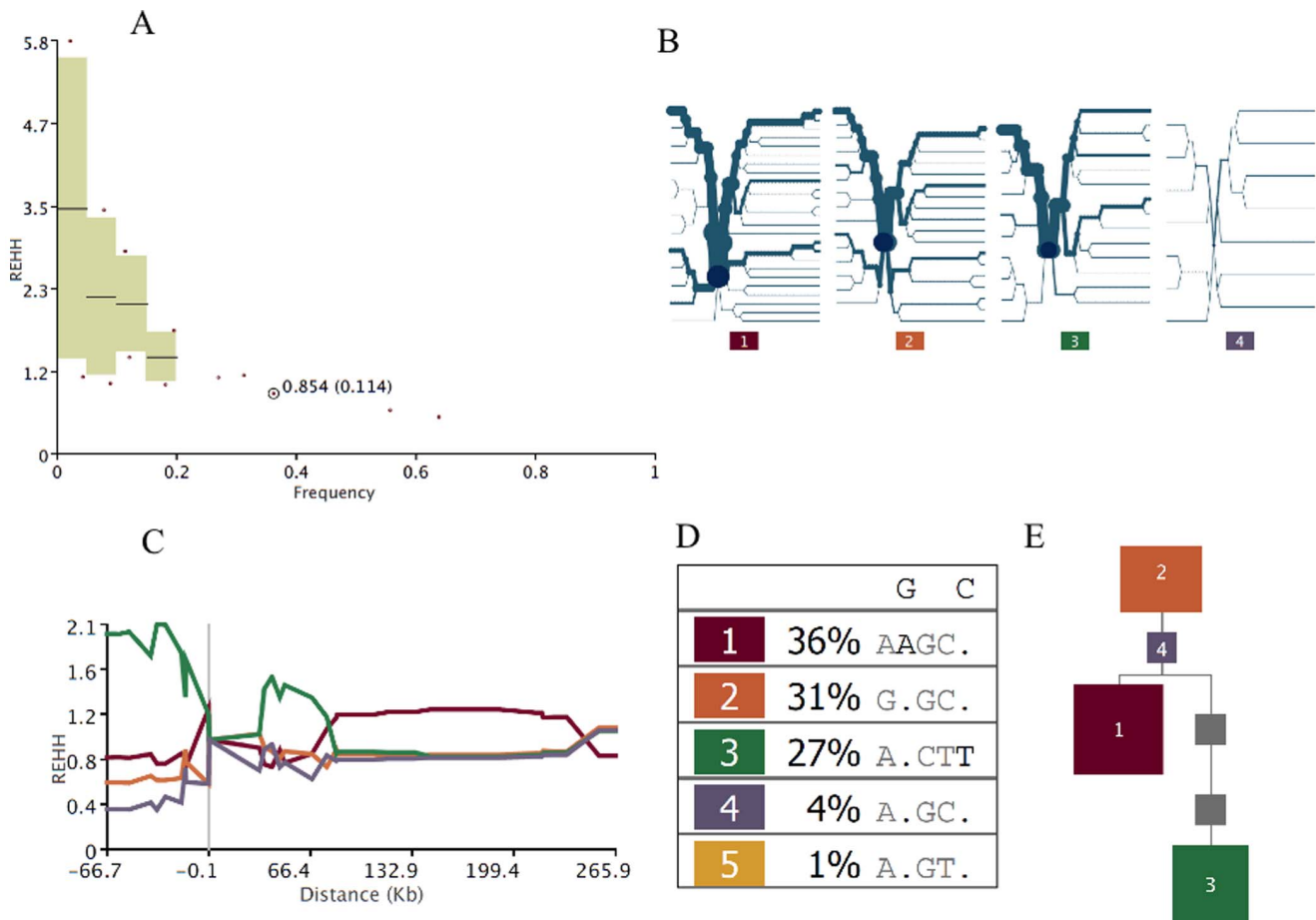


Figure 3. The analysis results based on the early-AFP group are shown. Panel A shows the scatter plot of REHH plotted against all core haplotype frequency (circled dot indicates the selected core haplotype “rs1074014-rs1072877-rs1564582-rs11212724-rs11211725”). Panel B shows the haplotype bifurcation diagram, which visualizes the breakdown of LD at increasing distances from core haplotypes at the selected core region. The root of each diagram is a core haplotype, identified by a dark blue circle. Panel C illustrates how the REHH value varies by the selected core haplotype. Panel D shows the table of core haplotype, and the dot in the observed haplotype sequence represents the allele that matches the ancestral. Panel E presents the theoretical phylogenetic tree of different core haplotypes. Gray squares represent haplotypes that are not present in the observed data, but are missing links in the phylogeny. The area of the squares is proportional to the frequency of the haplotype. doi:10.1371/journal.pone.0072056.g003

might not be identified in a case-only analysis of our study. It remains unclear if the molecular mechanisms of language impairment in individuals without autism differ from those in individuals with autism.

The current study has several limitations. First, the current study might not have sufficient power to detect variants of small to moderate effect on traits. This might at least partly explain the failure of our case-control association tests to replicate findings of previous studies. However, based on the parameters estimated in our case-control study, we achieved the statistical power of 30% given the α value = 0.0001. Second, the psychosocial factors that may influence language acquisition, such as parenting style and previous intervention, are not available in our samples. However, we did adjust for education levels of parents in the analysis and did not detect remarkable impact of parental education level on the genetic effect on clinical features. Nevertheless, parental education level might not fully reflect the quality of parenting and preschool education that may influence language acquisition. Third, the ROH based on the Affymetrix SNP 6.0 data might not consist of entirely homozygous SNPs, unless we have whole-genome sequencing data to verify these findings. Therefore, such a limitation might lead to the concern about the interpretation of

our findings. Additionally, since we could only perform the analysis of clinical features in cases, our findings might not be generalized to the genetic basis for speech delay. However, our study has yielded some insight into the molecular basis for clinical heterogeneity in autism.

In contrast to the continuous outcome, the analysis based on the dichotomous outcome yielded relatively less significant results for the same region on 11q22.3. This mild inconsistency might imply that the variant on 11q associated with speech delay might lead to a more extremely speech delay. Therefore, the comparison between relatively extremely late age-of-first-phrase group and extremely early age-of-first-phrase group might yield a more remarkable difference in ROH distributions between the two subgroups. However, in the replication study, we noticed that the dichotomous outcome yielded a slightly stronger association signal than the continuous outcome. These findings suggest that more research is needed to investigate how to define “speech delay” based on the age of first phrase and genomic data.

To sum up, the current study suggests that novel candidate genes may yield a greater impact on speech delay compared to autism per se. Untangling the mechanisms of speech delay may shed some light on molecular mechanisms underlying the

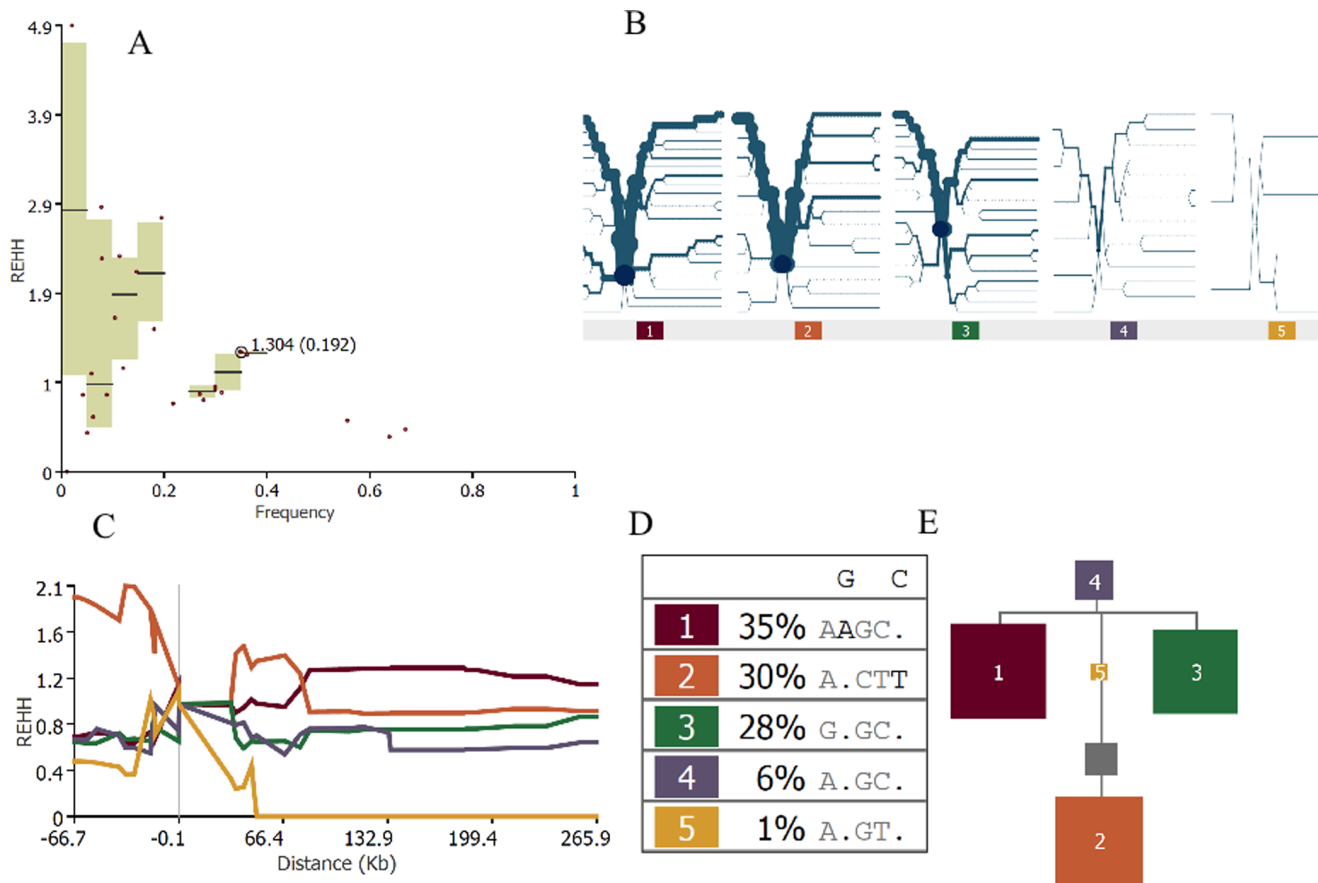


Figure 4. The analysis results based on the late age of first phrase (late-AFP) group are shown. Panel A shows the scatter plot of REHH plotted against all core haplotype frequency (circled dot indicates the selected core haplotype “rs1074014-rs1072877-rs1564582-rs11212724-rs11211725”). Panel B shows the haplotype bifurcation diagram, which visualizes the breakdown of LD at increasing distances from core haplotypes at the selected core region. The root of each diagram is a core haplotype, identified by a dark blue circle. Panel C illustrates how the REHH value varies by the selected core haplotype. Panel D shows the table of core haplotype, and the dot in the observed haplotype sequence represents the allele that matches the ancestral. Panel E presents the theoretical phylogenetic tree of different core haplotypes. Gray squares represent haplotypes that are not present in the observed data, but are missing links in the phylogeny. The area of the squares is proportional to the frequency of the haplotype. doi:10.1371/journal.pone.0072056.g004

development of autism. The extended homozygous haplotypes associated with speech delay may be more likely to be attributable to the recessive disease model than selection sweeps or consanguinity in our sample. Our findings also suggest that susceptibility genes may not necessarily contribute to clinical heterogeneity in autism. Taken together, these findings may lead to the evidence-based classification algorithm for clinical subgroups. Finally, our findings suggest that a few cerebellum-associated genes may play a role in speech delay in autism. Multiple adjacent loci of these genes may act in concert to cause speech delay in autism. More research is warranted to investigate if any cerebellum-related pathological changes could predispose to speech delay in autism.

Methods and Materials

Ethics Statement

The protocol entitled “Clinical and molecular genetic studies of autism spectrum disorder”, submitted by Principle Investigator Dr. Susan Shur-Fen Gau, Department of Psychiatry, National Taiwan University Hospital, Taiwan, has been approved by the 119th meeting of Research Ethics Committee of the National Taiwan University Hospital on September 26, 2006 (NTUH-REC ID: 9561709027) and the other two collaborating sites (Chang-

Gung Memorial Hospital in Taoyuan, CGMH ID: 93–6244 and Taoyuan Mental Hospital in Taoyuan, TYMH ID: C20060905). The committees of the three research sites were organized and operated according to GCP and the applicable laws and regulations. The Research Ethics Committee of three research sites approved this study [ClinicalTrials.gov number, NCT00494754]. Written informed consent was obtained from majority of the probands if they were able to give their signature after reading the informed consent and all their parents after the purposes and procedures of the study were fully explained and confidentiality was ensured. All subjects were Han Chinese. The data-sharing plan has been approved by all key investigators (SSG, YYW, and SKL) across three collaborating sites and approved by the Research Ethics Committee of the three sites. SSG, the principal investigator of this project, coordinated the research and managed all the clinical and genetic data. We reached the agreement that the de-identified data and key clinical variables will be released to investigators upon the request with relevant institutional approval documents.

Subject recruitment. The cases were selected from a sample of totally 1,164 subjects from 393 families (probands aged 9.1 ± 3.99 years, male 88.6%), recruited from the outpatient clinic of Psychiatric Department of three institutes (i.e., National Taiwan

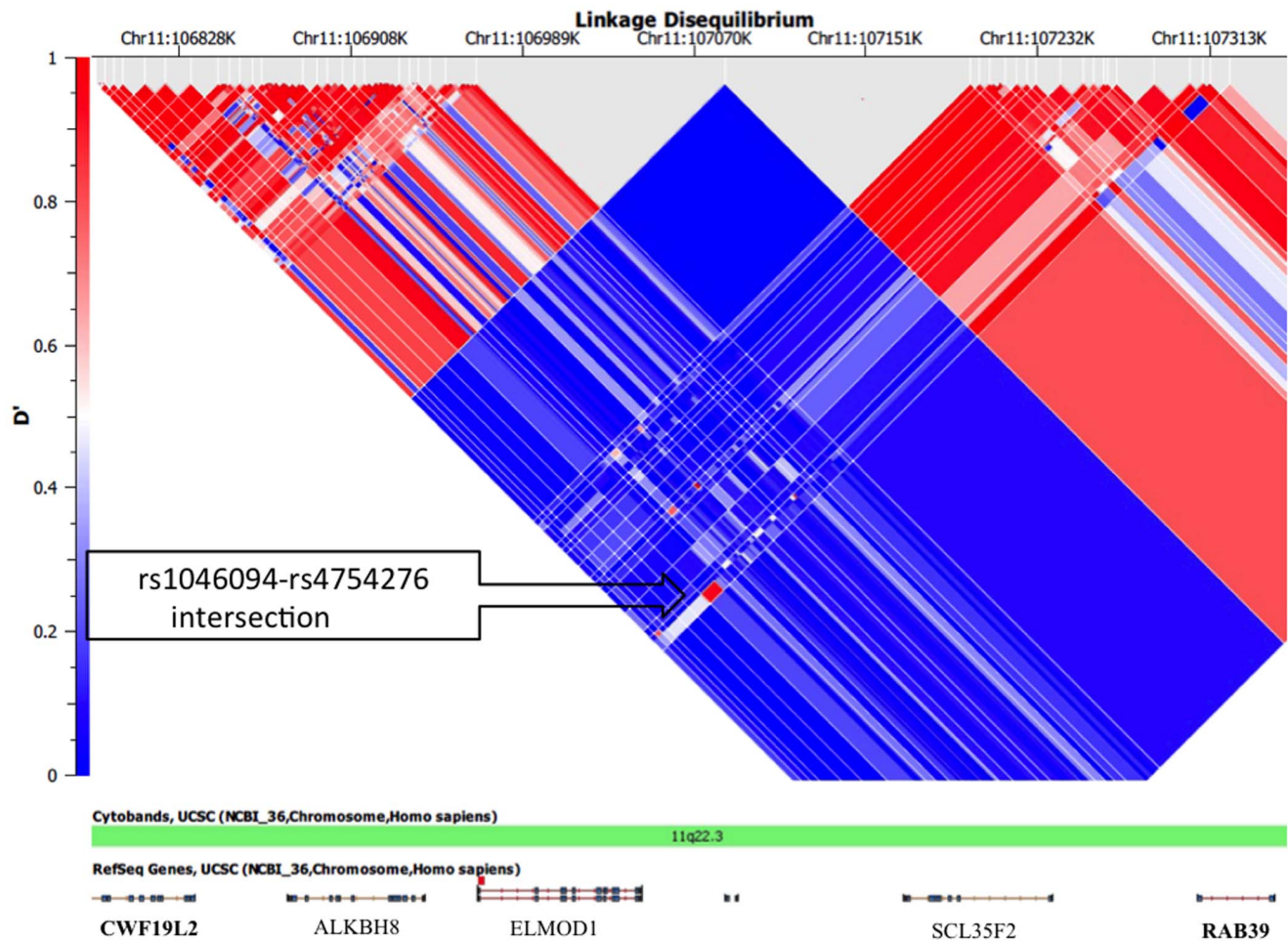


Figure 5. Linkage disequilibrium patterns in the 11q22.3 region are shown.
doi:10.1371/journal.pone.0072056.g005

University Hospital in Taipei, Chang-Gung Memorial Hospital in Taoyuan, and Taoyuan Mental Hospital in Taoyuan) in Northern Taiwan. Proband diagnosed with fragile X and Rett's disorder based on DNA testing or clinical features were excluded (unpublished data). Additionally, probands with previously identified chromosomal structural abnormality associated with autism, or had any other major neurological or medical conditions were also excluded. The initial diagnoses of probands were made by senior board-certified child psychiatrists based on the DSM-IV diagnostic criteria of autistic disorder or Asperger's disorder, and were further confirmed by interviewing the parents using the Chinese version of the Autism Diagnostic Interview-Revised (ADI-R) [54], adapted from the ADI-R [24]. The algorithm focuses on three domains based on the ICD-10 and DSM-IV diagnostic criteria, including reciprocal social interaction, verbal and non-verbal communication, as well as restricted, repetitive and stereotyped patterns of behaviors. We retrieved age of first phrase (AFP) to infer the presence of speech delay from the ADI-R assessment. AFP was treated as a continuous variable in the linear regression model that also controlled gender, SCQ, and parental education level. Additionally, we used k-means clustering algorithm with Euclidean distance to classify the sample into two subgroups, which were denoted as early-AFP group and late-AFP group.

The recruitment of controls was documented in detail elsewhere [55]. Briefly, the Institute of Biomedical Sciences, Academia Sinica and National Research Program for Genomic Medicine in Taiwan initiated the efforts to collect data to establish Han Chinese Cell and Genome Bank in Taiwan during 2002–2004. A three-stage sampling was implemented and complete bio-specimen and questionnaire data (with a focus on ethnicity and medical history) were collected for 3,380 individuals (gender ratio, 1:1; age range, 20–70 years). A total of 1,115 individuals with a Han Chinese ancestry that were found to have no definite diagnosis of major medical or mental illnesses were treated as the controls for the current study.

Genotyping. All cases and controls were genotyped on Affymetrix SNP array 6.0 platform that could generate a maximum of 906,600 SNPs and 946,000 probes for the detection of CNVs (Affymetrix Inc., Santa Clara, CA, USA). The DNA samples were extracted and purified from the peripheral lymphocytes according to the manufacturer's protocol. Genotype calls for SNPs were made based on the Birdseed algorithm that performs a multi-chip analysis to estimate a signal intensity for each allele of each SNP [56]. The average call rate was 99.86%. We also performed the Hardy-Weinberg Equilibrium (HWE) test, and excluded the SNPs with a HWE $P < 5 \times 10^{-5}$, so that the analysis would be less likely to be affected from genotyping or

calling errors. A total of 546,080 SNPs were thereby analyzed in the association tests.

Association analysis. We defined an ROH as a stretch of DNA spanning at least 500 kb or 50 consecutive SNPs without any heterozygous SNPs. Additionally, the maximum gap between SNPs could not exceed 100 kb. The overlapped region of multiple ROH regions shared by at least 10 individuals was regarded as a core ROH region. Furthermore, the prevalence rate of each common ROH marker should be at least 1% in the controls. We performed a case-control analysis based on cases ($n = 315$) and controls ($n = 1,115$) to identify risk ROHs. We also compared the difference in the length of ROHs of case and controls by t-test. To further clarify the role of ROHs in the heterogeneity of language developmental function in autism, we also assessed the associations between ROHs and the AFP. The continuous outcome variable was regressed against each ROH marker using the linear regression model. To adjust for the impacts of parenting and other confounders, we controlled for educational levels of parents, performance IQ, and gender in each linear regression model. To alleviate the problem of over-fitting due to intra-collinearity, we also performed step-wise regression analysis for the most significant trait-associated ROH marker. To determine the significance level, we took into account the number of ROH markers and outcome variables and applied the conservative Bonferroni method to correct inflated type-I errors due to multiple tests (corrected genome-wide significance threshold = $0.05/N$, N is the total number of OH regions). The ROH identification and association tests were performed using the software Golden Helix™ SNP and Variation Suite 7.6 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). Furthermore, we calculated the inbreeding coefficient F for sub-populations to assess if any spurious association arose from the difference in relatedness. Finally, we assessed if the size of gene might exert any impact on the association between trait-associated ROHs and traits by incorporating the gene size as a covariate in the regression model for the most significant finding.

References

- Muhle R, Trentacoste SV, Rapin I (2004) The genetics of autism. *Pediatrics* 113: e472–486.
- Wang K, Zhang H, Ma D, Bucan M, Glessner JT, et al. (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459: 528–533.
- Weiss LA, Arking DE, Daly MJ, Chakravarti A (2009) A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461: 802–808.
- Anney R, Klei L, Pinto D, Regan R, Conroy J, et al. (2010) A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 19: 4072–4082.
- Devlin B, Melhem N, Roeder K (2011) Do common variants play a role in risk for autism? Evidence and theoretical musings. *Brain Res* 1380: 78–84.
- Lin PI, Vance JM, Pericak-Vance MA, Martin ER (2007) No gene is an island: the flip-flop phenomenon. *Am J Hum Genet* 80: 531–538.
- Ma DQ, Rabionet R, Konidari I, Jaworski J, Cukier HN, et al. (2010) Association and gene-gene interaction of SLC6A4 and ITGB3 in autism. *Am J Med Genet B Neuropsychiatr Genet* 153B: 477–483.
- Singh AS, Chandra R, Guhathakurta S, Sinha S, Chatterjee A, et al. (2013) Genetic association and gene-gene interaction analyses suggest likely involvement of ITGB3 and TPH2 with autism spectrum disorder (ASD) in the Indian population. *Prog Neuropsychopharmacol Biol Psychiatry* 45C: 131–143.
- Ashley-Koch AE, Jaworski J, Ma de Q, Mei H, Ritchie MD, et al. (2007) Investigation of potential gene-gene interactions between APOE and RELN contributing to autism risk. *Psychiatr Genet* 17: 221–226.
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al. (2007) Strong association of de novo copy number mutations with autism. *Science* 316: 445–449.
- Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, et al. (2011) Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 70: 898–907.
- Keller MC, Simonson MA, Ripke S, Neale BM, Gejman PV, et al. (2012) Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS Genet* 8: e1002656.
- Lenz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, et al. (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A* 104: 19942–19947.
- Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, et al. (2009) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics* 10: 183–190.
- Casey JP, Magalhaes T, Conroy JM, Regan R, Shah N, et al. (2012) A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum Genet* 131: 565–579.
- Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, et al. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 91: 275–292.
- Chun S, Fay JC (2011) Evidence for hitchhiking of deleterious mutations within the human genome. *PLoS Genet* 7: e1002240.
- Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, et al. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5: e13996.
- Venter A, Lord C, Schopler E (1992) A follow-up study of high-functioning autistic children. *J Child Psychol Psychiatry* 33: 489–507.
- Rutter M (1970) Autistic children: infancy to adulthood. *Semin Psychiatry* 2: 435–450.
- Keller MC, Visscher PM, Goddard ME (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189: 237–249.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4: e72.
- Geschwind DH, Sowiński 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. *Am J Hum Genet* 69: 463–466.
- Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24: 659–685.

Selection sweep analysis. The phase of the haplotypes in the trait-associated ROH region were constructed using the program of PHASE v 2.1 [57]. We limited the search of core haplotypes to the brain-expressed genes. We then calculated extended EHH (i.e., the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent) to evaluate the evidence for selection sweeps. We further calculated the relative EHH value (REHH = core haplotype EHH divided by the decay of EHH on all other core haplotypes combined) to detect the signature of recent positive selection. We defined the evidence for selection sweep as REHH values ≥ 2 with long-range markers, radiating to distances greater than 200 kb from the core site, according to previous simulated data sets [58]. The phylogenetic relationship among all possible core haplotypes was inferred by ancestral alleles. All of the analyses were performed using the software Sweep [58].

Supporting Information

Figure S1 The distributions of age of first phrase (AFP) of the discovery population (Taiwan) and replication population (AGRE) are shown. (TIFF)

Acknowledgments

We greatly thank all the patients and families, who have made great contributions to this study. We also appreciate all of the research staff, especially Ms. Hui-Yi Huang and Ms. Mei-Hsin Su, for their efforts on data management and research coordination.

Author Contributions

Conceived and designed the experiments: PL SSG. Performed the experiments: CC JW. Analyzed the data: PL. Contributed reagents/materials/analysis tools: PK CC JW YW. Wrote the paper: PL SSG. SSG, YW, and SL conducted clinical diagnosis and helped recruit the patients. PK critically reviewed and revised the manuscript.

25. Lord C, Risi S, Lambrecht L, Cook EH, Jr., Leventhal BL, et al. (2000) The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 30: 205–223.
26. Wang LL, Yang AK, He SM, Liang J, Zhou ZW, et al. (2010) Identification of molecular targets associated with ethanol toxicity and implications in drug development. *Curr Pharm Des* 16: 1313–1355.
27. Vinck A, Verhagen MM, Gerven M, de Groot IJ, Weemaes CM, et al. (2011) Cognitive and speech-language performance in children with ataxia telangiectasia. *Dev Neurorehabil* 14: 315–322.
28. Thierry-Mieg D, Thierry-Mieg J (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* 7 Suppl 1: S12 11–14.
29. Gabis L, Wei H, Azizian A, DeVincent C, Tudorica A, et al. (2008) 1H-magnetic resonance spectroscopy markers of cognitive and language ability in clinical subtypes of autism spectrum disorders. *J Child Neurol* 23: 766–774.
30. Feng L, Allen NS, Simo S, Cooper JA (2007) Cullin 5 regulates Dab1 protein levels and neuron positioning during cortical development. *Genes Dev* 21: 2717–2730.
31. Simo S, Jossin Y, Cooper JA (2010) Cullin 5 regulates cortical layering by modulating the speed and duration of Dab1-dependent neuronal migration. *J Neurosci* 30: 5668–5676.
32. Zhang H, Liu X, Zhang C, Mundo E, Macciardi F, et al. (2002) Reelin gene alleles and susceptibility to autism spectrum disorders. *Mol Psychiatry* 7: 1012–1017.
33. Devlin B, Bennett P, Dawson G, Figlewicz DA, Grigorenko EL, et al. (2004) Alleles of a reelin CGG repeat do not convey liability to autism in a sample from the CPEA network. *Am J Med Genet B Neuropsychiatr Genet* 126B: 46–50.
34. Guerreiro MM, Hage SR, Guimaraes CA, Abramides DV, Fernandes W, et al. (2002) Developmental language disorder associated with polymicrogyria. *Neurology* 59: 245–250.
35. Sartori da Silva MA, Tee JM, Paridaen J, Brouwers A, Runtuwene V, et al. (2010) Essential role for the d-Asb11 cul5 Box domain for proper notch signaling and neural cell fate decisions in vivo. *PLoS One* 5: e14023.
36. Grossfeld PD, Mattina T, Lai Z, Favier R, Jones KL, et al. (2004) The 11q terminal deletion disorder: a prospective study of 110 cases. *Am J Med Genet A* 129A: 51–61.
37. Horelli-Kuitunen N, Gahmberg N, Eeva M, Palotie A, Jarvela I (1999) Interstitial deletion of bands 11q21–>22.3 in a three-year-old girl defined using fluorescence in situ hybridization on metaphase chromosomes. *Am J Med Genet* 86: 416–419.
38. Manolagos E, Orru S, Neroutsou R, Kefalas K, Louizou E, et al. (2009) Detailed molecular and clinical investigation of a child with a partial deletion of chromosome 11 (Jacobsen syndrome). *Mol Cytogenet* 2: 26.
39. Kini U, Hurst JA, Byren JC, Wall SA, Johnson D, et al. (2010) Etiological heterogeneity and clinical characteristics of metopic synostosis: Evidence from a tertiary craniofacial unit. *Am J Med Genet A* 152A: 1383–1389.
40. Perez Castillo A, Mardomingo Sanz MJ, Abrisqueta Zarrabe JA (1989) [Distal deletion at 11q and language delay]. *An Esp Pediatr* 30: 242–244.
41. Guerin A, Stavropoulos DJ, Diab Y, Chenier S, Christensen H, et al. (2012) Interstitial deletion of 11q-implicating the KIRREL3 gene in the neurocognitive delay associated with Jacobsen syndrome. *Am J Med Genet A* 158A: 2551–2556.
42. Burnside RD, Lose EJ, Dominguez MG, Sanchez-Corona J, Rivera H, et al. (2009) Molecular cytogenetic characterization of two cases with constitutional distal 11q duplication/triplication. *Am J Med Genet A* 149A: 1516–1522.
43. Lin PI, Chien YL, Wu YY, Chen CH, Gau SS, et al. (2012) The WNT2 gene polymorphism associated with speech delay inherent to autism. *Res Dev Disabil* 33: 1533–1540.
44. Spence SJ, Cantor RM, Chung L, Kim S, Geschwind DH, et al. (2006) Stratification based on language-related endophenotypes in autism: attempt to replicate reported linkage. *Am J Med Genet B Neuropsychiatr Genet* 141B: 591–598.
45. Cheung J, Petek E, Nakabayashi K, Tsui LC, Vincent JB, et al. (2001) Identification of the human cortactin-binding protein-2 gene from the autism candidate region at 7q31. *Genomics* 78: 7–11.
46. Poot M, Beyer V, Schwaab I, Damatova N, Van't Slot R, et al. (2010) Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder. *Neurogenetics* 11: 81–89.
47. Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH (2002) Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am J Hum Genet* 70: 60–71.
48. Berg JS, Brunetti-Pierri N, Peters SU, Kang SH, Fong CT, et al. (2007) Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region. *Genet Med* 9: 427–441.
49. Depienne C, Heron D, Betancur C, Benyahia B, Trouillard O, et al. (2007) Autism, language delay and mental retardation in a patient with 7q11 duplication. *J Med Genet* 44: 452–458.
50. Buxbaum JD, Silverman JM, Smith CJ, Kilifarski M, Reichert J, et al. (2001) Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *Am J Hum Genet* 68: 1514–1520.
51. Ramoz N, Cai G, Reichert JG, Silverman JM, Buxbaum JD (2008) An analysis of candidate autism loci on chromosome 2q24–q33: evidence for association to the STK39 gene. *Am J Med Genet B Neuropsychiatr Genet* 147B: 1152–1158.
52. 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 disorders. *Hum Mol Genet* 21: 4781–4792.
53. Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, et al. (2008) A functional genetic link between distinct developmental language disorders. *N Engl J Med* 359: 2337–2345.
54. Gau SS, Chou MC, Lee JC, Wong CC, Chou WJ, et al. (2010) Behavioral problems and parenting style among Taiwanese children with autism and their siblings. *Psychiatry Clin Neurosci* 64: 70–78.
55. Pan WH, Fann CS, Wu JY, Hung YT, Ho MS, et al. (2006) Han Chinese cell and genome bank in Taiwan: purpose, design and ethical considerations. *Hum Hered* 61: 27–30.
56. Rabbee N, Speed TP (2006) A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 22: 7–12.
57. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68: 978–989.
58. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419: 832–837.