Annals of human genetics

doi: 10.1111/ahg.12121

Common Regulatory Variants of *CYFIP1* Contribute to Susceptibility for Autism Spectrum Disorder (ASD) and Classical Autism

Jian Wang[1,2,5], Yu Tao[3], Fan Song[3], Yue Sun[1,2], Jurg Ott[6] and David Saffen[1,2,3,4]*

[1]*Institutes of Brain Science, Fudan University, Shanghai, China*
[2]*School of Life Sciences, Fudan University, Shanghai, China*
[3]*Department of Cellular and Genetic Medicine, Fudan University, Shanghai, China*
[4]*State Key Laboratory of Medical Neurobiology, Fudan University, Shanghai, China*
[5]*Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Shanghai Ocean University, Shanghai, Ministry of Education, China*
[6]*Institute of Psychology, Chinese Academy of Science, Beijing, China*

Summary

Based on the analysis of mRNA expression and genotype data from the "Brain Cloud" database, we identified seven SNPs within or near the autism candidate gene *CYFIP1* that show nominally significant correlations between genotype and *CYFIP1* mRNA expression in human dorsolateral prefrontal cortex. Analysis of transmission disequilibrium test (TDT) odds ratios (ORs) for these SNPs in a large Autism Genome Project (AGP) trio-based association study revealed the high-expression alleles of four of these SNPs (rs8028440, rs2289823, rs7403800 and rs3751566) to be susceptibility alleles. Correlations between the regression coefficients for mRNA expression and $\log_{10}$-transformed TDT ORs were statistically significant [$P = 0.008$ (ASD); $P = 0.002$ (classical autism)]. Similarly, statistically significant correlations were obtained between levels of *CYFIP1* mRNA expression predicted using the regression equations obtained from multiple linear regression analysis and $\log_{10}$-transformed TDT ORs for specific combinations of genotypes for both ASD (rs2289823 + rs3751566: $P = 0.008$) and classical autism (rs2289823 + rs3751566: $P = 0.008$; rs2289823 + rs3751566 + rs765763: $P = 0.0006$) diagnoses. Together, these results support the hypothesis that high expression of *CYFIP1* mRNA increases susceptibility for both ASD and classical autism.

Keywords: *CYFIP1*, autism spectrum disorder, regulatory genetic variant, expression quantitative locus (eQTL)

Introduction

Human cytoplasmic fragile X mental retardation protein (FMRP) interacting protein 1 (CYFIP1), also known as "specific Rac1-activated protein" SRA1 (Kobayashi et al., 1998), has been shown to function in two independent multiprotein complexes that are essential for the maintenance of synaptic morphology and function (De Rubeis et al., 2013): (1) CYFIP1-FMRP-eIF4E, which represses translation of mRNAs encoding postsynaptic proteins (Napoli et al., 2008) and (2) CYFIP1-NCKAP1-WAVE1-ABI2, which regulates actin remodeling in dendritic spines (Chen et al., 2010). Together, these observations suggest that functional impairment of *CYFIP1* may contribute to synaptic dysfunction observed in neurological disorders such as autism and schizophrenia.

The *CYFIP1* gene is located at Chromosome 15q11.2, a hot spot for autism and autism spectrum disorders (ASD) due to frequent structural abnormalities at this locus. Rearrangements of 15q11-q13, which include *CYFIP1* and many additional genes, are frequently detected in autistic patients (Vorstman et al., 2005; Depienne et al., 2009). Deletions of this region produce Angelman syndrome or Prader-Willi syndrome (Knoll et al., 1989; Sahoo et al., 2006), both of which have a high incidence of autism. Children with Chr15 duplications, including inv dup (15) or idic (15) syndrome,

*Corresponding author: David Saffen, Department of Cellular and Genetic Medicine, School of Basic Medical Sciences, 130 Dong'an Road, Building 13, Rm 226, Shanghai 200032, China. Tel: +86 215 423 7816; Fax: +81-21-54237643. E-mail: saffen@fudan.edu.cn

have been widely described as displaying autistic or autistic-like behaviors (Battaglia et al., 1997; Caglayan, 2010). Together, these observations suggest the presence of one or more dosage-sensitive genes within Chromosome 15q that contribute to autism and ASD when duplicated or deleted.

In addition to rare chromosomal rearrangements and copy number variants, common regulatory variants are hypothesized to contribute to human disorders by influencing the expression of dosage-sensitive genes (Villani et al., 2008; Dubois et al., 2010). For example, Direk et al. (2014) reported that an SNP associated with type 2 diabetes is an expression quantitative locus (eQTL) for *ABCC5*, and eQTLs for *SLC1A1* were reported to correlate with obsessive-compulsive disorder (Wendland et al., 2009).

Recently, evidence for the contribution of common regulatory variants to ASD and autism has also been reported. Davis et al. (2012) observed that many nominally significant loci in an autism GWAS study are also brain eQTLs. Smith et al. (2014) found that risk for autism correlated with a SNP that affects *HTR2A* mRNA expression and Xia et al. (2014) reported that several common genetic variants on 1q13.2 that associate with autism correlate with *CSDE1* and *TRIM33* expression in human brain.

To date, a small number of SNPs within or near the *CYFIP1* gene have been reported to associate with psychiatric disorders. Zhao *et al.* reported a nominal association between the *CYFIP1* intronic SNP rs1009153 and schizophrenia in Han Chinese (Zhao et al., 2013). Waltes et al. (2014) reported evidence that the nonsynonymous coding region SNP rs7170637 is associated with ASD in a German population. Contributions to autism and ASD of regulatory variants that influence the *CYFIP1* expression in brain, however, have not yet been reported.

In this study, we used high-quality mRNA expression data from the "BrainCloud" database (Colantuoni et al., 2011) and the Autism Genome Project (AGP) trio study (Anney et al., 2010) to investigate possible contributions of *CYFIP1* regulatory variants to ASD and classical autism.

Methods

"Braincloud" mRNA Expression and Genotype Data

CYFIP1 mRNA expression data obtained from 109 independent samples of dorsolateral prefrontal cortex (DLPFC) of Caucasian origin in the "BrainCloud for Expression" project (Colantuoni et al., 2011) were downloaded from the Gene Expression Omnibus Web site: http://www.ncbi.nlm.nih.gov/geo (Accession Number GSE30272; File name: GSE30272_series_matrix.txt.gz). Samples had been obtained from individuals aged 0.3–78 years and the quality of the RNA was good: average RNA Integrity Number (RIN) = 8.1. Normalized *CYFIP1* expression data were based on array hybridization to an oligonucleotide probe (ID number 6699) that binds to a sequence in the *CYFIP1* mRNA 3'-untranslated region (Fig. 1) and detects the expression of all major *CYFIP1* mRNA transcripts. Array expression data provided for analysis by the BrainCloud had passed quality control criteria, had been $\log_2$-transformed and normalized across mean $\log_2$ fluorescent intensities using loess correction (Colantuoni et al., 2011). Genotype data for the 109 samples were downloaded from the database for Genotypes and Phenotypes (dbGaP): http://www.ncbi.nlm.nih.gov/gap (accession number: phs000417.v1.p1). The genotype data set was "cleaned" by eliminating SNPs with: (1) minor allele frequencies (MAF) <0.01 (2) genotype missing rates $>5\%$, or (3) Hardy-Weinberg equilibrium $P < 0.001$.

Identification of *CYFIP1* eQTLs

To eliminate possible "environmental" influences on *CYFIP1* mRNA expression, normalized *CYFIP1* mRNA expression values were corrected for the covariates age, sex, postmortem interval, RIN, smoking status, sample source, and batch number in a multiple linear regression model. The residuals from this analysis (= "cleaned" mRNA levels) closely matched a normal distribution (Fig. S1) and were used for linear regression analysis of correlations between SNP genotypes and *CYFIP1* mRNA expression. Quantitative trait association studies for individual SNPs were carried out using the—assoc command in PLINK (Purcell et al., 2007) and the "lm" function in R.

Analysis of the contributions of multiple SNPs to *CYFIP1* mRNA expression was carried out by "backwards" stepwise linear regression using the "lm" and "step" functions in R. Prediction of *CYFIP1* mRNA levels based on SNP genotypes were calculated using regression equations obtained from multiple linear regression analysis of "cleaned" mRNA levels versus genotypes

$$\textit{predicted mRNA level} = \beta_0 + \sum_{j=1}^{n} \beta_j X_j,$$

(n = total number of SNPs analyzed),

where X_j = genotype of the j^{th} SNP (coded 0 for homozygous minor allele; 1 for heterozygous; 2 for homozygous major allele) and β_j = regression coefficients.

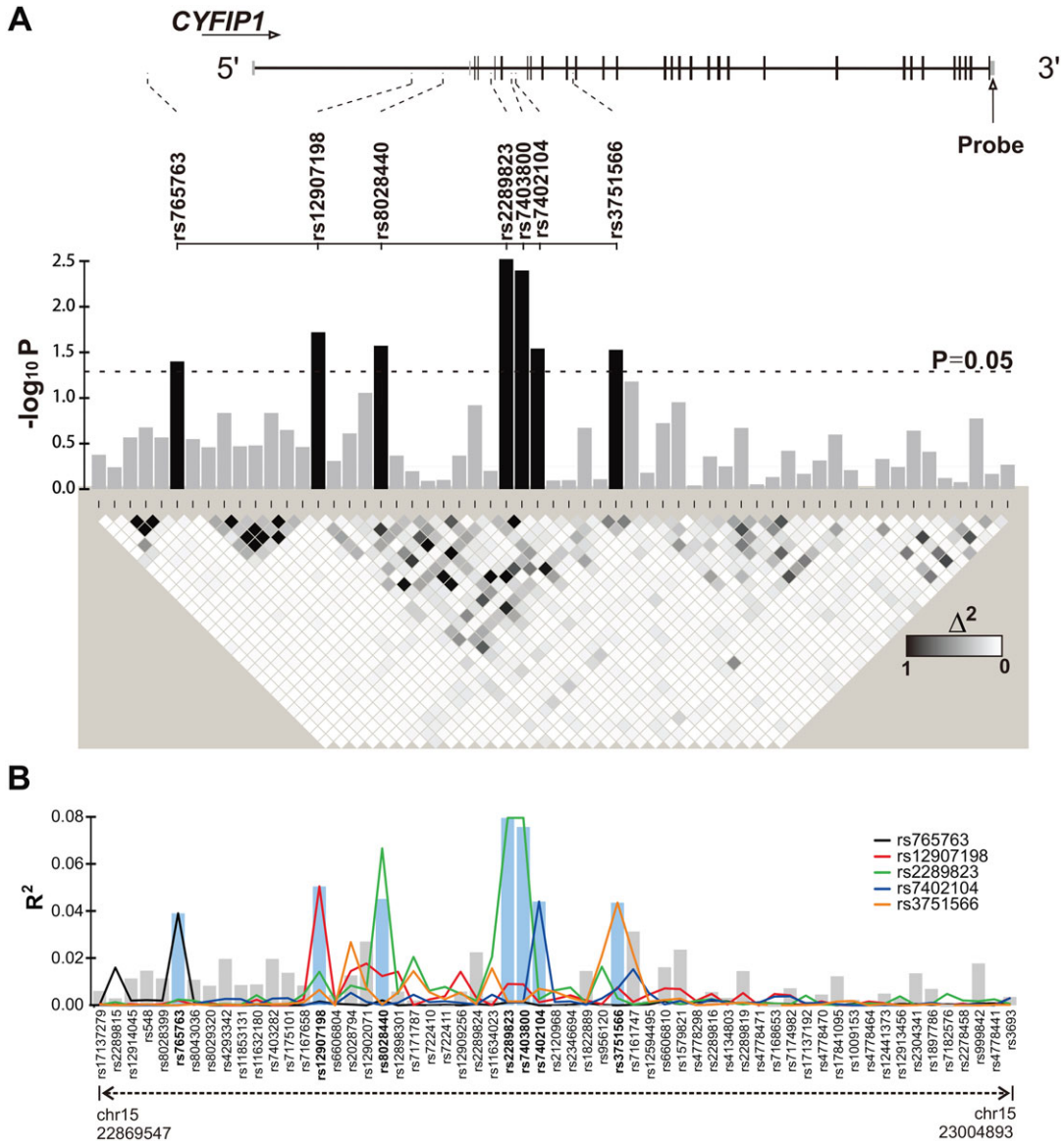

Figure 1 Identification of *CYFIP1* region SNPs that correlate with *CYFIP1* mRNA expression in 109 Caucasian "BrainCloud" DLPFC samples. (A; Top) Structure of the major prespliced *CYFIP1* RNA transcript, with exons represented by vertical bars, noncoding exons represented by short, gray bars and coding exons represented by long, black bars. Light dotted lines indicate the locations of seven SNPs with nominally significant correlations between genotype and *CYFIP1* mRNA levels. The location of the hybridization site of the oligonucleotide probe used to quantify *CYFIP1* mRNA is shown ("Probe"). (Middle) Plot of negative $\log_{10}$-tranformed P-values obtained from linear regression analysis of correlations between genotype and normalized *CYFIP1* mRNA expression for a set of 59 genotyped SNPs. Seven SNPs with nominally significant correlations (P-value < 0.05; not corrected for multiple testing) are represented by black bars. (Bottom) Plot of pairwise Δ^2 ($= r^2$) LD constants based on genotypes of 59 SNPs in the BrainCloud Caucasian sample, generated using Haploview [http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview]. (B) A "R^2-Δ^2" plot identifies five independent "index" SNPs among the seven correlating SNPs identified above. The bar plot shows R^2 values of individual SNPs obtained from the linear regression analysis, with nominally significant SNPs represented by blue bars. Colored lines plot Δ^2 values calculated with respect to the five unlinked index SNPs, with the height ($\Delta^2 = 1$) of each index SNP set at the same height as the corresponding R^2 value. The chromosome locations of the first and last SNP within this interval are listed (GRCh37/hg19 coordinates).

Linkage Disequilibrium (LD) Analysis

The LD structure of a ~135 kb fragment of Chromosome 15 containing *CYFIP1* for the Caucasian population and pairwise LD constants, Δ^2 (=r^2), for SNPs in the same region were generated using Haploview 4.2 (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview), based on genotype data from the 109 Caucasian BrainCloud samples.

Genotype Imputation

Imputation of SNP genotypes into the "BrainCloud" samples was carried out using IMPUTE version 2.2.2. The reference panel was the 1000 Genomes Project Phase I integrated haplotypes, which were produced using SHAPEIT2 and released in June 2014 (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html#reference). All imputed SNPs met the quality control criteria described above.

"R^2-Δ^2" Analysis

This method is based on the hypothesis that genotyped or imputed SNPs correlate with tissue levels of mRNA in proportion to their degree of linkage to the genetic variants that directly regulate mRNA expression. Because the identities of the actual regulatory variants are generally unknown, SNPs (singly or in combination) that show the highest correlations with mRNA expression (termed "index" SNPs) are selected and the contributions of all additional SNPs in a chromosome region of interest (ROI) are assumed to be proportional to the degree of linkage to one or more of the index SNPs, with additive contributions from multiple SNPs.

We take the coefficient of determination (R^2) obtained from linear regression analysis of the correlation between normalized $\log_2$-transformed tissue levels of mRNA and SNP genotype as the measure of the contribution of an individual SNP to mRNA expression. The degree of linkage between an "index" SNP and other SNPs in the chromosome ROI is the standard pairwise r^2 LD constant, which we designate as "Δ^2" to avoid confusion with "R^2."

In mathematical terms, we use the "lm" function in R to carry out multiple linear regression analysis of the set of measured R^2 values for SNPs within a chromosome ROI (the dependent variable) versus sets of pairwise Δ^2 values calculated between an "index" SNP and all other SNPs in the chromosome ROI, with one set for every index SNP (the independent variables). This analysis yields a regression equation

$$predicted\, R_i^2 = \alpha_0 + \sum_{j=1}^{n} \alpha_j X_{ij},$$

where n is total number of "index" SNPs included in the analysis, X_{ij} is the pairwise LD constant (Δ^2_{ij}) for the i^{th} SNP and j^{th} index SNP and α_j is regression coefficient reflecting the relative contributions of SNPs linked to the j^{th} index SNP to measured R^2 values, with the values of the coefficients (α_j) determined by the regression.

Because combinations of potential "index" SNPs often include two or more linked SNPs, it is essential to check each combination of SNPs for excessive influence of collinearity among sets of Δ^2 values (Δ^2_{ij}). For this purpose, we screen each SNP combination for the sizes of its associated variance inflation factors (VIFs) and select only those with a maximum VIF <5. We also select only SNP combinations that produce positive regression coefficients for all terms, since the meaning of a negative contribution to R^2 is difficult to interpret. Finally, the goodness of fit for each set of candidate "index" SNPs is determined by the adjusted R^2 value (termed adjusted R^2_{model}) obtained from the "lm" function, which also yields a P-value for each SNP combination. The combination of SNPs yielding the highest adjusted R^2_{model} is selected as the best set of "index" SNPs. In this study, we identified the best set of "index SNPs" among the seven candidate *CYFIP1* eQTLs with nominally significant correlations with *CYFIP1* mRNA expression.

AGP data

Phenotype and genotype data from an AGP trio study (Anney et al., 2010) were downloaded from dbGaP (Study Accession: phs000267.v4.p2). Individuals that self-described as having European ancestry (phenotype ID: phv00161302.v2.p2.c1) were used in this study. A total of 2147 trios with probands diagnosed with ASD (phenotype ID: phv00161303.v2.p2.c1) and a subset of 1266 trios with probands diagnosed with classical autism (phenotype ID: phv00161304.v2.p2.c1) were investigated.

AGP Family-Based Association Study Data

SNP genotype-disorder associations in AGP trio samples were investigated using the transmission disequilibrium test (TDT) as implemented in PLINK (Purcell et al., 2007). As described in the AGP study, individuals diagnosed with autism on the Autism Diagnostic Interview^{TM}-Revised (ADI-R) and ASD on Autism Diagnostic Observation Schedule (ADOS) criteria were classified as classical autism. Individuals, who were

diagnosed with ASD on both the ADI-R and ADOS or as autism on the other instruments, are classified as ASD. Allelic odds ratios (ORs) for transmission versus nontransmission to ASD probands and P-values were obtained for single candidate expression quantitative trait loci (eQTLs) in both ASD and autism studies. Putative ASD risk alleles were identified based on the observed allelic ORs. Numbers of transmitted and nontransmitted genotype combinations for multiple SNPs were obtained using UNPHASED for both ASD and classical autism trios. The OR for each genotype combination was calculated using the "fmsb" package in R. Details concerning the calculation of ORs are provided in Supplementary Methods.

Correlations between $\log_{10}$-transformed ORs and predicted mRNA expression for specific genotype patterns present in the AGP data set were analyzed by weighted linear regression using the R function "lm," with weights based on the number of occurrences of specific genotype combinations (transmitted + untransmitted) within the population. Specifically, $\log_{10}$-transformed observed ORs for transmission versus nontransmission of specific genotype combinations to probands ($y = log_{10} OR$: calculated as described in Supplementary Methods) were regressed against predicted *CYFIP1* mRNA levels for the same genotype combinations ($x = predicted\ mRNA\ level$: calculated as described in Methods: "Identification of *CYFIP1* eQTLs"), yielding regression equations, $y = \alpha + \beta x$. The values of correlation coefficients β were estimated by minimizing the weighted sum of squares for errors (WSSE):

$$\mathrm{WSSE} = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2,$$

where n is the total number of genotype combinations detected in all samples, w_i is the total number of the i^{th} genotype combination (transmitted + untransmitted) and $y_i - \hat{y}_i$ is the vertical distance of the i^{th} sample point (x_i, y_i) from the regression line.

Results

Identification of eQTLs for *CYFIP1* mRNA expression in human DLPFC

Single variable linear regression analysis of *CYFIP1* mRNA levels and genotype data for 109 Caucasian samples from the "BrainCloud" study identified seven SNPs that correlated with *CYFIP1* mRNA expression in DLPFC with nominal statistical significance ($P < 0.05$, without multiple corrections) within a ~135 kb segment of Chromosome 15 containing the *CYFIP1* gene (Fig. 1A). Six of these candidate eQTLs are located within 5'-introns and one is located approximately 16 kb upstream from the transcriptional start site.

The SNP that accounts for the largest fraction of the variance of mRNA expression, rs2289823 [adjusted $R^2 = 0.071$; $P = 0.003$; note: unless specifically stated, P-values are nominal (not corrected for multiple testing)], is tightly linked with the SNPs rs8028440 (adjusted $R^2 = 0.037$; $P = 0.027$) and rs7403800 (adjusted $R^2 = 0.067$; $P = 0.004$) in the BrainCloud Caucasian sample, with pairwise LD constants ($\Delta^2 = r^2$), of 0.83 and 1.0, respectively. The remaining four candidate eQTLs, rs765763 (adjusted $R^2 = 0.030$; $P = 0.04$), rs12907198 (adjusted $R^2 = 0.041$; $P = 0.019$), rs7402104 (adjusted $R^2 = 0.035$; $P = 0.029$), and rs3751566 (adjusted $R^2 = 0.035$; $P = 0.030$), are independent of each other (Fig. 1B). Graphs of *CYFIP1* mRNA expression versus genotype for the seven candidate eQTLs are shown in Figure S2 and the high-expression alleles are listed in Table 1.

To determine how well the seven candidate eQTLs account for the pattern of measured coefficients of determination (R^2) for *CYFIP1* mRNA expression for the 59 genotyped SNPs in this region, we carried out multiple linear regression analysis of measured R^2 values as a function of the linkage of individual SNPs to the candidate eQTLs ($\Delta^2 = r^2$ = pairwise LD constants between candidate eQTLs and genotyped SNPs in the *CYFIP1* region). This "R^2-Δ^2" analysis confirmed that the candidate eQTLs belong to five independent SNP families. Because three of the candidate eQTLs are tightly linked, only three 5-candidate eQTL combinations had a VIF <5, and among these the combination rs765763, rs12907198, rs2289823, rs7402104, and rs3751566 (VIF = 1.176) yielded the largest adjusted R^2_{model} (=0.660) and was selected as the best model. Together, these SNPs account for approximately 66% of the variance in measured R^2 values for the 59 genotyped SNPs (Fig. S3A).

Analysis of Imputed SNPs Revealed No Additional Independent Candidate *CYFIP1* eQTLs

To obtain a more detailed map of correlations between *CYFIP1* mRNA expression and SNP genotypes, we imputed SNPs within the same segment of Chromosome 15 analyzed above into the BrainCloud samples using data from the 1000 Genomes Project as a reference. Genotypes for 213 additional SNPs were successfully imputed, after excluding SNPs with low imputation quality (info score <0.3), minor allele frequency (MAF) <0.05 or a genotype missing rate >5%.

Interestingly, although a total of 35 additional imputed SNPs with nominally significant ($P < 0.05$) correlations with *CYFIP1* mRNA expression were identified, each of these SNPs was linked to one of the five independent candidate

Table 1 Correlations between *CYFIP1* SNP genotypes and mRNA expression for ASD or classical autism risk.

			CYFIP1 mRNA expression			ASD risk		Classical Autism risk		
SNP	Genotype*	MAF	Regression coefficient (95% CI)	P	High-expression allele	OR** (95% CI)	P	OR** (95% CI)	P	Risk allele
rs765763	T/C	0.326	−0.09 (-0.17 ~ −0.01)	**0.040**	C	0.99 (0.91 ~ 1.09)	0.870	0.97 (0.86 ~ 1.09)	0.563	-
rs12907198	G/A	0.436	0.1 (0.02 ~ 0.18)	**0.019**	G	1.02 (0.94 ~ 1.12)	0.613	1.06 (0.94 ~ 1.18)	0.343	-
rs8028440	T/C	0.252	0.1 (0.01 ~ 0.19)	**0.027**	T	1.12 (1.01 ~ 1.24)	**0.029**	1.13 (0.99 ~ 1.29)	**0.074**	T
rs2289823	T/C	0.220	0.15 (0.05 ~ 0.24)	**0.003**	T	1.12 (1.01 ~ 1.25)	**0.039**	1.14 (0.99 ~ 1.31)	**0.067**	T
rs7403800	G/A	0.220	0.14 (0.05 ~ 0.24)	**0.004**	G	1.14 (1.02 ~ 1.27)	**0.019**	1.17 (1.01 ~ 1.35)	**0.033**	G
rs7402104	T/C	0.101	−0.14 (-0.26 ~ −0.02)	**0.029**	C	0.99 (0.85 ~ 1.14)	0.851	0.96 (0.79 ~ 1.16)	0.661	-
rs3751566	T/C	0.362	−0.09 (-0.18 ~ −0.01)	**0.030**	C	0.94 (0.86 ~ 1.04)	0.220	0.9 (0.79 ~ 1.01)	**0.069**	C

*Minor/major allele; **OR = odds ratio defined with respect to minor allele; MAF = minor allele frequency; 95% CI = 95% confidence interval.

eQTLs described above and none of the imputed SNPs showed higher levels of association with *CYFIP1* expression than the original candidate eQTLs. The R^2-Δ^2 analysis showed that these five independent eQTLs account for approximately 69% of the variance in measured R^2 values for 272 genotyped and imputed SNPs (Fig. S3B).

High-Expression Alleles of Candidate *CYFIP1* eQTLs Associate with Risk for ASD and Classical Autism

To explore possible contributions of *CYFIP1* mRNA expression to autism susceptibility, the candidate *CYFIP1* eQTLs described above were used as genetic markers in a European ancestry family-based association study for ASD and classical autism using data obtained from the AGP (Anney et al., 2010).

The TDT results revealed that the three linked candidate eQTLs with the highest correlation with *CYFIP1* mRNA expression, rs2289823, rs8028440, and rs7403800, associated with ASD ($P = 0.019$–0.039) and classical autism ($P = 0.033$–0.074; Table 1). For each SNP, the *CYFIP1* mRNA high-expression allele was identified as the ASD and classical autism risk allele. Similarly, the high-expression allele of rs3751566 showed a trend for association ($P = 0.069$) with classical autism. By contrast, no significant disease association was detected for the remaining three candidate eQTLs. Figure 2 summarizes the results of our investigations of the contributions of individual SNPs to *CYFIP1* mRNA expression and susceptibility for ASD and classical autism. A statistically significant correlation was obtained between the $\log_{10}$-transformed OR for diagnosis and the size of the mRNA expression linear regression coefficients for both ASD (adjusted $R^2 = 0.743$; $P = 0.008$) and classical autism (adjusted $R^2 = 0.845$; $P = 0.002$) diagnoses.

Combinations of Multiple *CYFIP1* eQTLs Explain a Larger Fraction of the mRNA Expression Variance and Associate with a Wider Range of Disease ORs Compared to Individual eQTLs

The contribution of multiple *CYFIP1* eQTLs to mRNA expression was analyzed by stepwise linear regression (see Methods section). This analysis showed that the combined contributions of three eQTLs, rs2289823, rs3751566, and rs765763, account for approximately 12% of the variance in *CYFIP1* mRNA expression in DLPFC ($P = 0.000125$), compared to 3%–7.1% for individual SNPs ($P = 0.003$–0.04; Figs 3A and S2). Two additional independent eQTLs,

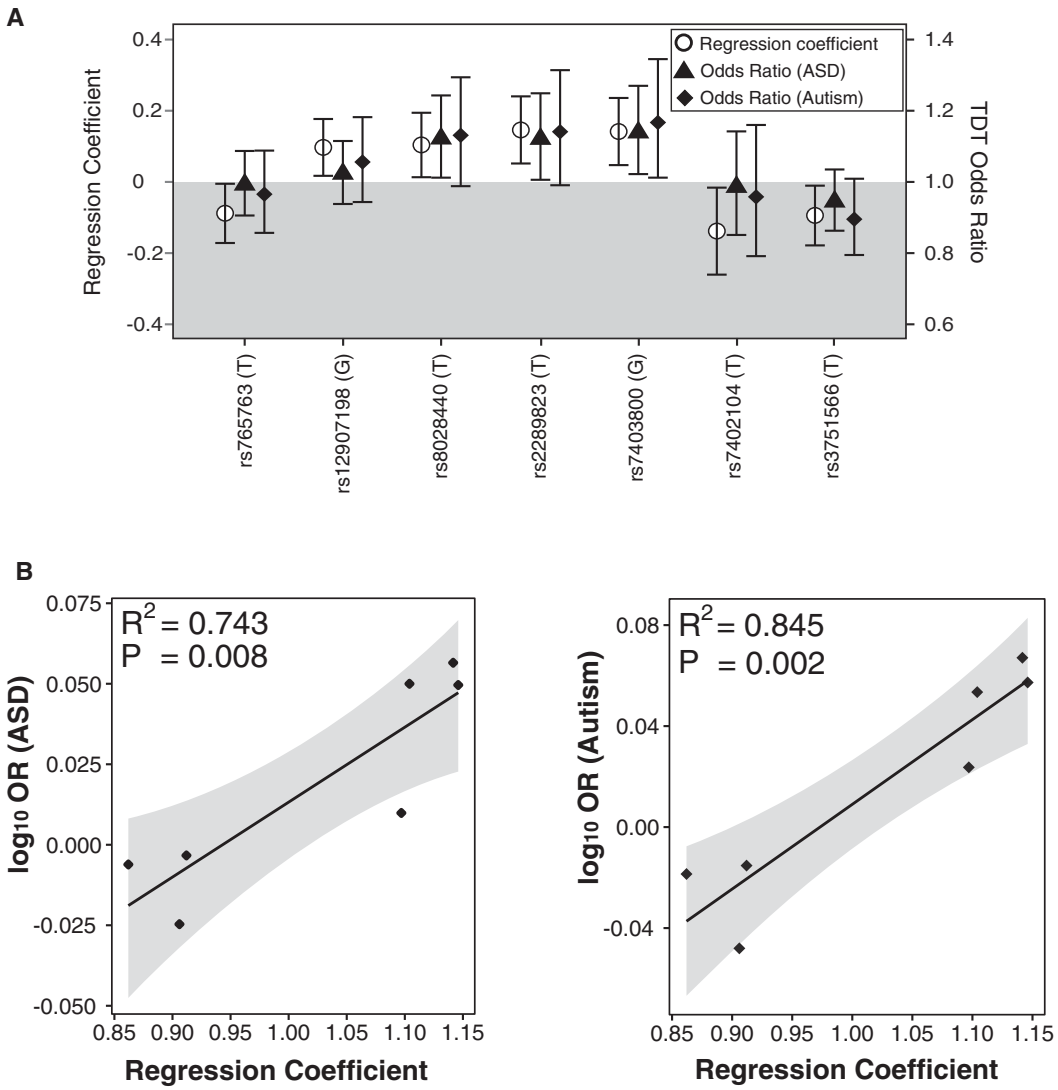

Figure 2 Analysis of individual SNPs shows that high *CYFIP1* mRNA expression correlates with ASD and classical autism in the Caucasian population. (A) Plot of linear regression beta coefficients for *CYFIP1* mRNA expression in DLPFC (○, left *y*-axis) and TDT ORs for ASD (▲, right *y*-axis) or classical autism (◆, right *y*-axis) diagnosis for Caucasian trios in the AGP study: 2147 trios with a proband diagnosed with ASD and 1266 trios with a proband diagnosed with classical autism. Error bars indicate the 95% confidence intervals (95% CI) of the corresponding regression coefficients and ORs (calculated for the minor allele of each SNP). (B) Correlations between mRNA expression regression coefficients and ORs for candidate eQTLs in ASD trios (left) and classical autism trios (right). The adjusted R^2 and P-value for the correlation obtained from the linear regression analysis are shown within each plot. The shaded areas indicate the 95% CI for the regression lines.

rs12907198 and rs7402104, did not increase the percentage of mRNA expression variance beyond that explained by the other three eQTLs (data not shown).

To investigate the combined contributions of the three eQTLs to ASD and autism, we calculated relative ORs for transmission versus nontransmission of specific genotype combinations to probands based on the results obtained using UNPHASED (Horvath et al., 2004). Regression analysis of $\log_{10}$-transformed ORs for ASD or classical autism versus the level of predicted *CYFIP1* mRNA expression for each genotype pattern, revealed statistically significant linear correlations between disorder risk and predicted mRNA expression, with higher risk associated with higher predicted *CYFIP1* mRNA levels, both for combinations of the two SNPs rs2289823 and

rs3751566 [adjusted $R^2 = 0.673$, $P = 0.008$ (ASD); adjusted $R^2 = 0.673$, $P = 0.008$ (classical autism)] and the three SNPs rs2289823, rs3751566, and rs765763 [adjusted $R^2 = 0.426$, $P \sim 0.0006$ (classical autism)] (Figs 3B and C).

Discussion

Using data provided by the "BrainCloud" Project (Colantuoni et al., 2011), we identified seven SNPs in the *CYFIP1* gene region with genotypes that correlate with *CYFIP1* mRNA expression in 109 Caucasian samples of prefrontal cortex (Fig. 1). The R^2-Δ^2 analysis showed that five of these SNPs (designated "index" SNPs) are independent and together account for 66% of the variation in measured R^2 values for 59 genotyped SNPs in this region (Fig. S3A). No new families of unlinked SNPs were found among 272 genotyped and imputed SNPs within 135 kb segment of chromosome 15 containing the *CYFIP1* gene (Fig. S3B). Analysis of transmission/nontransmission of alleles of the seven correlating SNPs in more than 2000 ASD/autism trios using the TDT revealed that the high-expression alleles of four SNPs (rs8028440, rs2289823, rs7403800, and rs3751566) associated with ASD and/or autism susceptibility (Fig. 2). Analysis of the combined contributions of two unlinked SNPs in this set (rs2289823 and rs3751566) to *CYFIP1* mRNA expression in brain and association with ASD or classical autism revealed a linear increase in susceptibility for ASD and classical autism with increasing levels of *CYFIP1* mRNA expression. Similar results were obtained for the combined contributions of the three unlinked SNPs rs2289823, rs3751566, and rs765763 for classical autism (Fig. 3).

These observations are consistent with a report that microduplication of this region cosegregates with autism in a Dutch pedigree (van der Zwaag et al., 2010). Overexpression of *CYFIP1* has recently been reported to activate mTOR signaling and promote morphological abnormalities of neurons (Oguro-Ando et al., 2014). Together, these results support the hypothesis that high expression of *CYFIP1* mRNA increases susceptibility for ASD and autism.

The identification of five independent families of SNPs close to or within the *CYFIP1* gene that associate with *CYFIP1* mRNA expression in human DLPFC (Figs 1 and S3) implies the existence of at least five independent *cis*-acting regulatory variants within the region of the *CYFIP1* gene. Molecular mechanisms by which *cis*-acting regulatory variants influence mRNA expression include the alteration of the binding of transcription factors to sites within promoters or enhancers (Szalai et al., 2005; Kasowski et al., 2010), alteration of the binding of proteins required for RNA splicing (Byrne et al., 2009), and alteration of mRNA stability by changing RNA secondary structure (Chamary & Hurst, 2005; Wang et al., 2005).

SNPs that regulate mRNA expression themselves or through strong LD (Reich et al., 2001) with common regulatory variants are termed eQTLs (Doerge, 2002, Gibson & Weir, 2005). eQTLs often correlate with mRNA expression in specific tissues (Heinzen et al., 2008) and many are associated with disease, especially with complex genetic diseases (Nica & Dermitzakis, 2008; Cookson et al., 2009; Nicolae et al., 2010).

Using genotype data from the 1000 Genome Project CEU population (http://www.broadinstitute.org/mpg/snap/ldsearch.php), we identified 26 SNPs that are closely linked ($\Delta^2 > 0.8$) to at least one of our five "index" SNPs (data not shown). Each of the linked SNPs, except one, is located within the 5'-upstream region or within *CYFIP1* introns. We hypothesize that one or more of these 26 SNPs are the regulatory variants responsible for the correlation of the index SNPs with mRNA expression.

Based on information from the Encode project (ENCODE Project Consortium, 2012) accessed via the RegulomeDB (http://regulomedb.org; Boyle et al., 2012), 5 of the 26 linked SNPs (rs8026390, rs11636046, rs11633285, rs7402672, rs11637146) are likely to affect the binding of a transcription factor (RegulomeDB score >3). For example, rs8026390 is found within a consensus binding motif for NeuroD, a transcription factor essential for neuronal differentiation and development (Miyata et al., 1999; Schwab et al., 2000) and reported to be expressed at lower levels in a mouse model for autism (Stephenson et al., 2011). Another example is rs11637146, which is located within a consensus binding motif for Tal1::Gata1 and may affect *CYFIP1* mRNA expression by altering the binding efficiency of this heterodimeric transcription factor. Mutations in Tal1 have been reported to associate with primary microcephaly (Kumar et al., 2009), a neurodevelopmental disorder that includes intellectual disability (Woods et al., 2005).

Each of the five SNPs with RegulomeDB scores >3, are located within introns near the 5'-end of the *CYFIP1* gene and are closely linked with the index SNP rs2289823, which exhibited the highest correlation with *CYFIP1* mRNA expression in our study. Although the remaining 21 SNPs were not associated with specific transcription factor binding sites in RegulomeDB, it is possible that they influence the binding of transcription factors not included in this database or influence *CYFIP1* mRNA expression by other mechanisms.

In addition to providing evidence that regulatory variants in *CYFIP1* contribute to susceptibility for ASD and classical autism, our study introduces a novel method, R^2-Δ^2 analysis, for visualizing independent families of SNPs within specific chromosome regions that we believe will be generally useful for estimating the number of independent *cis*-acting regulatory variants within the neighborhood of genes. Our study also demonstrates how correlations between predicted mRNA expression and ORs for disorder diagnosis for specific

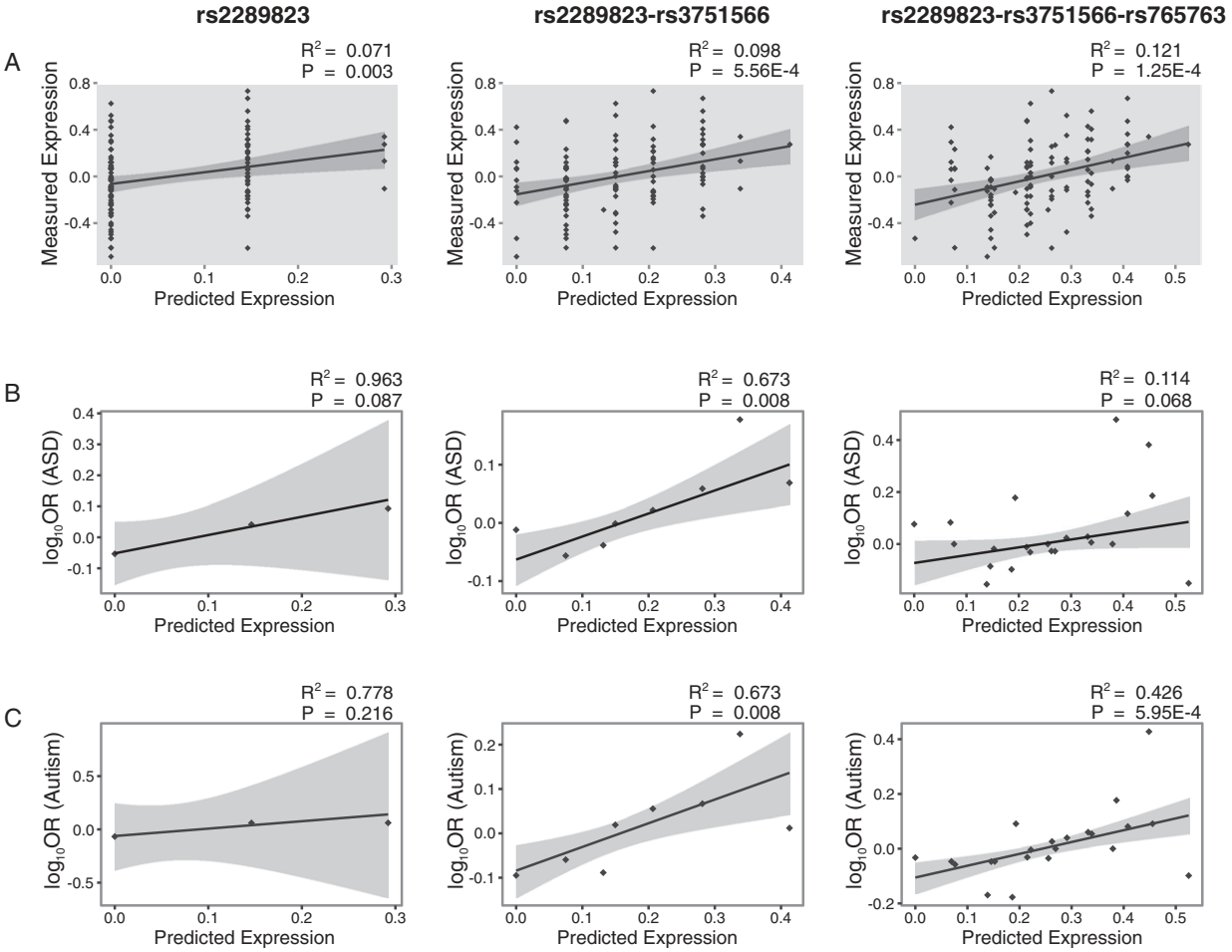

Figure 3 Analysis of multiple SNPs also shows correlations between *CYFIP1* mRNA expression and risk for ASD and classical autism. (A) Plots of measured normalized *CYFIP1* mRNA levels in DLPFC (*y*-axis) versus *CYFIP1* mRNA levels predicted from SNP genotypes (*x*-axis; details in Methods). Each point in the plots represents one individual (n = 109). (B) Plots of $\log_{10}$-transformed OR for transmission of genotypes of a single SNP (left) or genotype combinations of multiple SNPs (middle, right) to ASD probands (*y*-axis) versus predicted mRNA levels for each genotype or genotype combination (*x*-axis). Each dot represents one genotype (left) or genotype combination (middle, right). (C) Corresponding analysis for classical autism probands. The adjusted R^2 and P-values obtained from the weighted linear regression analysis of the data shown are listed above the respective plots. Shaded regions indicate the 95% CI for regression lines.

genotype combinations can be used to estimate the contribution of common regulatory variants to disorder susceptibility. We suggest that this method may also be generally useful for elucidating the combined contribution of common regulatory variants to complex genetic traits and disorders.

Acknowledgments

This work was supported by a grant from the National Natural Science Foundation of China (NSFC) grant numbers 81070908 (DS) and 31470070 (JO). We are also thankful for a grant from College Youth Scholars Supporting Plan of the Shanghai Municipal Education Commission (JW).

Author Contributions

J. Wang performed the data analysis and wrote the manuscript, Y. Tao carried out the genotype imputation, F. Song wrote the computer program for and assisted with the R^2-Δ^2 analysis, Y. Sun assisted with data analysis, J. Ott oversaw the statistical analysis, and D. Saffen developed the R^2-Δ^2 method, coordinated the project and revised the manuscript.

References

Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T. R., Correia, C., Abrahams, B. S., Sykes, N., Pagnamenta, A. T., Almeida, J., Bacchelli, E., Bailey, A. J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bölte, S., Bolton, P. F., Bourgeron, T., Brennan, S., Brian, J., Carson, A. R., Casallo, G., Casey, J., Chu, S. H., Cochrane, L., Corsello, C., Crawford, E. L., Crossett, A., Dawson, G., DeJonge, M., Delorme, R., Drmic, I., Duketis, E., Duque, F., Estes, A., Farrar, P., Fernandez, B. A., Folstein, S. E., Fombonne, E., Freitag, C. M., Gilbert, J., Gillberg, C., Glessner, J. T., Goldberg, J., Green, J., Guter, S. J., Hakonarson, H., Heron, E. A., Hill, M., Holt, R., Howe, J. L., Hughes, G., Hus, V., Igliozzi, R., Kim, C., Klauck, S. M., Kolevzon, A., Korvatska, O., Kustanovich, V., Lajonchere, C. M., Lamb, J. A., Laskawiec, M., Leboyer, M., LeCouteur, A., Leventhal, B. L., Lionel, A. C., Liu, X.-Q., Lord, C., Lotspeich, L., Lund, S. C., Maestrini, E., Mahoney, W., Mantoulan, C., Marshall, C. R., Mcconachie, H., Mcdougle, C. J., Mcgrath, J., Mcmahon, W. M., Melhem, N. M., Merikangas, A., Migita, O., Minshew, N. J., Mirza, G. K., Munson, J., Nelson, S. F., Noakes, C., Noor, A., Nygren, G., Oliveira, G., Papanikolaou, K., Parr, J. R., Parrini, B., Paton, T., Pickles, A., Piven, J., Posey, D. J., Poustka, A., Poustka, F., Prasad, A., Ragoussis, J., Renshaw, K., Rickaby, J., Roberts, W., Roeder, K., Roge, B., Rutter, M. L., Bierut, L. J., Rice, J. P., Salt, J., Sansom, K., Sato, D., Segurado, R., Senman, L., Shah, N., Sheffield, V. C., Soorya, L., Sousa, I., Stoppioni, V., Strawbridge, C., Tancredi, R., Tansey, K., Thiruvahindrapduram, B., Thompson, A. P., Thomson, S., Tryfon, A., Tsiantis, J., VanEngeland, H., Vincent, J. B., Volkmar, F., Wallace, S., Wang, K., Wang, Z., Wassink, T. H., Wing, K., Wittemeyer, K., Wood, S., Yaspan, B. L., Zurawiecki, D., Zwaigenbaum, L., Betancur, C., Buxbaum, J. D., Cantor, R. M., Cook, E. H., Coon, H., Cuccaro, M. L., Gallagher, L., Geschwind, D. H., Gill, M., Haines, J. L., Miller, J., Monaco, A. P., Nurnberger, J. I. Jr, Paterson, A. D., Pericak-Vance, M. A., Schellenberg, G. D., Scherer, S. W., Sutcliffe, J. S., Szatmari, P, Vicente AM, Vieland VJ, Wijsman EM, Devlin B, Ennis S. & Hallmayer J. (2010) A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* **19**, 4072–4082.

Battaglia, A., Gurrieri, F., Bertini, E., Bellacosa, A., Pomponi, M., Paravatou-Petsotas, M., Mazza, S. & Neri, G. (1997) The inv dup(15) syndrome: A clinically recognizable syndrome with altered behaviour, mental retardation and epilepsy. *Neurology* **48**, 1081–1086.

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., Park, J., Hitz, B. C., Weng, S., Cherry, J. M. & Snyder, M. (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790–1797.

Byrne, J. A., Strautnieks, S. S., Ihrke, G., Pagani, F., Knisely, A., Linton, K. J., Mieli-Vergani, G. & Thompson, R. J. (2009) Missense mutations and single nucleotide polymorphisms in ABCB11 impair bile salt export pump processing and function or disrupt pre-messenger RNA splicing. *Hepatology* **49**, 553–567.

Caglayan, A. O. (2010) Genetic causes of syndromic and non-syndromic autism. *DevMed Child Neurol* **52**, 130–138.

Chamary, J. & Hurst, L. D. (2005) Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol* **6**, R75.

Chen, Z., Borek, D., Padrick, S. B., Gomez, T. S., Metlagel, Z., Ismail, A. M., Umetani, J., Billadeau, D. D., Otwinowski, Z. & Rosen, M. K. (2010) Structure and control of the actin regulatory WAVE complex. *Nature* **468**, 533–538.

Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., Colantuoni, E. A., Elkahloun, A. G., Herman, M. M., Weinberger, D. R. & Kleinman, J. E. (2011) Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523.

Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. (2009) Mapping complex disease traits with global gene expression. *Nat Rev Genet* **10**, 184–194.

Davis, L. K., Gamazon, E. R., Kistner-Griffin, E., Badner, J. A., Liu, C., Cook, E. H., Sutcliffe, J. S. & Cox, N. J. (2012) Loci nominally associated with autism from genome-wide analysis show enrichment of brain expression quantitative trait loci but not lymphoblastoid cell line expression quantitative trait loci. *Mol Autism* **3**, 3.

DeRubeis, S., Pasciuto, E., Li, K. W., Fernandez, E., Di Marino, D., Buzzi, A., Ostroff, L. E., Klann, E., Zwartkruis, F. J., Komiyama, N. H., Grant, S. G., Poujol, C., Choquet, D., Achsel, T., Posthuma, D., Smit, A. B. & Bagni, C. (2013) CYFIP1 coordinates mRNA translation and cytoskeleton remodeling to ensure proper dendritic spine formation. *Neuron* **79**, 1169–1182.

Depienne, C., Moreno-De-Luca, D., Heron, D., Bouteiller, D., Gennetier, A., Delorme, R., Chaste, P., Siffroi, J.-P., Chantot-Bastaraud, S., Benyahia, B., Trouillard, O., Nygren, G., Kopp, S., Johansson, M., Rastam, M., Burglen, L., Leguern, E., Verloes, A., Leboyer, M., Brice, A., Gillberg, C. & Betancur, C. (2009) Screening for genomic rearrangements and methylation abnormalities of the 15q11-q13 region in autism spectrum disorders. *Biol Psychiatry* **66**, 349–359.

Direk, K., Lau, W., Small, K. S., Maniatis, N. & Andrew, T. (2014) ABCC5 transporter is a novel type 2 diabetes susceptibility gene in European and African American populations. *Ann Hum Genet* **78**, 333–344.

Doerge, R. W. (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* **3**, 43–52.

Dubois, P. C., Trynka, G., Franke, L., Hunt, K. A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G. A., dány, R. & Aromaa, A. (2010) Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* **42**, 295–302.

ENCODE Project Consortium, (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

Gibson, G. & Weir, B. (2005) The quantitative genetics of transcription. *Trends Genet* **21**, 616–623.

Heinzen, E. L., Ge, D., Cronin, K. D., Maia, J. M., Shianna, K. V., Gabriel, W. N., Welsh-Bohmer, K. A., Hulette, C. M., Denny, T. N. & Goldstein, D. B. (2008) Tissue-specific genetic control of splicing: Implications for the study of complex traits. *PLoS Biol* **6**, e1000001.

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

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S. M., Habegger, L., Rozowsky, J., Shi, M., Urban, A. E., Hong, M.-Y., Karczewski, K. J., Huber, W., Weissman, S. M., Gerstein, M. B., Korbel, J. O. & Snyder, M. (2010) Variation in transcription factor binding among humans. *Science* **328**, 232–235.

Knoll, J. H. M., Nicholls, R. D., Magenis, R. E., Graham, J. M., Lalande, M., Latt, S. A., Opitz, J. M. & Reynolds, J. F. (1989) Angelman and Prader-Willi syndromes share a common

chromosome 15 deletion but differ in parental origin of the deletion. *Am J Med Genet* **32**, 285–290.

Kobayashi, K., Kuroda, S., Fukata, M., Nakamura, T., Nagase, T., Nomura, N., Matsuura, Y., Yoshida-Kubomura, N., Iwamatsu, A. & Kaibuchi, K. (1998) p140Sra-1 (specifically Rac1-associated protein) is a novel specific target for Rac1 small GTPase. *J Biol Chem* **273**, 291–295.

Kumar, A., Girimaji, S. C., Duvvari, M. R. & Blanton, S. H. (2009) Mutations in< i>STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. *Am J Hum Genet* **84**, 286–290.

Miyata, T., Maeda, T. & Lee, J. E. (1999) NeuroD is required for differentiation of the granule cells in the cerebellum and hippocampus. *Genes Dev* **13**, 1647–1652.

Napoli, I., Mercaldo, V., Boyl, P. P., Eleuteri, B., Zalfa, F., DeRubeis, S., Di Marino, D., Mohr, E., Massimi, M., Falconi, M., Witke, W., Costa-Mattioli, M., Sonenberg, N., Achsel, T. & Bagni, C. (2008) The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP. *Cell* **134**, 1042–1054.

Nica, A. C. & Dermitzakis, E. T. (2008) Using gene expression to investigate the genetic basis of complex disorders. *Hum Mol Genet* **17**, R129–R134.

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E. & Cox, N. J. (2010) Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet* **6**, e1000888.

Oguro-Ando, A., Rosensweig, C., Herman, E., Nishimura, Y., Werling, D., Bill, B. R., Berg, J. M., Gao, F., Coppola, G., Abrahams, B. S. & Geschwind, D. H. (2014) Increased CYFIP1 dosage alters cellular and dendritic morphology and dysregulates mTOR. *Mol Psychiatry*, 1–10.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., DeBakker, P. I., Daly, M. J. & Sham, P. C. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. & Lander, E. S. (2001) Linkage disequilibrium in the human genome. *Nature* **411**, 199–204.

Sahoo, T., Peters, S. U., Madduri, N. S., Glaze, D. G., German, J. R., Bird, L. M., Barbieri-Welge, R., Bichell, T. J., Beaudet, A. L. & Bacino, C. A. (2006) Microarray based comparative genomic hybridization testing in deletion bearing patients with Angelman syndrome: Genotype-phenotype correlations. *J Med Genet* **43**, 512–516.

Schwab, M. H., Bartholomae, A., Heimrich, B., Feldmeyer, D., Druffel-Augustin, S., Goebbels, S., Naya, F. J., Zhao, S., Frotscher, M. & Tsai, M.-J. (2000) Neuronal basic helix-loop-helix proteins (NEX and BETA2/Neuro D) regulate terminal granule cell differentiation in the hippocampus. *J Neurosci* **20**, 3714–3724.

Smith, R. M., Banks, W., Hansen, E., Sadee, W. & Herman, G. E. (2014) Family-based clinical associations and functional characterization of the serotonin 2A receptor gene (HTR2A) in autism spectrum disorder. *Autism Res* **7**, 459–467.

Stephenson, D. T., O'neill, S. M., Narayan, S., Tiwari, A., Arnold, E., Samaroo, H. D., Du, F., Ring, R. H., Campbell, B. & Pletcher, M. (2011) Histopathologic characterization of the BTBR mouse model of autistic-like behavior reveals selective changes in neurodevelopmental proteins and adult hippocampal neurogenesis. *MolAutism* **2**, 1–22.

Szalai, A. J., Wu, J., Lange, E. M., Mccrory, M. A., Langefeld, C. D., Williams, A., Zakharkin, S. O., George, V., Allison, D. B., Cooper, G. S., Xie, F., Fan, Z., Edberg, J. C. & Kimberly, R. P. (2005) Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *J Mol Med* **83**, 440–447.

VanDer Zwaag, B., Staal, W. G., Hochstenbach, R., Poot, M., Spierenburg, H. A., DeJonge, M. V., Verbeek, N. E., Van'T Slot, R., VanEs, M. A., Staal, F. J., Freitag, C. M., Buizer-Voskamp, J. E., Nelen, M. R., VanDen Berg, L. H., VanAmstel, H. K. P., VanEngeland, H. & Burbach, J. P. H. (2010) A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder. *Am J Med Genet B: Neuropsychiatr Genet* **153B**, 960–966.

Villani, A.-C., Lemire, M., Fortin, G., Louis, E., Silverberg, M. S., Collette, C., Baba, N., Libioulle, C., Belaiche, J. & Bitton, A. (2008) Common variants in the NLRP3 region contribute to Crohn's disease susceptibility. *Nat Genet* **41**, 71–76.

Vorstman, J., Staal, W. G., VanDaalen, E., VanEngeland, H., Hochstenbach, P. F. R. & Franke, L. (2005) Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol Psychiatry* **11**, 18–28.

Waltes, R., Duketis, E., Knapp, M., Anney, R. J., Huguet, G., Schlitt, S., Jarczok, T. A., Sachse, M., Kampfer, L. M., Kleinbock, T., Poustka, F., Bolte, S., Schmotzer, G., Voran, A., Huy, E., Meyer, J., Bourgeron, T., Klauck, S. M., Freitag, C. M. & Chiocchetti, A. G. (2014) Common variants in genes of the postsynaptic FMRP signalling pathway are risk factors for autism spectrum disorders. *Hum Genet* **133**, 781–792.

Wang, D., Johnson, A. D., Papp, A. C., Kroetz, D. L. & Sadee, W. (2005) Multidrug resistance polypeptide 1 (MDR1, ABCB1) variant 3435C>T affects mRNA stability. *Pharmacogenet Genomics* **15**, 693–704.

Wendland, J. R., Moya, P. R., Timpano, K. R., Anavitarte, A. P., Kruse, M. R., Wheaton, M. G., Ren-Patterson, R. F. & Murphy, D. L. (2009) A haplotype containing quantitative trait loci for SLC1A1 gene expression and its association with obsessive-compulsive disorder. *Arch Gen Psychiatry* **66**, 408–416.

Woods, C. G., Bond, J. & Enard, W. (2005) Autosomal recessive primary microcephaly (MCPH): A review of clinical, molecular, and evolutionary findings. *Am J Hum Genet* **76**, 717–728.

Xia, K., Guo, H., Hu, Z., Xun, G., Zuo, L., Peng, Y., Wang, K., He, Y., Xiong, Z., Sun, L., Pan, Q., Long, Z., Zou, X., Li, X., Li, W., Xu, X., Lu, L., Liu, Y., Hu, Y., Tian, D., Long, L., Ou, J., Liu, Y., Li, X., Zhang, L., Pan, Y., Chen, J., Peng, H., Liu, Q., Luo, X., Su, W., Wu, L., Liang, D., Dai, H., Yan, X., Feng, Y., Tang, B., Li, J., Miedzybrodzka, Z., Xia, J., Zhang, Z., Luo, X., Zhang, X., St Clair, D., Zhao, J. & Zhang, F. (2014) Common genetic variants on 1p13.2 associate with risk of autism. *Mol Psychiatry* **19**, 1212–1219.

Zhao, Q., Li, T., Zhao, X., Huang, K., Wang, T., Li, Z., Ji, J., Zeng, Z., Zhang, Z., Li, K., Feng, G., St Clair, D., He, L. & Shi, Y. (2013) Rare CNVs and tag SNPs at 15q11.2 are associated with schizophrenia in the Han Chinese population. *Schizophr Bull* **39**, 712–719.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplementary Methods: Calculation of ORs for transmission versus nontransmission of genotype combinations to probands.

Figure S1: Quantile-quantile plots for *CYFIP1* mRNA expression data used in this study.

Figure S2: Identification of high- and low-expression alleles for seven genotyped candidate *CYFIP1* eQTLs.

Figure S3: R^2-Δ^2 analysis shows that five unlinked, genotyped CYFIP1 SNPs identified as independent contributors to CYFIP1 mRNA expression in DLPFC account for most of the variance in linear regression R^2 values obtained for 59 genotyped SNPs (A) or 272 genotyped or imputed SNPs (B) within a 135 kb segment of Chromosome 15 containing the CYFIP1 gene.

Received: 19 February 2015
Accepted: 8 April 2015

