UCLA
UCLA Previously Published Works

Title

A rare functional noncoding variant at the GWAS-implicated MIR137/MIR2682 locus might confer risk to schizophrenia and bipolar disorder.

Permalink

https://escholarship.org/uc/item/61f7f9j7

Journal

American journal of human genetics, 95(6)

ISSN

0002-9297

Authors

Duan, Jubao
Shi, Jianxin
Fiorentino, Alessia
et al.

Publication Date

2014-12-01

DOI

10.1016/j.ajhg.2014.11.001

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

REPORT

A Rare Functional Noncoding Variant at the GWAS-Implicated *MIR137/MIR2682* Locus Might Confer Risk to Schizophrenia and Bipolar Disorder

Jubao Duan,[1,2,*] Jianxin Shi,[3] Alessia Fiorentino,[4] Catherine Leites,[1] Xiangning Chen,[5] Winton Moy,[1] Jingchun Chen,[5] Boian S. Alexandrov,[6,7] Anny Usheva,[6] Deli He,[1] Jessica Freda,[1] Niamh L. O'Brien,[4] MGS,[11] GPC,[12] Andrew McQuillin,[4] Alan R. Sanders,[1,2] Elliot S. Gershon,[2] Lynn E. DeLisi,[8] Alan R. Bishop,[7] Hugh M.D. Gurling,[4,13] Michele T. Pato,[9] Douglas F. Levinson,[10] Kenneth S. Kendler,[5] Carlos N. Pato,[9] and Pablo V. Gejman[1,2]

Schizophrenia (SZ) genome-wide association studies (GWASs) have identified common risk variants in >100 susceptibility loci; however, the contribution of rare variants at these loci remains largely unexplored. One of the strongly associated loci spans *MIR137* (*miR137*) and *MIR2682* (*miR2682*), two microRNA genes important for neuronal function. We sequenced ~6.9 kb *MIR137/MIR2682* and upstream regulatory sequences in 2,610 SZ cases and 2,611 controls of European ancestry. We identified 133 rare variants with minor allele frequency (MAF) <0.5%. The rare variant burden in promoters and enhancers, but not insulators, was associated with SZ ($p = 0.021$ for MAF < 0.5%, $p = 0.003$ for MAF < 0.1%). A rare enhancer SNP, 1:g.98515539A>T, presented exclusively in 11 SZ cases (nominal $p = 4.8 \times 10^{-4}$). We further identified its risk allele T in 2 of 2,434 additional SZ cases, 11 of 4,339 bipolar (BP) cases, and 3 of 3,572 SZ/BP study controls and 1,688 population controls; yielding combined p values of 0.0007, 0.0013, and 0.0001 for SZ, BP, and SZ/BP, respectively. The risk allele T of 1:g.98515539A>T reduced enhancer activity of its flanking sequence by >50% in human neuroblastoma cells, predicting lower expression of *MIR137/MIR2682*. Both empirical and computational analyses showed weaker transcription factor (YY1) binding by the risk allele. Chromatin conformation capture (3C) assay further indicated that 1:g.98515539A>T influenced *MIR137/MIR2682*, but not the nearby *DPYD or LOC729987*. Our results suggest that rare noncoding risk variants are associated with SZ and BP at *MIR137/MIR2682* locus, with risk alleles decreasing *MIR137/MIR2682* expression.

MicroRNA (miRNA) dysfunction has been hypothesized to play an important role in neurodevelopmental disorders such as schizophrenia (SZ) (MIM 181500).[1–3] Recent SZ genome-wide association studies (GWASs) further strengthen an etiological role for miRNAs. Among >100 genome-wide significant (GWS) SZ risk loci, the *MIR137/MIR2682* locus at 1p21.3 is one of the most strongly associated.[4–10] The GWS SZ risk variants are also associated with the impaired dorsolateral prefrontal cortex hyperactivation[11] and prefrontal-hippocampal functional connectivity.[12] Common GWS ($p \leq 5 \times 10^{-8}$) variants cluster around *MIR137* (MIM 614303) and *MIR2682*, with much weaker association extending to *dihydropyrimidine dehydrogenase* (*DPYD* [MIM 612779]) (Figure 1A). Large-scale exome sequencing[13,14] did not identify SZ-associated variants in coding regions of *DPYD* or within *MIR137/MIR2682* that could explain the association, thus implying the importance of noncoding variants in conferring SZ risk at this locus. *MIR137* is abundantly expressed in brain, enriched at neuronal synapses,[15] and regulates neuronal differentiation, migration, and dendritogenesis.[16–20] Interestingly, ~25% of SZ GWAS loci contain *MIR137* targets (predicted by TargetScan),[4,7,9,10] including several empirically validated targets *CACNA1C* (MIM 114205), *ZNF804A* (MIM 612282), *TCF4* (MIM 602272), *CSMD1* (MIM 608397), and *C10orf26* (MIM 611129),[21,22] suggesting a central hub role for *MIR137* in a SZ susceptibility gene network. *MIR137* has also been shown to target a large number of genes associated with autism spectrum disorders (ASD [MIM 209850]).[23] Although *MIR2682* has no known function, it is predicted (TargetScan) to target *ankyrin 3* (*ANK3* [MIM 600465]), a gene previously found to be associated with BP (MIM 125480) in GWAS.[24–26] *MIR137/MIR2682* thus represents a SZ risk locus with strong biology relevant to SZ. Rare deletions of genomic segments flanking *MIR137/MIR2682* have been reported in individuals with intellectual disability (ID)[15] and ASD.[27,28] Although we previously ruled out rare and large copy-number variants (CNVs) at this locus in our SZ GWAS sample,[29] it remained to be explored whether there were any rare SNPs or small

[1]Center for Psychiatric Genetics, Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL 60201, USA; [2]Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago, IL 60637, USA; [3]Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA; [4]Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London WC1E 6JJ, UK; [5]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA 23298, USA; [6]Harvard Medical School, Boston, MA 02115, USA; [7]Los Alamos National Laboratory, Los Alamos, NM 87544, USA; [8]VA Boston Healthcare System, Harvard Medical School, Brockton, MA 02301, USA; [9]Department of Psychiatry and Behavioral Sciences, Keck School of Medicine at USC, Los Angeles, CA 90033, USA; [10]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, CA 94305, USA
[11]Members of the Molecular Genetics of Schizophrenia collaboration are listed in the Supplemental Data
[12]Members of the Genomic Psychiatric Cohort consortium are listed in the Supplemental Data
[13]Deceased
*Correspondence: jduan@uchicago.edu
http://dx.doi.org/10.1016/j.ajhg.2014.11.001. ©2014 by The American Society of Human Genetics. All rights reserved.

CrossMark

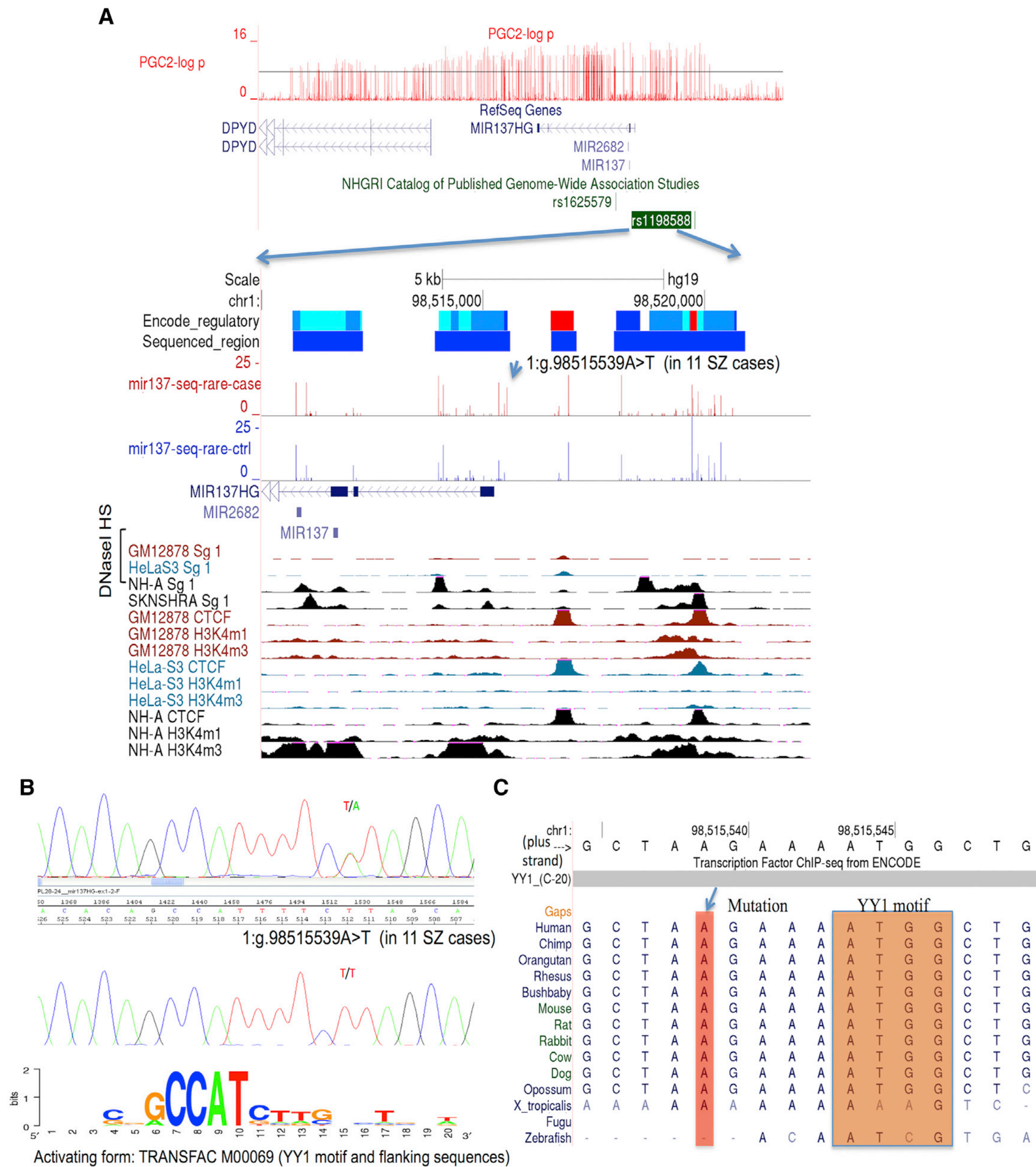

Figure 1. Genomic Features of the Sequenced *MIR137*/*MIR2682* Locus

(A) Associations at the *MIR137*/*MIR2682* SZ GWAS locus:[9] regions selected for sequencing (light blue/blue, promoter/enhancer; red, insulator) and counts of the identified sequence variants in cases (red bar) and controls (blue bar), as well as the ENCODE chromatin modification marks in LCL (GM12878), neuronal cell lines (NH-A and SKNSHRA), and HeLa cells. DNaseI HS indicate regulatory regions. CTCF track indicates transcriptional insulators. Histone methylation mark tracks for H3K4m1 and H3K4m3 indicate enhancers and promoters, respectively. Different tracks were overlaid with physical positions on UCSC genome browser. GWAS association p values were downloaded from Ricopili and expressed as –log *P*.

(B) Sequencing chromatograms at the rare enhancer SNP 1:g.98515539A>T (T/A shown on DNA minus strand) and a consensus DNA sequence motif for YY1 binding (activation form).

(c) Cross-species sequence conservation around the SNP 1:g.98515539A>T and at the YY1 motif shown in UCSC genome browser (Note: DNA plus strand sequence is shown).

indels of strong effect that could explain additional SZ risk and help to inform the functionality of common risk variants at the same GWAS loci.[30–33]

We first sequenced ~6.9 kb of *MIR137* and *MIR2682* and their upstream regulatory sequences (Figure 1A and Table S1 available online) in 2,610 SZ cases and 2,611 controls from the Molecular Genetics of SZ (MGS) EA GWASs.[4,7] NorthShore University HealthSystem's IRB approved the human subjects protocol, and proper informed consent was obtained. The selection of the region for sequencing was based on the DNaseI hypersensitive site (DHS) mapping data from ENCODE (Encyclopedia of DNA Elements[34,35])

Table 1. SZ Association of Rare Variants Grouped by Promoter/Enhancer and Insulator in 2,610 SZ Cases and 2,611 Controls from the MGS EA GWASs

	MAF < 0.5%				MAF < 0.1%			
	Minor Allele Count				Minor Allele Count			
Variant Location	SZ Cases	Controls	OR	p Value	SZ Cases	Controls	OR	p Value
Promoter/ Enhancer	223	176	1.27	0.021	105	65	1.62	0.003
Insulator (CTCF)	62	81	0.77	0.132	26	32	0.81	0.435
All	285	257	1.11	0.246	131	97	1.35	0.029

ENCODE-annotated transcriptional promoters (H3K4me3), enhancers (H3K4me1), or insulators (CTCF-binding sites) were based on chromatin histone methylation patterns[34] in neuronal NH-A cells (ChIP-seq Signal > 10). MAF, minor allele frequency in MGS controls or cases. OR, odds ratio. p value is the nominal significance calculated by two-sided Fisher's exact test. Using MAF < 1% as a cut-off (not tabulated) did not yield significant association. A total of 5,262 DNAs (2,634 SZ cases and 2,628 controls) from the MGS EA GWAS sample[4,7,9,29] were used for Sanger sequencing. 5,223 subjects (2,610 cases and 2,611 controls) remained for association analysis after sample quality control.

from neuronal cells (SK-N-SH and NH-A; Figure 1A) and in fetal brain. We further classified these putative regulatory sequences as ENCODE-annotated transcriptional promoters (H3K4me3), enhancers (H3K4me1), or insulators (CTCF-binding sites)[34] (Figure 1A and Figure S1). The PCR-amplified genomic DNA amplicons were sequenced on an ABI 3730 DNA Analyzer. The automatically (SeqScape 2.5; ABI) called SNPs and indels were manually verified, followed by extensive sequencing quality control metrics including genotype call rate (>90%), genotype concordance rate (>99.9%) between sequencing data and known GWAS genotypes,[4] and absence of Hardy–Weinberg equilibrium (HWE) departures ($p < 0.001$ in controls).

We identified 143 SNPs and Indels (Table S2), of which 133 (~93%) were rare (MAF < 0.5%). The variant density, proportion of singletons, and MAF distribution of the identified variants were all similar compared to whole exome or genome sequencing results (NHLBI-Exome Sequencing Project [ESP] and UK10K-TwinsUK)[36,37] for the same targeted region (Figure S2, Figure S3, and Table S3). The proportion of rare variants identified from our sequencing of putative regulatory sequences was similar to that of missense SNPs and frameshift or nonsense SNPs, but higher than that of intronic SNPs (Fisher's exact test $p = 0.0007$; Figure S3B) uncovered by exome sequencing of 4,298 EA subjects (NHLBI-ESP).[36] The enrichment of rare variants in the sequenced noncoding regions relative to introns was consistent with the expected functionality of the sequenced region and possible purifying selection acting on deleterious variants.[38]

We identified two common (MAF > 5%) SNPs (rs2660304 and rs2660302) in linkage disequilibrium (LD) ($r^2 = 0.74$ and 0.50, respectively, based on 1000 Genomes) with the reported SZ-associated rs1198588.[4–10] As expected, they showed nominally significant association to SZ in the sequenced MGS sample (Table S2), providing evidence of accuracy of our sequencing data. We found no association of SZ to a 15 bp Sequence Tandem Repeat (STR; rs58335419) 54 bp upstream from *MIR137* (Table S1 and Figure S4), although the STR was recently found to reduce in vitro biogenesis of mature *MIR137*.[39] As in a previous report of no association of this functional *MIR137* STR with SZ in a Japanese sample,[40] our nonreplication of the originally reported weak association ($p = 0.049$)[39] with SZ might be due to insufficient power of our MGS sample. Alternatively, because cellular RNA level is ultimately a result of multilayer of complex gene regulation predominately at transcriptional level,[41] the in vitro posttranscriptional effect of the *MIR137* STR[39] might be masked by other variant(s) with opposite functional effect and thus does not manifest in vivo phenotypic changes. A burden test aggregating all 133 rare variants did not produce evidence for association. As promoter and enhancer mutations likely reduce transcription while insulator mutations likely increase transcription, we reasoned that simply collapsing two sets of rare variants expected to have opposite directions might reduce the power. We thus performed burden tests separately for these two sets of rare variants. We found that the promoter and enhancer rare variants as a whole were significantly associated with SZ: $p = 0.021$ for variants with MAF < 0.5% in either cases or controls (to capture both risk and protective variants), and $p = 0.003$ for variants with MAF < 0.1% (in cases or controls); while insulator variants showed no case-control difference (Table 1). This highlights the importance of considering the direction of putative functional effects in association testing of rare variants in aggregate.

The association was primarily driven by a single SNP, 1:g.98515539A>T, that presented in 11 SZ cases versus 0 controls ($p = 4.8 \times 10^{-4}$, Fisher's exact test) (Table 2). We repeated the DNA sequencing for 1:g.98515539A>T and confirmed all 11 heterozygous SZ cases. Although we had a higher proportion of blood DNAs in MGS SZ cases (25%) than in controls (1%), the rare variant 1:g.98515539A>T was not overrepresented in DNAs from bloods (27%) or from LCLs (Table S4). Furthermore, we confirmed the presence of 1:g.98515539A>T by repeated sequencing of DNAs re-extracted from available blood samples of two heterozygous subjects and from an induced pluripotent stem cell (iPSC) line derived at RUCDR (Rutgers University Cell and DNA Repository) from a heterozygous subject's cryopreserved peripheral lymphocytes (Figure S5). These results suggested that the presence of 1:g.98515539A>T in 11 SZ cases was not due to technical artifacts resulted from sequencing or LCL derivation. Please note that, with 1:g.98515539A>T left out, the rare variants with MAF < 0.1% remained excess in SZ cases ($p = 0.026$; Table 1), and the association with GWAS-implicated common variants in this region did not change much in MGS sample (from $p = 0.0005$ to 0.0006). To estimate the empirical significance of the observed association with 1:g.98515539A>T, we performed 1 million random

Table 2. Single Variants Nominally Associated with SZ in 2,610 SZ Cases and 2,611 Controls from the MGS EA GWASs

Variant Name	Position (hg19)	Function	Alleles	MAF_case	MAF_ctrl	OR	p Value	Perm_*P*
1:g.98515539A>T	98,515,539	Enhancer	A:T	0.0021	0.0000	NA	0.00048	0.004
1:g.98520090G>C	98,520,090	Insulator (CTCF)	G:C	0.0002	0.0019	0.10	0.00627	NS
1:g.98519714C>G	98,519,714	Insulator (CTCF)	C:G	0.0023	0.0048	0.48	0.04627	NS
rs2660302	98,520,219	Promoter/enhancer	A:T	0.1573	0.1803	0.85	0.00807	NT
rs2660304	98,512,127	Promoter/enhancer	T:G	0.1695	0.1933	0.85	0.01314	NT

Functional regions were defined as ENCODE-annotated transcriptional promoters (H3K4me3), enhancers (H3K4me1), or insulators (CTCF-binding sites) based on chromatin histone methylation patterns[34] in neuronal NH-A cells (ChIP-seq Signal > 10). Alleles are coded on plus strand. Alleles listed as major:minor. OR, Odds ratio. p value is the nominal significance calculated by two-sided Fisher's exact test. Perm_*P* is the region-wise significance after multiple testing correcting based on one million random permutations. NS, not significant; NT, not tested due to common SNPs.

permutations for all rare SNPs in the region and determined the region-wise p value was 0.004. Interestingly, two rare insulator variants (1:g.98520090G>C and 1:g.98519714C>G) were more frequent in controls than in cases ($p = 0.006$ and 0.046, respectively, Fisher's exact test); however, the associations did not survive multiple testing correction (Table 2), and the more strongly associated 1:g.98520090G>C was not functional (see functional studies below). We further genotyped 1:g.98515539A>T in three additional EA SZ samples: (1) the Irish Case-Control Study of SZ (ICCSS) sample (1,034 cases and 658 controls)[42,43] by a customized TaqMan assay (Life Technologies), (2) the University College London (UCL) sample (900 cases and 1,314 controls) by allele-specific PCR using KASPar reagents (LGC Genomics) on a LightCycler 480 (Roche), and (3) the Genomic Psychiatric Cohort (GPC; 500 cases and 200 controls)[44] by extracting the genotype of the 1:g.98515539A>T from available whole-genome sequencing data sets and found the rare allele in 2 SZ cases and 1 control (Table 3).

Table 3. Association of 1:g.98515539A>T with SZ and BP in the Combined Samples

	Cases		Controls		
	Minor Allele T	Major Allele A	Minor Allele T	Major Allele A	Fisher p Value
MGS-SZ	11	5,177	–	5,192	
ICCSS-SZ	1	2,067	–	1,316	
UCL-SZ	1	1,799	1	2,627	
GPC-SZ	–	1,000	–	400	
NIMH-BP	2	2,090			
UCL-BP	8	3,778			
GPC-BP	1	2,799	1	2,799	
TwinsUK			1	3,205	
1000 Genome			–	170	
Combined-SZ	13	10,043	3	15,709	0.0007
Combined-BP	11	8,667			0.0013
Combined-SZ/BP	24	18,710			0.0001

MGS[4,7,9] were Sanger sequenced. ICCSS[42,43] were genotyped by TaqMan. UCL were genotyped by allele-specific PCR using KASPar reagents. GPC genotypes were extracted from whole genome sequenced data. NIMH-BP were genotyped by TaqMan. Two public whole-genome sequence (WGS) data sets were used as population controls (TwinsUK and 1000 Genomes). Fisher's Exact Test p values (single-sided) for the combined analysis were shown, and the combined control sample was used. UCL samples. The case sample included 1,917 BP and 932 SZ cases. All SZ or BP samples, except for UCL sample, have been previously described. UCL sample included 1,917 BP and 932 SZ cases, as well as 1,348 control subjects comprised of 868 screened subjects who had no first degree family or personal history of psychiatric illness and an additional 480 unscreened normal British subjects obtained from the European Collection of Cell Cultures (ECACC). BP cases had been given an NHS clinical diagnosis of BP by the International Classification of Disease 10 (ICD-10) and then needed to fulfill RDC for BP with clinical data collected by the lifetime version of the Schizophrenia and Affective Disorder Schedule (SADS-L).[76] The SZ cases were diagnosed in an analogous manner. National Health Service (NHS) multicenter research ethics approval was obtained for UCL sample.

Because SZ and BP share genetic architecture,[45] we also genotyped 1:g.98515539A>T in three EA BP samples: (1) NIMH-BP (1,046 cases)[46,47] by a customized TaqMan assay (Life Technologies), (2) UCL (1,893 cases) by allele-specific PCR using KASPar reagents (LGC Genomics), and (3) GPC (1,400 cases and 1,400 controls) by extracting the genotype of the 1:g.98515539A>T from available whole-genome sequencing data sets. We found 1:g.98515539A>T in 11 BP cases and 1 control (Table 3). Furthermore, we identified 1 heterozygous subject in whole-genome sequenced population controls (publically available; TwinsUK: 1,603 controls, 1000 Genomes: 85 controls) (Table S3). In the combined case (i.e., case-control) samples, the association of 1:g.98515539A>T with SZ, BP, and SZ and BP combined was 0.0007, 0.0013, and 0.0001, respectively (Table 3). We also examined 1:g.98515539A>T, in two family collections: (1) Clinical Neurogenetics (CNG) SZ families (307 subjects in 67 pedigrees)[48,49] by Sanger sequencing and (2) Irish Study of High-Density SZ Families (ISHDSF) (1,400 subjects in 274 pedigrees)[42,43] by a customized TaqMan assay (Life Technologies). We identified one SZ proband and an unaffected father in a CNG family to be heterozygous for 1:g.98515539A>T. The presence of the rare risk allele in an unaffected parent is consistent with incomplete penetrance, which is even commonly seen for SZ-associated rare CNVs of large effect[50] and for causal variants in most complex disorders.[51]

Next, we investigated whether allele T of 1:g.98515539A>T in the 11 SZ MGS EA cases originated from the same haplotypic background. We merged the genotypes of 1:g.98515539A>T with MGS GWAS genotypes[4]

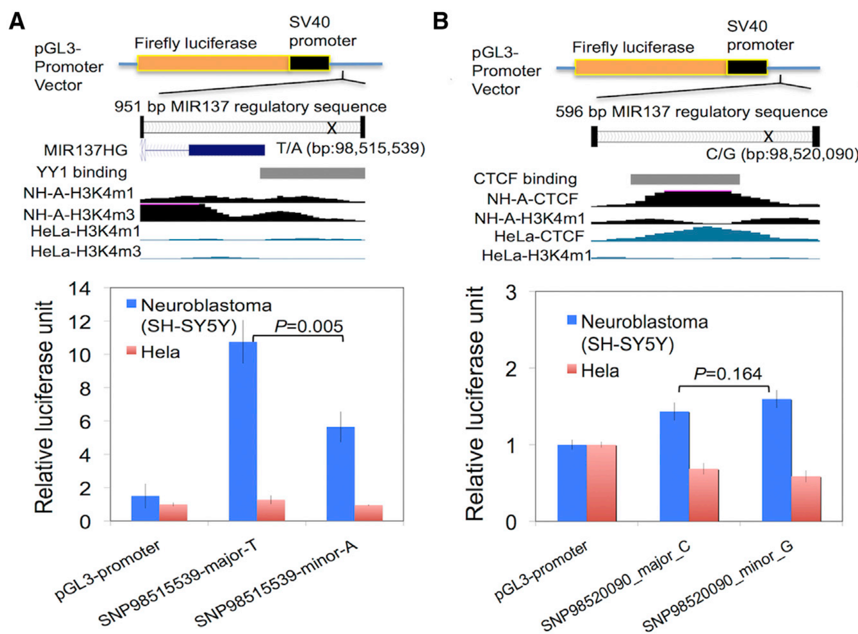

Figure 2. Reporter Gene Assay for Variant Functionality in Neuroblastoma (SH-SY5Y) and HeLa Cell Lines
Putative regulatory sequence flanking the rare enhancer (A) SNP (bp: 98,515,539; hg19) or the CTCF (B) SNP (bp: 98,520,090; hg19) was cloned in a configuration of minus strand upstream of the SV40 promoter of pGL3-promoter vector. The resultant reporter gene vectors carrying regulatory sequences with different alleles of a SNP were transiently transfected into neuroblastoma (SH-SY5Y) and HeLa cell lines, and the relative luciferase activity was measured for each reporter gene construct. Allele-specific effects on enhancer activity or insulator activity of the cloned sequence were expressed as the relative luciferase unit, with the reporter gene expression of pGL3-promoter (no insert) as a control for basal SV40 promoter activity. Note that in (B), the cloned sequence shows an enhancer activity in SH-SY5Y cells and repressor activity in HeLa cells, which is consistent with ENCODE functional annotations (CTCF and H3K4m1 peaks in NH-A, but only CTCF peak in HeLa). For normalizing the transfection efficiencies between different constructs, the firefly luciferase constructs were cotransfected with the *Renilla* luciferase pRL-CMV Vector (Promega). We used Lipofectamine 2000 Transfection Reagent (Invitrogen) for the transfection when the cultured cells are at 60% confluence for SH-SY5Y and 90% confluence for HeLa cells on 96-well plates. 48 hr after transfection, we measured the firefly and *Renilla* luciferase activities using Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol. Data were from 3 independent experiments and expressed as mean ± SD. Statistical significance was generated by Student's t test.

and phased the *MIR137* local region using BEAGLE[52] for all subjects. The T alleles were on the same haplotype, suggesting that they were inherited from the same common ancestor. The shared common-SNP haplotype spans ~77.2 kb (from rs1938567 to rs1198588) (Table S5) with a frequency of 0.787 in MGS controls, 0.792 in TwinsUK controls, and 0.790 in a publically available GWAS sample with 9,562 subjects of European ancestry (The Atherosclerosis Risk in Communities-ARIC Study data set downloaded from dbGAP). These results suggest that the rare risk allele is not just from an unusual ancestry that happens to carry this rare SNP on the haplotype, and our MGS controls are not an atypical EA subpopulation. To further assess any confounding effect of population stratification on our observed association with 1:g.98515539A>T, we analyzed the genotypic ancestry principal components (PCs)[4] of the MGS cases carrying 1:g.98515539A>T and other MGS subjects (Figure S6). MGS cases and controls appeared to be matched very well, with northern European population as a predominate cluster. The MGS cases carrying 1:g.98515539A>T spread all over the main northern European cluster (Figure S6), indicating that these individuals with the rare variant were not preferentially from a small regional subpopulation. For the SZ subjects carrying 1:g.98515539A>T, we also calculated kinship coefficients using genome-wide SNPs with SNP frequencies estimated based on the whole MGS sample. Out of all possible pairs of individuals with the rare variant 1:g.98515539A>T, only one was greater than 0.02, nine were close to 0.01, and the rest of them were zero. This result demonstrated that the SZ cases carrying 1:g.98515539A>T are unlikely from a localized region, which is consistent with ancestry principal components analysis. From all of these analyses, we concluded that the observed disease association with 1:g.98515539A>T was unlikely due to population stratification.

SNP 1:g.98515539A>T is within an ENCODE-annotated neuronal cell (NH-A and SKNSH) specific enhancer and promoter ~3.6 kb upstream of *MIR137* (Figure 1A). It is 5 bp away from the binding motif (CCAT) of TF YY-1 (gene activating form; Figure 1B); the flanking sequence of the YY1 motif with the major allele of 1:g.98515539A>T is highly conserved in mammals (Figure 1C). We hypothesized that the risk allele of 1:g.98515539A>T reduced enhancer activity by interfering with YY1 binding. To test the hypothesis, we cloned the putative enhancer sequence (the transcribed minus strand) flanking 1:g.98515539A>T into a luciferase reporter gene vector (pGL3-promoter) (Figure 2A; Table S6). As expected from the ENCODE functional annotation, the cloned sequence exhibited robust enhancer activity as indicated by luciferase expression in a human neuroblastoma cell line (SH-SY5Y), but not in HeLa cells. The enhancer activity of the cloned sequence was reduced ~50% by the risk allele A relative to the major allele T (we refer to the nucleotide on the minus strand hereinafter for the functional study section) of 1:g.98515539A>T (Figure 2A). However, 1:g.98520090G>C (nominally overrepresented in controls; $p = 0.006$), near an insulator (CTCF binding-site), did not display any effect on transcription in the reporter gene assay (Figure 2B) or on open chromatin (regulatory elements) in a formaldehyde-assisted isolation

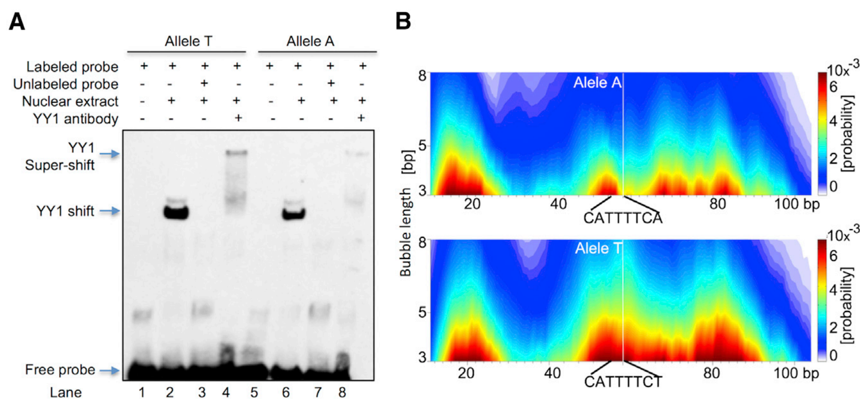

Figure 3. The Rare Allele A of 1:g.98515539A>T Reduces TF YY1 Binding
(A) EMSA for YY1 binding. A double-stranded oligonucleotide (29 bp; on minus strand) flanking the YY1-binding site and 1:g.98515539A>T (allele T or A; AGAGGTGCTGTGAACACACAGC-CATTTTC t/a TAGCAGCTTTTTGACTG TATGTTACCATA) was incubated with the nuclear extracts of neuroblastoma (SH-SY5Y) cells. Oligonucleotide probe (20 fmol) with allele T (lane 2) produced a much denser band of specific DNA-protein binding complex than the risk allele A (lane 6). The DNA-protein complex was recognized by antibody of YY1 (C-20; sc-281; rabbit polyclonal; Santa Cruz Biotechnology), showing up as a super-shift band (lanes 4 and 8), but weaker for allele A. The specificity of the YY1 binding is shown by the abolishment of the DNA-YY1 binding complex in the presence of excess (50 pmol) unlabeled probe (lanes 3 and 7). The EMSA was performed using the LightShift Chemiluminescent EMSA kit (Thermo Scientific). Binding reactions were separated by electrophoresis on 6% polyacrylamide gels on XCell SureLock Mini-Cell Electrophoresis System (Life Technologies).
(B) Computing models of the effect of 1:g.98515539A>T on DNA breathing dynamics (~100 bp flanking the SNP site) that is necessary for binding of TFs. Langevin molecular dynamic (LMD) simulations, based on the Extended Peyrard-Bishop-Dauxois (EPBD) nonlinear model of DNA, assessed bubble-formation probability and average strand separation as the mechanistic parameters characterizing the TF binding activity of the sequence.[54,75] The sequences are simulated in 1,000 separate realizations. Sequence containing allele A in close proximity to YY1 binding site predicts weak DNA breathing activity (top), in contrast to the much stronger breathing potential of the sequence with allele T (bottom). The white horizontal lines mark the SNP sites. The YY1 binding sequence ending at the SNP site is shown below the SNP line. Vertical axis, bubble length in bp; color axis, bubble probability at specific nucleotide positions (horizontal axis) predicted for DNA openings with amplitude > 2Å at temperature of 310°K.

of regulatory elements (FAIRE) assay[53] in lymphoblastoid cell lines (LCLs; Figure S7).

We further carried out an electrophoresis mobility shift assay (EMSA) to examine whether the reduced enhancer activity by allele A of 1:g.98515539A>T correlated with altered DNA binding to TFs in SH-SY5Y cells. We found that the YY1 motif-flanking DNA probe carrying the risk allele A had a much weaker YY1 binding capacity than the probe with the major allele T (Figure 3A). Consistent with the EMSA result, computational analysis of local DNA breathing dynamics also predicted an effect of 1:g.98515539A>T on differential local DNA breathing associated with specific TF binding[54,55] (Figure 3B). Major allele T features a conformational DNA dynamics (probability for bubble formation $p = 5.8 \times 10^{-4}$) that favors strong YY1 binding, while the risk allele A nearly silences the DNA breathing in the vicinity of the YY1 binding sequence (Figure 3B). This effect could explain the weak YY1 binding associated with allele A in EMSA (Figure 3A). Allele T distinctively activates bubble formation with significantly higher probability that favors strong specific YY1 binding.[56]

Although 1:g.98515539A>T at the *MIR137/MIR2682* locus is proximal to the transcription start of the *MIR137 host gene* (*MIR137HG*), the enhancer sequence surrounding the SNP could still affect other adjacent genes (*DPYD* and *LOC729987*) (Figure 4A) through long-range regulation.[57,58] We therefore performed the 3C assay[58–60] in SH-SY5Y cells to examine whether the 1:g.98515539A>T site can physically interact with core promoters of *DPYD* and *LOC729987* (Figure 4A). We identified specific physical interaction of the enhancer sequence flanking 1:g.98515539A>T with other putative regulatory sequences upstream of *MIR137/MIR2682*, but not with the core promoters of *DPYD* or *LOC729987* (Figure 4B). This suggests the functional 1:g.98515539A>T might influence expression of *MIR137/MIR2682*, but not of *DPYD* or *LOC729987*.

Together, our reporter gene assay, EMSA, and the 3C experiment on the rare SNP 1:g.98515539A>T suggest a mechanism of reduced expression of *MIR137/MIR2682* as contributing to SZ risk at this locus. This is consistent with a recent report showing an association of the SZ risk allele of the GWAS-implicated rs1625579 with decreased *MIR137* expression in SZ postmortem brain tissue.[61] The regulatory effect of SNP 1:g.98515539A>T is also consistent with the possible functional mechanism underlying the reported microdeletions involving MIR137 in ASD[27,28] and ID individuals.[15] Although we have confirmed the transcriptional effect of the SZ-associated rare SNP 1:g.98515539A>T, we cannot rule out other functional variants at *MIR137* locus.[39] Neurons differentiated from induced pluripotent stem cells (iPSCs) generated from SZ cases carrying the rare risk allele would be an appropriate experimental model to test the functional effect of 1:g.98515539A>T on *MIR137/MIR2682* expression and on SZ-relevant cellular and physiological phenotypes.[62,63] We found a dramatic increase of *MIR137/MIR2682* expression during the dopaminergic neuronal differentiation and maturation from iPSCs[64] (Figure S8). Such isogenic human iPSC-derived neurons differing only at the functional SNP site will also be an invaluable cellular model to study the downstream molecular target of *MIR137* in a more disease-relevant physiological condition, complementing the current knowledge of genome-wide transcriptional effect

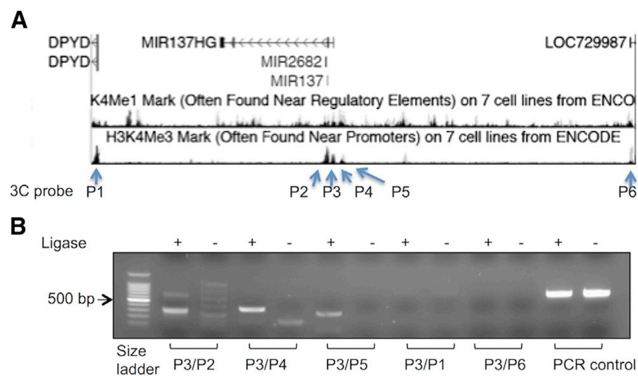

Figure 4. Chromatin Conformation Capture (3C) for Detecting the Physical Interaction between the Regulatory Sequences of *MIR137/MIR2682* Locus in SH-SY5Y Cells

(A) Genomic locations of 3C probes (P1 to P6) on UCSC genome browser tracks along ~300 kb region (chr1:98,378,000-98,688,000). Each probe is a plus-strand sequence ~150 bp upstream of a HindIII cutting site. P3 is the bait probe adjacent to the enhancer sequence where 1:g.98515539A>T resides. P2, P4, and P5 target the other sequenced regulatory regions upstream *MIR137/MIR2682*. P1 and P5 target the core promoters of the flanking genes *DPYD* and *LOC729987*, respectively.

(B) PCR products from different pairs of 3C probes on an agarose gel (2%). The interacting genomic regions in SH-SY5Y cells were captured and enriched by sequentially cross-linking of chromatin, HindIII cutting of cross-linked chromatins, and religation of HindIII-digested DNAs. Only physically interacting genomic regions were enriched and produced specific amplification (corresponding to each pair of 3C probes) in PCR. The specific PCR products were verified by the expected size (Table S7) and by direct DNA sequencing. The HindIII-digested but unligated (i.e., no ligase) DNAs served as negative controls. PCR controls are an amplicon within two consecutive HindIII cutting sites and thus this shows specific amplification for both ligated and unligated chromatins.

of *MIR137* obtained from overexpressing *MIR137* in human neuronal stem cell line.[65,66]

In summary, we have identified a functional rare noncoding risk variant for SZ at one of the most strongly associated SZ susceptibility loci from GWASs, highlighting the importance of rare noncoding variants in SZ genetics. Rare functional variants identified at GWAS-implicated loci not only explain additional genetic risk but also can provide unparalleled direct links to causal variants or mechanisms given their relatively larger effects compared to common risk variants at the same locus.[30,31,67–70] The existence of rare noncoding SZ-risk variants at the SZ GWAS-implicated *MIR137/MIR2682* locus is in line with recently reported polygenic burden of rare disruptive (nonsense, essential splicing site, and frameshift) variants in genes implicated by SZ GWAS and CNV studies.[13] Our study also demonstrates an approach of analyzing rare noncoding variants based on a priori knowledge of directional functionality of a variant; a higher burden of rare noncoding variants in SZ cases was only observed in putative promoter and enhancer regions, but not in transcriptional insulators. Moreover, we have shown that a rare functional variant at *MIR137/MIR2682* locus might confer risk for developing both SZ and BP, although common variants at this locus have been suggested to confer shared risk effect across major psychiatric disorders.[71] Indeed, rare CNVs associated with a psychiatric disorder frequently show variable phenotypic expressivity, i.e., with shared risk to other psychiatric disorders or neurodevelopmental conditions.[72] We acknowledge that although the association of 1:g.98515539A>T with SZ in the MGS sample remained significant after multiple-testing correction ($p = 0.004$), it did not reach genome-wide significance, and additional confirmation in independent samples will be necessary to definitively establish the association. It is however noteworthy that replicating the association with rare variants can be challenging, even in larger samples.[31,73,74] For SZ, there has been no report of a single rare variant showing genome-wide significant association, even in recent large-scale exome-sequencing studies.[13,14] Given that we have provided evidence that technical artifacts or population stratification are unlikely to explain our results, our study warrants further resequencing efforts with more comprehensive coverage of putative regulatory sequences at the *MIR137/MIR2682* locus, confirmation of the observed association in larger independent samples, and a deeper understanding of the causal link between SZ risk alleles and disease phenotypes.

Supplemental Data

Supplemental Data include eight figures and seven tables and can be found with this article online at http://www.cell.com/ajhg.

Acknowledgments

We thank the study participants of MGS, CNG, ICCSS, ISHDSF, UCL, GPC, NIMH-BP collections. This study also makes use of whole-genome sequencing data (TwinsUK) generated by the UK10K Consortium. A full list of the investigators who contributed to the generation of the data is available from UK10K Project homepage. Wellcome Trust award WT091310 provided funding for UK10K. We also thank K. Fang and N. Park (Illinois Mathematics and Science Academy) for their technical help with the 3C experiment. This work was primarily supported by R01MH059571, R01MH081800, and U01MH079469 (to P.V.G.) and other NIH grants for MGS (R01MH067257 to N.G.B., R01MH059588 to B.J.M., R01MH059565 to R.F., R01MH059587 to F.A., R01MH060870 to W.F.B., R01MH059566 to D.W.B., R01MH059586 to J.M.S., R01MH061675 to D.F.L., R01MH060879 to C.R.C., U01MH046276 to C.R.C., and U01MH079470 to D.F.L). GPC was supported by MH085548 and MH085542 (to C.N.P and M.T.P). UCL genotyping was supported by MRC grant G1000708 (to H.M.D.G. and A.M.). The computational modeling of transcription factor binding at Los Alamos National Laboratory was carried out under the auspices of the National Nuclear Security Administration of the US Department of Energy and was supported by the LANL, LDRD, 20110516ECR grant (to B.S.A.). This work was also partially supported by National Institutes of Health (NIH) grant R21MH102685 and NorthShore University HealthSystem Research Career Development Award (to J.D.).

Received: August 25, 2014
Accepted: November 3, 2014
Published: November 26, 2014

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
ANNOVAR, http://www.openbioinformatics.org/annovar/
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
ENCODE, http://genome.ucsc.edu/ENCODE/
GATK, http://www.broadinstitute.org/gatk/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
Ricopili, www.broadinstitute.org/mpg/ricopili/
UCSC Genome Browser, http://genome.ucsc.edu
UK10K Consortium, http://www.uk10k.org/

Accession Numbers

The accession number for SNP 1:g.98515539A>T is ss1425684360. Previously unreported SNPs or indels identified in our study were submitted to dbSNP with accession numbers from ss1425684307 to ss1425684424.

References

1. Im, H.I., and Kenny, P.J. (2012). MicroRNAs in neuronal function and dysfunction. Trends Neurosci. *35*, 325–334.
2. Guarnieri, D.J., and DiLeone, R.J. (2008). MicroRNAs: a new class of gene regulators. Ann. Med. *40*, 197–208.
3. Sun, E., and Shi, Y. (2014). MicroRNAs: Small molecules with big roles in neurodevelopment and diseases. Exp. Neurol. Published online August 13, 2014. http://dx.doi.org/10.1016/j.expneurol.2014.08.005.
4. Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J., et al. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature *460*, 753–757.
5. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P.; International Schizophrenia Consortium (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752.
6. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P., Mors, O., Mortensen, P.B., et al.; Genetic Risk and Outcome in Psychosis (GROUP) (2009). Common variants conferring risk of schizophrenia. Nature *460*, 744–747.
7. Consortium, S.P.G.-W.A.S.G. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat Genet *43*, 969–976.
8. Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R.; Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of Schizophrenia Collaboration (MGS) (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat. Genet. *44*, 247–250.
9. Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al.; Multicenter Genetic Studies of Schizophrenia Consortium; Psychosis Endophenotypes International Consortium; Wellcome Trust Case Control Consortium 2 (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. *45*, 1150–1159. http://dx.doi.org/10.1038/ng.2742.
10. Consortium., S.W.G.o.t.P.G. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.
11. van Erp, T.G., Guella, I., Vawter, M.P., Turner, J., Brown, G.G., McCarthy, G., Greve, D.N., Glover, G.H., Calhoun, V.D., Lim, K.O., et al. (2013). Schizophrenia miR-137 Locus Risk Genotype is Associated with Dorsolateral Prefrontal Cortex Hyperactivation. Biol Psychiatry *75*, 398–405.
12. Liu, B., Zhang, X., Hou, B., Li, J., Qiu, C., Qin, W., Yu, C., and Jiang, T. (2014). The impact of MIR137 on dorsolateral prefrontal-hippocampal functional connectivity in healthy subjects. Neuropsychopharmacology *39*, 2153–2160.
13. Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kähler, A., et al. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. Nature *506*, 185–190.
14. Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., et al. (2014). De novo mutations in schizophrenia implicate synaptic networks. Nature *506*, 179–184.
15. Willemsen, M.H., Vallès, A., Kirkels, L.A., Mastebroek, M., Olde Loohuis, N., Kos, A., Wissink-Lindhout, W.M., de Brouwer, A.P., Nillesen, W.M., Pfundt, R., et al. (2011). Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. J. Med. Genet. *48*, 810–818.
16. Sun, G., Ye, P., Murai, K., Lang, M.F., Li, S., Zhang, H., Li, W., Fu, C., Yin, J., Wang, A., et al. (2011). miR-137 forms a regulatory loop with nuclear receptor TLX and LSD1 in neural stem cells. Nat Commun *2*, 529.
17. Silber, J., Lim, D.A., Petritsch, C., Persson, A.I., Maunakea, A.K., Yu, M., Vandenberg, S.R., Ginzinger, D.G., James, C.D., Costello, J.F., et al. (2008). miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. BMC Med. *6*, 14.
18. Smrt, R.D., Szulwach, K.E., Pfeiffer, R.L., Li, X., Guo, W., Pathania, M., Teng, Z.Q., Luo, Y., Peng, J., Bordey, A., et al. (2010). MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. Stem Cells *28*, 1060–1070.
19. Szulwach, K.E., Li, X., Smrt, R.D., Li, Y., Luo, Y., Lin, L., Santistevan, N.J., Li, W., Zhao, X., and Jin, P. (2010). Cross talk between microRNA and epigenetic regulation in adult neurogenesis. J. Cell Biol. *189*, 127–141.
20. Volvert, M.L., Rogister, F., Moonen, G., Malgrange, B., and Nguyen, L. (2012). MicroRNAs tune cerebral cortical neurogenesis. Cell Death Differ. *19*, 1573–1581.
21. Kim, A.H., Parker, E.K., Williamson, V., McMichael, G.O., Fanous, A.H., and Vladimirov, V.I. (2012). Experimental validation of candidate schizophrenia gene ZNF804A as target for hsa-miR-137. Schizophr. Res. *141*, 60–64.
22. Kwon, E., Wang, W., and Tsai, L.H. (2013). Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. Mol. Psychiatry *18*, 11–12.
23. Devanna, P., and Vernes, S.C. (2014). A direct molecular link between the autism candidate gene RORa and the schizophrenia candidate MIR137. Scientific Reports *4*, 3994.
24. Chen, D.T., Jiang, X., Akula, N., Shugart, Y.Y., Wendland, J.R., Steele, C.J., Kassem, L., Park, J.H., Chatterjee, N., Jamain, S.,

et al.; BiGS (2013). Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. Mol. Psychiatry *18*, 195–205.

25. Schulze, T.G., Detera-Wadleigh, S.D., Akula, N., Gupta, A., Kassem, L., Steele, J., Pearl, J., Strohmaier, J., Breuer, R., Schwarz, M., et al.; NIMH Genetics Initiative Bipolar Disorder Consortium (2009). Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. Mol. Psychiatry *14*, 487–491.
26. Ferreira, M.A., O'Donovan, M.C., Meng, Y.A., Jones, I.R., Ruderfer, D.M., Jones, L., Fan, J., Kirov, G., Perlis, R.H., Green, E.K., et al.; Wellcome Trust Case Control Consortium (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat. Genet. *40*, 1056–1058.
27. Carter, M.T., Nikkel, S.M., Fernandez, B.A., Marshall, C.R., Noor, A., Lionel, A.C., Prasad, A., Pinto, D., Joseph-George, A.M., Noakes, C., et al. (2011). Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder. Clin. Genet. *80*, 435–443.
28. Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., et al. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am. J. Hum. Genet. *94*, 677–694.
29. Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F., et al. (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am. J. Psychiatry *168*, 302–316.
30. Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., et al.; National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC); United Kingdom Inflammatory Bowel Disease Genetics Consortium; International Inflammatory Bowel Disease Genetics Consortium (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat. Genet. *43*, 1066–1073.
31. Bonnefond, A., Clément, N., Fawcett, K., Yengo, L., Vaillant, E., Guillaume, J.L., Dechaume, A., Payne, F., Roussel, R., Czernichow, S., et al.; Meta-Analysis of Glucose and Insulin-Related Traits Consortium (MAGIC) (2012). Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat. Genet. *44*, 297–301.
32. Blain, H., Jaussent, A., Thomas, E., Micallef, J.P., Dupuy, A.M., Bernard, P.L., Mariano-Goulart, D., Cristol, J.P., Sultan, C., Rossi, M., and Picot, M.C. (2010). Appendicular skeletal muscle mass is the strongest independent factor associated with femoral neck bone mineral density in adult and older men. Exp. Gerontol. *45*, 679–684.
33. Diogo, D., Kurreeman, F., Stahl, E.A., Liao, K.P., Gupta, N., Greenberg, J.D., Rivas, M.A., Hickey, B., Flannick, J., Thomson, B., et al.; Consortium of Rheumatology Researchers of North America; Rheumatoid Arthritis Consortium International (2013). Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. Am. J. Hum. Genet. *92*, 15–27.
34. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
35. Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.
36. Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64–69.
37. Muddyman, D., Smee, C., Griffin, H., and Kaye, J. (2013). Implementing a successful data-management framework: the UK10K managed access model. Genome Med *5*, 100.
38. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
39. Strazisar, M., Cammaerts, S., van der Ven, K., Forero, D.A., Lenaerts, A.S., Nordin, A., Almeida-Souza, L., Genovese, G., Timmerman, V., Liekens, A., et al. (2014). MIR137 variants identified in psychiatric patients affect synaptogenesis and neuronal transmission gene sets. Mol. Psychiatry. Published online June 3, 2014. http://dx.doi.org/10.1038/mp.2014.53.
40. Egawa, J., Nunokawa, A., Shibuya, M., Watanabe, Y., Kaneko, N., Igeta, H., and Someya, T. (2013). Resequencing and association analysis of MIR137 with schizophrenia in a Japanese population. Psychiatry Clin. Neurosci. *67*, 277–279.
41. Duan, J., Shi, J., Ge, X., Dolken, L., Moy, W., He, D., Shi, S., Sanders, A.R., Ross, J., and Gejman, P.V. (2013). Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. Scientific Reports *3*, 1318.
42. Chen, X., Wang, X., Hossain, S., O'Neill, F.A., Walsh, D., Pless, L., Chowdari, K.V., Nimgaonkar, V.L., Schwab, S.G., Wildenauer, D.B., et al. (2006). Haplotypes spanning SPEC2, PDZ-GEF2 and ACSL6 genes are associated with schizophrenia. Hum. Mol. Genet. *15*, 3329–3342.
43. Chen, X., Wang, X., Chen, Q., Williamson, V., van den Oord, E., Maher, B.S., O'Neill, F.A., Walsh, D., and Kendler, K.S. (2008). MEGF10 association with schizophrenia. Biol. Psychiatry *63*, 441–448.
44. Pato, M.T., Sobell, J.L., Medeiros, H., Abbott, C., Sklar, B.M., Buckley, P.F., Bromet, E.J., Escamilla, M.A., Fanous, A.H., Lehrer, D.S., et al.; Genomic Psychiatry Cohort Consortium (2013). The genomic psychiatry cohort: partners in discovery. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *162B*, 306–312.
45. Lichtenstein, P., Yip, B.H., Björk, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F., and Hultman, C.M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet *373*, 234–239.
46. Smith, E.N., Bloss, C.S., Badner, J.A., Barrett, T., Belmonte, P.L., Berrettini, W., Byerley, W., Coryell, W., Craig, D., Edenberg, H.J., et al. (2009). Genome-wide association study of bipolar disorder in European American and African American individuals. Mol. Psychiatry *14*, 755–763.
47. Smith, E.N., Koller, D.L., Panganiban, C., Szelinger, S., Zhang, P., Badner, J.A., Barrett, T.B., Berrettini, W.H., Bloss, C.S., Byerley, W., et al. (2011). Genome-wide association of bipolar disorder suggests an enrichment of replicable associations in regions near genes. PLoS Genet. *7*, e1002134.

48. Cao, Q., Martinez, M., Zhang, J., Sanders, A.R., Badner, J.A., Cravchik, A., Markey, C.J., Beshah, E., Guroff, J.J., Maxwell, M.E., et al. (1997). Suggestive evidence for a schizophrenia susceptibility locus on chromosome 6q and a confirmation in an independent series of pedigrees. Genomics *43*, 1–8.
49. Gejman, P.V., Sanders, A.R., Badner, J.A., Cao, Q., and Zhang, J. (2001). Linkage analysis of schizophrenia to chromosome 15. Am. J. Med. Genet. *105*, 789–793.
50. Levinson, D.F., Shi, J., Wang, K., Oh, S., Riley, B., Pulver, A.E., Wildenauer, D.B., Laurent, C., Mowry, B.J., Gejman, P.V., et al.; Schizophrenia Psychiatric GWAS Consortium (2012). Genome-wide association study of multiplex schizophrenia pedigrees. Am. J. Psychiatry *169*, 963–973.
51. MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., et al. (2014). Guidelines for investigating causality of sequence variants in human disease. Nature *508*, 469–476.
52. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.
53. Simon, J.M., Giresi, P.G., Davis, I.J., and Lieb, J.D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nat. Protoc. *7*, 256–267.
54. Alexandrov, B.S., Gelev, V., Yoo, S.W., Alexandrov, L.B., Fukuyo, Y., Bishop, A.R., Rasmussen, K.O., and Usheva, A. (2010). DNA dynamics play a role as a basal transcription factor in the positioning and regulation of gene transcription initiation. Nucleic Acids Res. *38*, 1790–1795.
55. Jablensky, A., Angelicheva, D., Donohoe, G.J., Cruickshank, M., Azmanov, D.N., Morris, D.W., McRae, A., Weickert, C.S., Carter, K.W., Chandler, D., et al. (2012). Promoter polymorphisms in two overlapping 6p25 genes implicate mitochondrial proteins in cognitive deficit in schizophrenia. Mol. Psychiatry *17*, 1328–1339.
56. Alexandrov, B.S., Fukuyo, Y., Lange, M., Horikoshi, N., Gelev, V., Rasmussen, K.O., Bishop, A.R., and Usheva, A. (2012). DNA breathing dynamics distinguish binding from nonbinding consensus sites for transcription factor YY1 in cells. Nucleic Acids Res. *40*, 10116–10123.
57. Hwang, Y.C., Zheng, Q., Gregory, B.D., and Wang, L.S. (2013). High-throughput identification of long-range regulatory elements and their target promoters in the human genome. Nucleic Acids Res. *41*, 4835–4846.
58. Meyer, K.B., Maia, A.T., O'Reilly, M., Ghoussaini, M., Prathalingam, R., Porter-Gill, P., Ambs, S., Prokunina-Olsson, L., Carroll, J., and Ponder, B.A. (2011). A functional variant at a prostate cancer predisposition locus at 8q24 is associated with PVT1 expression. PLoS Genet. *7*, e1002165.
59. Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat. Protoc. *2*, 1722–1733.
60. Cope, N.F., and Fraser, P. (2009). Chromosome conformation capture. Cold Spring Harbor protocols 2009, pdb prot5137.
61. Guella, I., Sequeira, A., Rollins, B., Morgan, L., Torri, F., van Erp, T.G., Myers, R.M., Barchas, J.D., Schatzberg, A.F., Watson, S.J., et al. (2013). Analysis of miR-137 expression and rs1625579 in dorsolateral prefrontal cortex. J. Psychiatr. Res. *47*, 1215–1221.
62. Brennand, K.J., Simone, A., Tran, N., and Gage, F.H. (2012). Modeling psychiatric disorders at the cellular and network levels. Mol. Psychiatry *17*, 1239–1253.
63. Penzes, P., Cahill, M.E., Jones, K.A., VanLeeuwen, J.E., and Woolfrey, K.M. (2011). Dendritic spine pathology in neuropsychiatric disorders. Nat. Neurosci. *14*, 285–293.
64. Shi, S., Leites, C., He, D., Schwartz, D., Moy, W., Shi, J., and Duan, J. (2014). MicroRNA-9 and microRNA-326 regulate human dopamine D2 receptor expression, and the microRNA-mediated expression regulation is altered by a genetic variant. J. Biol. Chem. *289*, 13434–13444.
65. Collins, A.L., Kim, Y., Bloom, R.J., Kelada, S.N., Sethupathy, P., and Sullivan, P.F. (2014). Transcriptional targets of the schizophrenia risk gene MIR137. Translational Psychiatry *4*, e404.
66. Hill, M.J., Donocik, J.G., Nuamah, R.A., Mein, C.A., Sainz-Fuertes, R., and Bray, N.J. (2014). Transcriptional consequences of schizophrenia candidate miR-137 manipulation in human neural progenitor cells. Schizophr. Res. *153*, 225–230.
67. Khetarpal, S.A., Edmondson, A.C., Raghavan, A., Neeli, H., Jin, W., Badellino, K.O., Demissie, S., Manning, A.K., DerOhannessian, S.L., Wolfe, M.L., et al. (2011). Mining the LIPG allelic spectrum reveals the contribution of rare and common regulatory variants to HDL cholesterol. PLoS Genet. *7*, e1002393.
68. Momozawa, Y., Mni, M., Nakamura, K., Coppieters, W., Almer, S., Amininejad, L., Cleynen, I., Colombel, J.F., de Rijk, P., Dewit, O., et al. (2011). Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. Nat. Genet. *43*, 43–47.
69. Raychaudhuri, S., Iartchouk, O., Chin, K., Tan, P.L., Tai, A.K., Ripke, S., Gowrisankar, S., Vemuri, S., Montgomery, K., Yu, Y., et al. (2011). A rare penetrant mutation in CFH confers high risk of age-related macular degeneration. Nat. Genet. *43*, 1232–1236.
70. Helgason, H., Sulem, P., Duvvari, M.R., Luo, H., Thorleifsson, G., Stefansson, H., Jonsdottir, I., Masson, G., Gudbjartsson, D.F., Walters, G.B., et al. (2013). A rare nonsynonymous sequence variant in C3 is associated with high risk of age-related macular degeneration. Nat. Genet. *45*, 1371–1374.
71. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371–1379.
72. Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell *148*, 1223–1241.
73. Flannick, J., Thorleifsson, G., Beer, N.L., Jacobs, S.B., Grarup, N., Burtt, N.P., Mahajan, A., Fuchsberger, C., Atzmon, G., Benediktsson, R., et al.; Go-T2D Consortium; T2D-GENES Consortium (2014). Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. Nat. Genet. *46*, 357–363.
74. van de Ven, J.P., Nilsson, S.C., Tan, P.L., Buitendijk, G.H., Ristau, T., Mohlin, F.C., Nabuurs, S.B., Schoenmaker-Koller, F.E., Smailhodzic, D., Campochiaro, P.A., et al. (2013). A functional variant in the CFI gene confers a high risk of age-related macular degeneration. Nat. Genet. *45*, 813–817.
75. Alexandrov, B.S., Gelev, V., Yoo, S.W., Bishop, A.R., Rasmussen, K.O., and Usheva, A. (2009). Toward a detailed description of the thermally induced dynamics of the core promoter. PLoS Comput. Biol. *5*, e1000313.
76. Spitzer, R., and Endicott, J. (1977). The Schedule for Affective Disorder and Schizophrenia, Lifetime Version (New York: New York State Psychiatric Institute).