ARTICLES
https://doi.org/10.1038/s41588-018-0092-1
nature genetics

Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights

Alexander Gusev[1,2,3*], Nicholas Mancuso[4], Hyejung Won[5,15], Maria Kousi[6], Hilary K. Finucane[1,2,7], Yakir Reshef[8], Lingyun Song[9,10], Alexias Safi[9,10], Schizophrenia Working Group of the Psychiatric Genomics Consortium[11], Steven McCarroll[12,13], Benjamin M. Neale[2,13,14], Roel A. Ophoff[15,16], Michael C. O'Donovan[17], Gregory E. Crawford[9,10], Daniel H. Geschwind[5,15,18,19], Nicholas Katsanis[6], Patrick F. Sullivan[20,21], Bogdan Pasaniuc[4,18,22*] and Alkes L. Price[1,2,22*]

Genome-wide association studies (GWAS) have identified over 100 risk loci for schizophrenia, but the causal mechanisms remain largely unknown. We performed a transcriptome-wide association study (TWAS) integrating a schizophrenia GWAS of 79,845 individuals from the Psychiatric Genomics Consortium with expression data from brain, blood, and adipose tissues across 3,693 primarily control individuals. We identified 157 TWAS-significant genes, of which 35 did not overlap a known GWAS locus. Of these 157 genes, 42 were associated with specific chromatin features measured in independent samples, thus highlighting potential regulatory targets for follow-up. Suppression of one identified susceptibility gene, *mapk3*, in zebrafish showed a significant effect on neurodevelopmental phenotypes. Expression and splicing from the brain captured most of the TWAS effect across all genes. This large-scale connection of associations to target genes, tissues, and regulatory features is an essential step in moving toward a mechanistic understanding of GWAS.

GWAS have yielded thousands of robustly associated variants for schizophrenia and many other complex traits, but relatively few of these associations have implicated specific biological mechanisms[1,2], because GWAS signals often span many putative target genes, may affect gene expression through regulatory[3] or structural elements[4], and may affect genes at considerable genomic distances via chromatin looping[5,6]. A growing body of research has demonstrated the enrichment of schizophrenia GWAS risk variants and heritability within regulatory elements, as identified through maps of chromatin modifications and accessibility[1,7–13]. Because chromatin modifications are themselves under genetic control[6,14–19], a plausible causal mechanism for most schizophrenia loci is genetic variation leading to regulatory changes marked by chromatin, to gene expression, and finally to disease risk. Indeed, quantitative trait loci (QTLs) for chromatin (and other molecular phenotypes) are enriched within GWAS associations, thus further supporting this hypothesis[6,18,20,21].

In this work, we leveraged gene-expression and splicing data from schizophrenia and bipolar cases and controls in the brain, as well as gene-expression data from controls in other tissues, to perform a TWAS[22–24] in a large schizophrenia GWAS dataset[1] and to identify genes whose expression is genetically correlated with schizophrenia. We subsequently performed a TWAS for a diverse set of ~~chromatin phenotypes~~ to connect putative schizophrenia-susceptibility genes with specific regulatory elements. To our knowledge, this is the first TWAS for any disease to integrate analysis of gene expression, splicing, and chromatin variation, and it has moved beyond top SNPs to implicate schizophrenia-associated molecular phenotypes across the regulatory cascade (Fig. 1).

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [4]Department of Pathology and Lab Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [5]Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [6]Center for Human Disease Modeling, Duke University Medical Center, Durham, NC, USA. [7]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Department of Computer Science, Harvard University, Cambridge, MA, USA. [9]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. [10]Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, NC, USA. [11]A full list of members and affiliations appears in the Supplementary Note. [12]Department of Genetics, Harvard Medical School, Boston, MA, USA. [13]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [14]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [15]Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [16]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands. [17]MRC Centre for Psychiatric Genetics and Genomics, Cardiff University, Cardiff, UK. [18]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, CA, USA. [19]Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [20]Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA. [21]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [22]These authors jointly supervised this work: Bogdan Pasaniuc and Alkes L. Price. *e-mail: alexander_gusev@dfci.harvard.edu; pasaniuc@ucla.edu; aprice@hsph.harvard.edu

Results

TWAS for schizophrenia identifies new susceptibility genes. We analyzed gene-expression and genome-wide SNP-array data in 3,693 individuals across four expression reference panels: RNA-seq from the dorsolateral prefrontal cortex of 621 individuals (including 283 schizophrenia cases, 47 bipolar cases, and 291 controls) collected by the CommonMind Consortium (CMC)[25], expression array data measured in peripheral blood from 1,245 unrelated control individuals from the Netherlands Twin Registry (NTR)[26], expression array data measured in blood from 1,264 control individuals from the Young Finns Study (YFS)[23], and RNA-seq data measured in adipose tissue from 563 control individuals from the Metabolic Syndrome in Men study (METSIM)[23]; pre-computed weights from ref. [23] were used for the YFS/METSIM studies. We further characterized splicing events[27] in the CMC brain RNA-seq data (Methods). The average cis and trans estimates of the SNP heritability of expression (h_g^2, Methods) were highly significant in each panel, with a total of a total of 18,084 genes summed across the four panels (10,819 unique genes; Supplementary Table 1), as well as an additional 9,009 splicing events in the brain (in 3,908 unique genes; Supplementary Table 1) exhibiting nominally significant cis-h_g^2 ($P<0.01$ by likelihood ratio test).

We performed a TWAS using each of the four gene-expression reference panels and summary-level data from the Psychiatric Genomics Consortium (PGC) schizophrenia GWAS of 79,845 individuals[1] to identify genes associated with schizophrenia (Fig. 1 and Supplementary Fig. 1a). Briefly, this approach integrated information from expression reference panels (SNP–expression correlation), GWAS summary statistics (SNP–schizophrenia correlation), and linkage disequilibrium (LD) reference panels (SNP–SNP correlation) to assess the association between the cis-genetic component of expression and phenotype (expression–schizophrenia correlation)[23]. In practice, the expression reference panel was used as the LD reference panel, and cis-SNP-expression effect sizes were estimated with a sparse mixed linear model[28] (Methods). Because schizophrenia is a highly polygenic trait, we expected these control reference samples to carry disease-affecting regulatory variants. By leveraging genetic predictors of expression, our approach was not affected by reverse causality (disease → expression), but pleiotropic effects on expression and trait could not be ruled out without additional analyses (Discussion)[23].

The TWAS identified 247 transcriptome-wide-significant gene–schizophrenia and intron–schizophrenia associations (summed across expression reference panels) for a total of 157 unique genes, including 49 genes that were significant in more than one expression panel (Fig. 2, Table 1, Supplementary Fig. 2 and Supplementary Tables 2 and 3). We observed no significant differences when performing the TWAS by using brain expression data from schizophrenia/bipolar cases or controls separately, thus confirming that the presence of cases in the reference panel did not affect our results (Supplementary Note and Supplementary Table 4). We observed hotspots[29] of multiple TWAS-associated genes at 33 loci (defined by genes <500 kb apart). However, only 6/33 loci exhibited evidence of statistically independent genetic effects in a summary-based joint test[30], thus suggesting that most of these loci could be explained by a single underlying genetic effect (Methods and Supplementary Table 3). Across all TWAS associations, the implicated gene was the nearest gene to the top SNP at the locus in only 56% of instances (with the 10,819 cis-heritable genes used as background; this value decreased to 24% of instances when all 26,469 known RefSeq genes were used), thus underscoring previous findings[23,24,29,31]. We confirmed that the summary-based approach was consistent with individual-level predictions by using individual-level PGC data, and we replicated the associations in aggregate by using out-of-sample schizophrenia plus bipolar phenotypes (Supplementary Note, Supplementary Tables 5 and 6, and Supplementary Figs. 1a and 3–5).

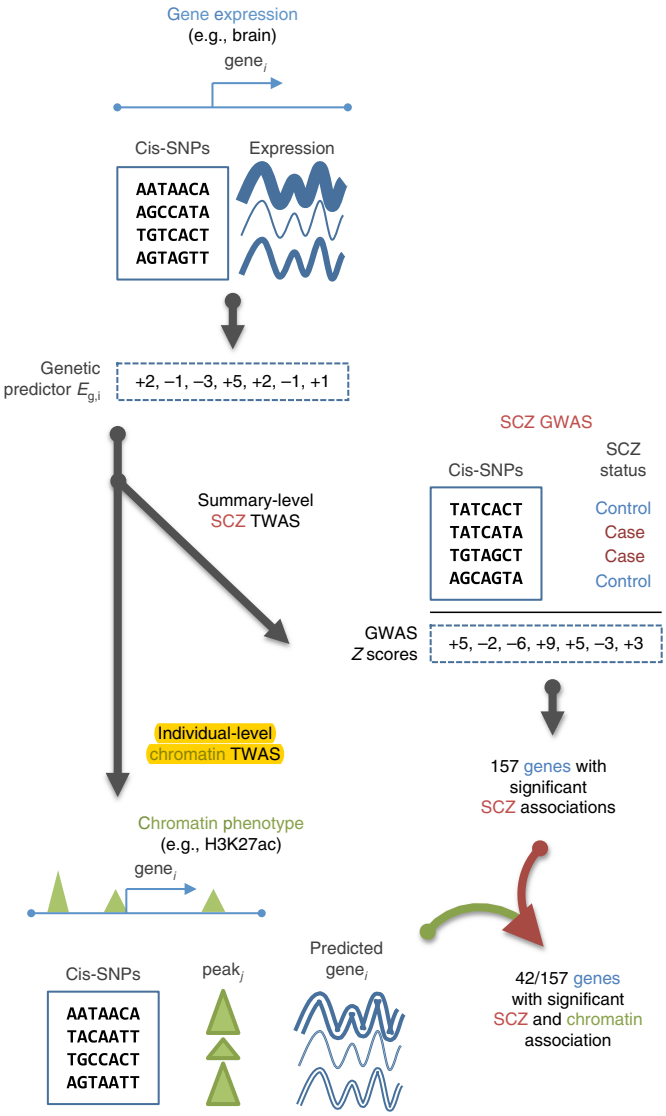

Fig. 1 | Schematic of the TWAS approach. Illustration of the TWAS approach: the genetic predictor of gene expression (E_g) is learned in a reference panel (top), integrated with schizophrenia GWAS association statistics to infer schizophrenia-E_g association (middle), and further integrated with individual-level chromatin phenotypes to infer genes with schizophrenia and chromatin-E_g associations (bottom). Detailed analysis flowchart in Supplementary Fig. 1. SCZ, schizophrenia.

Of the 108 published PGC GWAS regions[1], 47 regions were located near (±500 kb) at least one TWAS gene (accounting for 122/157 genes), and the remaining 35/157 genes implicated novel targets. The GWAS association statistics at novel TWAS loci were often well below genome-wide significance (Supplementary Fig. 6), and we hypothesized that some of the new discoveries might be driven by the TWAS aggregating partially independent effects on schizophrenia that operate through a single gene. As evidence of this model, the TWAS association was stronger than the lead SNP for 27% of TWAS associations that did not overlap a genome-wide-significant SNP, but for only 3% of TWAS associations that did overlap a genome-wide-significant SNP (Fisher's exact $P=8.1\times10^{-7}$). Across all TWAS associations, 21/247 were more significant than the lead GWAS SNP, and the percentage of cis expression heritability that was explained by the top expression QTL (eQTL) for these 21 genes was significantly lower than that for the rest (56% versus 88%, t-test $P=9.6\times10^{-5}$), a result indicative of secondary QTL effects. We excluded the major

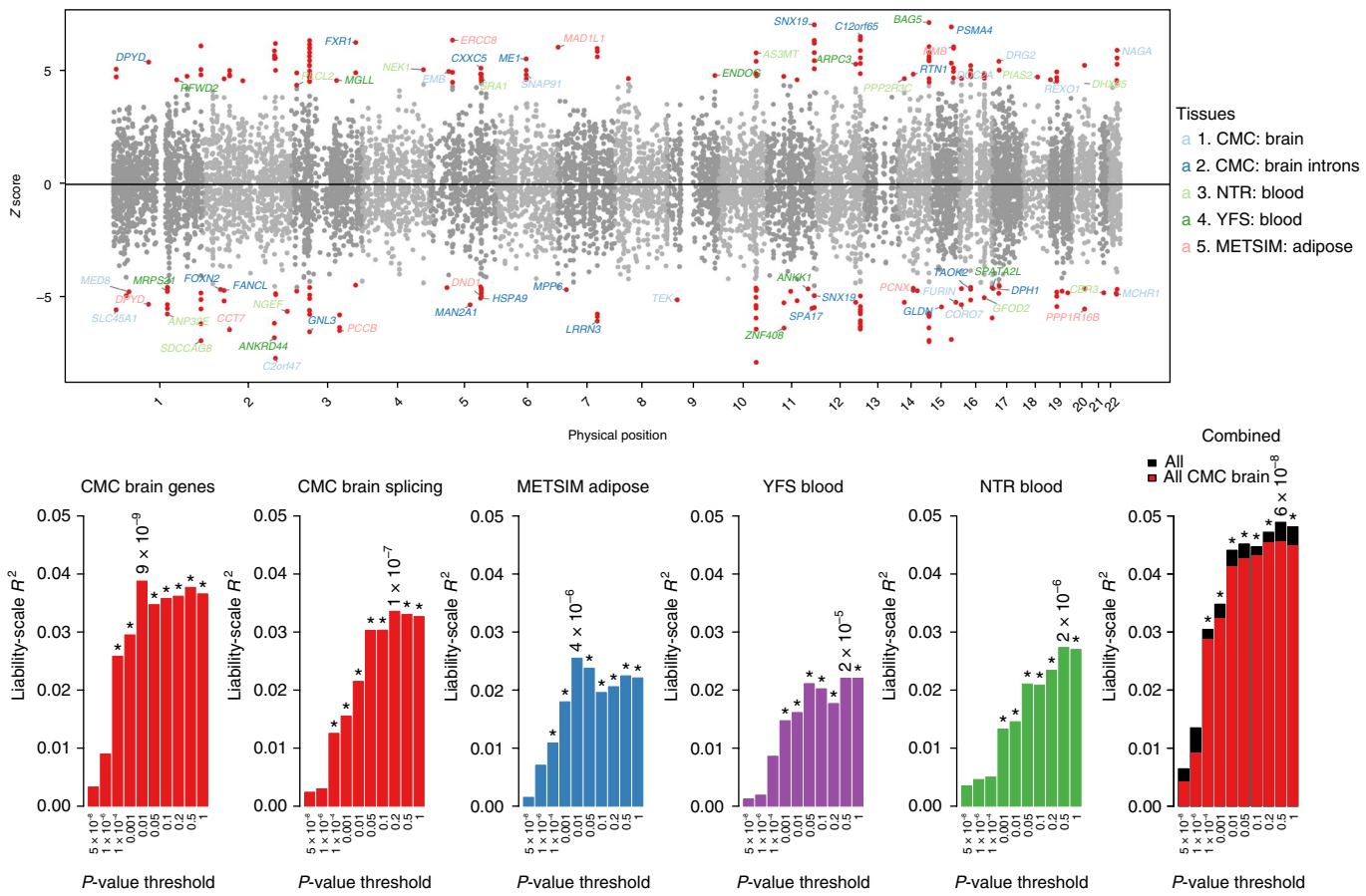

Fig. 2 | Schizophrenia TWAS associations and polygenic effects. Top, Manhattan plot of all TWAS associations. Each point represents a single gene tested, with physical position plotted on the *x* axis and *Z* score of association between gene and schizophrenia plotted on the *y* axis. Transcriptome-wide-significant associations are highlighted as red points, with jointly significant independent associations (Methods) labeled with gene names and color coded according to expression reference (red, CMC; blue, METSIM; purple, YFS; green, NTR; black, all). Bottom, polygenic TWAS effects across reference tissues. Out-of-sample schizophrenia prediction R^2 for GE-PRS as a function of significance cutoff. Significant correlations (after Bonferroni correction for number of thresholds tested) are indicated with an asterisk, and the most significant *P* value is reported. Rightmost panel shows prediction from all tissues jointly (black) and from CMC brain genes plus splicing events jointly (red). R^2 was computed after subtraction of ancestry principal components and conversion to liability scale with a population prevalence of 1%.

histocompatibility complex region (chromosome (chr) 6: 28–34 Mb) from our primary analyses because of its complex haplotype and LD structure. However, as a positive control, we specifically tested the *C4A* gene, which has recently been fine mapped for schizophrenia[4] and lies inside the major histocompatibility complex region, and we confirmed a highly significant TWAS association between *C4A* expression in brain tissue and schizophrenia ($P = 1.8 \times 10^{-18}$).

Splicing events in the brain accounted for 46 transcriptome-wide-significant gene associations (of which ten were at novel loci), a number comparable to the 44 significant gene associations from the brain (Table 1 and Supplementary Table 3), although splicing events accounted for 30% fewer significantly cis-heritable genes than total expression (Supplementary Table 1). Overall, 20/46 associations corresponded to genes that were not tested in the analysis of total gene expression, owing to nonsignificant expression heritability, and 19 of the remaining 26 associations did not have a transcriptome-wide-significant association for total gene expression. This result was consistent with the recent observation that splicing

Table 1 | Number of TWAS-associated genes across all phenotypes and tissues

	CMC brain splicing[a]	CMC brain	NTR blood	YFS blood	METSIM adipose	Total[b]
Heritable	(9,009) 3,890	5,514	2,743	5,418	4,654	11,749
Schizophrenia associated	(80) 46	44	35	48	39	157
Schizophrenia associated (novel[c])	(12) 10	9	6	6	7	35
Chromatin associated	(224) 125	244	182	346	232	806
Schizophrenia and chromatin associated	(10) 8	11	10	13	7	42

[a]Number of unique genes reported, with number of splicing events reported in parentheses. [b]Total number of unique gene associations. [c]Novel is defined as not overlapping (±500 kb) with 108 published PGC schizophrenia GWAS regions.

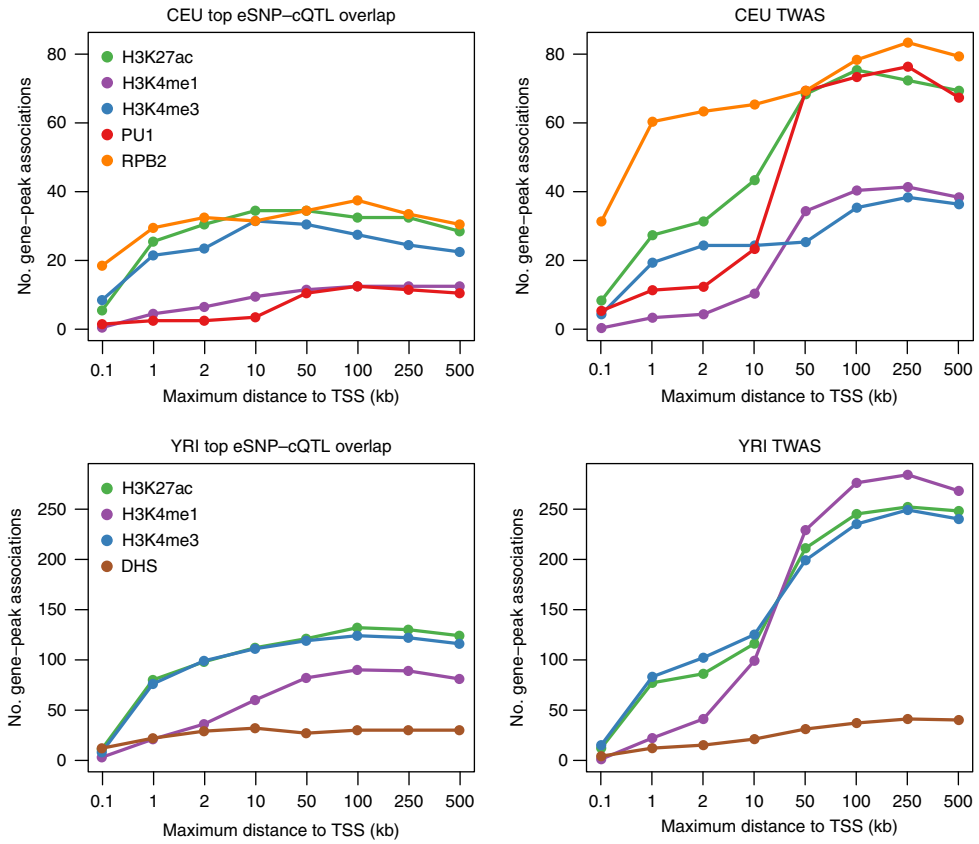

Fig. 3 | Chromatin TWAS associations compared with top eSNP–cQTL associations. Number of unique genes significantly associated with a chromatin peak after Bonferroni correction for a given distance from the gene (*x* axis), determined by using the top eSNP in the chromatin cohort (left) or using chromatin TWAS from all reference panels (right). Results from CEU and YRI populations are shown at top and bottom, respectively.

QTLs are typically independent of eQTLs at the same gene[27]. We caution that effect direction for splicing events is difficult to interpret because alternatively spliced exons are often negatively correlated (Supplementary Note and Supplementary Fig. 7). Although the largest number of associations came from the brain, the enrichment was not striking after the total number of heritable genes was accounted for (Table 1), thus suggesting that expression-data quality and sample size currently are more important than tissue specificity in finding significant associations.

TWAS associations may be caused by coincidental overlap between eQTLs and noncausal disease variants at a GWAS locus, a possibility that we investigated through formal colocalization and conditional analyses. First, we used the COLOC method[32] to estimate the posterior probability of a single shared causal variant for TWAS-implicated genes and schizophrenia by using the marginal association statistics. We calibrated a 5% false-discovery threshold for considering a gene 'colocalized', using randomly selected heritable genes in the same schizophrenia GWAS regions (Methods). Colocalization between eQTLs and schizophrenia was observed for 55% of the TWAS-implicated genes (Supplementary Fig. 8 and Supplementary Table 3). We note that COLOC's posterior is highly dependent on the prior probability of a single shared causal variant (Supplementary Fig. 9) and is conservative when multiple causal variants mediate the effects on expression and trait[23], so that colocalization at the remaining loci may be underestimated. For the 45% genes that did not significantly colocalize, the percentage of cis expression heritability explained by the top eQTL was lower than that explained by the rest (79% versus 89%), thus suggesting secondary effects; however, the difference was not statistically significant. Second, conditioning on the predicted expression of a TWAS-associated gene (using summary-level data[30]; Methods) reduced the χ^2 of the lead GWAS SNP at the locus (including genome-wide-significant and nonsignificant loci) from 42 to 10 on average, and explained more of the association signal than did conditioning on the corresponding top eQTL (Supplementary Table 7). For the 43 lead GWAS SNPs at genome-wide-significant loci that were correlated ($r^2>0.05$) with the predicted expression of at least one TWAS-significant gene (out of 47 overlapping index SNPs), joint conditioning on the predicted expression of all such genes decreased the median SNP *P* value from $P=1.2\times10^{-10}$ to $P=0.028$ (Methods and Supplementary Table 8). Given that the expression predictor typically captures only 60–80% of the cis component of gene expression at the expression-panel sample sizes used here[23], the complete elucidation of the cis component may potentially explain the entire GWAS signal at these loci.

This schizophrenia GWAS dataset[1] has recently been evaluated in a TWAS with gene expression in blood through summary-based Mendelian randomization (SMR)[24], which identified 16 transcriptome-wide-significant associated genes (in contrast to 157 identified here). Of the 16 gene associations identified by SMR, 12 were tested in our study in blood; all replicated at nominal $P<0.05$ (with consistent sign), and 9 were transcriptome-wide significant—a striking concordance given the different methods and independent expression panels used.

Functional validation of TWAS-associated genes by using chromatin-interaction data. We leveraged recently published chromatin-interaction (Hi-C) data in the developing human brain[33] to investigate whether TWAS-associated genes were supported by physical chromatin interactions that occur during brain

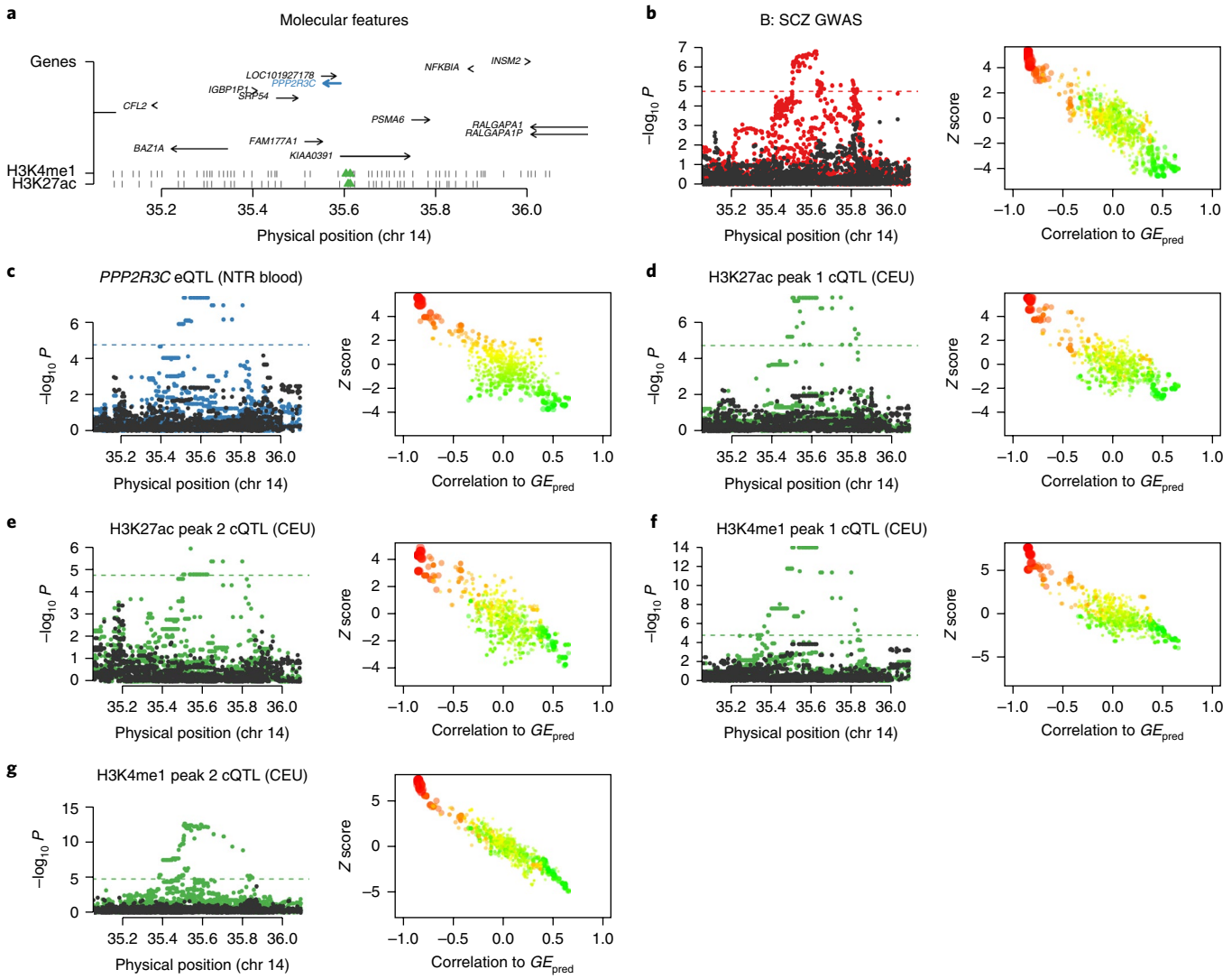

Fig. 4 | Chromatin and schizophrenia TWAS association at *PPP2R3C*. Example association of *PPP2R3C* gene expression and schizophrenia and four nearby chromatin peaks. **a**, Locus schematic showing all nearby genes and chromatin peaks; TWAS-associated features are highlighted in blue and green. **b**–**g**, Left, Manhattan plots of marginal association statistics before and after conditioning on the TWAS-predicted expression (colored and dark dots, respectively). Dashed line shows the local significance threshold after Bonferroni correction for the number of SNPs. Right, relationship between marginal GWAS–QTL association (*y* axis) and the correlation (*x* axis) between TWAS-predicted expression (GE_{pred} estimated in the 1000 Genomes reference) and marginal GWAS–QTL association. The color of each point reflects the eQTL effect size of the expression used for GE_{pred}, and the size of each point reflects the absolute significance of the eQTL. **b**, Schizophrenia GWAS association. **c**, *PPP2R3C* expression phenotype used for TWAS prediction and associated with schizophrenia/chromatin. **d**, First TWAS-associated H3K27ac peak in CEU. **e**, Second TWAS-associated H3K27ac peak in CEU. **f**, First TWAS-associated H3K4me1 peak in CEU. **g**, Second TWAS-associated H3K4me1 peak in CEU. Additional examples and simulations in Supplementary Note and Supplementary Figs. 32–34.

development (Supplementary Fig. 1b). We used the Hi-C data to construct a set of comparison schizophrenia-risk genes on the basis of 3D chromatin interactions between gene transcription start sites (TSSs) and SNPs in the fine-mapped 95%-causal credible set (Methods). This procedure yielded a set of 59 loci with both TWAS and fine-mapped Hi-C data, containing 474 Hi-C-predicted schizophrenia-risk genes. The 474 Hi-C-predicted genes overlapped with 105/157 TWAS-associated genes (Supplementary Fig. 10; Fisher's exact test $P = 1.03 \times 10^{-18}$, odds ratio = 4.68 compared with random heritable genes at these loci), thus indicating that most of the TWAS-associated genes were supported by 3D chromatin interactions with a schizophrenia SNP in the developing brain. The TWAS associations were also significantly correlated with higher expression during mid-fetal developmental in independent samples ($P < 0.05/19$; Supplementary Note and Supplementary Figs. 11 and 12), thus further underscoring the etiological relevance of mechanisms active during brain development.

Polygenic TWAS signal largely explained by expression in the brain. To assess the full polygenic architecture of the TWAS associations, we relaxed the transcriptome-wide-significance threshold and constructed gene-based polygenic risk scores (GE-PRS) from their predicted expression in the CMC (schizophrenia plus bipolar) case–control samples (Supplementary Fig. 1c). For each out-of-sample individual, the GE-PRS was the sum of predicted expression weighted by its signed schizophrenia TWAS *Z* score (Methods). The GE-PRS was significantly associated with schizophrenia status (conditioned on ancestry) across the full spectrum of TWAS association *P* values (Fig. 2), as seen with SNP-based polygenic scores[1,34,35]. Although the prediction was significant in

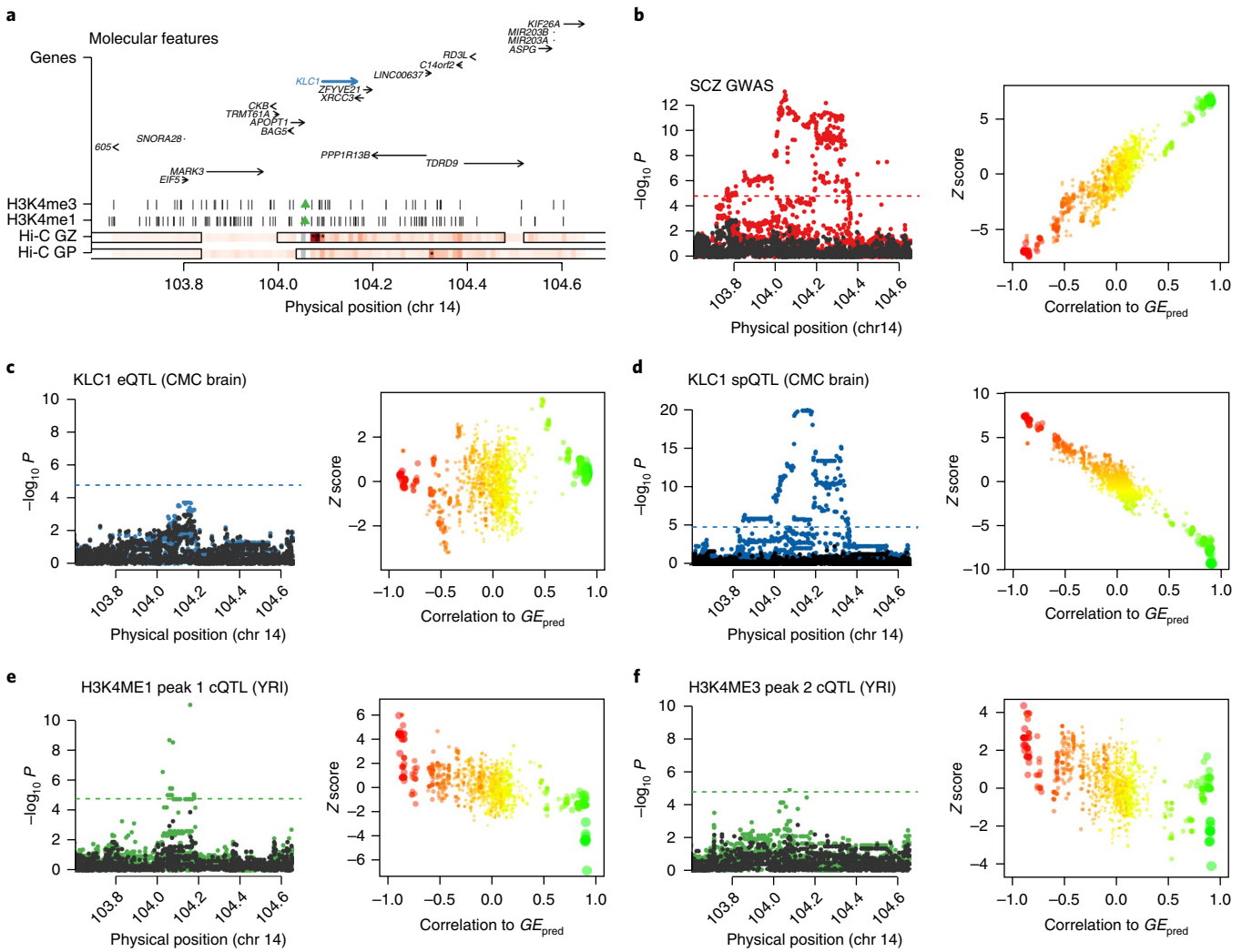

Fig. 5 | Chromatin and schizophrenia TWAS association at *KLC1*. Example association of *KLC1* splice event and schizophrenia, with evidence of chromatin interaction in Hi-C from the developing brain. **a**, Locus schematic with all nearby genes and chromatin peaks; TWAS-associated features are highlighted in blue and green. Hi-C germinal zone (GZ) and cortical and subcortical plate (CP) rows show the significance of the Hi-C chromatin interaction between the 10-kb block containing the associated chromatin peaks (gray, with neighboring white blocks not tested) and every other 10-kb block in the region (with 10 kb being the highest resolution for this Hi-C data). Darker-red shading indicates higher significance, and interactions significant at 0.01 FDR are labeled with asterisks. The most significant interaction in the locus overlaps the *KLC1* promoter. The interactions are shown for fetal-brain data from CP and GZ, and corresponding topological domains are outlined with solid black lines. **b**–**f**, Left, Manhattan plots of marginal association statistics before and after conditioning on the TWAS-predicted expression (colored and dark dots, respectively). Dashed line shows local significance threshold after Bonferroni correction for number of SNPs. Right, relationship between the marginal GWAS–QTL association (*y* axis) and the correlation (*x* axis) between TWAS-predicted expression (GE_{pred} estimated in the 1000 Genomes reference) and marginal GWAS–QTL association. The color of each point reflects the eQTL effect size of the expression used for GE_{pred}, and the size of each point reflects the absolute significance of the eQTL. **b**, Schizophrenia GWAS association. **c**, *KLC1* total expression. Both panels show independence from the TWAS-predicted expression. **d**, *KLC1* splicing-event phenotype used for TWAS prediction and associated with schizophrenia/chromatin. spQTL, splicing QTL. **e**, TWAS-associated H3K4me1 chromatin peak in YRI. **f**, TWAS-associated H3K4me3 chromatin peak in YRI. Additional examples and simulations in Supplementary Note and Supplementary Figs. 32–34.

all tissues individually, there was clear evidence of an increased effect in the brain (in contrast to the transcriptome-wide-significant results), and the prediction from the brain (genes and splicing events) captured 92% of the joint prediction from all tissues (Fig. 2 and Supplementary Fig. 13). A GE-PRS from actual measured expression and differential splicing in the brain was significant but substantially less so than the genetic GE-PRS (Supplementary Fig. 13). According to polygenic theory[36,37], the best TWAS GE-PRS was estimated to account for 26% of the total schizophrenia SNP heritability, thus providing an upper bound on the amount of trait variance that could be mediated by the steady-state expression in these tissues (Supplementary Note).

Chromatin TWAS identifies specific regulatory features associated with expression. We next sought to identify relationships between the expression of TWAS genes and cis regulatory elements marked by chromatin activity. We used population-level chromatin immunoprecipitation–DNA sequencing (ChIP–seq) chromatin phenotypes measured in 76 HapMap Yoruba in Ibadan, Nigeria (YRI) lymphoblastoid cell lines (LCLs) for acetylated histone H3 Lys27 (H3K27ac; marking active enhancers), methylated H3 Lys4 (H3K4me1; enhancers), trimethylated H3 Lys4 (H3K4me3; promoters), and DNase I–hypersensitive sites (DHS; open chromatin)[6], and in 45 HapMap Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) LCLs for

H3K27ac, H3K4me1, H3K4me3, the regulatory transcription factor PU1, and RNA polymerase II (RPB2, associated with active transcription)[18]. For each of the nine chromatin phenotypes, regions with an excess of ChIP–seq reads were segmented into local peaks, and the chromatin abundance within each peak was treated as a quantitative trait[6,18]. Both cohorts additionally had gene expression measured by RNA-seq in the same samples, and we confirmed that the genetic correlation was highly significant between expression and each chromatin mark (as well as between different chromatin marks) and persisted as far as 500 kb from the TSS (Supplementary Figs. 14–16, Supplementary Table 9 and Methods).

We applied individual-level TWAS methods[23] to predict expression of the 10,819 significantly heritable genes and 9,009 differentially spliced introns into samples with chromatin phenotypes and searched for expression–chromatin associations (Fig. 1 and Supplementary Fig. 1d). Prediction was performed from expression to chromatin-phenotype samples (instead of from chromatin-phenotype to expression samples) because of higher prediction accuracy in the larger expression panels, but this choice was agnostic to the direction of causality (Supplementary Note). Our approach yielded an average of 2.4× more Bonferroni-significant expression–chromatin associations than the conventional approach using in-sample lead cis expression-associated SNP (eSNP)–chromatin QTL (cQTL) overlap[38,39], primarily because of associations >10 kb from the TSS (Fig. 3 and Supplementary Fig. 17). We obtained similar results when overlapping all cis eQTLs[6,18] and in simulation (Supplementary Note, Supplementary Figs. 18–20 and Supplementary Table 10). Across all tissues, 806 unique genes had a transcriptome-wide-significant association (Methods) with at least one chromatin phenotype (Supplementary Fig. 18b and Supplementary Table 11), and 4,294 genes were significant at the 10% (per-phenotype) false discovery rate (FDR) used in previous studies[6,18] (Supplementary Table 12). In contrast, only 224 of 9,009 splicing events in the CMC had a transcriptome-wide-significant chromatin association, corresponding to two- to three-times-fewer associations than we identified by using total CMC gene expression (depending on the chromatin phenotype; Supplementary Table 13). Half of the chromatin associations were distal (10–500 kb from the TSS), and these were significantly enriched in Hi-C interactions in LCLs[6] relative to random (distance-matched) gene–peak pairs (Supplementary Figs. 1g and 21–24). No other differences in chromatin-mark usage or mark–gene distance were observed across the expression reference panels. However, we found that genes with associations with multiple chromatin peaks were more likely to be driven by a single eQTL (Supplementary Table 14), thus suggesting that multiple chromatin TWAS peaks were typically related by a single genetic mechanism.

We used the measured RNA-seq expression in the chromatin individuals to confirm these associations. Across the 806 chromatin TWAS-associated genes, the correlation between measured expression and an associated chromatin phenotype was highly significant when compared against a distance-matched background null (Supplementary Figs. 1e and 14b), and the average TWAS-associated chromatin peak explained a striking 20% of the variance in expression of its target gene in CEU (Supplementary Figs. 25–28 and Supplementary Table 16). For the three chromatin phenotypes that were measured in both CEU and YRI, chromatin TWAS peaks implicated in one population were predictive of a correlation with measured expression in the other (Supplementary Figs. 1f, 29 and 30, and Supplementary Table 17), thus supporting our use of chromatin phenotypes from multiple populations.

Putative regulatory mechanisms for schizophrenia-associated genes. Focusing on the 157 transcriptome-wide-significant genes from the schizophrenia TWAS, we identified 42 genes (including seven genes at novel loci) that also had Bonferroni-significant chromatin TWAS associations (to a total of 78 individual chromatin peaks) in analyses using the same expression reference panel (Tables 1 and 2, Supplementary Fig. 1h and Supplementary Tables 3, 18 and 19). Only 8 of the 78 chromatin peaks underlying joint schizophrenia TWAS and chromatin TWAS associations were within the promoter (±2 kb from the TSS) of their associated gene, thus suggesting that most regulatory elements affecting schizophrenia are distally located, as previously observed in other traits[6,8,20]. Schizophrenia TWAS genes were nominally enriched in chromatin TWAS associations (odds ratio = 1.53, Fisher's exact $P = 4 \times 10^{-4}$), but the effect was largely dampened after matching on the cis-genetic properties of genes ($P = 0.01$; Supplementary Table 20) and may potentially be explained by other unknown properties.

We observed significant evidence of chromatin–schizophrenia association and colocalization for most of the identified peaks by using independent statistical methods (Supplementary Fig. 1h). We analyzed the subset of schizophrenia TWAS loci with expression–chromatin associations by applying COLOC to (i) SNP–expression and SNP–chromatin association data to investigate expression–chromatin colocalization and (ii) SNP–chromatin and SNP–schizophrenia association data to investigate chromatin–schizophrenia colocalization. Colocalization was observed for 100% of the expression–chromatin associations and 97% of the chromatin–schizophrenia associations in CEU (Supplementary Fig. 8 and Supplementary Table 19). The chromatin associations in YRI pose a model violation for COLOC, owing to differences in LD structure between populations, but colocalization still remained much higher than background, and 70% (43%) of expression–chromatin (chromatin–schizophrenia) associations colocalized (Supplementary Fig. 8). Estimating pleiotropic associations between chromatin activity and schizophrenia by using SMR[24] (which tests only the best cQTL) or predicting chromatin activity by using a TWAS-like test (testing all SNPs in the Bayesian sparse linear mixed model (BSLMM) predictor) replicated >60% of the associations at Bonferroni significance and >90% at $P < 0.05$ (Supplementary Note and Supplementary Tables 3, 19 and 21). However, the chromatin sample size was insufficient to robustly estimate genetic predictors of chromatin and carry out a full chromatin-wide association study.

Examples of schizophrenia and chromatin TWAS loci. We highlight three examples of TWAS associations with both schizophrenia and chromatin phenotypes. We visualized these loci by using a 'TWAS scatter plot' of the relationship between each marginal GWAS–QTL association (Z score, y axis) and the correlation (x axis) between TWAS-predicted expression (GE_{pred}) and the marginal GWAS–QTL association. This relationship was expected to be linear and without outliers under the TWAS model (Figs. 4 and 5, Supplementary Figs. 32–34 and Supplementary Note).

First, the total expression of *PPP2R3C* in NTR blood was associated with schizophrenia (TWAS $P = 3.4 \times 10^{-6}$)—despite no genome-wide-significant SNPs at the locus—as well as four distal chromatin peaks (minimum $P = 1.0 \times 10^{-9}$; Fig. 4). Conditioning each GWAS SNP on the predicted expression of *PPP2R3C* explained all significant marginal associations for the implicated phenotypes, and formal colocalization was supported between all features and schizophrenia (average posterior = 92%; Supplementary Table 24). *PPP2R3C* was the nearest gene to the most significantly associated SNP at the locus and to the implicated chromatin peaks. However, because the locus was not genome-wide significant, this association would not have been identified in a conventional analysis of known GWAS loci. *PPP2R3C* has recently been identified by SMR analysis of schizophrenia in an independent expression panel[24], and our findings pinpoint specific regulatory elements for experimental follow-up.

Second, a splicing event at *KLC1* in CMC had a schizophrenia TWAS $P = 6.7 \times 10^{-12}$ and overlapping H3K4me1/me3 chromatin

Table 2 | TWAS genes with association with schizophrenia and chromatin phenotypes

Gene	Chromosome	Position	YFS blood	METSIM adipose	NTR blood	CMC brain	DHS	H3K27ac	H3K4me1	H3K4me3	PU1	RPB2
RERE	1	8483747	4×10^{-7}	2×10^{-6}	2×10^{-6}							3
SLC45A1	1	8378144	–	–	–	4×10^{-8}	–	–	–	–	–	1
MAP7D1[a]	1	36621565	6×10^{-4}	–	1×10^{-6}	–	–	–	–	–	–	1
MED8	1	43855483	5×10^{-1}	–	–	2×10^{-6}	–	–	1	–	–	–
ANP32E	1	150207026	–	–	1×10^{-8}	–	–	–	–	1	–	–
MRPS21	1	150266261	3×10^{-6}	3×10^{-3}	6×10^{-3}	2×10^{-2}	–	–	1	–	–	–
COP1[a,b]	1	176176380	4×10^{-6}	–	–	–	–	–	1	–	–	–
C2orf69	2	200775978	–	6×10^{-10}	–	–	–	–	–	1	–	–
GLT8D1[b]	3	52737714	–	–	5×10^{-8}	3×10^{-8}	–	1	–	–	–	–
GLYCTK	3	52321835	2×10^{-8}	–	–	–	–	1	–	–	–	–
GNL3	3	52719935	7×10^{-9}	6×10^{-7}	–	5×10^{-2}	–	–	–	1	–	–
NEK4[b]	3	52804965	–	–	–	2×10^{-9}	–	–	–	1	–	–
NT5DC2[b]	3	52567793	6×10^{-6}	6×10^{-6}	–	7×10^{-1}	–	1	–	1	–	–
PPM1M	3	52279808	2×10^{-7}	2×10^{-7}	–	2×10^{-3}	–	1	–	–	–	–
TMEM110	3	52931597	1×10^{-2}	4×10^{-1}	1×10^{-8}	6×10^{-6}	–	1	1	2	–	–
PCCB	3	135969166	1×10^{-8}	1×10^{-10}	–	3×10^{-10}	1	–	3	–	–	–
RP11-53O19.3	5	44826178	–	6×10^{-6}	–	–	–	–	1	–	–	–
DND1	5	140053171	–	8×10^{-7}	1×10^{-2}	–	–	1	–	1	–	–
IK[b]	5	140027383	4×10^{-6}	1×10^{-6}	–	5×10^{-5}	–	1	–	1	–	–
NDUFA2	5	140027370	2×10^{-6}	–	–	4×10^{-6}	–	1	–	2	–	–
PCDHA2	5	140174443	–	–	–	7×10^{-6}	–	1	–	1	–	–
ZMAT2	5	140080031	5×10^{-6}	1×10^{-3}	–	3×10^{-6}	–	–	–	1	–	–
AS3MT	10	104629209	–	6×10^{-8}	7×10^{-9}	1×10^{-5}	–	1	–	–	–	–
MPHOSPH9[b]	12	123717785	4×10^{-9}	1×10^{-5}	–	2×10^{-8}	–	1	–	–	1	–
KIAA0391[a]	14	35591526	7×10^{-1}	2×10^{-7}	5×10^{-1}	–	–	2	1	–	–	–
PPP2R3C[a]	14	35591748	6×10^{-5}	1×10^{-1}	3×10^{-6}	2×10^{-2}	–	2	2	–	–	–
MAPK3	16	30134630	5×10^{-5}	–	–	1×10^{-6}	–	1	–	–	–	1
GFOD2	16	67753273	–	–	6×10^{-7}	2×10^{-5}	–	1	–	2	–	–
TSNAXIP1	16	67840780	–	–	–	2×10^{-6}	–	–	1	2	–	–
DUS2	16	68038024	1×10^{-6}	–	3×10^{-6}	4×10^{-4}	–	–	–	2	–	–
PRMT7[b]	16	68344876	1×10^{-5}	8×10^{-4}	–	8×10^{-6}	–	–	1	1	–	–
GRAP[a]	17	18950336	–	–	5×10^{-7}	–	–	–	–	–	–	1
RNF112[a]	17	19314490	8×10^{-6}	–	–	–	–	–	–	–	–	1
ACTR5[b]	20	37377096	2×10^{-7}	2×10^{-4}	–	7×10^{-1}	1	–	1	–	–	–
CBR3	21	37507262	6×10^{-3}	2×10^{-3}	2×10^{-6}	5×10^{-4}	1	–	2	–	–	–
CMC brain splicing												
TBC1D5	3	17255862-17279655	–	–	–	3×10^{-6}	–	–	1	–	–	–
NEK4[b]	3	52800010-52800194	–	–	–	1×10^{-6}	–	–	1	–	–	–
CCDC90B	11	82985783-82991184	–	–	–	3×10^{-7}	–	1	–	–	–	–
SBNO1[b]	12	123821038-123825535	–	–	–	4×10^{-10}	–	–	–	–	1	–
KLC1	14	104145855-104151323	–	–	–	7×10^{-12}	–	–	1	1	–	–
RTN1[a]	14	60074210-60193637	–	–	–	1×10^{-6}	–	–	1	–	–	–
TAOK2[b]	16	29997825-29998165	–	–	–	4×10^{-6}	–	–	–	–	–	1
PPP4C[b]	16	30094168-30094715	–	–	–	2×10^{-6}	–	–	–	–	–	1

[a]Novel, not overlapping with 108 PGC schizophrenia GWAS loci. [b]No significant schizophrenia colocalization posterior in any reference (excluding chromatin features in YRI). Forty-two genes (including the seven genes at novel loci[a]) had a significant TWAS association with schizophrenia and chromatin phenotypes. For each significant TWAS association with schizophrenia, the number of significant gene–chromatin associations (family-wise error rate 5% among TWAS gene-mark associations, by Bonforroni correction) are reported. In the middle columns, dashes represent genes that were not heritable in the study and therefore not TWAS associated. In the right columns, dashes represent no identified association; genes with no chromatin associations are not shown. Top, results from genes, with TSS listed as position; bottom, results from splicing events in CMC with exon–exon junction listed as position (details in Supplementary Table 18). For loci without additional evidence of colocalization of cQTL/eQTL with schizophrenia[b], full numerical results are shown in Supplementary Table 3.

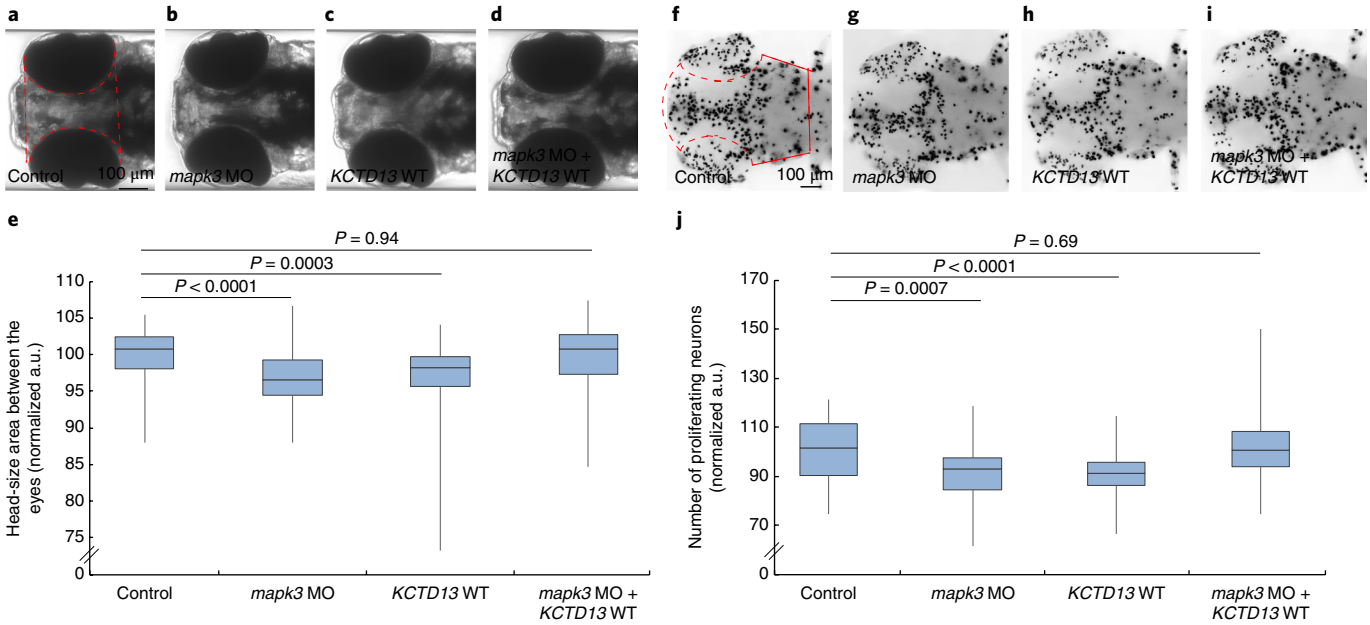

Fig. 6 | Suppression of endogenous *mapk3* rescues the microcephaly and neuronal-proliferation phenotypes induced by overexpression of wild-type *KCTD13*. **a**–**d**, Dorsal views of control larvae (**a**) and embryos injected with morpholino (MO) against endogenous *mapk3* (**b**), human capped wild-type (WT) *KCTD13* mRNA (**c**) or *mapk3* MO plus WT human *KCTD13* mRNA (**d**) at 4 days postfertilization (dpf). **e**, Quantification of the head-size phenotype across the four conditions. **f**–**i**, Dorsal view of 3-dpf embryos stained with an antibody to phospho-histone 3 (PH3), a marker of neuronal proliferation of control larvae (**f**), or embryos injected with MO against *mapk3* (**g**), human capped WT *KCTD13* mRNA (**h**) or both (**i**). **j**, Graph showing quantification of the proliferating neuronal count across the four conditions. Student's *t* test (two tailed) was used to determine statistical significance. The sample size for the head-size assay consisted of control = 67, *mapk3* MO = 59, *KCTD13* WT = 61 and *mapk3* MO + *KCTD13* WT = 60; for PH3, it consisted of control = 37, *mapk3* MO = 40, *KCTD13* WT = 39 and *mapk3* MO + *KCTD13* WT = 40. All experiments were repeated in duplicate and were scored by investigators blinded to injection cocktail. For box plots (**e**,**j**), the horizontal line drawn along the box in each evaluated condition marks the median. The boxes above and below the median line represent the first and third quartiles of the numerical values graphed, respectively. The whisker outside the first quartile marks the maximum values, and the whisker outside the third quartile marks the minimum values for each condition.

TWAS associations (minimum $P = 2.5 \times 10^{-7}$) (Fig. 5). Conditioning on the top splicing QTL explained all significant schizophrenia GWAS signal at the locus, whereas conditioning on the most significant eQTL had a negligible effect, thus highlighting an effect on schizophrenia that was explained by splicing and was independent of total expression. Notably, both chromatin TWAS associations were supported by Hi-C interactions with the *KLC1* promoter in the developing brain[33] (FDR 0.01 significant, and the most significant interaction in the locus), thus providing a functional validation of coordinated activity (Fig. 5 and Supplementary Fig. 35). We performed a TWAS-like test for chromatin–schizophrenia association, which was highly significant for both peaks (best $P = 2.6 \times 10^{-13}$; Supplementary Table 3). Evidence for colocalization was high for *KLC1* splicing and schizophrenia (posterior = 58%) as well as for the chromatin phenotypes and both *KLC1* splicing and schizophrenia (posterior > 80%), even though the chromatin phenotypes were identified in YRI and may exhibit LD differences across populations (Supplementary Table 24). Differential DNA methylation[40] and expression at *KLC1* in schizophrenia cases versus controls has recently been identified in two independent analyses of brain tissue, thus further supporting a cis-regulatory effect on schizophrenia.

Third, total expression of *MAPK3* in CMC brain data was associated with schizophrenia ($P = 1.3 \times 10^{-6}$) as well as two chromatin peaks near the TSS: H3K27ac ($P = 7 \times 10^{-6}$) and RPB2 ($P = 1 \times 10^{-11}$). In the CEU chromatin phenotype samples, in which *MAPK3* expression was also measured in LCLs, the H3K27ac and RPB2 peaks explained 36% ($P = 7 \times 10^{-6}$) and 23% ($P = 5 \times 10^{-4}$) of the variance in measured expression, respectively, but only the H3K27ac peak was significant in a joint model. Formal colocalization analysis supported a single shared causal variant across all eQTL–cQTL–GWAS combinations for the implicated features (posterior probabilities 54–97%; Supplementary Table 24). We confirmed that the associated peaks were observed in epigenetic data from H3K27ac, H3K4me3 and assay for transposase-accessible chromatin using sequencing (ATAC-seq) measured in brain tissues[41] and contained two SNPs with significant allele-specific effects[42] on *MAPK3* (Supplementary Note and Supplementary Figs. 36–39). Strikingly, these peaks overlapped two recently identified human gained neurodevelopmental enhancers in independent fetal cortex tissues[43] (Supplementary Fig. 36). This class of enhancers clusters with genes important for cortical development and neuronal differentiation and has been hypothesized to play a key role in human cortical evolution.

Functional interrogation of *mapk3* in zebrafish. *MAPK3* maps within the 16p11.2 600-kb copy number variant that has been associated with both schizophrenia and autism[44–48]. Previous studies have shown that dosage perturbation of another transcript in that region, *KCTD13* can induce reciprocal head-size and neuronal proliferative defects, characteristics consistent with the anatomical pathology in patients[44]. Critically, pairwise dosage analyses have shown a genetic interaction of *KCTD13* with *MAPK3* (as well as a third locus, *MVP*)[44], whereas independent transcriptional studies in human cells and mouse models have highlighted a functional 'cassette' composed of *KCTD13*, *MVP*, and *MAPK3*, a set of co-regulated genes associated with the head-size phenotype[47]. Together with our TWAS observations, these data implicate a transcriptional relationship between these genes in the 16p11.2 region and suggest that *MAPK3* (and its expression) might be a functional trigger. If so, suppression of *MAPK3* should rescue the pathology induced by

increased expression of *KCTD13*. To test this hypothesis, we performed an experimental assay in zebrafish embryos (Methods). In agreement with findings from prior studies, overexpression of human *KCTD13* (associated with microcephaly in humans) induced both a decrease in head size and a concomitant decrease in the number of cycling cells in the brain (Fig. 6). However, suppression of endogenous *mapk3* in *KCTD13*-overexpressing embryos rescued both phenotypes reproducibly (Fig. 6).

Discussion

The landmark PGC schizophrenia GWAS paper has concluded that "if most risk variants are regulatory, available eQTL catalogues do not yet provide power, cellular specificity, or developmental diversity to provide clear mechanistic hypotheses for follow-up experiments"[1]. In this work, we integrated data from GWAS, expression, splicing, and chromatin activity to identify mechanistic hypotheses. We found 157 unique genes with transcriptome-wide-significant associations with schizophrenia, which were significantly supported by chromatin contact measured during brain development. Genes below the transcriptome-wide-significance threshold continued to be strongly associated with schizophrenia and exhibited enrichment for expression and splicing in the brain (though this result may also reflect expression-data quality). Associations for splicing events that were independent of total expression highlighted an important source of disease-relevant variation[27] with potential therapeutic implications[49,50]. Notably, 42 of the 157 schizophrenia-associated genes were significantly associated with nearby chromatin phenotypes, thus implicating specific regulatory features for functional follow-up. We interrogated one TWAS association, *MAPK3*, in zebrafish embryos and observed a significant effect on neurodevelopmental phenotypes with consistent direction; thus, we prioritized this as a candidate for further follow-up.

We conclude with several limitations and future directions of this study. First, although TWAS is not confounded by reverse causality (disease →expression independent of SNP), instances of pleiotropy (in which a SNP or linked SNPs influence schizophrenia and expression independently) are statistically indistinguishable from truly causal susceptibility genes. As more molecular studies are performed, and the chance of incidental QTL–GWAS overlap increases, experimental causal inference is necessary to validate these findings. Second, the chromatin phenotypes analyzed here were measured in LCLs (because population-level chromatin data from other tissues are currently unavailable), thus preventing us from identifying brain-specific expression–chromatin associations. Third, the use of summary-based data necessitates linear predictors of expression, which may lead to misinterpretation of relationships between expression and disease/chromatin, if, for example, the weaker/secondary eQTLs/cQTLs have stronger effects on the trait because of context specificity. Finally, although we did not observe significant pathway/ontology enrichment for the identified susceptibility genes, we posit that these genes and chromatin features may serve as anchors for network-based analyses of genome-wide coexpression and co-regulation; we view this direction as an intriguing prospect for future investigation.

Because tissue acquisition may pose the greatest hurdle for producing larger datasets, methods that do not depend on measurements from the same samples will remain critical. Beyond specific mechanistic findings for schizophrenia, this work outlines a systematic approach to identify functional mediators of complex disease.

URLs. BRAINSPAN transcriptomes, http://www.brainspan.org/static/download.html/; CommonMind consortium, https://www.synapse.org/cmc; YRI chromatin data, http://chromovar3d.stanford.edu/; PGC summary data, https://www.med.unc.edu/pgc/downloads/; PLINK, https://www.cog-genomics.org/plink2/; PsychENCODE knowledge portal, https://www.synapse.org/#!Synapse:syn4921369/wiki/235539/; SNPWeights for principal component analysis, http://www.hsph.harvard.edu/alkes-price/software/.

Received: 12 December 2016; Accepted: 9 February 2018;
Published online: 9 April 2018

References

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
2. Price, A. L., Spencer, C. C. & Donnelly, P. Progress and promise in understanding the genetic basis of common diseases. *Proc. R. Soc. B* **282**, 20151684 (2015).
3. Soldner, F. et al. Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* **533**, 95–99 (2016).
4. Sekar, A. et al. Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
5. Claussnitzer, M. et al. FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
6. Grubert, F. et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* **162**, 1051–1065 (2015).
7. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
8. Trynka, G. et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
9. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
10. Gusev, A. et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
11. Kichaev, G. et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
12. Won, H.-H. et al. Disproportionate contributions of select genomic compartments and cell types to genetic risk for coronary artery disease. *PLoS Genet.* **11**, e1005622 (2015).
13. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
14. Degner, J. F. et al. DNaseI sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012).
15. McVicker, G. et al. Identification of genetic variants that affect histone modifications in human cells. *Science* **342**, 747–749 (2013).
16. Kasowski, M. et al. Extensive variation in chromatin states across humans. *Science* **342**, 750–752 (2013).
17. Kilpinen, H. et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342**, 744–747 (2013).
18. Waszak, S. M. et al. Population variation and genetic control of modular chromatin architecture in humans. *Cell* **162**, 1039–1050 (2015).
19. Taudt, A., Colomé-Tatché, M. & Johannes, F. Genetic sources of population epigenomic variation. *Nat. Rev. Genet.* **17**, 319–332 (2016).
20. Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
21. Moyerbrailean, G. A. et al. Which genetics variants in DNase-seq footprints are more likely to alter binding? *PLoS Genet.* **12**, e1005875 (2016).
22. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
23. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
24. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
25. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
26. Wright, F. A. et al. Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* **46**, 430–437 (2014).
27. Li, Y. I. et al. RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).
28. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).
29. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
30. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012). S1–S3.
31. Nica, A. C. et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).

32. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
33. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
34. Purcell, S. M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
35. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
36. Palla, L. & Dudbridge, F. A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am. J. Hum. Genet.* **97**, 250–259 (2015).
37. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3**, e3395 (2008).
38. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
39. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
40. Ryan, J. & Saffery, R. Crucial timing in schizophrenia: role of DNA methylation in early neurodevelopment. *Genome Biol.* **15**, 495 (2014).
41. Akbarian, S. et al. The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
42. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063 (2015).
43. Reilly, S. K. et al. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).
44. Golzio, C. et al. *KCTD13* is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363–367 (2012).
45. Maillard, A. M. et al. The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Mol. Psychiatry* **20**, 140–147 (2015).
46. McCarthy, S. E. et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat. Genet.* **41**, 1223–1227 (2009).
47. Migliavacca, E. et al. A potential contributory role for ciliary dysfunction in the 16p11.2 600kb BP4–BP5 pathology. *Am. J. Hum. Genet.* **96**, 784–796 (2015).
48. Föcking, M. et al. Proteomic and genomic evidence implicates the postsynaptic density in schizophrenia. *Mol. Psychiatry* **20**, 424–432 (2015).
49. Sibley, C. R., Blazquez, L. & Ule, J. Lessons from non-canonical splicing. *Nat. Rev. Genet.* **17**, 407–421 (2016).
50. Nelson, C. E. et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**, 403–407 (2016).

Acknowledgements

We acknowledge M. Gandal, B. van de Geijn, A. Ko, P.-R. Loh, L. O'Connor, P. Pajukanta, and N. Zaitlen for helpful discussions. This research was funded by NIH grants F32GM106584 (A.G.), R01GM105857 (A.L.P.), R01MH109978 (A.L.P.), R01MH107649 (B.M.N.), R01MH105472 (G.E.C. and P.F.S.), R01HG009120 (B.P.), U01 MH103339-03S2 (D.H.G.), and R01 MH110927-02 (D.H.G.). H.K.F. was supported by the Fannie and John Hertz Foundation. The project described was also supported by award no. T32GM007753 from the National Institute of General Medical Sciences. This study was supported by a P50MH094268 grant (to N.K.). N.K. is suported as a distinguished Jean and George Brumley Professor. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. We are grateful to the CommonMind Consortium and the PsychENCODE Consortium for making data publicly and readily available. Data were generated as part of the CommonMind Consortium supported by funding from Takeda Pharmaceuticals Company Limited; F. Hoffman-La Roche Ltd.; and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881, R37MH057881S1, HHSN271201300031C, AG02219, AG05138, and MH06692. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimers Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories, and the NIMH Human Brain Collection Core. CMC Leadership: P. Sklar, J. Buxbaum (Icahn School of Medicine at Mount Sinai), B. Devlin, D. Lewis (University of Pittsburgh), R. Gur, C.-G. Hahn (University of Pennsylvania), K. Hirai, H. Toyoshiba (Takeda Pharmaceuticals Company Limited), E. Domenici, L. Essioux (F. Hoffman-La Roche Ltd.), L. Mangravite, M. Peters (Sage Bionetworks), T. Lehner, and B. Lipska (NIMH). Data were generated as part of the PsychENCODE Consortium, supported by U01MH103339, U01MH103365, U01MH103392, U01MH103340, U01MH103346, R01MH105472, R01MH094714, R01MH105898, R21MH102791, R21MH105881, R21MH103877, and P50MH106934 awarded to S. Akbarian (Icahn School of Medicine at Mount Sinai), G. Crawford (Duke), S. Dracheva (Icahn School of Medicine at Mount Sinai), P. Farnham (USC), M. Gerstein (Yale), D. Geschwind (UCLA), T. M. Hyde (LIBD), A. Jaffe (LIBD), J. A. Knowles (USC), C. Liu (UIC), D. Pinto (Icahn School of Medicine at Mount Sinai), N. Sestan (Yale), P. Sklar (Icahn School of Medicine at Mount Sinai), M. State (UCSF), P. Sullivan (UNC), F. Vaccarino (Yale), S. Weissman (Yale), K. White (University of Chicago), and P. Zandi (JHU).

Author contributions

A.G., B.P., and A.L.P. designed the study. A.G., N.M., H.W., H.K.F., and Y.R. conducted analyses. M.K., L.S., A.S., G.E.C., D.H.G., N.K., and P.F.S. conducted and supervised experiments. The Psychiatric Genomics Consortium, S.M., B.M.N., R.A.O., M.C.O., and P.F.S. collected the data. A.G., B.P., and A.L.P. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0092-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.G. or B.P. or A.L.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Data and quality control. Genotypes and expression from the NTR[26], YFS[23], and METSIM[23] were previously analyzed as described in ref. [23] (and below), and the corresponding expression weights were downloaded directly from the TWAS website (URLs). Genotypes and expression data from the CMC[25] were processed according to the GTEx Consortium guidelines for eQTL analysis of RNA-seq data. Specifically, RNA-seq RPKM values were quantile-normalized across samples; genes with more than ten individuals having zero reads were removed; and each gene was rank-normalized, 15 PEER factors were computed, and the residual expression was used.

For alternative splice variants in the brain, we used the LeafCutter algorithm[27,51] to quantify de novo intron excision in the CMC RNA-seq data by clustering reads that spanned intron junctions. These clusters corresponded to individual isoforms and enabled an estimate of differential intron splicing computed from the ratio of reads spanning an intron relative to the total isoform read count. Splice variants were quantified with default parameters: a minimum of 50 reads per cluster, and a maximum intron length of 500 kb. According to the guidelines in ref. [51], the following quality controls were applied to the inferred isoform clusters: clusters with more than ten individuals having zero reads were removed; clusters with < 100 individuals having >20 reads were removed; and introns with fewer than five individuals having nonzero counts were removed. The inferred per-sample abundance for each intron was then treated as a molecular phenotype, normalized, and PEER-corrected as described for total expression above. This process identified 123,480 splicing events, of which 99,562 mapped to canonical gene introns. We treated the differential splicing of these 99,562 splicing events as quantitative traits in the same manner as total expression.

For genotype data in the above studies, individuals failing a sex check or having 5% missing SNPs were removed. Additionally, SNPs were removed if they had >5% missing calls; $P<0.05$ case–control missing association; $P<5\times10^{-6}$ Hardy–Weinberg disequilibrium; $P<5\times10^{-3}$ association to batch; $P<5\times10^{-8}$ missing haplotype association; or frequency <1%. Principal components (PCs) were computed by using all samples for the NTR, YFS, and METSIM data directly and using SNPweights (v2.1)[52] for the CMC data; outliers (samples more than six s.d. away from the mean along any top component) were removed; and PCs were included as fixed effects in estimating h_g^2. For all datasets, related individuals with GRM values >0.05 were also removed before estimation of h_g^2.

For chromatin data, we used population-level ChIP–seq chromatin phenotypes measured in 76 HapMap YRI LCLs for H3K27ac (marking active enhancers), H3K4me1 (enhancers), H3K4me3 (promoters), and DHS (open chromatin)[6], and in 45 HapMap CEU LCLs for H3K27ac, H3K4me1, H3K4me3, PU1, and RPB2 (associated with active transcription)[18]. We did not perform any additional QC of the functional features, which were previously adjusted for PEER/covariates and normalized[6,18].

h_g^2 estimation. Cis- and trans-h_g^2 were estimated by using variance components, modeling the phenotype as a multivariate normal $y\sim\sigma^2_{g,cis}K_{cis}+\sigma^2_{g,trans}K_{trans}+\sigma^2_e I$, where the K values are the standard genetic-relatedness matrices from SNPs in the cis locus (K_{cis}) and in the rest of the genome (K_{trans}). The σ^2 parameters were fit for each gene with AI-REML, as implemented in GCTA software[53], with principal components and sex included as fixed effects. For h_g^2 of splicing events, the intron ratios conditioned out isoform abundance, but total gene expression was also included as a covariate to account for any residual correlation. As in previous studies[26], individual estimates outside the plausible 0–1 range were allowed to achieve unbiased mean estimates. The standard error of each estimate was approximated as the s.d. divided by the square root of the number of genes tested; however, significant differences were confirmed by permutation tests (described below).

To evaluate the contribution of low-frequency variants, we imputed the NTR data to the Haplotype Reference Consortium reference, thus yielding high-quality imputed SNPs down to a MAF of 0.001. On average, we did not observe a significantly nonzero contribution of imputed rare variants to cis-h_g^2, nor did we observe a significant change in common cis-h_g^2 as a result of denser imputation relative to array SNPs (Supplementary Table 1). Although recent work has identified biases in estimates of h_g^2 from rare variants[54], we expect these biases to be small in the cis region and to be largely mitigated by the two-component model. We did not further evaluate the contribution of rare variants to trans-h_g^2. No differences were observed when dosages were used to construct the cis-GRM.

In the CMC data, in which schizophrenia/bipolar and control status were also available, the average cis-genetic correlation of expression between (schizophrenia/bipolar) cases and controls was 1.00 (s.e. 0.02), thus indicating consistent direction of eQTL effect sizes between cases and controls and motivating us to use the full cohort as a TWAS reference panel (Supplementary Table 25).

Schizophrenia TWAS. We performed TWAS, using publicly available summary statistics from the PGC GWAS of 79,845 individuals[1] and four gene-expression reference panels in independent samples (URLs). For a given gene, SNP-expression weights in the 1-Mb cis locus were first computed with the BSLMM[28], which models effects on expression as a mixture of normal distributions to account for the sparse expression architecture. Given weights w, schizophrenia Z scores Z, and SNP-correlation (LD) matrix D; the association between predicted expression and schizophrenia (i.e., the TWAS statistic) was estimated as $Z_{TWAS}=w'Z/(w'Dw)^{1/2}$ (methodological details in ref. [23]). We computed TWAS statistics by using either the SNPs genotyped in each expression reference panel or imputed HapMap3 SNPs (which typically represent well-imputed SNPs). To account for multiple hypotheses, we applied Bonferroni correction within each expression panel used. This threshold was chosen to maximize consistency with previous published results and to not penalize for additional (and often highly correlated) expression panels tested. Specifically, we report 'transcriptome-wide' significance after correcting for the number of genes tested within each of the five reference panels (CMC, CMC-splicing, NTR, YFS, and METSIM; 5,419 tests on average). This procedure is consistent with the correction applied in previous TWAS results of multiple expression references[23].

Summary-based joint/conditional tests and figures. Conditional and joint analysis was performed with the summary-statistic-based method described in ref. [30], which was adapted to genes instead of SNPs. This joint test aims to distinguish genes with independent genetic predictors (that are also schizophrenia associated) from those that are merely coexpressed with a shared genetic predictor. This procedure requires marginal association statistics (i.e., the main TWAS results) and a correlation/LD matrix to evaluate the joint/conditional model. The correlation matrix was estimated by predicting the cis-genetic component of expression for each TWAS gene into the 1000 Genomes genotypes and computing Pearson correlations across all pairs of genes as well as between all gene–SNP pairs (with correlations below 0.01 set to zero, owing to sampling noise). The 247 transcriptome-wide-significant TWAS associations across four reference panels were then added to the model one at a time in decreasing order of significance and were retained if their conditional TWAS association remained significant after Bonferroni correction for 247 tests. To quantify strongly independent gene associations at hotspot loci, this procedure was repeated with the additional constraint that genes were added to the model at each step only if they had $r^2<0.30$ with all genes already in the model. To assess how much of the schizophrenia GWAS association signal remained after the TWAS signal was removed, each GWAS SNP association was conditioned on the joint gene model, one SNP at a time. For the Manhattan plots in Figs. 4 and 5, each GWAS SNP was conditioned on the predicted expression of the single target TWAS gene. For TWAS scatter plots in Figs. 4 and 5, the correlation of each SNP to the TWAS-predicted phenotype was computed by predicting expression in the 1000 Genomes reference.

Colocalization analyses. We used COLOC software[32] to estimate the posterior probability of two phenotypes sharing a causal variant (which we refer to as 'colocalization'). For a locus and pair of traits (for example, chromatin and schizophrenia), the corresponding SNP-trait QTL effect sizes (and standard errors) were tested with the coloc.abf function, with molecular phenotypes treated as quantitative traits and schizophrenia input treated as a case–control trait with 43% cases. The posterior probability of one shared causal variant ($H4$) was reported. The default prior on sharing ($P=1\times10^{-5}$) was used for all primary analyses, with priors $P=1\times10^{-4}$ and $P=1\times10^{-3}$ evaluated separately with all other parameters unchanged.

Functional validation of the TWAS-associated genes with chromatin-interaction data. The TWAS genes were validated according to the presence of Hi-C interaction with fine-mapped schizophrenia GWAS SNPs in the locus.

Schizophrenia GWAS loci were fine mapped as previously described[33]. First, independent genome-wide-significant SNPs and all nearby SNPs in LD ($r^2>0.6$) having $P<1\times10^{-5}$ were selected. The CAVIAR fine-mapping algorithm[55] was then applied in each locus (allowing for a maximum of two causal variants) to identify the 95%-credible set of causal SNPs. Functional SNPs (those causing nonsense and missense variation and residing within gene promoters) were directly assigned to their target genes, whereas nonannotated SNPs were assigned to the genes of action on the basis of chromatin interactions in developing human cortices[33].

Hi-C interactions were defined as previously described[33]. We fit background interaction profiles as a function of distance from all 9,444,230 imputed PGC GWAS SNPs by using a Weibull distribution. The significance for a given Hi-C contact was then measured by calculating the probability of observing a stronger contact under this null. Fine-mapped GWAS SNPs were assigned to 10-kb bins (which is the highest resolution available for the fetal-brain Hi-C data), and the significance of interactions for every bin within a 1-Mb flanking region (500 kb upstream to downstream) was calculated. Significant Hi-C interacting regions (FDR <0.01) were then overlapped with Gencode v19 gene coordinates to identify the potential gene targets, and only genes with significant SNP heritability estimates (cis-$h_g^2P<0.01$) were evaluated, thus resulting in 474 genes. These Hi-C-defined schizophrenia-risk genes were then overlapped with the genes and splice variants identified by TWAS. Overrepresentation analysis was performed with Fisher's exact test with a background gene list of 1,392 genes residing within a ±500-kb window to any credible SNP and having nominally significant cis-$h_g^2(P<0.01)$.

Polygenic TWAS signal from gene-based polygenic risk scores (GE-PRS). We extended the SNP-based polygenic risk score[34] to genes to evaluate the TWAS

predictive accuracy and validation. Given a 1-by-M vector z of signed association statistics in the discovery study (for example, PGC) and an N-by-M matrix X of predicted expression for the corresponding M genes in the replication study, we constructed a GE-PRS $S = Xz$. The M genes were either all transcriptome-wide-significant genes (Supplementary Note) or all genes passing relaxed P-value thresholds. This risk score was then tested with ancestry against case/control status with the standard linear model $y \sim S + P + e$, where S is the risk score, and P is a matrix of PCs accounting for ancestry. Risk-score performance was measured as the linear R^2 from the above model minus the R^2 from the model $y \sim P + e$ to account for ancestry and was converted to the liability scale assuming a prevalence of 1%.

For the TWAS using METSIM, YFS, and NTR expression reference panels, the cis-genetic component of expression was predicted in CMC samples. For the TWAS using the CMC expression panel, either the total expression was used (Supplementary Fig. 13) or the cis-genetic component of expression was estimated directly with BSLMM (equivalent up to a scaling factor to estimate genetic values by dropping each individual in turn). We stress that the case/control label from the CMC data was never used to identify the TWAS associations, and that the GePRS from the CMC expression panel were thus evaluated against an independent CMC case/control phenotype. Ascertaining cases in the CMC expression panel may increase the frequency of causal variants and may consequently make the prediction more accurate than using a randomly ascertained expression panel; however, we observed little difference when performing the TWAS using an expression panel consisting of CMC controls only (Supplementary Table 4).

Individual-level chromatin TWAS. We used cis SNP-expression effect sizes computed according to BSLMM scores in the four expression reference panels (including splicing events) to predict individual-level expression for the 45 CEU[18] and 76 YRI[6] individuals with measured chromatin phenotypes. We retained only post-QC SNPs that were typed in both studies and removed strand-ambiguous SNPs. We note that even though the YRI target samples are of different ancestry, this prediction does not require an LD reference panel and is therefore expected to suffer loss in only power (but not increased type I error), owing to the differences in LD. For each predicted gene, we identified all chromatin peaks within a given window of the TSS (primary results used ±500 kb) and tested each mark for association with predicted expression by linear regression.

Top eSNP–cQTL overlap analysis. We compared the chromatin TWAS to a traditional approach of identifying SNPs that are significant both as cQTLs and eQTLs in real data. For each population and given distance to TSS, we performed this analysis in two stages. In stage 1, we used fastQTL[39] to identify the most significantly associated eSNP for each gene by permuting the expression and retesting the cis locus, and restricted analysis to those genes with eSNP $P < 0.01$ (for consistency with the TWAS gene selection). In stage 2, each significant eSNP from stage 1 was then tested for association with all nearby chromatin peaks by standard linear regression, and those passing Bonferroni correction for all gene–peak pairs for each chromatin phenotype (for example, H3K27ac in CEU) were reported. This analysis was compared with the chromatin TWAS analysis in which each gene was tested against any peak within the given distance (by standard linear regression), and the number of significant results was reported after Bonferroni correction for the total number of gene–peak pairs tested in that phenotype. We separately considered an approach in which all significant eQTLs in a gene were evaluated for overlap (as in refs [6] and [18]); however, this approach underperformed the permuted top eSNP–cQTL approach here (Supplementary Note).

Multiple-hypothesis correction for chromatin TWAS. The large number of correlated phenotypes analyzed (expression from five experiments and chromatin from nine experiments in two populations) allowed for several approaches to multiple-testing correction. For the chromatin TWAS, we corrected for the number of gene–peak pairs tested within a single expression reference and chromatin phenotype experiment (for example, the number of gene–peak pairs in evaluation of predicted CMC expression with the CEU H3K27ac chromatin phenotype). This procedure is directly comparable to the experiment-wide corrections applied in previous eQTL–cQTL analyses[6,18]. The same correction was applied for the schizophrenia–chromatin TWAS overlap: for example, the 44 schizophrenia TWAS genes identified on the basis of CMC expression were within 500 kb of 1,528 total peaks in the CEU H3K27ac experiment, and 'overlap' was reported for any peak with a chromatin TWAS association $P < 0.05/1{,}528$.

For comparison, we separately calculated the number of associations that were significant at 5% FDR across all molecular experiments. This process yielded approximately 3.5× more chromatin TWAS associations and 1.2× more schizophrenia and chromatin TWAS associations (Supplementary Table 2), thus demonstrating that the above experiment-wide Bonferroni correction strategy corresponded to a conservative study-wide FDR.

Estimating support for mediation by expression/chromatin. We sought to evaluate the evidence in support of two models of mediation: M_{CH}, in which SNP → chromatin → expression → disease, and M_{EX}, in which SNP → expression → chromatin → disease. Under the assumption of linear, additive variance across molecular phenotypes, this parameter can be estimated via the ratio of genetic covariance (cov_g) between chromatin–schizophrenia and expression–schizophrenia. Conceptually, the genetic effect of a given molecular phenotype on schizophrenia will be attenuated by environmental noise, which will manifest as lower cov_g to schizophrenia for phenotypes further along the molecular cascade. The fraction of environmental variance on expression (env_{EX}) under each model of mediation can then computed from the following equation (derivation in Supplementary Note):

$$cov_{g,CH}/cov_{g,EX} = 1/\sqrt{1-env_{EX}^2}$$

We inferred these quantities from both CEU and YRI, using cov_g estimates from cross-trait LD-score regression[56] and computed significance by comparing against randomly sampled gene–peak pairs within 500 kb of the TSS. We separately considered a partial correlation approach using residuals of expression and chromatin in turn (Supplementary Note). We caution that the estimate of *env* in the above equation was computed from an average across all loci and may also be consistent with confounding from different levels of measurement error for ChIP–seq and RNA-seq, a mixture of models M_{CH} and M_{EX} that favors model M_{CH}, or mediation by other unobserved molecular phenotypes.

In vivo complementation in zebrafish embryos and whole-mount immunostaining. The human WT mRNA of *KCTD13* (NM_178863) was cloned into the pCS2+ vector and transcribed in vitro with aSP6 Message Machine kit (Ambion), as previously described[44]. To suppress endogenous *mapk3*, we identified ENSDART00000103746.5 with 93% similarity and 88% identity to human *MAPK3* (NM_002746) as the sole zebrafish ortholog, against which we designed a splice-blocking MO targeting exon 2 (*mapk3*sb: CTGTGAGTGTTTAAGGATACACATC). We injected 10 ng of *mapk3* MO and 150 pg of WT *KCTD13* RNA alone and in combination into WT zebrafish embryos at the one- to four-cell stage. For the evaluation of neuronal proliferation, anti–phosphor histone 3 (PH3) (Ser10)-R (sc-8656-R, rabbit, Santa Cruz; 1:500) was used, and experiments were performed as previously described[57]. Proliferating neurons were quantified by counting all positive cells on a dorsal view of 3-dpf embryos, excluding the eyes from the scored area, by using the ITCN ImageJ plugin, which considers cells with 20-pixel width and 5-pixel minimum distance between them to be separate cells. Statistical significance for this assay was established with two-tailed Student's t test. For the head-size assay, injected larvae were grown to 4 dpf and imaged live on dorsal view. The area of the head, excluding the eyes, was traced from the measurements, and statistical significance was calculated with two-tailed Student's t test. Zebrafish embryos and adults were maintained and mated according to standard procedures, and all experiments were carried out with the approval of the Institutional Animal Care and Use Committee (IACUC).

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. The data and methods that support the findings of this study are available through the following links: TWAS results (http://www.gusevlab.org/projects/chromatinTWAS/) and TWAS methods (http://www.gusevlab.org/projects/fusion/).

References

51. Li, Y. I. et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
52. Chen, C.-Y. et al. Improved ancestry inference using weights from external reference panels. *Bioinformatics* **29**, 1399–1406 (2013).
53. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
54. Yang, J. et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).
55. Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014).
56. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
57. Jordan, D. M. et al. Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* **524**, 225–229 (2015).

nature research

Corresponding Author: ALEXANDER GUSEV

Date: 3/20/2018

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   No data was generated for this study, the largest sample sizes from publicly available data were used.

2. Data exclusions

   Describe any data exclusions.

   Pre-established quality control criteria were used for expression data as defined in previous publications from the GTEx Consortium.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   No individual variant replication was carried. Gene prediction accuracy was established by cross-validation and reported for all genes. Aggregate replication of significant TWAS Associations was performed using risk scores and shown to be significant. Individual and summary-based predictions were shown to be highly correlated.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Randomization was not necessary for the statistical analyses, and the functional analyses were performed with blinding.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   For functional interrogation of zebrafish, all experiments were repeated in duplicate and were scored by investigators blinded to injection cocktail.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	TWAS/FUSION software was developed and used for analyses. The software + documentation is available on the web: http://gusevlab.org/projects/fusion/ and the source is available in a GitHub repository: https://github.com/gusevlab/fusion_twas

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	no restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	a phosphor histone 3 (PH3) antibody (ser10)-R (sc-8656-R, rabbit, Santa Cruz; 1:500) was used and experiments were performed as previously described in reference 57: Jordan, D. M. et al. Identification of cis-suppression of human disease mutations by comparative genomics. Nature 524, 225–229 (2015).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	n/a
b. Describe the method of cell line authentication used.	n/a
c. Report whether the cell lines were tested for mycoplasma contamination.	n/a
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	n/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Zebrafish embryos and adults were maintained and mated according to standard procedures and all experiments were carried out with the approval of the Institutional Animal Care and Use Committee (IACUC).

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Human data was previously collected and de-identified. Sex, age, genetic ancestry, and multiple measures of assay quality were always included as covariates.