

ARTICLE OPEN



Transcriptional-regulatory convergence across functional MDD risk variants identified by massively parallel reporter assays

Bernard Mulvey¹ and Joseph D. Dougherty¹✉

© The Author(s) 2021

Family and population studies indicate clear heritability of major depressive disorder (MDD), though its underlying biology remains unclear. The majority of single-nucleotide polymorphism (SNP) linkage blocks associated with MDD by genome-wide association studies (GWASes) are believed to alter transcriptional regulators (e.g., enhancers, promoters) based on enrichment of marks correlated with these functions. A key to understanding MDD pathophysiology will be elucidation of which SNPs are functional and how such functional variants biologically converge to elicit the disease. Furthermore, retinoids can elicit MDD in patients and promote depressive-like behaviors in rodent models, acting via a regulatory system of retinoid receptor transcription factors (TFs). We therefore sought to simultaneously identify functional genetic variants and assess retinoid pathway regulation of MDD risk loci. Using Massively Parallel Reporter Assays (MPRAs), we functionally screened over 1000 SNPs prioritized from 39 neuropsychiatric trait/disease GWAS loci, selecting SNPs based on overlap with predicted regulatory features—including expression quantitative trait loci (eQTL) and histone marks—from human brains and cell cultures. We identified >100 SNPs with allelic effects on expression in a retinoid-responsive model system. Functional SNPs were enriched for binding sequences of retinoic acid-receptive transcription factors (TFs), with additional allelic differences unmasked by treatment with all-*trans* retinoic acid (ATRA). Finally, motifs overrepresented across functional SNPs corresponded to TFs highly specific to serotonergic neurons, suggesting an *in vivo* site of action. Our application of MPRAs to screen MDD-associated SNPs suggests a shared transcriptional-regulatory program across loci, a component of which is unmasked by retinoids.

Translational Psychiatry (2021)11:403; <https://doi.org/10.1038/s41398-021-01493-6>

INTRODUCTION

Major depressive disorder (MDD) is a common but debilitating psychiatric disorder affecting hundreds of millions worldwide [1], exacting substantial tolls on both individuals [2] and societies [3]. Despite the global burden of MDD, nearly half of patients do not clinically respond to treatment [4], in part due to limited understanding of its biological underpinnings. Family studies have demonstrated that MDD risk is at least 30% heritable [5, 6]. More recently, genome-wide association studies (GWASes) have demonstrated similar estimates for severe MDD [7], and have helped narrow in on associated single-nucleotide polymorphisms (SNPs) [8–12], a tantalizing entry point for understanding the biology of MDD. However, GWAS studies do not identify causal variants, but rather implicate wider co-inherited regions consisting of many SNPs in linkage disequilibrium (LD). Most disease-associated SNPs are found outside of protein-coding sequences, suggesting probable roles in transcriptional regulation (TR) [13–16]. Which linked SNPs have functional allelic impacts on TR—and how they act together across loci to result in disease—remains unresolved.

It is thought that undetected, small-effect SNPs acting across the genome—including conditional SNPs within GWAS-significant loci [17]—contribute to the substantial heritability *not* caught by GWAS-significant SNPs alone [18]. Early support for multiple linked variants underlying GWAS signals came from examination of cell

line histone marks in loci from six autoimmune disorder GWASes; all six showed enrichment of TR-suggestive marks at LD SNPs only in a pertinent cell type (B lymphocytes). Strikingly, 65% of the loci with ≥ 1 SNP overlapping lymphocyte histone marks contained multiple SNP-mark pairs, and over half of these loci contained at least three such SNPs [19]. Altogether, these findings implied that GWAS regions likely affect several TR features via several linked variants, especially in relevant cell types. More recently, GWASes have identified what are now called “conditional SNPs” associated with MDD [20]. However, despite predictions of multiple TR SNPs within GWAS loci, functional demonstration of this phenomenon has been sparse to date. The largest functional TR assay of MDD-associated variants examined 34 SNPs using luciferase assays [21], representing successful but low-throughput identification of functional MDD SNPs. However, in terms of broad linkage, these loci constitute well over 10,000 SNPs, which will ultimately require higher-throughput approaches.

Furthermore, how functional SNPs—even once identified—biologically result in disease remains unclear, given their individually small effects on risk. The polygenic [22] and omnigenic [18] models were conceived of to address these aspects of complex disease genetics, establishing a guiding principle for GWAS interpretation. In brief, these theories posit that consistent emergence of a specific phenotype via widespread genomic variation necessarily requires common biological

¹Departments of Genetics and Psychiatry, Washington University in St. Louis, St. Louis, MO, USA. ✉email: jdougherty@wustl.edu

Received: 15 March 2021 Revised: 2 June 2021 Accepted: 16 June 2021

Published online: 22 July 2021

endpoints of those variants' effects. At the molecular level, these points of convergence could be either upstream (shared regulation across loci) [23] or downstream (common biological pathways across loci). For downstream analyses, myriad approaches have been developed to nominate gene targets of putative TR SNPs using proximity [24], chromatin structure [25–27], or expression quantitative trait loci (eQTLs) [28–30], yielding gene sets tested for enrichment in biological pathways [28, 29, 31] and cell types [32]. However, no analogous approaches exist for identifying convergent upstream (i.e., TR) molecular effects of genetic risk, in part because a prerequisite is defining the functional SNPs.

One possible point of upstream TR convergence of MDD risk variants is retinoic acid and related compounds (retinoids). Retinoids drive transcriptional responses via several retinoid-binding nuclear receptor transcription factors (TFs) and heterodimerizing partners [33, 34]. Besides their critical role in neurodevelopment, including of depression-implicated limbic structures [35], retinoids have been associated with MDD onset and suicidality by epidemiological studies of the retinoid agonist isotretinoin [36]. Moreover, thyroid hormone is often used as an adjunctive treatment in MDD, and thyroid receptor TR effects are frequently carried out cooperatively with RXR family retinoid receptors [37]. Additional evidence for retinoid pathway activity in the adult brain—and its overactivity as a risk factor for depression—comes from rodent pharmacology and genetic models. For example, knockdown of *Cyp26b1*—which metabolizes retinoids—in adult mouse anterior insula suppresses interest in social novelty by reducing spontaneous activity of excitatory neurons [38]. Likewise, depressive symptoms have been observed in rats after intracerebroventricular all-*trans* retinoic acid (ATRA) administration [39]. In addition, RARA is more abundant in CRH neurons of affective disorder hypothalami [40], where it both upregulates corticotropin-releasing hormone (CRH) expression and blocks glucocorticoid negative feedback on CRH [41], suggesting a link between retinoid TFs and elevated hypothalamic-pituitary-adrenal axis activity in MDD. Finally, given the substantial shared genetic risk across psychiatric disorders [42], it is notable that schizophrenia GWAS loci show enrichment for retinoid TR [43], and that circulating retinoids are dysregulated in schizophrenia patients [44]. Similarly, retinoid pathway genes, including *CYP26B1*, are dysregulated in postmortem brain from autism spectrum disorders, bipolar disorder, and schizophrenia patients [45]. Interestingly, retinoid deficiencies have been associated with these diseases, including recent observations of reduced serum levels of retinoic acid and its precursor, retinol, in schizophrenia [44]; similarly, reductions in serum retinol and expression of all three *RAR* genes were shown in autism spectrum disorders [46]. These findings led us to speculate that a component of MDD-associated genetic risk may likewise demonstrate an upstream convergence via recurrent retinoid-mediated TR disruptions across loci.

Massively parallel reporter assays (MPRAs) provide a solution to both experimentally identify functional variants and, consequently, their shared TR features. MPRAs assess thousands of DNA elements for transcriptional-regulatory functions and allelic differences simultaneously by pairing each short genomic sequence element of interest to several unique barcodes, with a constant promoter and reporter gene placed in between [47–50]. Delivery of a library of DNA elements to cells, followed by RNA collection and sequencing, enables quantitative estimation of the expression driven by each element as a ratio of expressed RNA barcode to delivered DNA barcode. These assays have recently been adapted to systematically identify SNPs with functional allelic TR differences from GWAS loci for several diseases [51–59]. Two key features make MPRAs advantageous for identifying both functional SNPs and their TR interactions. First, the assay is carried out via transfection and targeted RNA sequencing, meaning it can be executed in unmodified cell lines appropriate to the

application. Second, MPRAs can be conducted to define TR effects of experimental manipulations in these systems, such as drug administration [60, 61].

Therefore, we sought to experimentally identify functional TR SNPs from 39 GWAS loci associated with MDD, neuroticism, and broader psychiatric disease risk, with the hypothesis that functional SNPs converge at the level of retinoid-mediated TR. From broad linkage regions, we selected over 1000 SNPs based on overlapping human brain and neural epigenomic signals suggestive of TR activity. Critically, selection of neither the loci nor the SNPs was predicated on retinoid involvement, allowing for unbiased functional screening of a cross-section of MDD GWAS loci. To ensure we could detect SNPs subject to retinoid-mediated TR, we used neuroblastoma (N2a) cells, as they are strongly and rapidly retinoid-responsive [62, 63]. Our initial assay identified over 75 functional SNPs from 29 GWAS regions, confirming that GWAS loci contain several functional SNPs. We then examined whether these functional SNPs possessed shared upstream TR features—namely, transcription factor (TF) binding motifs. Remarkably, there was indeed enrichment of retinoic acid binding TFs among the MPRA-functional vs. -non-functional SNPs, supporting our hypothesis. To further characterize retinoid effects on TR at MDD-associated SNPs, we performed a second assay using all-*trans* retinoic acid (ATRA), known to stall division of N2a and other neuroblastoma cells by inducing neuronal-like differentiation [62]. First, we found that functional SNPs containing retinoid receptor motifs had increased magnitudes of effect in the presence of ATRA, consistent with bona fide retinoid receptor TR activity. More broadly, ATRA led to striking rearrangements of the baseline regulatory landscape, including altered magnitude and reversed direction of allelic effects. In addition, it revealed new SNPs with allelic TR differences unmasked by ATRA treatment. Significant ATRA-allele interaction SNPs largely overlapped RXRA binding sites from chromatin immunoprecipitation (ChIP)-seq, as well as motifs of several known retinoid-induced TFs, indicating broad roles of both retinoid TFs and their downstream TR systems at functional MDD-associated SNPs.

Finally, we explored the cell type-specificity of TFs predicted to regulate our functionally identified SNPs. Strikingly, we found TFs highly specific to serotonin neurons were strongly enriched among those we predicted to be recurrently involved in retinoid-dependent SNP function. These findings suggest that the broad transcriptional-regulatory systems engaged by retinoids—and as we illustrate, the genetic component of MDD risk they engage—may converge on serotonergic neurons. In summary, we identify MDD-associated functional SNPs with both baseline and ATRA-mediated allelic differences in TR, and these disproportionately show upstream convergent regulation by retinoid receptors and TFs they induce. This highlights a striking potential point of convergence between genetic risk loci and an environmental risk factor for MDD.

METHODS

Identifying candidate psychiatric GWAS regulatory variants

To prioritize putative regulatory variants from neuropsychiatric disease GWAS regions (predominantly MDD; Fig. 1A), SNPs in linkage disequilibrium (LD) with GWAS tag variants at $R^2 > 0.1$ were collected and intersected with histone modification, eQTL, Hi-C, and enhancer segmentation datasets from human postmortem tissue and neural lineage cell lines (see Supplemental Methods, Fig. 1B). SNPs were manually selected based on diversity and density of annotation overlap within each locus (Supplemental Methods). As a negative control, we identified candidates from one additional locus associated with several anthropomorphic traits [64], in a trait-blinded manner. Altogether, 1453 SNPs were selected. Final LD of selected SNPs was distributed similarly to starting SNPs (Fig. 1D). To confirm that we could detect CNS-relevant regulatory SNPs, a positive control TR SNP functionally demonstrated in mouse retina and brain [57] was also included.

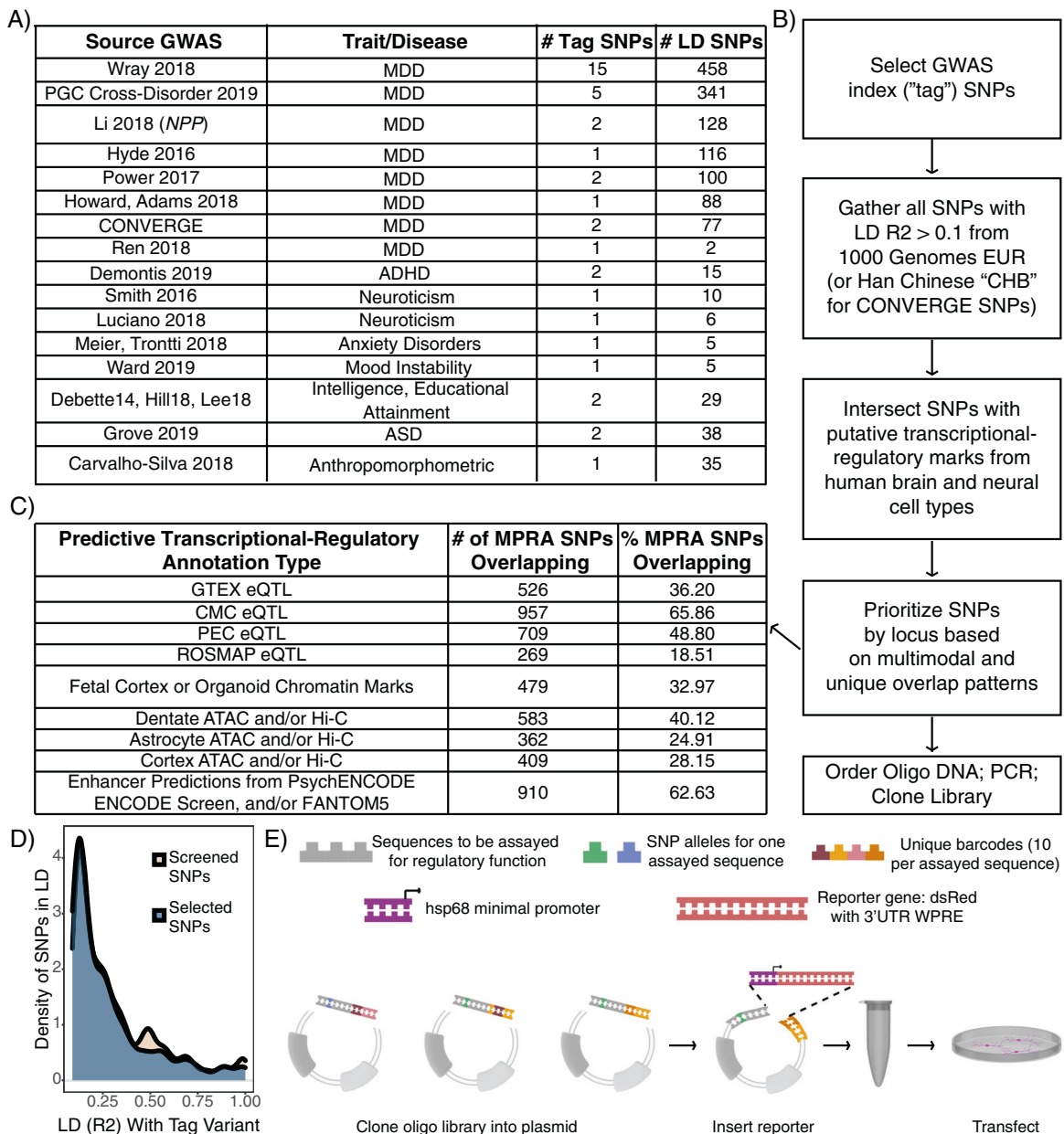


Fig. 1 Design of an MPRA library to identify candidate functional SNPs in MDD loci. A Table of GWAS studies and number of loci covered in the MPRA library. **B** Flow chart of design and prioritization process. **C** Brain and neural transcriptional-regulatory predictive annotation overlap with SNPs included in MPRA library. Fraction and number of SNPs in designed MPRA library intersecting each transcriptional-regulatory predictive annotation type. **D** The manual prioritization process was not LD biased. The subset of prioritized SNPs are spread over LD space similarly to the full set of screened SNPs. **E** Schematic of library construction and delivery. Panel adapted from [50].

Human genomic sequence (hg19) tiles up to 126 bp were taken centered on the 1454 candidate enhancer SNPs, each paired to ten unique 10 bp barcode sequences per allele and ordered as an oligonucleotide (oligo) pool from Twist Bioscience (San Francisco, CA). Also in the pool were 110 “basal” barcodes (no human genomic sequence cloned upstream of the minimal promoter), such that the only variable sequence between reporter clones was the barcode itself. The oligos were PCR amplified, then cloned into plasmid (Fig. 1E); subsequently, a reporter cassette containing a minimal promoter (*hsp68*) driving the dsRed reporter gene [65] and the untranslated “woodchuck” element (for RNA stabilization, to improve signal) [66] was cloned in.

Massively parallel reporter assays

N2A cells were grown in uncoated 6-well plates in medium consisting of 0.1 μ M vacuum-filtered DMEM with 10% Fetal Bovine Serum (2% fetal

bovine serum for the ATRA assay, based on media conditions from the literature [67, 68]). For transfection, cells were reverse transfected by plating in antibiotic-free medium onto pre-plated 400 μ L mixtures of 2.5 μ g plasmid with Lipofectamine 2000. In the first assay, $n = 6$ replicate wells were transfected and co-prepared for sequencing. A power analysis of these results using the 25th, 50th, and 75th percentile standard deviation of sequence expression measurements indicated we were $\geq 80\%$ powered to detect Bonferroni-corrected $p < 0.1$ variant effects as low as $\text{abs}(\log_2\text{FC}) 1.1$ (Supplemental Methods). In the drug MPRA experiment, $n = 12$ wells were transfected, harvested, and prepared for sequencing together, with $n = 6$ ATRA-treated and $n = 6$ vehicle-treated.

After transfection, cells incubated for 7 h at 37 $^{\circ}$ C and 5% CO_2 . Medium was replaced with the respective medium containing antibiotics, and in the second assay, a final concentration of 20 μ M ATRA dissolved in DMSO, or equivalent volume of vehicle (DMSO). Medium was not replaced before RNA collection in the first assay; in the second assay, it was refreshed every 24 h.

Seventy-two hours after transfection, cells were collected and RNA extracted using the Zymo (Irvine, CA) Clean-and-Concentrator 5 kit per manufacturer instructions. Eluted RNA was treated with Turbo DNA-free kit to remove any residual plasmid to prevent contaminating DNA reads during sequencing, and cleaned a second time using the Zymo kit as above.

Targeted sequencing of RNA and input plasmid

Briefly, equal amounts of RNA (1 µg) from each sample were prepared for sequencing by targeted cDNA synthesis using a primer against the distal 3' UTR of the reporter. These, along with input plasmid, were subjected to PCR, enzymatic digestion, ligation of Illumina sequencing adapters, and a final PCR to add sample indexes for sequencing. Enzymes, and size-selection cleanup steps used in this process are fully detailed in Supplemental Methods. No reverse-transcriptase controls utilizing sample RNA were co-prepared for both experiments and did not generate detectable product, indicating sequencing amplicons generated from RNA samples were exclusively representative of RNA content. Samples were sequenced to an average depth of ~8 million reads (first assay) or ~20 million reads (second assay).

Analysis

Allelic SNP effects on expression in the first assay and in single-condition analyses of the second assay were assessed by *t*-testing the element's expression of each allele across replicates. In the first assay, over 90% of SNPs had normally distributed expression values (Shapiro–Wilk test, $p > 0.05$). Uncorrected *t*-test *p*-values and Mann–Whitney U test *p*-values were well-correlated for the 89 non-normally distributed SNPs (Pearson's $r = 0.825$). Nonetheless, for *t*-test significant SNPs P_{emp} not passing the Shapiro–Wilk test, we verified the result by checking for a nominally significant Mann–Whitney U test at $p < 0.05$. No SNPs were excluded from analysis on this basis. dbSNP-assigned reference (“ref”) and alternative (“alt”) alleles for each SNP were used to define comparison direction (the difference of activity under the alt allele vs. the ref allele). For the first MPRA and single-condition analysis of vehicle samples from the second assay, empirical *p* values (P_{emp}) were calculated via simulated allelic comparisons between random subsets of “basal” barcodes (see Supplemental Methods) following an analogous procedure from a multiplex CRISPR study [69], with significance defined as $P_{\text{emp}} < 0.05$ unless specified otherwise. This ensures that a representative cross-section of expression variability driven by barcode sequences is accounted for when assessing TR differences. Single-condition analysis of ATRA samples utilized standard Benjamini–Hochberg FDR correction, as primary effects of interest in these samples were ascertained by linear modeling. For analysis of ATRA effects, we verified that variances were similar between the drug and vehicle conditions; indeed, the median barcode expression level standard deviation was 0.1216 in ATRA-treated and 0.1226 in vehicle-treated samples (with respective 25th and 75th percentile standard deviations also matched within 0.005 expression units). We calculated samplewise barcode-level expression values passing the “single-condition” filtering steps used for *t*-testing (Supplemental Methods) were fitted using a linear mixed model (LMM) requiring a minimum of 40% (96) of the 240 possible expression measurements per SNP. The LMM included a random term for replicate (to account for well-specific effects), expressed as: barcode expression ~ allele + drug + allele:drug + (1|replicate). Empirical *p* value calculation from LMM *F* statistics was performed in an analogous manner to the prior experiment, generating a vector of *F* statistics for each coefficient from 20,000 randomized basal-only comparisons. All SNPs with an interaction $P_{\text{emp}} < 0.05$ also had a likelihood ratio test (LRT) $p < 0.051$ comparing a maximum-likelihood (ML) interaction model to an ML LMM with additive terms only, indicating that the interactive model was more predictive but not overfit compared to an additive model. For SNPs with significant allele and interaction coefficients, a meaningful allele main effect was considered present if the single-condition vehicle and ATRA analyses showed the same allelic direction of effect, with a vehicle $P_{\text{emp}} < 0.1$ and ATRA FDR < 0.1 (i.e., near-significant within each condition of $n = 6$, thus reasonably capable of achieving significance in the LMM analysis of the two conditions combined).

MotifbreakR analysis and functional SNP enrichment for perturbed motifs

The motifbreakR [70] package was used to identify TF binding motifs significantly different between alleles of each SNP. Briefly, the number of MPRA-identified functional SNPs matching a given TF's motif(s) for at least

one allele was compared to the number of non-functional SNPs matching across 10,000 random draws of n (number of significant) SNPs. A second version of this analysis focused on the concordance rate—that is, whether the frequency of functional variants experiencing concurrent strengthening of motif and expression or vice versa—was significant compared to 10,000 draws of n random SNPs from the analyzed set. Analysis of the first assay's SNPs defined functionality based on a P_{emp} threshold of 0.05. We performed two motif analyses of the second assay results, one comparing allele main effect SNPs ($P_{\text{emp}} < 0.1$) to those with $P_{\text{emp}} > 0.1$ for allele, drug, and interaction effects, representing the breadth of functional variant-susceptible *cis*-regulators. The second analysis compared interaction SNPs ($P_{\text{emp}} < 0.05$) to SNPs with an allele main effect ($P_{\text{emp}}^{\text{allele}} < 0.1$) but no interaction ($P_{\text{emp}}^{\text{interaction}} > 0.1$).

Analyses of functional-SNP enriched TF expression in human brain and chromatin immunoprecipitation (ChIP)-seq

We utilized outside ChIP-seq datasets to validate motif-based predictions of retinoid receptor binding and refine prediction of involved TFs. We intersected our functional SNPs to 25 tracks of ChIP-seq for retinoid receptors (19 human [71, 72], 6 mouse (3 ATRA-treated, 3 vehicle-treated) converted to hg19 coordinates using UCSC's LiftOver [73]); 11 tracks of RXR-heterodimerization partners (10 human THRA/THRB [71, 72] and one aggregate analysis of human VDR [74]); and human genome-wide predictions of DR5 [75], a canonical RAR-RXR heterodimer binding sequence. For functional SNPs implicated at an RAR, RXR, VDR, or THRA/B site by either motifbreakR or ChIP, we identified potential target genes using chromatin-conformation [76] and eQTL [77–80] data. We performed broad-scope gene enrichment analysis of this gene set using Enrichr [81]. To examine shared biology of TFs implicated by motifbreakR enrichment at functional variants, we utilized PantherDb [82]. We finally examined TFs for enrichment among highly-expressed genes in adult and developing human brain using the ABAEnrichment package's Wilcoxon approach [83], effectively weighting TFs by the number of functional SNPs implicated by motifbreakR (see Supplemental Methods).

RESULTS

Many MDD loci contain more than one functional SNP

We identified >1000 SNPs from MDD-associated GWAS loci, prioritizing SNPs overlapping with epigenetic data from neural samples, and cloned them into an MPRA library (Fig. 1). We included one positive control SNP, shown to alter neural tissue gene expression, and one control locus near *CDKAL1* not a priori associated with psychiatric disease. To identify functional variants from these SNPs, the library was transfected into N2a cells ($n = 6$ replicates, Fig. 1E). Variant activity was assessed by RNA sequencing and barcode counts compared to input plasmid barcode counts. After filtering for read depth and barcode representation, 1013 SNPs spanning all 40 LD regions remained for analysis. Results were highly replicable across samples (Pearson r 0.63–0.85 for barcode expression; 0.90–0.96 for elements, Supplemental Figure S1). We use “element” to signify the set of barcodes corresponding to one unique sequence of interest (1 SNP = 2 elements).

Of 1013 SNPs analyzed, we identified significant allelic TR ($P_{\text{emp}} < 0.05$) at 76 SNPs (65 from MDD loci; 1 from the control *CDKAL1* locus) across 27 of the 40 analyzed GWAS regions, with effects ranging 0.1 to 0.63 (median 0.2) log₂ fold-change (Fig. 2B). Interestingly, the functional variant from the control locus is suggestively associated (GWAS $p < 5 \times 10^{-6}$) with “Poisoning by analgesics, antipyretics, and antirheumatics” in UK Biobank [84]. As this likely includes attempted suicides, the SNP was retained for analyses. The positive control SNP, which we utilized to confirm our ability to detect small-effect sizes expected of regulatory SNPs, showed the expected lower expression of the T allele at a P_{emp} of < 0.051 (Fig. 2A) [57].

While our assay was designed to broadly examine wide LD regions around GWAS index variants, we did identify one functional variant, rs11209952, in a fine-mapped credible set of variants for seeking general practitioner care for depression in UK

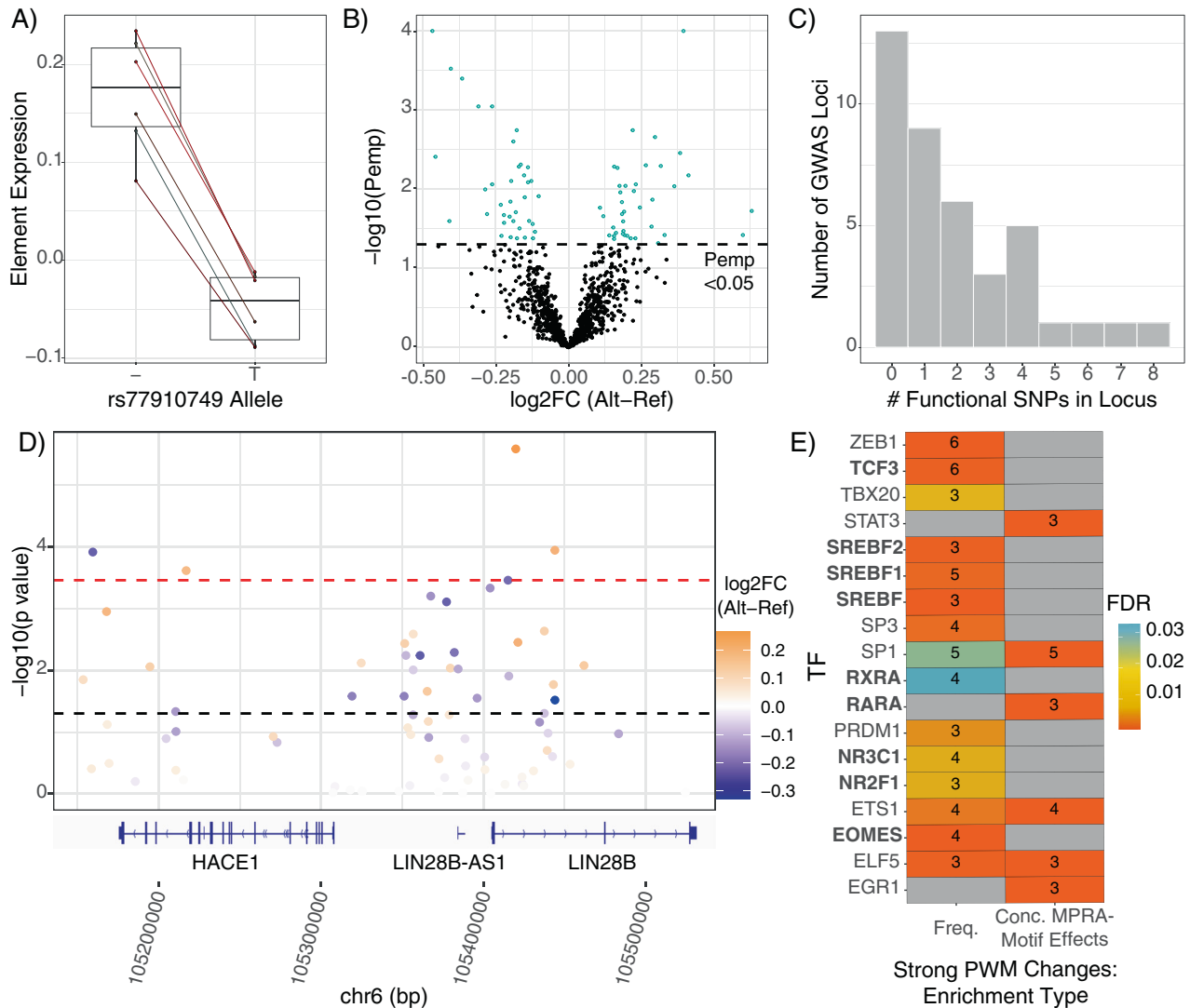


Fig. 2 MPRA defines SNPs with a functional effect on gene expression. **A** MPRA results of positive control SNP. Shen et al. found that the T allele drove decreased expression relative to the deletion (“-”) allele, which was robustly reproduced in the present assay. **B** Volcano plot of allelic differences in reporter expression. Points represent one SNP’s composite \log_2 allelic fold change (alt vs. ref), determined as the mean of samplewise alternative allele barcode expression minus the matched mean of reference allele barcodes. The dotted line indicates the statistically corrected significance threshold. **C** Number of functional SNPs (MPRA-significant SNPs) per GWAS locus in the assay. Number of loci (y-axis) containing a given number of MPRA-significant ($P_{\text{emp}} < 0.05$) SNPs (x-axis). **D** The LIN28B locus harbors several functional SNPs. SNPs are plotted according to their chromosomal position (hg19) and colored based on their composite \log_2 allelic fold change. Refseq genes are visualized by the Integrative Genomics Viewer [116]. **E** TF binding motifs involved in retinoid signaling, steroid synthesis and response, and neural activity are enriched among functional SNPs. Boxes are colored by FDR-corrected significance of enrichment for motifbreakR-defined “strong” allelic perturbations to binding motifs among functional SNPs; the number of functional SNPs perturbing (left column) and/or with concordant motif and MPRA effects (right column) are shown. Concordant effects were defined by greater MPRA expression driven by the allele better-matched to the corresponding TF motif and vice versa—the expected behavior of strictly activating TFs.

Biobank [85]. Moreover, consistent with prior studies of “conditional” or “secondary” SNP associations—wherein additional LD SNPs have associations independent of their linked, larger-effect variant [20, 86, 87]—we identified several loci with multiple functional SNPs (Fig. 2C) (range 1–8, mean 2.8, median 2). Notably, we identified as functional rs1806153, a recently defined “conditional SNP” for MDD [20]. Our findings support models predicting multiple functional SNPs in GWAS loci, and directly validate one such finding from association analysis.

One notable TR SNP we identified, rs314267, comes from a “LIN28B” (nearest gene) GWAS locus repeatedly linked to MDD

[9, 12] as well as cross-psychiatric disorder risk [42]. MPRA significance and effect size are illustrated for the region, showing that this locus contains several functional SNPs (Fig. 2D). All significant MPRA SNPs in the locus had effect directions consistent with brain eQTLs. rs314267 is the most significant LIN28B eQTL SNP (eSNP) in the region in PsychENCODE [78], and is a CommonMind Consortium (CMC) eSNP for both LIN28B and HACE1 [77]. HACE1 is also downregulated in postmortem MDD hippocampal CA1 [88]. Hi-C data from human neural cell cultures suggest rs314267 is within a neuron-specific LIN28B regulator, with promoter chromatin contacts found in dentate and cortical

neurons, but not astrocytes [76]. *LIN28B* plays broad roles in neurodevelopment [89] and has potentially sex-differentiated functions [90–93]; considering sex differences in MDD prevalence and severity [94, 95], *LIN28B* constitutes an especially interesting gene target from this locus. Finally, we examined potential upstream TR mechanisms for SNP activity using VARAdb [96]. Query of rs314267 revealed a two order of magnitude allelic difference in the motif match p -value for *TCF4*—a gene itself implicated in cross-psychiatric-disorder risk [42, 97]. Overall, the identification of functional SNPs implicated in regulation of *HACE1* and *LIN28B* exemplifies the ability of MPRA to identify functional variants involving sensible TR mechanisms and target genes.

Shared regulatory architecture across distinct loci

We next sought to test our hypothesis that functional MDD risk variants shared retinoic acid-related TR architecture. If so, functional SNPs should disproportionately disrupt binding sites of retinoid-binding TFs compared to SNPs without an allelic effect on TR. Such data would indicate that MDD risk is mediated in part through perturbations of specific upstream transcriptional circuits and may highlight how risk conferred through retinoids converges with risk conferred through genetics to perturb downstream gene expression.

To take an unbiased approach to our retinoid hypothesis, we broadly analyzed all motifs showing enrichment at TR SNPs. Motifs for several dozen TFs were perturbed by the functional SNPs more frequently than expected, often with ‘strong’ perturbations to motifs and/or overrepresentation of concordant expression effects (Fig. 2E, Supplemental Figure S2). This included several TFs aligned with biological processes relevant to psychiatric disease. For example, several TFs are involved in steroid pathways, from regulating biogenesis (*SREBF* family, 6 SNPs) to conveying downstream TR effects—most notably, via glucocorticoid receptor (*NR3C1*, 5 SNPs; Supplemental Figure S3), a central component of the stress response. Functional SNP overrepresentation of *SREBF* motifs is consistent with high expression of these TFs in N2as and related neuroblastomas [98, 99]. A second group of transcription factors included three TFs involved in neural lineage commitment/development: *TCF3* [100, 101], *EOMES*, and *NR2F1* [102] (6, 4, and 3 SNPs, respectively). Altogether, functional SNP enrichment for these TFs’ motifs bolster our confidence in this approach, as (a) detected variation involves TFs known to be expressed in N2As (*SREBF*); (b) functional variation involves TFs with roles in developing CNS, where disease variants likely act; and (c) that the single-best characterized trigger of MDD (stress) is reflected in enrichment of alterations to *NR3C1* motifs.

Finally, consistent with our hypothesis of convergence on retinoid-mediated TR, functional variants were enriched for “strong” perturbations of retinoid receptor TF motifs (Fig. 3), including *RARA*, *RARB*, and *RXRA* (5 SNPs from 4 MDD loci, Fig. 3). Especially notable is the motif configuration at SNP rs34416841, which falls within three partially overlapping motifs for retinoid TFs. In addition, the elements overlapping rs489591 and rs13330178 appear to be functional human retinoid TF binding sites *in vivo* based on DNase hypersensitivity footprinting [103].

Retinoids unmask additional functional SNPs in MDD loci

Our findings supported the hypothesis that MDD-associated variants across multiple loci converge on TR, including that modulated by retinoids. We thus designed a pharmacological follow-up with two goals in mind. The first goal was to functionally verify that retinoids were involved in TR at SNPs where their motifs were found (in *cis*), and potentially unmask additional retinoid-targeted alleles. Our second goal was to further assess retinoid signaling *trans* (i.e., indirect) effects on variants from these same GWAS regions, e.g., via non-retinoid TF induction, co-regulation, or

repression [33]. Therefore, we performed a second MPRA with an *all-trans* retinoic acid (ATRA) condition.

After 48 h, cultures were imaged to verify drug activity (as ATRA is light-sensitive) based on known morphologic responses of N2as to ATRA, which include neurite outgrowth and mitotic arrest [62, 104, 105]. Indeed, drug-treated cells had a qualitatively lower cell density and produced neurite-like processes (Fig. 4A) in comparison to vehicle-treated cells (Fig. 4B). After RNA sequencing, we first analyzed vehicle-treated replicates alone to ensure replicability of the assay. Element expression levels in the vehicle condition strongly correlated to the first experiment (Pearson $r = 0.91$; Fig. 4C), and replicated the functional variants (Fig. 4D); all 31 shared significant SNPs showed consistent directions of effect.

We next applied a linear mixed model (LMM) to identify SNPs responding to ATRA (that is, allele-drug interactions). A total of 1079 SNPs were analyzed after filtering for read and barcode depth. In part due to the effective doubling in power to detect allelic effects with 12 replicates and the LMM approach, we now identified 137 variants with a main effect of allele (129 from MDD loci). Four of the five retinoid receptor motif-perturbing variants from the first assay passed filtering; all four of these variants again showed allelic main effects (all $P_{\text{main}} < 0.01$), as did many other functional variants identified in the previous experiment (Fig. 4D). To our surprise, more variants showed a significant drug-allele interaction effect: a total of 128 SNPs (122 from MDD loci) (Fig. 4E, F). Among the drug-allele interaction SNPs were one of the four retinoid-related SNPs identified from the first assay (rs4801117; $P_{\text{interaction}} < 0.025$, Fig. 4G), while another trended toward interaction (rs489591; $P_{\text{interaction}} = 0.117$). This strongly supports a role of retinoid TF activity at rs4801117 as predicted by the motif analysis. More broadly, comparison of changes between the two conditions reveals the striking extent to which the regulatory landscape of the N2As was altered by ATRA (Fig. 4F and H). Notably, several additional functional variants were identified in the previously highlighted *LIN28B* locus, further illustrating multi-variant and context-dependent aspects of GWAS loci (Fig. 4E and I). In all, this experiment highlights the ability of MPRA to detect contextual influences such as cell states and signaling on functional noncoding variation, and to unmask distinct, context-dependent functional SNPs.

Retinoids reveal additional axes of convergent regulation at functional MDD-associated SNPs at levels of TF and cell type

As the ATRA-based assay provided improved power to identify allelic variant effects on expression, we again employed our motifbreakR-based analyses to assess convergent transcriptional mechanisms underlying identified regulatory variants. When examining SNPs with allelic effects in comparison to SNPs with no allelic, drug, or interaction effects, several retinoid receptor motifs were again overrepresented, including those of *RXRA*, *RXRB*, *RARA*, and *RARG* (Fig. 5A, Supplemental Figure S4), totaling 11 of the 92 allele main effect SNPs analyzed, spanning 10 MDD GWAS loci. These findings further support retinoid receptor binding sites as an upstream regulatory system recurrently involved in MDD risk genetics.

As retinoids resulted in stark changes across the transcriptional-regulatory landscape, we further sought to predict TFs potentially underlying allelic effects following retinoid exposure. Therefore, we also analyzed the interaction SNPs in comparison to allelic SNPs that were *not* subject to interactions. This revealed a novel set of TFs not observed in the preceding analyses, including TFs with roles in neural differentiation and maturation (Fig. 5A, Supplemental Figure S4), as well thyroid hormone receptor *THRB*, an *RXR* binding partner. We compared the overrepresented motifs to TFs recently demonstrated to be upregulated in human neuroblastoma lines (*KCNR*, *LAN5*) by ATRA. Of the 26 TFs identified as ATRA-induced in these

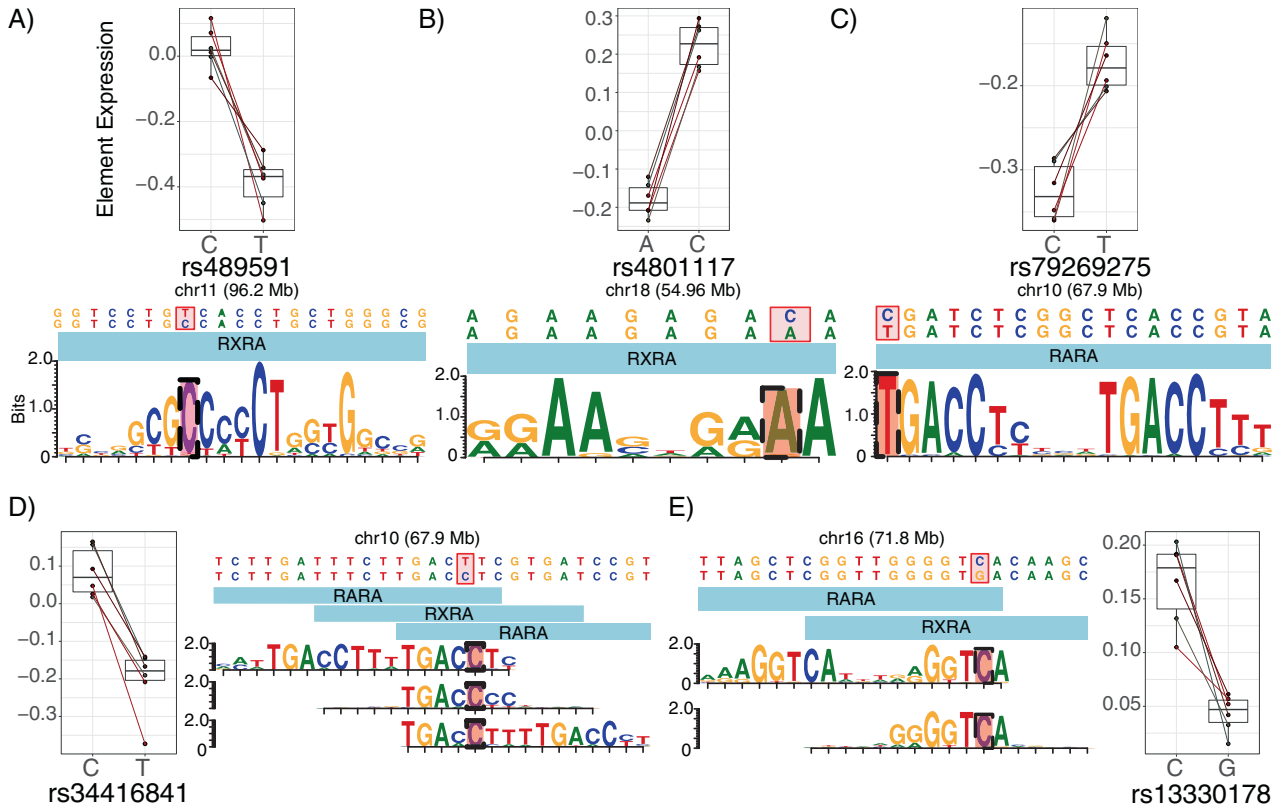


Fig. 3 MPRA signal at SNPs disrupting putative retinoid TF motifs. The five functional SNPs driving enrichment signals for retinoid TF motif perturbations are shown, along with the MPRA results for each variant. Each motif diagram shows only distinct position-weight matrices (PWMs) (motifbreakR uses a large meta-collection of motifs, which were often identical or nearly identical across retinoid TFs; such redundant motifs are not shown). Among functional-SNP enriched retinoid TFs, **A** rs489591 and **B** rs4801117 exclusively perturb RXRA motifs. **C** rs79269275 perturbs an RARA (or RARB, identical but not shown) motif. **D** rs34416841 alters several similar retinoid motifs across multiple positions and TFs. Not shown: near-identical motifs for RARB along the same sequence as the 5' RARA motif; near-identical RXRB and RARB motifs along the same sequence as the center RXRA motif; and a near-identical RARB motif along the same sequence as the 3' RARA motif. **E** rs13330178 disrupts an RARA or RXRA binding site. Given that RXRA and RARs are known to heterodimerize, it is possible that this SNP disrupts the RXRA component of such a heteromer's binding sequence.

two lines, motifs were available for 18 in our analysis. Of these, 6 of the TFs were enriched among allele main-effect-only SNPs, while 12 of these TFs were enriched among the retinoid-allele interaction variants [104] (Fig. 5A, Supplemental Figure S4), supporting our predictions of TFs playing ATRA-dependent roles at functional SNPs.

Integrative analysis of TF sets at functional variants: TF binding, spatiotemporal brain enrichment, and putative target genes

As retinoid receptors have highly redundant binding motifs, we sought to both validate motif-based implication of retinoid receptors and more finely identify the particular TFs binding at functional SNPs. We aggregated ChIP-seq data for RAR, RXR, and RXR-heterodimerization partners (VDR, THRA, THRB) and identified functional SNPs overlapping peaks for each TF. Altogether, 35 of our 277 functional SNPs from across the two assays were in at least one such binding site (Supplemental Table 1). 15/17 of the allele-ATRA interaction SNPs overlapped a ChIP peak for RXRA, suggesting RXRA may be the common mediator of the observed retinoid-dependent SNP effects.

We also performed Gene Ontology analysis of functional variant enriched TFs against a background of all TFs in the motifbreakR tool using PANTHERdb but found no detailed Biological Processes of note (Supplemental Table 2). We next sought to examine whether TFs enriched at functional variants in our motif analyses corresponded to particular spatiotemporal expression patterns in

the brain. To favor the most broadly-implicated TFs, we utilized the ABAEnrichment package's Wilcoxon analysis approach on the TF sets from the ATRA experiment using the number of motifbreakR SNPs as the TF gene "scores". In this analysis, several brain regions across developmental stages were nominally enriched (family-wide error rates <0.05) in ATRA-dependent and -independent TF expression, with especially broad enrichment at high expression thresholds (≥ 90 th percentile) in adolescent brain (Supplemental Table 3). This does not appear to be an artifact of the cell model, considering that neuroblastomas are arrested in a neural crest progenitor (i.e., pre-/peri-natal cell type) stage. If replicated in future studies with larger adolescent sample numbers, this may suggest that retinoid-mediated aspects of MDD genetic risk are especially active in the adolescent brain, perhaps contributing to frequent emergence of the disorder around this time.

We additionally utilized these sets of TFs as gene sets to investigate whether retinoid-dependent or -independent regulatory variants might be particularly active in certain cell types of the brain. We screened for enrichment of these TFs among genes with strong cell type-specific expression in brain as previously defined for over 20 cell type transcriptomes [32]. Three TFs (spanning 8 ATRA-interacting SNPs) were discovered to be highly specific to serotonin neurons (Fig. 5B): *GATA2*, *GATA3*, and *FEV*, while no cell type enrichments under FDR < 0.1 were noted for TFs linked to ATRA-independent variants. Supporting these findings, an enrichment analysis of putative target genes

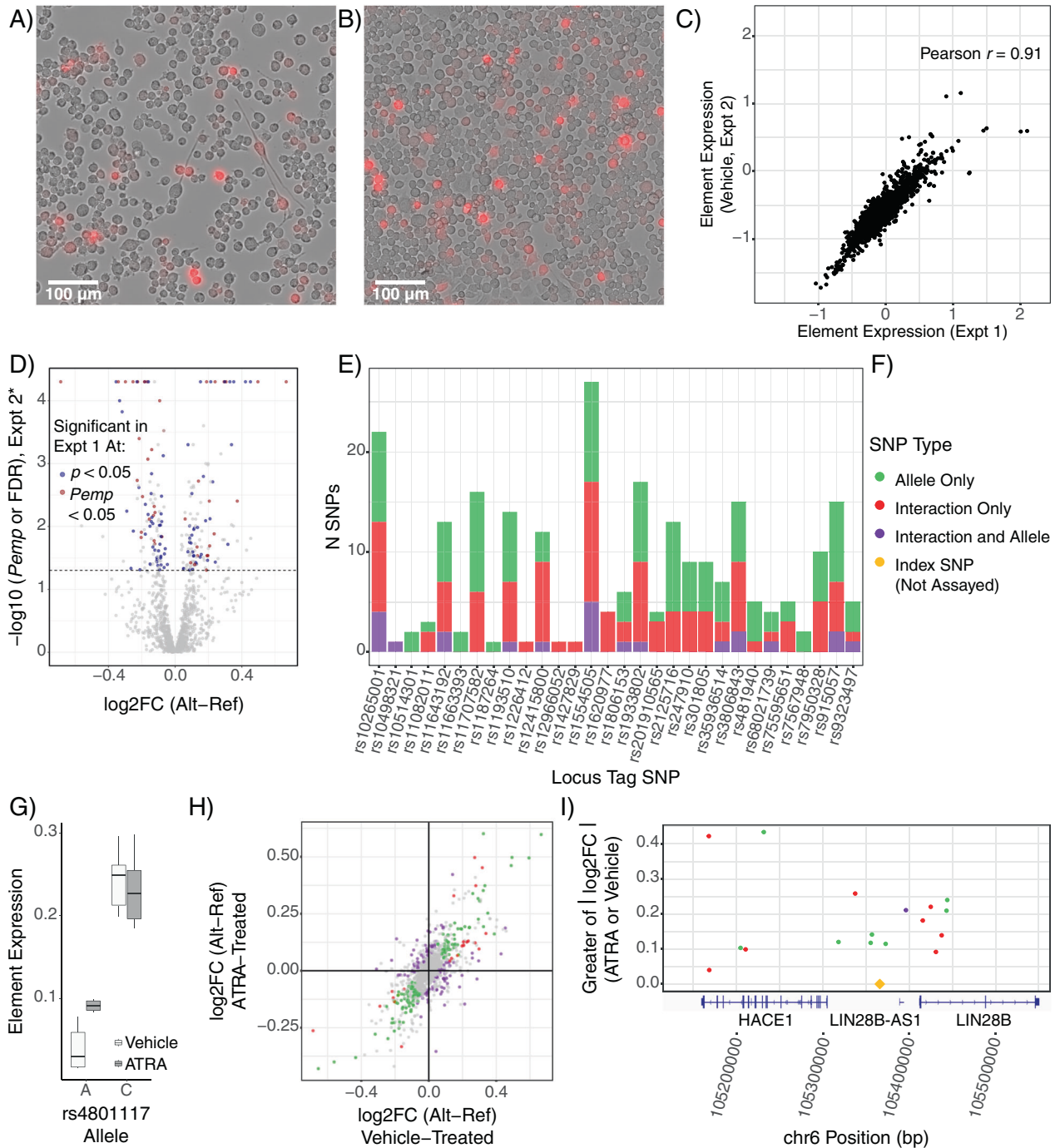


Fig. 4 Retinoid treatment alters transcriptional regulation and unmask additional functional variants. **A** ATRA-treated cells show growth arrest and neurite growth, demonstrating effective ATRA treatment, while **B** vehicle-treated cells continued to proliferate in a de-differentiated state. **C** Results of the vehicle treatment replicate the initial MPRA findings. Element (single-allele) expression values for each sequence assessed in both assays is plotted. **D** Significant and marginally significant functional SNPs from the first assay showed effects in the second assay. The larger allelic difference value from the ATRA and vehicle single-condition analyses is plotted for each SNP on the x-axis; the y-axis value is the corresponding corrected p -value (FDR correction for the ATRA-only analysis or P_{emp} for vehicle-only analysis). **E** Retinoids unmask functional SNPs with additional or exclusive retinoid-mediated effects. **F** SNP effect(s) color key for panels (**E**), (**H**), and (**I**). SNPs with both effects were those with LMM interaction $P_{emp} < 0.05$, LMM allele $P_{emp} < 0.05$, and both single-condition analyses showing the same allelic effect directionality at ATRA $FDR < 0.1$ AND vehicle P_{emp} . **G** rs4801117-A shows greater activity with ATRA treatment while the C allele is unaffected. The ATRA having an expression effect only on the A allele is consistent with the A allele matching the RXRA motif as shown in Fig. 3B. **H** Transcriptional-regulatory SNPs show a wide range of altered and unaltered effects with ATRA treatment. Single-condition \log_2FC values are shown. **I** Several additional SNPs with retinoid-dependent function (i.e., allele-ATRA interaction) in the *LIN28B* locus are illustrated. Notably, there are several functional SNPs clustered around the GWAS index SNP, suggesting association signal at this locus may be driven by multiple functional/conditionally functional variants.

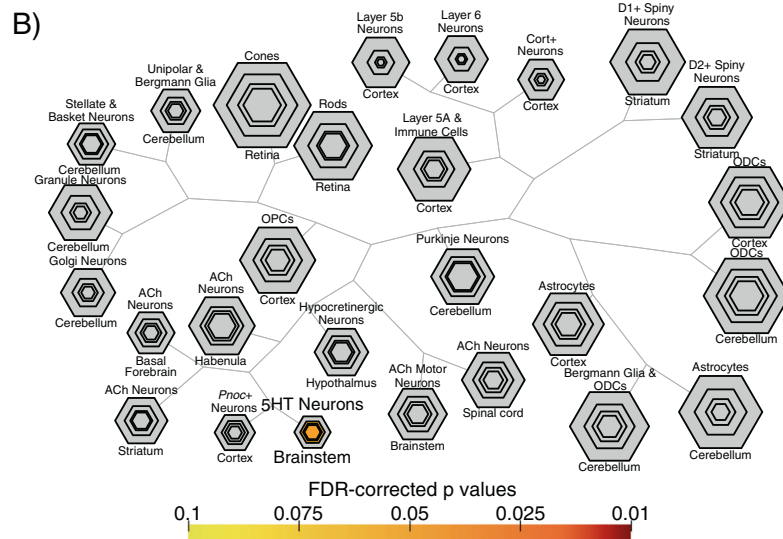
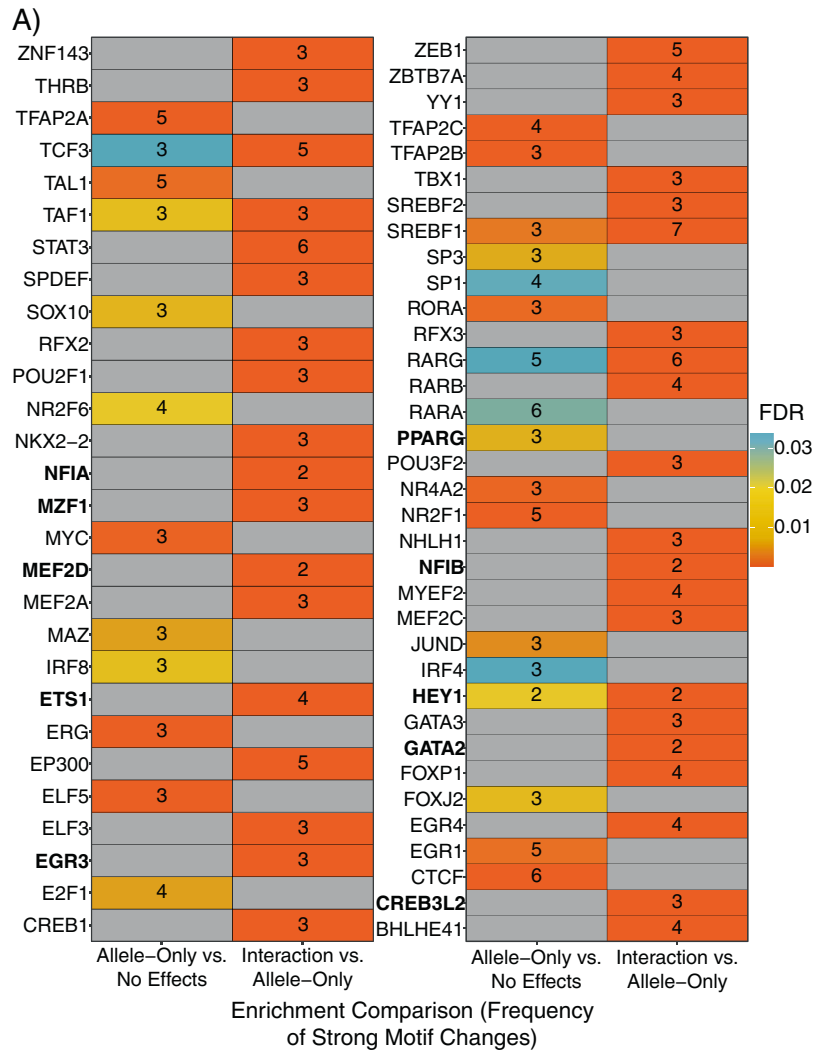


Fig. 5 Distinct TFs underlying retinoid-dependent functional SNPs and implication of serotonergic neurons. A Motifs overrepresented among ATRA-independent (allele effect without interaction) SNPs (left columns) or among ATRA-dependent (interaction) SNPs (right column). The heatmap is shown in halves for visibility. TFs identified as ATRA-upregulated in human neuroblastoma lines [104] are in bold font. **B** TFs implicated by ATRA-interacting SNPs significantly overlap TFs enriched in serotonergic neurons. Plot generated using the cell-specific expression analysis (CSEA) tool (<http://genetics.wustl.edu/jdlab/csea-tool-2/>) [32]. 5HT serotonin, ACh acetylcholine, ODC oligodendrocyte, OPC oligodendrocyte progenitor.

(implicated by brain eQTL or neural Hi-C) of SNPs in retinoid TF motifs or ChIP peaks (Supplemental Table 4) revealed 5 genes nominally enriched for high regional expression in rhombomere 9, which gives rise to medullary populations of serotonin neurons [106] (The full results can be explored at <https://maayanlab.cloud/Enrichr/enrich?dataset=27d6db2a8510a90ed0d78e6b60c59287>).

Using the R2 database (<http://r2.amc.nl>), we examined expression of *FEV*, *GATA2*, and *GATA3* TFs in 24 human neuroblastoma lines (GEO accession GSE28019), retinoic acid-treated human SH-SY5Y neuroblastoma cells [107], alongside human neural progenitors [108] and melanoma lines as comparators [109], confirming neuroblastomas strongly express all three of these TFs (Supplemental Figure S5). Single-cell RNA sequencing data from mouse brain confirms the specificity of these TFs, revealing that these TFs are only expressed in serotonergic, noradrenergic, peripheral autonomic, and midbrain inhibitory neurons—with all three expressed in serotonin neurons [110]. Furthermore, exogenous retinoids have been shown to lower circulating serotonin in humans [46] and to alter morphology of rat raphe neurons in slice culture [111], suggesting these neurons are retinoid responsive. We do not believe our finding is an artifact of the N2a system, as we could identify no evidence in the literature suggesting a serotonin-like identity of N2a cells with or without ATRA treatment. Altogether, these findings suggest serotonin neurons and closely related cell types [93] may be cellular points of convergence for several retinoid-mediated functional SNP effects on MDD risk.

DISCUSSION

To date, most functional investigations of SNPs in the context of psychiatric disorders have taken place in a low-throughput manner, such as single-variant classical reporter assays [21] or using CRISPR-Cas9 technology to edit limited positions for deep phenotyping [112]. Here, we leveraged MPRA to screen over 1000 SNPs from loci associated with MDD, related phenotypes, and broader psychiatric disease, demonstrating the utility of this technique for dissecting the functional regulatory architecture of psychiatric GWAS loci, and defining shared upstream regulatory features across loci.

In doing so, we identify over 100 SNPs with allelic effects on expression, with most coming from loci containing ≥ 2 functional SNPs. These data provide experimental support for the prediction that multiple SNPs with allelic effects exist within GWAS loci as put forth in polygenic/omnigenic theory literature. We further examined the omnigenic hypothesis' more central prediction of regulatory convergence across loci. By examining the shared regulatory features (TF binding motifs) based on enrichment at functional SNPs, we were able to predict several TFs with TR activity recurrently altered across MDD-associated SNPs, highlighting retinoid receptors in particular.

Retinoids are encountered both exogenously (e.g., as ATRA in oncology, and as isotretinoin, carrying a black-box warning for suicidality) and endogenously, including during brain development. To investigate how SNP functions may be altered by retinoids, we repeated the assay with an ATRA condition. ATRA drastically rearranged the TR landscape of N2a cells, resulting in altered and novel allelic effects at over 100 SNPs and revealing ATRA-dependent mechanisms of function across 122 SNPs from 22 of 26 MDD GWAS loci assessed. Of 17 ATRA-interacting functional SNPs overlapping ChIP peaks for retinoid receptors, 15 overlapped ChIP sites of RXRA, (Supplemental Table 1b) suggesting it may be central in functional SNP activity at retinoid receptor binding sites in this system. Interestingly, single-cell epigenomics of human cortical cell types recently found RXRA motifs to be uniquely enriched in open chromatin of SST interneurons [113], a strong candidate cell type for MDD [114].

These findings suggest that retinoid receptors—RXRA in particular—merit mechanistic follow-up regarding TR differences at MDD-associated SNPs. Future work may be able to leverage biobank-level datasets to ascertain whether retinoid-interacting SNPs are overrepresented in retinoid-treated patients experiencing adverse psychiatric side effects. While data on endogenous retinoids, e.g., plasma values, are not currently available in large genotype-phenotype-health record cohorts like UK Biobank, future datasets may enable investigation of circulating retinoids and their interaction with genotype in cognitive and psychiatric phenotypes.

The methodologic requirements of high-throughput assays such as MPRA bring inherent limitations to their results. The primary precaution in interpreting these results concerns cell type relevance. MPRA are subject to the TR landscape of the cell type used. Neuroblastoma cells, including N2As, are derived from peripheral neural crest progenitors—though they can be differentiated into dopaminergic neurons [105] and commit to neuronal differentiation with ATRA [62, 68]—and were selected for these assays based on intact retinoid signaling rather than representing a disease cell type per se. On the other hand, the neural crest-derived autonomic nervous system has received little consideration (relative to brain) in psychiatric genetics of MDD despite the well-appreciated role of stress in depression. These data may form an interesting foundation for future study of autonomic effects of MDD genetic risk.

Still, we can broadly speculate on brain cell types implicated by our findings. A notable prior pharmacology MPRA cleverly tested gDNA fragments for regulatory activity over a time course of dexamethasone treatment, while collecting epigenomic data in the same cell type over the same time course to compare MPRA signal and endogenous genomic marks. They found that endogenous genomic regulatory elements with repressive marks or depleted of glucocorticoid receptor binding were oftentimes active and/or dexamethasone-differentially active when assayed on the MPRA plasmids. This suggests that the transcriptional-regulatory capacity of an MPRA is not constrained by the epigenome of the model cell, but rather by its expressed TFs [61]. As such, retinoid receptor-mediated SNP functions observed are not limited to sequences that would be active in the N2a genome; as such, it is entirely plausible that the observed effects also occur in retinoid-receptor expressing brain populations. Mouse nervous system single-cell RNA-seq suggests retinoid receptor expression is absent in brain glia, but robust in many neuron types [110]. Thus, we suspect the directly-mediated retinoid receptor SNP effects we observe may be neuron-specific. Future studies may be able to address the interesting question of differences in neuronal subtypes exhibiting functional SNP effects.

We find that principles of the omnigenic model appear to hold true for MDD risk genetics, including the presence of far more functional variants (a total of 277 SNPs with allelic and/or interaction effects of 1178 assessed across the two assays; Supplemental Table 1) than there were GWAS loci (i.e., tag SNPs). We find, interestingly, that functional SNPs form convergent subsets of upstream (transcription-regulatory) sequences and systems, which in turn have shared retinoid dependence and are collectively enriched in serotonin neurons via 8 ATRA-interacting functional SNPs in binding motifs of *GATA2*, *GATA3*, and *FEV*. It has previously been demonstrated that systemic administration of ATRA depletes serotonin by over 40% in the rat brain [115], supporting the serotonin system as a convergent target of retinoid-regulated pathways. As GWAS of MDD begins to explore severe, treatment-refractory cases [7], it will be interesting to see whether associated variation still shows such convergence, as treatment-resistant depression (generally, non-response to two or more classes of antidepressant) effectively signifies non-response to multiple serotonergic agents.

In all, we assessed the architecture of *cis*-regulatory variation in psychiatric disease risk loci, identifying at least one functional SNP in the majority of the 40 GWAS loci examined, largely corresponding to MDD-associated SNPs. Strikingly, retinoid receptor binding sites and TR systems subject to regulation by ATRA have a substantial impact on whether and how MDD-associated SNPs are functional. These findings constitute a robust experimental demonstration of the influence of physiological and environmental states on the molecular activities of disease-associated SNPs, and constitute a high-confidence set of MDD SNPs meriting deeper functional characterization of both their TR mechanisms and their environmental interactions.

CODE AVAILABILITY

A summary spreadsheet of all significant SNPs identified in one or both assays, along with full analysis results, including barcode-wise expression in each sample, single-condition allelic effect tests, linear modeling results, and significantly enriched TFs in each of the comparisons executed, along with the code utilized to execute these analyses, is available at https://bitbucket.org/jdlabteam/n2a_atra_mdd_mpra_paper/src.

DATA AVAILABILITY

A summary spreadsheet of all significant SNPs identified in one or both assays, along with full analysis results, including barcode-wise expression in each sample, single-condition allelic effect tests, linear modeling results, and significantly enriched TFs in each of the comparisons executed is available at https://bitbucket.org/jdlabteam/n2a_atra_mdd_mpra_paper/src. Raw sequencing files are deposited in GEO under accession GSE167519. Supplemental information is available at Translational Psychiatry's website.

REFERENCES

- Smith K. Mental health: a world of depression. *Nature*. 2014;515:180–1.
- Üstün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL. Global burden of depressive disorders in the year 2000. *Brit J Psychiatr*. 2004;184:386–92.
- Greenberg PE, Fournier A-A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry*. 2015;76:155–62.
- Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol J Eur Coll Neuropsychopharmacol*. 2009;19:34–40.
- Flint J, Kendler KS. The genetics of major depression. *Neuron*. 2014;81:484–503.
- Corfield EC, Yang Y, Martin NG, Nyholt DR. A continuum of genetic liability for minor and major depression. *Transl Psychiatr*. 2017;7:e1131–e1131.
- Clements CC, Karlsson R, Lu Y, Juréus A, Rück C, Andersson E, et al. Genome-wide association study of patients with a severe major depressive episode treated with electroconvulsive therapy. *Mol Psychiatr*. 2021;1–11.
- Cai N, et al. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*. 2015;523:588–91.
- Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50:668–81.
- Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22:343–52.
- Levey DF, Stein MB, Wendt FR, Pathak GA, Zhou H, Aslan M, et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat Neurosci* 2021.
- Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet*. 2016;48:1031–6.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190–5.
- Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–30.
- Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012;22:1748–59.
- Chen J, Rozowsky J, Galeev TR, Harmanci A, Kitchen R, Bedford J, et al. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat Commun*. 2016;7:11101.
- Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am J Hum Genet*. 2016;99:139–53.
- Boyle EA, Li Yi, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177–86.
- Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Salari R, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res*. 2014;24:1–13.
- Byrne EM, Zhu Z, Qi T, Skene NG, Bryois J, Pardinas AF, et al. Conditional GWAS analysis to identify disorder-specific SNPs for psychiatric disorders. *Mol Psychiatr*. 2020;1–12.
- Li S, Li Y, Li X, Liu J, Huo Y, Wang J, et al. Regulatory mechanisms of major depressive disorder risk variants. *Mol Psychiatr*. 2020;1–20.
- Furlong LI. Human diseases through the lens of network biology. *Trends Genet*. 2013;29:150–9.
- Liu X, Li Yi, Pritchard JK. Trans effects on gene expression can drive omnigenic inheritance. *Cell*. 2019;177:1022–34.
- Leeuw CA, de, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11:e1004219.
- Sey NY, Hu B, Mah W, Fauni H, McAfee JC, Rajarajan P, et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci*. 2020;23:583–93.
- Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 2016;538:523–7.
- de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, et al. The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell*. 2018;172:289–304.
- Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48:245–52.
- Gamazon ER, Zwinderman AH, Cox NJ, Denys D, Derks EM. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat Genet*. 2019;51:933–40.
- Gerring ZF, Mina-Vargas A, Gamazon ER, Derks EM. E-MAGMA: an eQTL-informed method to identify risk genes using genome-wide association study summary statistics. *Bioinformatics*. 2021;btab115.
- Gerring ZF, Gamazon ER, Derks EM, MDD Working Group of the Psychiatric Genetics Consortium. A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression. *PLoS Genet*. 2019;15:e1008245.
- Xu X, Wells AB, O'Brien DR, Nehorai A, Dougherty JD. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci*. 2014;34:1420–31.
- Ghosh JC, Yang X, Zhang A, Lambert MH, Li H, Xu HE, et al. Interactions that determine the assembly of a retinoid X receptor/corepressor complex. *Proc Natl Acad Sci USA*. 2002;99:5842–7.
- Mangelsdorf DJ, Evans RM. The RXR heterodimers and orphan receptors. *Cell*. 1995;83:841–50.
- Liao W-L, Tsai HC, Wang HF, Chang J, Lu KM, Wu HL, et al. Modular patterning of structure and function of the striatum by retinoid receptor signaling. *Proc Natl Acad Sci USA*. 2008;105:6765–70.
- Bremner JD, Shearer KD, McCaffery PJ. Retinoic acid and affective disorders: the evidence for an association. *J Clin Psychiatry*. 2011;73:37–50.
- Grøntved L, Waterfall JJ, Kim DW, Baek S, Sung MH, Zhao L, et al. Transcriptional activation by the thyroid hormone receptor through ligand-dependent receptor recruitment and chromatin remodelling. *Nat Commun*. 2015;6:7048.
- Kim SH, An K, Namkung H, Rannals MD, Moore JR, Cash-Padgett T, et al. Anterior insula-associated social novelty recognition: orchestrated regulation by a local retinoic acid cascade and oxytocin signaling. *bioRxiv*: 2021.01.15.426848 [Preprint]. 2021. Available from: <https://doi.org/10.1101/2021.01.15.426848>.
- Hu P, Wang Y, Liu J, Meng FT, Qi XR, Chen L, et al. Chronic retinoic acid treatment suppresses adult hippocampal neurogenesis, in close correlation with depressive-like behavior. *Hippocampus*. 2016;26:911–23.
- Chen X-N, Meng QY, Bao AM, Swaab DF, Wang GH, Zhou JN. The involvement of retinoic acid receptor- α in corticotropin-releasing hormone gene expression and affective disorders. *Biol Psychiatr*. 2009;66:832–9.
- Hu P, Liu J, Zhao J, Qi XR, Qi CC, Lucassen PJ, et al. All-trans retinoic acid-induced hypothalamus-pituitary-adrenal hyperactivity involves glucocorticoid receptor dysregulation. *Transl Psychiatr*. 2013;3:e336.

42. Cross-Disorder Group of the Psychiatric Genetics Consortium. et al. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*. 2019;179:1469–82.
43. Reay WR, Atkins JR, Quidé Y, Carr VJ, Green MJ, Cairns MJ. Polygenic disruption of retinoid signalling in schizophrenia and a severe cognitive deficit subtype. *Mol Psychiatr*. 2018;25:719–31.
44. Regen F, Cosma NC, Otto LR, Clemens V, Saksone L, Gellrich J, et al. Clozapine modulates retinoid homeostasis in human brain and normalizes serum retinoic acid deficit in patients with schizophrenia. *Mol Psychiatr*. 2020;1–12.
45. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*. 2018;359:693–7.
46. Guo M, Zhu J, Yang T, Lai X, Liu X, Liu J, et al. Vitamin A improves the symptoms of autism spectrum disorders and decreases 5-hydroxytryptamine (5-HT): a pilot study. *Brain Res Bull*. 2018;137:35–40.
47. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol*. 2012;30:271–7.
48. Kwasnieski JC, Mogno I, Myers CA, Corbo JC, Cohen BA. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc Natl Acad Sci USA*. 2012;109:19498–503.
49. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol*. 2009;27:1173–5.
50. Mulvey B, Lagunas T, Dougherty JD. Massively parallel reporter assays: defining functional psychiatric genetic variants across biological contexts. *Biol Psychiatr*. 2020. <https://doi.org/10.1016/j.biopsych.2020.06.011>.
51. Choi J, Zhang T, Vu A, Ablain J, Makowski MM, Colli LM, et al. Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. *Nat Commun*. 2020;11:2718.
52. Bourges C, Groff AF, Burren OS, Gerhardinger C, Mattioli K, Hutchinson A, et al. Resolving mechanisms of immune-mediated disease in primary CD4 T cells. *Embo Mol Med*. 2020;12:e12112.
53. Ulirsch JC, Nandakumar SK, Wang L, Giani FC, Zhang X, Rogov P, et al. Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*. 2016;165:1530–45.
54. Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, et al. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*. 2016;165:1519–29.
55. Myint L, Wang R, Boukas L, Hansen KD, Goff LA, Avramopoulos D. A screen of 1,049 schizophrenia and 30 Alzheimer's-associated variants for regulatory potential. *Am J Med Genet Part B Neuropsychiatr Genet*. 2020;183:61–73.
56. Shen SQ, Myers CA, Hughes AE, Byrne LC, Flannery JG, Corbo JC. Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res*. 2016;26:238–55.
57. Shen SQ, Kim-Han JS, Cheng L, Xu D, Gokcumen O, Hughes AE, et al. A candidate causal variant underlying both higher intelligence and increased risk of bipolar disorder. *bioRxiv*: 10.1101/580258 [Preprint]. 2019. Available from: <https://doi.org/10.1101/580258>.
58. Vockley CM, Guo C, Majoros WH, Nodzinski M, Scholtens DM, Hayes MG, et al. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res*. 2015;25:1206–14.
59. Lu X, Chen X, Forney C, Donmez O, Miller D, Parameswaran S, et al. Global discovery of lupus genetic risk variant allelic enhancer activity. *Nat Commun*. 2021;12:1611.
60. Vockley CM, D'Ippolito AM, McDowell IC, Majoros WH, Safi A, Song L, et al. Direct GR binding sites potentiate clusters of TF binding across the human genome. *Cell*. 2016;166:1269–81. e19
61. Johnson GD, Barrera A, McDowell IC, D'Ippolito AM, Majoros WH, Vockley CM, et al. Human genome-wide measurement of drug-responsive regulatory activity. *Nat Commun*. 2018;9:5317.
62. Shea TB, Fischer I, Sapirstein VS. Effect of retinoic acid on growth and morphological differentiation of mouse NB2a neuroblastoma cells in culture. *Dev Brain Res*. 1985;21:307–14.
63. Evangelopoulos ME, Weis J, Krüttgen A. Signalling pathways leading to neuroblastoma differentiation after serum withdrawal: HDL blocks neuroblastoma differentiation by inhibition of EGFR. *Oncogene*. 2005;24:3309–18.
64. Ochoa D, et al. Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res*. 2020;49:gkaa1027.
65. White MA, Myers CA, Corbo JC, Cohen BA. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc Natl Acad Sci USA*. 2013;110:11952–7.
66. Donello JE, Loeb JE, Hope TJ. Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *J Virol*. 1998;72:5085–92.
67. Wu P-Y, Lin YC, Chang CL, Lu HT, Chin CH, Hsu TT, et al. Functional decreases in P2X7 receptors are associated with retinoic acid-induced neuronal differentiation of Neuro-2a neuroblastoma cells. *Cell Signal*. 2009;21:881–91.
68. Chowanadisai W, Graham DM, Keen CL, Rucker RB, Messerli MA. Neurulation and neurite extension require the zinc transporter ZIP12 (slc39a12). *Proc Natl Acad Sci USA*. 2013;110:9903–8.
69. Tian R, Gachechiladze MA, Ludwig CH, Laurie MT, Hong JY, Nathaniel D, et al. CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived. *Neurons Neuron*. 2019;104:239–55. e12
70. Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinform Oxf Engl*. 2015;31:3847–9.
71. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
72. Davis CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2017;46:gkx1081.
73. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. 2006;34:D590–D598.
74. Tuoresmäki P, Väisänen S, Neme A, Heikkinen S, Carlberg C. Patterns of genome-wide VDR locations. *PLoS ONE*. 2014;9:e96105.
75. Lalevéé S, Anno YN, Chatagnon A, Samarut E, Poch O, Laudet V, et al. Genome-wide in silico identification of new conserved and functional retinoic acid receptor response elements (direct repeats separated by 5 bp). *J Biol Chem*. 2011;286:33322–34.
76. Song M, Yang X, Ren X, Maliskova L, Li B, Jones IR, et al. Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet*. 2019;51:1252–62.
77. Hauberg ME, Zhang W, Giambartolomei C, Franzén O, Morris DL, Vyse TJ, et al. Large-scale identification of common trait and disease variants affecting gene expression. *Am J Hum Genet*. 2017;100:885–94.
78. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro F, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362:eaat8464.
79. Aguet F, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550:204–13.
80. Ng B, White CC, Klein HU, Siebert's SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci*. 2017;20:1418–26.
81. Xie Z, Bailey A, Kuleshov MV, Clarke D, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. *Curr Protoc*. 2021;1:e90.
82. Mi H, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res*. 2020;49:gkaa1106.
83. Grote S, Prüfer K, Kelso J, Dannemann M. ABAEnrichment: an R package to test for gene set expression enrichment in the adult and developing human brain. *Bioinformatics*. 2016;32:3201–3.
84. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50:1335–41.
85. Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat Genet*. 2020;52:1355–63.
86. Dobbyn A, Huckins LM, Boocock J, Sloofman LG, Glicksberg BS, Giambartolomei C, et al. Landscape of conditional eQTL in dorsolateral prefrontal cortex and co-localization with schizophrenia GWAS. *Am J Hum Genet*. 2018;102:1169–84.
87. Walker RL, Ramaswami G, Hartl C, Mancuso N, Gandal MJ, de la Torre-Ubieta L, et al. Genetic control of expression and splicing in developing human brain informs disease mechanisms. *Cell*. 2019;179:750–71.
88. Ciuculete DM, Voisin S, Kular L, Jonsson J, Rask-Andersen M, Mwinyi J, et al. meQTL and ncRNA functional analyses of 102 GWAS-SNPs associated with depression implicate HACE1 and SHANK2 genes. *Clin Epigenetics*. 2020;12:99.
89. Cimadamore F, Amador-Arjona A, Chen C, Huang C-T, Tersikh AV. SOX2-LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors. *Proc Natl Acad Sci USA*. 2013;110:E3017–E3026.
90. Ong KK, Elks CE, Li S, Zhao JH, Luan J, Andersen LB, et al. Genetic variation in LIN28B is associated with the timing of puberty. *Nat Genet*. 2009;41:729–33.
91. Perry JR, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*. 2014;514:92–97.
92. Corre C, Shinoda G, Zhu H, Cousminer DL, Crossman C, Bellissimo C, et al. Sex-specific regulation of weight and puberty by the Lin28/let-7 axis. *J Endocrinol*. 2016;228:179–91.

93. Mulvey B, Bhatti DL, Gyawali S, Lake AM, Kriaucionis S, Ford CP, et al. Molecular and functional sex differences of noradrenergic neurons in the mouse locus coeruleus. *Cell Rep.* 2018;23:2225–35.
94. Marcus SM, Young EA, Kerber KB, Kornstein S, Farabaugh AH, Mitchell J, et al. Gender differences in depression: findings from the STAR*D study. *J Affect Disord.* 2005;87:141–50.
95. Brody DJ, Pratt LA, Hughes JP. Prevalence of depression among adults aged 20 and over: United States, 2013–6. *Nchs Data Brief.* 2018;1–8.
96. Pan Q, et al. VARDdb: a comprehensive variation annotation database for human. *Nucleic Acids Res.* 2020;49:gkaa922.
97. Teixeira JR, Szeto RA, Carvalho VMA, Muotri AR, Papes F. Transcription factor 4 and its association with psychiatric disorders. *Transl Psychiat.* 2021;11:19.
98. Liu M, Xia Y, Ding J, Ye B, Zhao E, Choi JH, et al. Transcriptional profiling reveals a common metabolic program in high-risk human neuroblastoma and mouse neuroblastoma sphere-forming cells. *Cell Rep.* 2016;17:609–23.
99. Korade Ž, Kenworthy AK, Mirnics K. Molecular consequences of altered neuronal cholesterol biosynthesis. *J Neurosci Res.* 2009;87:866–75.
100. Rao C, Malaguti M, Mason JO, Lowell S. The transcription factor E2A drives neural differentiation in pluripotent cells. *Development.* 2020;147:dev.184093.
101. Imayoshi I, Kageyama R. bHLH factors in self-renewal, multipotency, and fate choice of neural progenitor cells. *Neuron.* 2014;82:9–23.
102. Ypsilanti AR, Pattabiraman K, Catta-Preta R, Golonzka O, Lindtner S, Tang K, et al. Transcriptional network orchestrating regional patterning of cortical progenitors. *bioRxiv:* 2020.11.03.366914 [Preprint]. 2020. Available from: <https://doi.org/10.1101/2020.11.03.366914>.
103. Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, et al. Global reference mapping of human transcription factor footprints. *Nature.* 2020;583:729–36.
104. Banerjee D, Gryder B, Bagchi S, Liu Z, Chen HC, Xu M, et al. Lineage specific transcription factor waves reprogram neuroblastoma from self-renewal to differentiation. *bioRxiv:* 2020.07.23.218503 [Preprint]. 2020. Available from: <https://doi.org/10.1101/2020.07.23.218503>.
105. Tremblay RG, Sikorska M, Sandhu JK, Lanthier P, Ribocco-Lutkiewicz M, Bani-Yaghoub M. Differentiation of mouse Neuro 2A cells into dopamine neurons. *J Neurosci Meth.* 2010;186:60–67.
106. Alonso A, Merchán P, Sandoval JE, Sánchez-Arrones L, García-Cazorla A, Artuch R, et al. Development of the serotonergic cells in murine raphe nuclei and their relations with rhombomeric domains. *Brain Struct Funct.* 2013;218:1229–77.
107. Nishida Y, Adati N, Ozawa R, Maeda A, Sakaki Y, Takeda T. Identification and classification of genes regulated by phosphatidylinositol 3-kinase- and TRKB-mediated signalling pathways during neuronal differentiation in two subtypes of the human neuroblastoma cell line SH-SY5Y. *Bmc Res Notes.* 2008;1:95.
108. Marteyn A, Maury Y, Gauthier MM, Lecuyer C, Vernet R, Denis JA, et al. Mutant human embryonic stem cells reveal neurite and synapse formation defects in type 1 myotonic dystrophy. *Cell Stem Cell.* 2011;8:434–44.
109. Johansson P, Pavey S, Hayward N. Confirmation of a BRAF mutation-associated gene expression signature in melanoma. *Pigm Cell Res.* 2007;20:216–21.
110. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, et al. Molecular architecture of the mouse nervous system. *Cell.* 2018;174:999–1014. e22
111. Ishikawa J, Sutoh C, Ishikawa A, Kagechika H, Hirano H, Nakamura S. 13-cis-retinoic acid alters the cellular morphology of slice-cultured serotonergic neurons in the rat. *Eur J Neurosci.* 2008;27:2363–72.
112. Schrode N, Ho SM, Yamamuro K, Dobbyn A, Huckins L, Matos MR, et al. Synergistic effects of common schizophrenia risk variants. *Nat Genet.* 2019;51:1475–85.
113. Mich JK, Graybuck LT, Hess EE, Mahoney JT, Kojima Y, Ding Y, et al. Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep.* 2021;34:108754.
114. Fee C, Banas M, Sibille E. Somatostatin-positive gamma-aminobutyric acid interneuron deficits in depression: cortical microcircuit and therapeutic perspectives. *Biol Psychiat.* 2017;82:549–59.
115. Smazal AL, Schalsinske KL. Oral administration of retinoic acid lowers brain serotonin concentration in rats. *Faseb J.* 2013;27:635.6.
116. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29:24–26.

ACKNOWLEDGEMENTS

We thank Stephen Plassmeyer and Tomás Lagunas, Jr. for technical assistance; Barak Cohen, PhD, Michael A. White, PhD, Dana King, PhD, Brett Maricque, PhD, and Ryan Friedman for library design, cloning, and analysis advice; and Genome Technology Access Center@McDonnell Genome Institute (GTAC@MGI) and Jessica Hoisington-Lopez and MariaLynn Crosby from the DNA Sequencing Innovation Lab at The Edison Family Center for Genome Sciences and Systems Biology for sequencing support. This work was supported by the NIH (1F30MH1116654 to BM, and 1R01MH116999 to JDD) and The Simons Foundation (571009 to JDD). GTAC@MGI is supported by UL1 TR002345. SNP annotation data resources are detailed in the supplemental text. We would also like to thank Idoia Lahortiga, Ph.D. and Luk Cox, Ph.D. curators of Somersault1824 (<https://gumroad.com/somersault1824>), for their open-access, Creative Commons BY-NC-SA 4.0 licensed libraries of high-quality biomedicine graphics (especially those from Graphite Life Explorer, ePMV, and Eyewire), adapted for Fig. 1E. Color palettes for plots are from the R package wesanderson (<https://github.com/karthik/wesanderson>).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41398-021-01493-6>.

Correspondence and requests for materials should be addressed to J.D.D.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021