

TITLE: Transcriptional-Regulatory Convergence Across Functional MDD Risk Variants Identified by Massively Parallel Reporter Assays

Bernard Mulvey¹, Joseph D. Dougherty¹

¹Depts. of Genetics and Psychiatry, Washington University in St. Louis

Corresponding Author:

Joseph D. Dougherty

Depts of Genetics and Psychiatry, Washington University in St. Louis

Address: 660 S. Euclid Ave, Campus Box 8232, St. Louis, MO 63110

Phone: (314) 286-0752

email: jdougherty@wustl.edu

Keywords (5-10): Depression, Massively Parallel Reporter Assay, Retinoid Signaling, Omnigenic, Functional Annotation

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 BJM, 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 Idoya 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 Figure 1E. Color palettes for plots are from the R package wesanderson (<https://github.com/karthik/wesanderson>).

ABSTRACT

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 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 1 000 SNPs prioritized from 39 neuropsychiatric trait/disease GWAS loci, with SNPs selected 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. Further, 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 subset of which are unmasked by retinoids.

INTRODUCTION

Major depressive disorder (MDD) is a common but debilitating psychiatric disorder affecting hundreds of millions worldwide¹, exacting substantial tolls on both individuals² and societies³. Despite the global burden of MDD, nearly half of patients do not clinically respond to treatment⁴, 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 MDD⁷, 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}. However, this leaves unresolved issues of identifying which linked SNPs have functional allelic impacts on TR, and how these act together across loci to result in disease.

In addition, it is generally thought that undetected, small-effect SNPs acting across the genome—including conditional SNPs within GWAS-significant loci¹⁷—contribute to the substantial heritability *not* caught by GWAS-significant SNPs alone¹⁸. A recent analysis of psychiatric disorder GWASes confirmed the existence of such conditional SNPs in MDD¹⁹. As such, multiple SNPs with allelic TR effects could exist per LD block, but functional demonstration has been sparse to date. The largest functional TR assay of MDD-associated variants examined 34 SNPs using luciferase assays²⁰, 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 biologically result in disease remains unclear, even once such SNPs are identified. Nonetheless, polygenic²¹ and omnigenic¹⁸ models provide a guiding principle for interpretation of complex disease genetics. In brief, these theories posit that consistent emergence of a specific phenotype via widespread genomic variation necessarily requires common biological endpoints of those variants' effects. At the molecular level, these points of convergence could be either upstream (shared regulation across loci)²² or downstream (common biological pathways across loci). For downstream analyses, myriad approaches have been developed to nominate gene targets of putative TR SNPs using proximity²³, chromatin structure²⁴⁻²⁶, or expression quantitative trait loci (eQTLs)²⁷⁻²⁹, yielding gene sets tested for enrichment in biological pathways^{27,28,30} and cell types³¹. 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 are steroid hormones, driving transcriptional responses via several retinoid-binding transcription factors (TFs) and heterodimerizing partners^{32,33}. Besides their critical role in neurodevelopment, including of depression-implicated limbic structures³⁴, retinoids have been associated with MDD onset and suicidality by epidemiological studies of the retinoid agonist isotretinoin³⁵. 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 neurons³⁶. Likewise, depressive symptoms have been observed in rats after intracerebroventricular all-*trans* retinoic acid (ATRA) administration³⁷. In addition, retinoid agonism of the TF RARA upregulates corticotrophin releasing hormone (*CRH*) expression, and RARA is more abundant in CRH neurons of affective disorder hypothalami³⁸; moreover, RARA prevents corticoid negative feedback on CRH expression³⁹, resulting in overactivity of stress signaling in the brain and body. Finally, given the substantial shared genetic risk across psychiatric disorders⁴⁰, it is notable that schizophrenia GWAS loci show enrichment for retinoid TR⁴¹, and that circulating retinoids are dysregulated in schizophrenia patients⁴². Similarly, retinoid pathway genes including *CYP26B1* are dysregulated in postmortem brain from autism spectrum disorders, bipolar disorder, and schizophrenia patients⁴³. These findings led us to speculate that a component of MDD risk genetics may, in fact, demonstrate an upstream convergence via recurrent retinoid-mediated TR disruptions by MDD-associated SNPs.

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 between⁴⁴⁻⁴⁷. 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 collected 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^{48–56}. Two key features make MPRA advantageous for identifying both functional SNPs and their TR interactions.⁵⁷ 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, MPRA can be conducted to define TR effects of experimental manipulations in these systems, such as steroid hormone administration⁵⁷.

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 1 000 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^{58,59}. 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. -nonfunctional 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 differentiation⁵⁸. First, we found that functional SNPs containing retinoid receptor motifs had increased magnitudes of effect in the presence of ATRA, consistent with bonafide 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. Additionally, it revealed new SNPs with allelic TR differences unmasked by ATRA treatment. Significant ATRA-allele interaction SNPs were enriched not only for binding sequences of the retinoid receptor RARG, but for 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 putatively regulatory variants from neuropsychiatric disease GWAS regions (predominantly MDD; **Figure 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*, **Figure 1B**). SNPs were manually selected based on diversity and density of annotation overlap within each locus (detailed in *Supplemental Methods*.) As a negative control, we identified candidates from one additional locus associated with several anthropomorphic traits⁶⁰, in a blinded manner. Altogether, 1453 SNPs were selected. Final LD of selected SNPs was distributed similarly to starting SNPs (**Fig 1D**).⁵⁹ Further included was a positive control TR SNP functionally demonstrated in mouse retina and brain⁵⁵.

Human genomic sequence (hg19) tiles up to 126bp were taken centered on the 1454 candidate enhancer SNPs, each paired to 10 unique 10 base pair barcode sequences per allele and ordered as an oligonucleotide pool. Further included in the pool were 110 “basal” barcodes, where no human genomic sequence was placed upstream of the minimal promoter (i.e., such that the only variable sequence was the barcode itself). PCR-amplified oligos were then cloned into plasmid (**Fig 1E**), with subsequent insertion of a reporter cassette containing a constant minimal promoter (*hsp68*) driving the dsRed reporter gene⁶¹, and the untranslated “woodchuck” element (for RNA stabilization, to improve signal)⁶².

Massively Parallel Reporter Assays

N2A cells were grown in uncoated 6-well plates in medium consisting of 0.1μM vacuum-filtered DMEM with 10% Fetal Bovine Serum (2% fetal bovine serum for the ATRA assay, based on media conditions from the literature^{63,64}). For transfection, cells were reverse transfected by plating in antibiotic-free medium onto pre-plated 400μL mixtures of 2.5μg plasmid with Lipofectamine 2000. After 7 hours of incubation at 37°C and 5% CO₂, medium was replaced with the respective medium containing antibiotics, and in the second assay, a final concentration of 20 μM 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 hours. 72 hours after transfection, cells were collected and RNA extracted using the Zymo 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 (1μg) 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. 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, P values were adjusted using empirical *q*-value correction via simulated allelic comparisons between random subsets of “basal” barcodes (see *Supplemental Methods*) following an analogous procedure from a multiplex CRISPR study⁶⁵, with significance defined as $q < 0.05$ 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, barcode level expression values for both alleles of a SNP from all samples were used in a linear mixed model (LMM) with a random variable for replicate (to account for well-specific effects) as: barcode expression \sim allele + drug + allele:drug + (1|replicate). Empirical *q*-value correction for 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. 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 $q < 0.1$ 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 motifbreakR⁶⁶ 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 motif for at least one allele was compared to the number of non-functional SNPs matching that motif 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. In the first assay, SNPs were defined as functional based on a *q* value threshold of 0.05. In the second assay, we performed two comparisons. One analysis compared allele main effect SNPs at $q < 0.1$ vs. those subject to neither main effects of allele or drug, nor their interaction (all $q > 0.1$), representing the breadth of functional variant-susceptible *cis*-regulators. The second analysis compared interaction SNPs at $q < 0.05$ vs. SNPs with an allele main effect ($q_{\text{allele}} < 0.1$) but no interaction ($q_{\text{interaction}} > 0.1$).

RESULTS

Many MDD loci contain more than one functional SNP

We identified >1 000 SNPs from MDD-associated GWAS loci, prioritizing SNPs overlapping with epigenetic data from neural samples, and cloned them into an MPRA library (**Figure 1**). We included one positive control

SNP, shown to alter neural tissue gene expression, and one control locus near *CDKALI* not *a priori* associated with psychiatric disease. To identify functional variants from these SNPs, the library was transfected into N2a (also known as Neuro2a) 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 1**). 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 ($q < 0.05$) at 76 SNPs (65 from MDD loci; 1 from the control *CDKALI* 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 $p < 5 \cdot 10^{-6}$) with “*Poisoning by analgesics, antipyretics, and antirheumatics*” in UK Biobank⁶⁷. As this likely includes attempted suicides, the SNP was retained for analyses. The positive control SNP showed the expected higher expression of the T allele at a q-value of < 0.051 (**Fig 2A**)⁵⁵.

Consistent with prior studies of “conditional” or “secondary” SNP associations—wherein additional LD SNPs have associations independent of their linked, larger-effect variant^{19,68,69}—we identified several loci with multiple functional SNPs (**Fig 2C**) (range 1-8, mean 2.8, median 2). Notably, we identified as functional rs1806153, which was recently defined as a “conditional SNP” for MDD¹⁸. 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 risk⁴⁰. 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 PsychENCODE⁷⁰, and is a CommonMind Consortium (CMC) eSNP for both *LIN28B* and *HACE1*⁶³. *HACE1* is also downregulated in postmortem MDD hippocampal CA1⁷¹. 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 astrocytes⁷². *LIN28B* plays broad roles in neurodevelopment⁷³ and has potentially sex-differentiated functions⁷⁴⁻⁷⁷—which, considering sex differences in MDD prevalence and severity^{78,79}—makes *LIN28B* an especially interesting gene target from this locus. Finally, we examined potential upstream TR mechanisms for SNP activity using VARAdb⁸⁰. 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^{40,81}. 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 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 first 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 2**). 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 3**), 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^{82,83}. A second group of transcription factors included three TFs involved in neural lineage commitment/development: TCF3^{84,85}, EOMES, and NR2F1⁸⁶ (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 footprinting⁸⁷.

Retinoids unmask additional functional SNPs in MDD loci.

Our findings supported the hypothesis that MDD-associated variants across multiple loci converge on TR that may be 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 repression³². Therefore, we performed a second MPRA adding an all-*trans* retinoic acid (ATRA) condition.

After 48 hours, cultures were imaged to verify drug delivery and activity (as ATRA is light-sensitive) based on known morphologic responses of N2as to ATRA, which include neurite outgrowth and mitotic arrest^{58,88,89}. 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 $q < 0.01$), 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; $q_{\text{interaction}} < 0.025$, **Fig 4G**), while another trended towards interaction (rs489591; $q_{\text{interaction}} = 0.117$). 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 4H**). 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 4I**). 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 sought to again assess convergent transcriptional effects of these regulatory variants by employing the same motifbreakR-based analyses. 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 S3**), totaling 11 of the 92 allele-main effect SNPs analyzed and 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 understand the TFs potentially underlying allelic effects with 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 S3**). We compared the overrepresented motifs to TFs recently demonstrated to be upregulated in neuroblastoma lines (KCNr, LAN5) by ATRA. Of the 26 TFs identified as ATRA-induced in these two lines, motifs were available for 18 in our analysis. Of these, motifs for 6 of the TFs were enriched among allele main-effect-only SNPs, while motifs for 12 of these TFs were enriched among the retinoid-allele

interaction variants⁸⁸ (**Fig. 5A, Supplemental Figure S3**), supporting our predictions of TFs playing ATRA-dependent roles at functional SNPs.

Comparison between ATRA-interacting and -independent functional SNPs results should properly control for motif matches that do not correspond to retinoid-bound TFs (as ATRA agonism of retinoid receptors should modify their activity). Indeed, in contrast to our preceding motif analyses of allelic effect variants, only two retinoid receptors were enriched among the retinoid-interacting functional variants: RARG (eight SNPs, five of which also matched RARB motifs) (**Fig 5A, Supplemental Figure S3**). This suggests that RARG and/or RARB is the particular retinoid receptor at the nexus of allele-ATRA interaction effects in this model, and may likewise be TFs responsible for ATRA-dependent allelic effects at MDD risk loci overall. Altogether, these findings support cross-locus involvement of retinoid receptor binding sites—including those of RARG—and of retinoid-regulated TFs in MDD-associated regulatory variants.

Finally, we 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. Using a collection of previously published transcriptome-wide brain cell type-specificity metrics³¹, we screened these TFs for enriched expression in specific cell types. 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. We believe this is not an artifact of the N2a system because we could identify no evidence in the literature suggesting a serotonin-like identity of N2a cells with or without ATRA treatment. However, exogenous retinoids have been shown to lower circulating serotonin in humans⁹⁰ and to alter morphology of rat raphe neurons in slice culture⁹¹, suggesting they are retinoid responsive. Single-cell RNA sequencing data from mouse brain reveals that these TFs are only expressed in serotonergic, noradrenergic, peripheral autonomic, and midbrain inhibitory neurons—with all three expressed in serotonin neurons⁹²—suggesting that these cell types in particular may be a cellular point 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 assays²⁰ or using CRISPR-Cas9 technology to edit limited positions for deep phenotyping⁹³. Here, we leveraged MPRA to screen over 1 000 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 assumptions of polygenic/omnigenic theories, which predict multiple SNPs with allelic effects within GWAS loci. Further, we examined convergence across loci, predicted by omnigenic theory. By examining the shared regulatory features (transcription factor binding site 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 receptor TFs 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 drug 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. Motif-SNP perturbation enrichment identified two retinoid receptors, RARG and RARB, as the likely TFs involved in functional SNP activity at retinoid receptor binding sites in this system. This could be because either a) in N2as, RARG/RARB are the predominant retinoid receptor, or b) that these cells express a broader array of retinoid-activated TFs, but only the RARG and RARB motifs are a high-fidelity representation of “true” retinoid receptor family-wide binding sites. Nonetheless, these findings suggest that RARB, and RARG 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 MPRAs, bring inherent limitations to their results. The primary precaution to be taken is with regard to cell type relevance. MPRAs are subject to the TR landscape of the cell type used. The cell type here was selected for its intact retinoid signaling, but is not likely central to MDD, as neuroblastomas are derived from peripheral neural crest progenitors (though they can be differentiated into dopaminergic neurons⁸⁹, and commit to neuronal differentiation with ATRA^{58,64}). On the other hand, the neural crest-derived autonomic nervous system has received little consideration (i.e., relative to brain) in psychiatric genetics of MDD despite the well-appreciated role of stress in depression. Some of the functional SNPs identified may thus play a role in the autonomic system as opposed—or in addition—to the brain, though this assay is not well designed to empirically address this, as SNPs were prioritized for assay based on brain epigenomics.

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 351 SNPs with allelic and/or interaction effects of 1178 assessed across the two assays) than there are GWAS loci per se. 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 eight of 113 SNPs spanning functional variants in binding motifs of TFs GATA2, GATA3, and FEV. It has previously been demonstrated that systemic administration of all-*trans* retinoic acid depletes serotonin by over 40% in the rat brain⁹⁴, supporting the serotonin system as a convergent target of retinoid-regulated pathways. As GWAS of MDD begins to explore severe, treatment-refractory cases⁷, 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.

Data and 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 https://bitbucket.org/jdlabteam/n2a_atra_mdd_mpra_paper/src. Raw sequencing files are deposited in GEO under accession GSE167519.

FIGURES

Figure 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 from⁴⁷.

Figure 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 assay⁵⁵. B) Volcano plot of allelic differences in reporter expression. Points represent one SNP’s composite log₂ 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 ($q < 0.05$) 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 log₂ allelic fold change. Refseq genes are visualized by the Integrative Genomics Viewer⁹⁵. 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.

Figure 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.

Figure 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 empirical q-value correction 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 $q < 0.05$, LMM allele $q < 0.05$, and both single-condition analyses showing the same allelic effect directionality at ATRA FDR < 0.1 AND vehicle

q<0.1. F) rs4801117-A shows greater activity with ATRA treatment while the G 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 log2FC values are shown. I) Several additional SNPs with retinoid-dependent function (i.e., allele-ATRA interaction) in the *LIN28B* locus. Only significant SNPs 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.

Figure 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 lines⁸⁸ 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/>)³¹.

REFERENCES

1. Smith, K. Mental health: A world of depression. *Nature* **515**, 180–181 (2014).
2. Üstün, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C. & Murray, C. J. L. Global burden of depressive disorders in the year 2000. *Brit J Psychiat* **184**, 386–392 (2004).
3. Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T. & Kessler, R. C. The Economic Burden of Adults With Major Depressive Disorder in the United States (2005 and 2010). *J Clin Psychiatry* **76**, 155–162 (2015).
4. Papakostas, G. I. & Fava, M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *European Neuropsychopharmacol J European Coll Neuropsychopharmacol* **19**, 34–40 (2009).
5. Flint, J. & Kendler, K. S. The genetics of major depression. *Neuron* **81**, 484–503 (2014).
6. Corfield, E. C., Yang, Y., Martin, N. G. & Nyholt, D. R. A continuum of genetic liability for minor and major depression. *Transl Psychiat* **7**, e1131–e1131 (2017).
7. Clements, C. C. *et al.* Genome-wide association study of patients with a severe major depressive episode treated with electroconvulsive therapy. *Mol Psychiatr* 1–11 (2021) doi:10.1038/s41380-020-00984-0.
8. Cai, N. *et al.* Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
9. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668–681 (2018).
10. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343–352 (2019).
11. Levey, D. F. *et al.* GWAS of Depression Phenotypes in the Million Veteran Program and Meta-analysis in More than 1.2 Million Participants Yields 178 Independent Risk Loci. *Medrxiv* 2020.05.18.20100685 (2020) doi:10.1101/2020.05.18.20100685.
12. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* **48**, 1031–1036 (2016).
13. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Sci New York N Y* **337**, 1190–5 (2012).
14. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–30 (2015).
15. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res* **22**, 1748–1759 (2012).
16. Chen, J. *et al.* A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat Commun* **7**, 11101 (2016).

17. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genetics* **99**, 139–153 (2016).
18. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
19. Consortium, B. W. G. of the P. G. *et al.* Conditional GWAS analysis to identify disorder-specific SNPs for psychiatric disorders. *Mol Psychiatr* 1–12 (2020) doi:10.1038/s41380-020-0705-9.
20. Li, S. *et al.* Regulatory mechanisms of major depressive disorder risk variants. *Mol Psychiatr* 1–20 (2020) doi:10.1038/s41380-020-0715-7.
21. Furlong, L. I. Human diseases through the lens of network biology. *Trends Genet* **29**, 150–159 (2013).
22. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022–1034.e6 (2019).
23. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *Plos Comput Biol* **11**, e1004219 (2015).
24. Sey, N. Y. A. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci* 1–11 (2020) doi:10.1038/s41593-020-0603-0.
25. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
26. Torre-Ubieta, L. de la *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289–304.e18 (2018).
27. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245–52 (2016).
28. Gamazon, E. R., Zwinderman, A. H., Cox, N. J., Denys, D. & Derks, E. M. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat Genet* **51**, 933–940 (2019).
29. Gerring, Z. F., Mina-Vargas, A., Gamazon, E. R. & Derks, E. M. E-MAGMA: an eQTL-informed method to identify risk genes using genome-wide association study summary statistics. *Bioinformatics* btab115- (2021) doi:10.1093/bioinformatics/btab115.
30. Gerring, Z. F., Gamazon, E. R., Derks, E. M. & Consortium, M. D. D. W. G. of the P. G. A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression. *Plos Genet* **15**, e1008245 (2019).
31. Xu, X., Wells, A. B., O'Brien, D. R., Nehorai, A. & Dougherty, J. D. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci Official J Soc Neurosci* **34**, 1420–31 (2014).
32. Ghosh, J. C. *et al.* Interactions that determine the assembly of a retinoid X receptor/corepressor complex. *Proc National Acad Sci* **99**, 5842–5847 (2002).

33. Mangelsdorf, D. J. & Evans, R. M. The RXR heterodimers and orphan receptors. *Cell* **83**, 841–850 (1995).
34. Liao, W.-L. *et al.* Modular patterning of structure and function of the striatum by retinoid receptor signaling. *Proc National Acad Sci* **105**, 6765–6770 (2008).
35. Bremner, J. D., Shearer, K. D. & McCaffery, P. J. Retinoic acid and affective disorders: the evidence for an association. *J Clin Psychiatry* **73**, 37–50 (2011).
36. Kim, S.-H. *et al.* Anterior insula-associated social novelty recognition: orchestrated regulation by a local retinoic acid cascade and oxytocin signaling. *Biorxiv* 2021.01.15.426848 (2021) doi:10.1101/2021.01.15.426848.
37. Hu, P. *et al.* Chronic retinoic acid treatment suppresses adult hippocampal neurogenesis, in close correlation with depressive-like behavior. *Hippocampus* **26**, 911–923 (2016).
38. Chen, X.-N. *et al.* The Involvement of Retinoic Acid Receptor- α in Corticotropin-Releasing Hormone Gene Expression and Affective Disorders. *Biol Psychiat* **66**, 832–839 (2009).
39. Hu, P. *et al.* All-trans retinoic acid-induced hypothalamus–pituitary–adrenal hyperactivity involves glucocorticoid receptor dysregulation. *Transl Psychiat* **3**, e336–e336 (2013).
40. Consortium, C.-D. G. of the P. G. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469–1482.e11 (2019).
41. Reay, W. R. *et al.* Polygenic disruption of retinoid signalling in schizophrenia and a severe cognitive deficit subtype. *Mol Psychiatr* **25**, 719–731 (2018).
42. Regen, F. *et al.* Clozapine modulates retinoid homeostasis in human brain and normalizes serum retinoic acid deficit in patients with schizophrenia. *Mol Psychiatr* 1–12 (2020) doi:10.1038/s41380-020-0791-8.
43. Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Sci New York N Y* **359**, 693–697 (2018).
44. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* **30**, 271–277 (2012).
45. Kwasnieski, J. C., Mogno, I., Myers, C. A., Corbo, J. C. & Cohen, B. A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *P Natl Acad Sci Usa* **109**, 19498–503 (2012).
46. Patwardhan, R. P. *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* **27**, 1173–1175 (2009).
47. Mulvey, B., Lagunas, T. & Dougherty, J. D. Massively Parallel Reporter Assays: Defining Functional Psychiatric Genetic Variants across Biological Contexts. *Biol Psychiat* (2020) doi:10.1016/j.biopsych.2020.06.011.
48. Choi, J. *et al.* Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. *Nat Commun* **11**, 2718 (2020).
49. Lu, X. *et al.* Genome-wide discovery of SLE genetic risk variant allelic enhancer activity. *Biorxiv* 2020.01.20.906701 (2020) doi:10.1101/2020.01.20.906701.

50. Bourges, C. *et al.* Resolving mechanisms of immune-mediated disease in primary CD4 T cells. *Embo Mol Med* **12**, e12112 (2020).
51. Ulirsch, J. C. *et al.* Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell* **165**, 1530–1545 (2016).
52. Tewhey, R. *et al.* Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell* **165**, 1519–1529 (2016).
53. Myint, L. *et al.* A screen of 1,049 schizophrenia and 30 Alzheimer's-associated variants for regulatory potential. *Am J Medical Genetics Part B Neuropsychiatric Genetics* **183**, 61–73 (2020).
54. Shen, S. Q. *et al.* Massively parallel cis -regulatory analysis in the mammalian central nervous system. *Genome Res* **26**, 238–255 (2016).
55. Shen, S. Q. *et al.* A candidate causal variant underlying both higher intelligence and increased risk of bipolar disorder. *Biorxiv* 580258 (2019) doi:10.1101/580258.
56. Vockley, C. M. *et al.* Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res* **25**, 1206–1214 (2015).
57. Vockley, C. M. *et al.* Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. *Cell* **166**, 1269-1281.e19 (2016).
58. Shea, T. B., Fischer, I. & Sapirstein, V. S. Effect of retinoic acid on growth and morphological differentiation of mouse NB2a neuroblastoma cells in culture. *Dev Brain Res* **21**, 307–314 (1985).
59. Evangelopoulos, M. E., Weis, J. & Krüttgen, A. Signalling pathways leading to neuroblastoma differentiation after serum withdrawal: HDL blocks neuroblastoma differentiation by inhibition of EGFR. *Oncogene* **24**, 3309–3318 (2005).
60. Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res* **47**, gky1133 (2018).
61. White, M. A., Myers, C. A., Corbo, J. C. & Cohen, B. A. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc National Acad Sci* **110**, 11952–11957 (2013).
62. Donello, J. E., Loeb, J. E. & Hope, T. J. Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *J Virol* **72**, 5085–92 (1998).
63. Wu, P.-Y. *et al.* Functional decreases in P2X7 receptors are associated with retinoic acid-induced neuronal differentiation of Neuro-2a neuroblastoma cells. *Cell Signal* **21**, 881–891 (2009).
64. Chohanadisai, W., Graham, D. M., Keen, C. L., Rucker, R. B. & Messerli, M. A. Neurulation and neurite extension require the zinc transporter ZIP12 (slc39a12). *Proc National Acad Sci* **110**, 9903–9908 (2013).
65. Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104**, 239-255.e12 (2019).

66. Coetzee, S. G., Coetzee, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinform Oxf Engl* **31**, 3847–9 (2015).
67. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335–1341 (2018).
68. Dobbryn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am J Hum Genet* **102**, 1169–1184 (2018).
69. Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771.e22 (2019).
70. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
71. Ciuculete, D. M. *et al.* meQTL and ncRNA functional analyses of 102 GWAS-SNPs associated with depression implicate HACE1 and SHANK2 genes. *Clin Epigenetics* **12**, 99 (2020).
72. Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* **51**, 1252–1262 (2019).
73. Cimadamore, F., Amador-Arjona, A., Chen, C., Huang, C.-T. & Terskikh, A. V. SOX2–LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors. *Proc National Acad Sci* **110**, E3017–E3026 (2013).
74. Ong, K. K. *et al.* Genetic variation in LIN28B is associated with the timing of puberty. *Nat Genet* **41**, 729–733 (2009).
75. Perry, J. R. B. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
76. Corre, C. *et al.* Sex-specific regulation of weight and puberty by the Lin28/let-7 axis. *J Endocrinol* **228**, 179–191 (2016).
77. Mulvey, B. *et al.* Molecular and Functional Sex Differences of Noradrenergic Neurons in the Mouse Locus Coeruleus. *Cell Reports* **23**, 2225–2235 (2018).
78. Marcus, S. M. *et al.* Gender differences in depression: Findings from the STAR*D study. *J Affect Disorders* **87**, 141–150 (2005).
79. Brody, D. J., Pratt, L. A. & Hughes, J. P. Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013-2016. *Nchs Data Brief* 1–8 (2018).
80. Pan, Q. *et al.* VARAdb: a comprehensive variation annotation database for human. *Nucleic Acids Res* **49**, gkaa922- (2020).
81. Teixeira, J. R., Szeto, R. A., Carvalho, V. M. A., Muotri, A. R. & Papes, F. Transcription factor 4 and its association with psychiatric disorders. *Transl Psychiat* **11**, 19 (2021).
82. Liu, M. *et al.* Transcriptional Profiling Reveals a Common Metabolic Program in High-Risk Human Neuroblastoma and Mouse Neuroblastoma Sphere-Forming Cells. *Cell Reports* **17**, 609–623 (2016).

83. Korade, Ž., Kenworthy, A. K. & Mirnics, K. Molecular consequences of altered neuronal cholesterol biosynthesis. *J Neurosci Res* **87**, 866–875 (2009).
84. Rao, C., Malaguti, M., Mason, J. O. & Lowell, S. The transcription factor E2A drives neural differentiation in pluripotent cells. *Development* **147**, dev.184093 (2020).
85. Imayoshi, I. & Kageyama, R. bHLH Factors in Self-Renewal, Multipotency, and Fate Choice of Neural Progenitor Cells. *Neuron* **82**, 9–23 (2014).
86. Ypsilanti, A. R. *et al.* Transcriptional Network Orchestrating Regional Patterning of Cortical Progenitors. *Biorxiv* 2020.11.03.366914 (2020) doi:10.1101/2020.11.03.366914.
87. Vierstra, J. *et al.* Global reference mapping of human transcription factor footprints. *Nature* **583**, 729–736 (2020).
88. Banerjee, D. *et al.* Lineage specific transcription factor waves reprogram neuroblastoma from self-renewal to differentiation. *Biorxiv* 2020.07.23.218503 (2020) doi:10.1101/2020.07.23.218503.
89. Tremblay, R. G. *et al.* Differentiation of mouse Neuro 2A cells into dopamine neurons. *J Neurosci Meth* **186**, 60–67 (2010).
90. Guo, M. *et al.* Vitamin A improves the symptoms of autism spectrum disorders and decreases 5-hydroxytryptamine (5-HT): A pilot study. *Brain Res Bull* **137**, 35–40 (2018).
91. Ishikawa, J. *et al.* 13-cis-retinoic acid alters the cellular morphology of slice-cultured serotonergic neurons in the rat. *Eur J Neurosci* **27**, 2363–2372 (2008).
92. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014.e22 (2018).
93. Schrode, N. *et al.* Synergistic effects of common schizophrenia risk variants. *Nat Genet* **51**, 1475–1485 (2019).
94. Smazal, A. L. & Schalinske, K. L. Oral Administration of Retinoic Acid Lowers Brain Serotonin Concentration in Rats. *Faseb J* **27**, 635.6-635.6 (2013).
95. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).

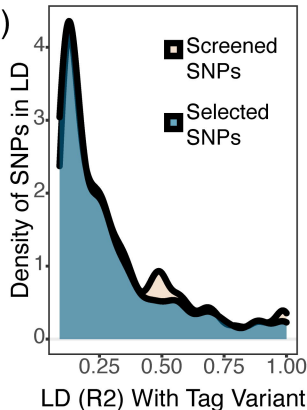
A)

Source GWAS	Trait/Disease	# Tag SNPs	# LD SNPs
Wray 2018	MDD	15	458
PGC Cross-Disorder 2019	MDD	5	341
Li 2018 (<i>NPP</i>)	MDD	2	128
Hyde 2016	MDD	1	116
Power 2017	MDD	2	100
Howard, Adams 2018	MDD	1	88
CONVERGE	MDD	2	77
Ren 2018	MDD	1	2
Demontis 2019	ADHD	2	15
Smith 2016	Neuroticism	1	10
Luciano 2018	Neuroticism	1	6
Meier, Tronetti 2018	Anxiety Disorders	1	5
Ward 2019	Mood Instability	1	5
DeBette14, Hill18, Lee18	Intelligence, Educational Attainment	2	29
Grove 2019	ASD	2	38
Carvalho-Silva 2018	Anthropomorphometric	1	35

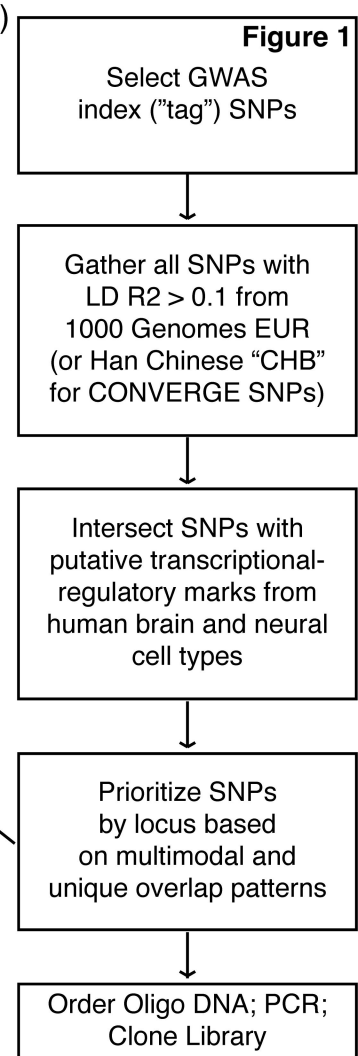
C)

Predictive Transcriptional-Regulatory Annotation Type	# of MPRA SNPs Overlapping	% MPRA SNPs Overlapping
GTEX eQTL	526	36.20
CMC eQTL	957	65.86
PEC eQTL	709	48.80
ROSMAP eQTL	269	18.51
Fetal Cortex or Organoid Chromatin Marks	479	32.97
Dentate ATAC and/or Hi-C	583	40.12
Astrocyte ATAC and/or Hi-C	362	24.91
Cortex ATAC and/or Hi-C	409	28.15
Enhancer Predictions from PsychENCODE ENCODE Screen, and/or FANTOM5	910	62.63

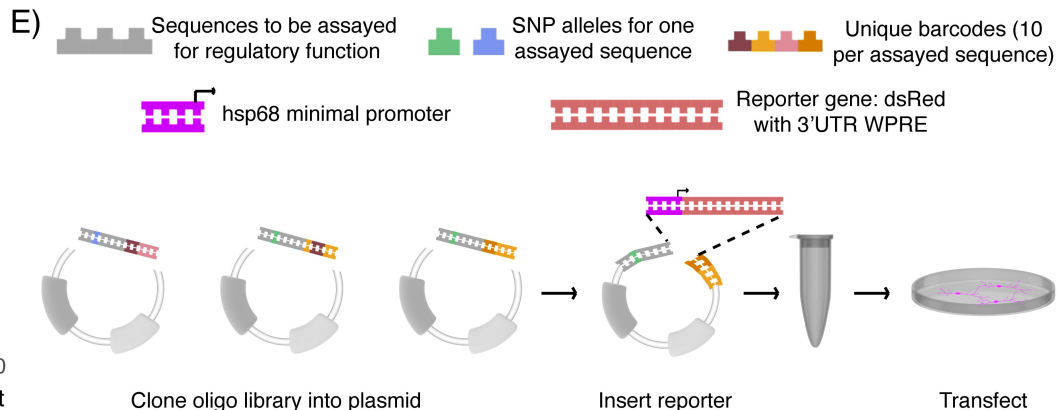
D)

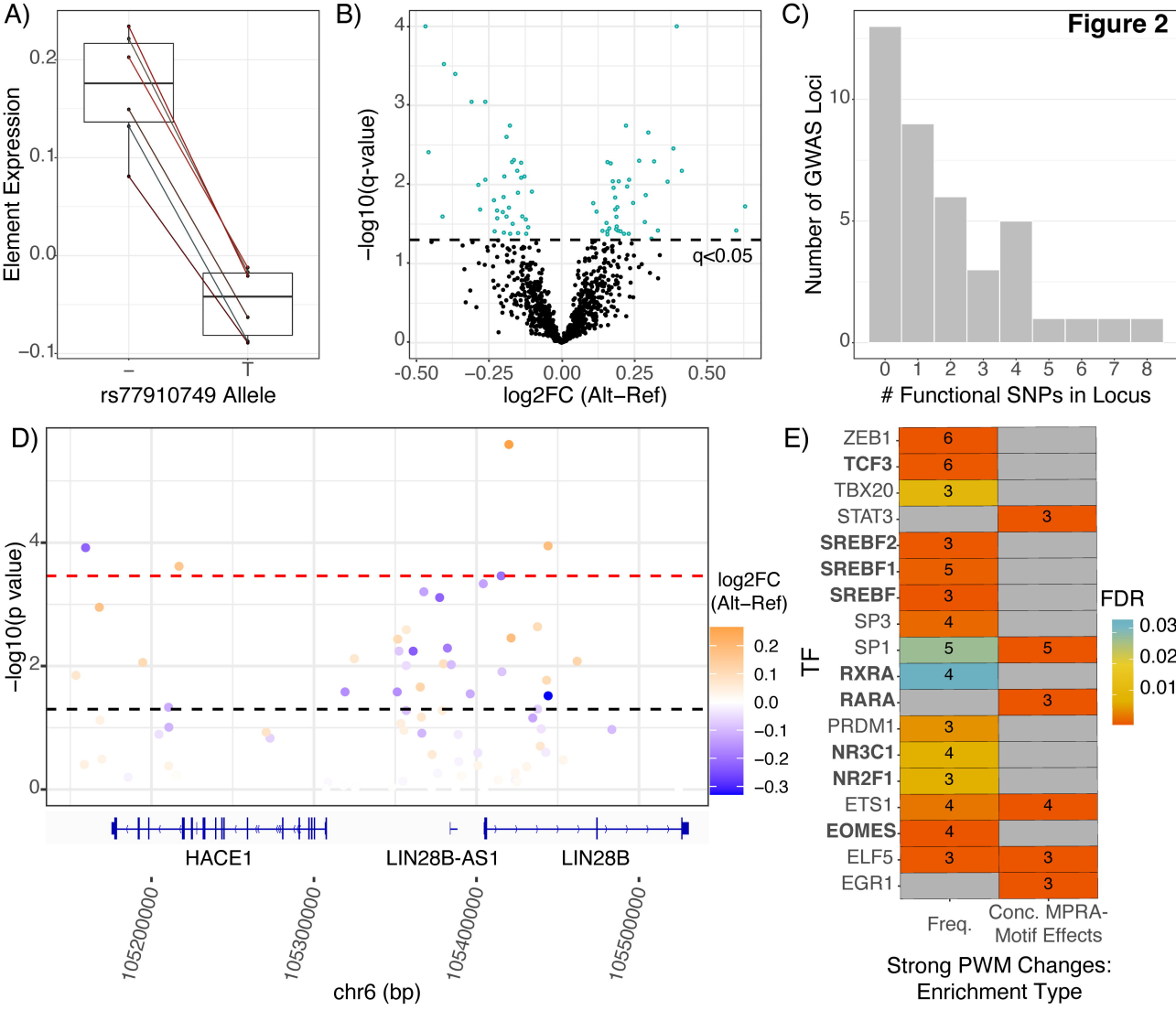


B)

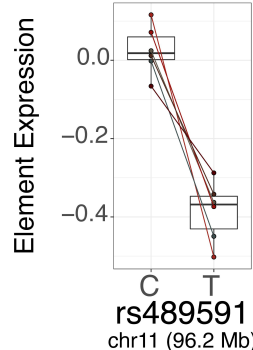


E)

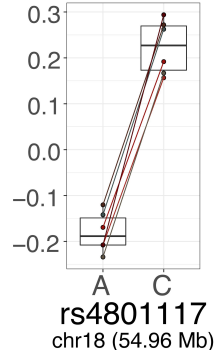




A)



B)



C)

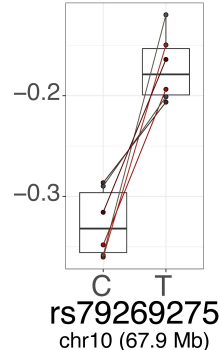
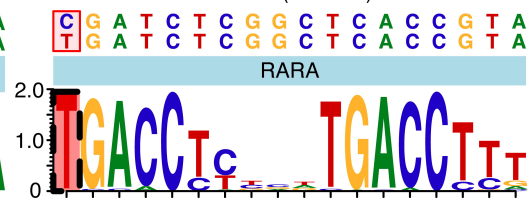
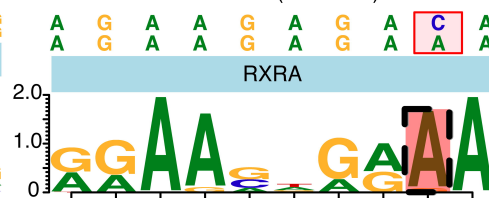
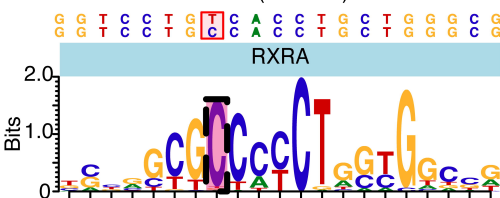
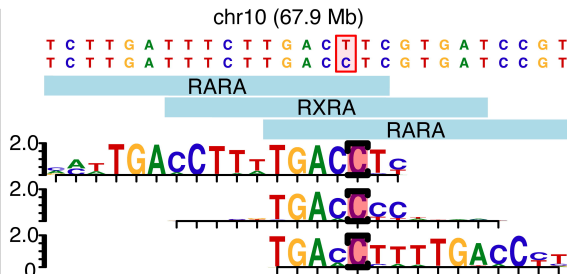
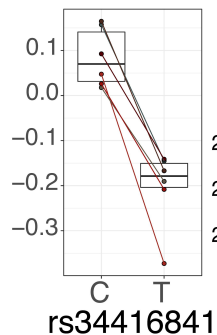


Figure 3



D)



E)

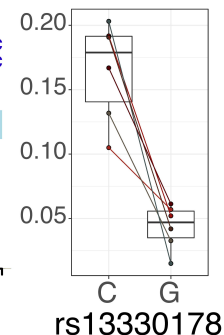
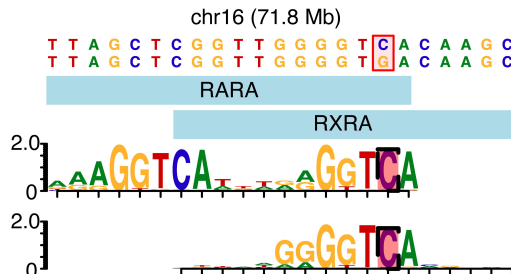
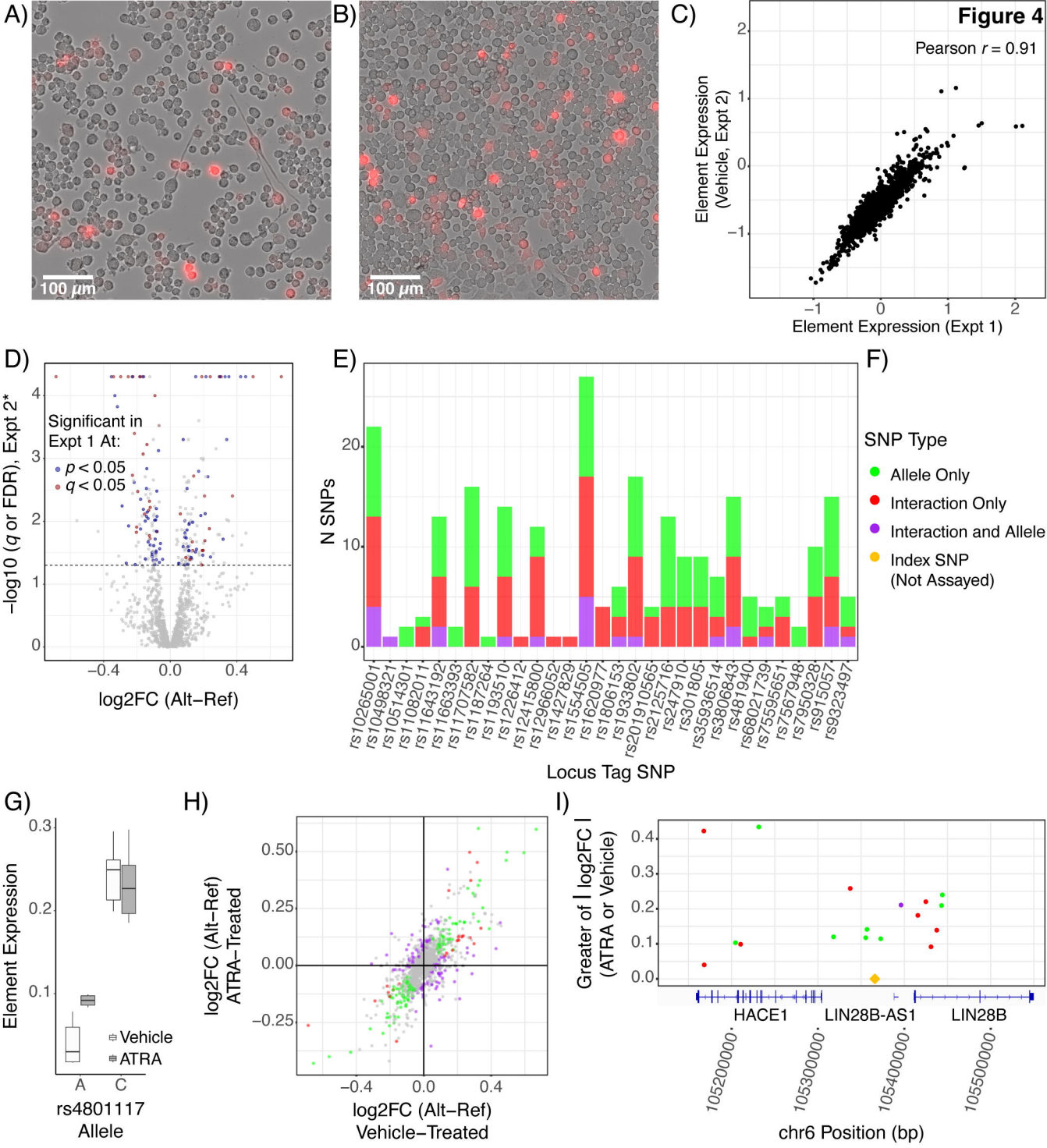


Figure 4

A)

