


REVIEW

Open Access

# Focus on your locus with a massively parallel reporter assay



Jessica C. McAfee<sup>1,2</sup>, Jessica L. Bell<sup>1,2</sup>, Oleh Krupa<sup>1,2</sup>, Nana Matoba<sup>1,2</sup>, Jason L. Stein<sup>1,2\*</sup> and Hyejung Won<sup>1,2\*</sup> 

## Abstract

A growing number of variants associated with risk for neurodevelopmental disorders have been identified by genome-wide association and whole genome sequencing studies. As common risk variants often fall within large haplotype blocks covering long stretches of the noncoding genome, the causal variants within an associated locus are often unknown. Similarly, the effect of rare noncoding risk variants identified by whole genome sequencing on molecular traits is seldom known without functional assays. A massively parallel reporter assay (MPRA) is an assay that can functionally validate thousands of regulatory elements simultaneously using high-throughput sequencing and barcode technology. MPRA has been adapted to various experimental designs that measure gene regulatory effects of genetic variants within cis- and trans-regulatory elements as well as posttranscriptional processes. This review discusses different MPRA designs that have been or could be used in the future to experimentally validate genetic variants associated with neurodevelopmental disorders. Though MPRA has limitations such as it does not model genomic context, this assay can help narrow down the underlying genetic causes of neurodevelopmental disorders by screening thousands of sequences in one experiment. We conclude by describing future directions of this technique such as applications of MPRA for gene-by-environment interactions and pharmacogenetics.

**Keywords:** MPRA, GWAS, Neurodevelopmental disorders, Noncoding genome, Gene regulation, Functional validation, Posttranscriptional regulation, Gene-environment interactions, WGS

## Introduction

Genome-wide association studies (GWAS) of neurodevelopmental and psychiatric disorders have demonstrated that the majority of common variation associated with these disorders is found in noncoding regions of the genome [1–8]. Similarly, whole-genome sequencing studies (WGS) are poised to discover rare noncoding genetic variation associated with neurodevelopmental disorders [9–11]. Whereas the functional impact of genetic variation in protein coding regions can be inferred through knowledge of the codon code, the impact of genetic variation in the noncoding genome is much more difficult

to understand as no such regulatory code is known. The noncoding genome contains cis-regulatory regulatory elements (CREs) such as enhancers, promoters, silencers, and insulators, which influence gene expression by serving as docking sites for DNA-binding proteins like transcription factors (TFs) [12, 13]. Variants within a regulatory element can alter TF binding and subsequently alter gene expression and cellular function [14, 15].

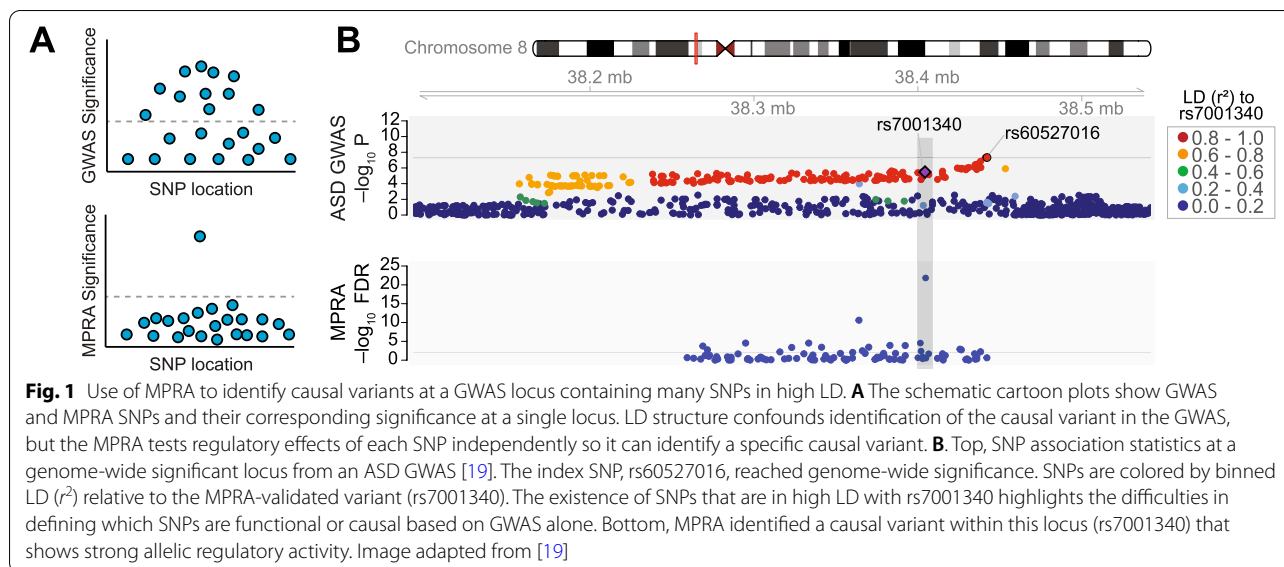
In addition to the lack of regulatory code, GWAS alone cannot pinpoint variants that are causing a disease because of linkage disequilibrium (LD), the non-random inheritance of nearby alleles on the genome. A genome-wide significant (GWS) locus typically contains tens to hundreds of single-nucleotide polymorphisms (SNPs) that are associated with a trait or disease. Only a subset of these SNPs are thought to be causal. While it is commonly thought that the index

\*Correspondence: [jason\\_stein@med.unc.edu](mailto:jason_stein@med.unc.edu); [hyejung\\_won@med.unc.edu](mailto:hyejung_won@med.unc.edu)

<sup>2</sup> UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Full list of author information is available at the end of the article





SNP, the SNP most significantly associated with the trait at a locus, is causal, growing evidence portrays a more complex picture [16]. The lead SNP is not always the causal allele when functionally validated, and a given locus can contain multiple causal variants [16]. Identifying the causal variant(s) at a locus can greatly facilitate our understanding of disease mechanisms by narrowing down the genetic underpinnings of a disease (Fig. 1). Moreover, causal variant identification provides the intriguing possibility of developing therapeutics by reversing pathological transcriptional mechanisms or genetically modifying causal variants [17, 18].

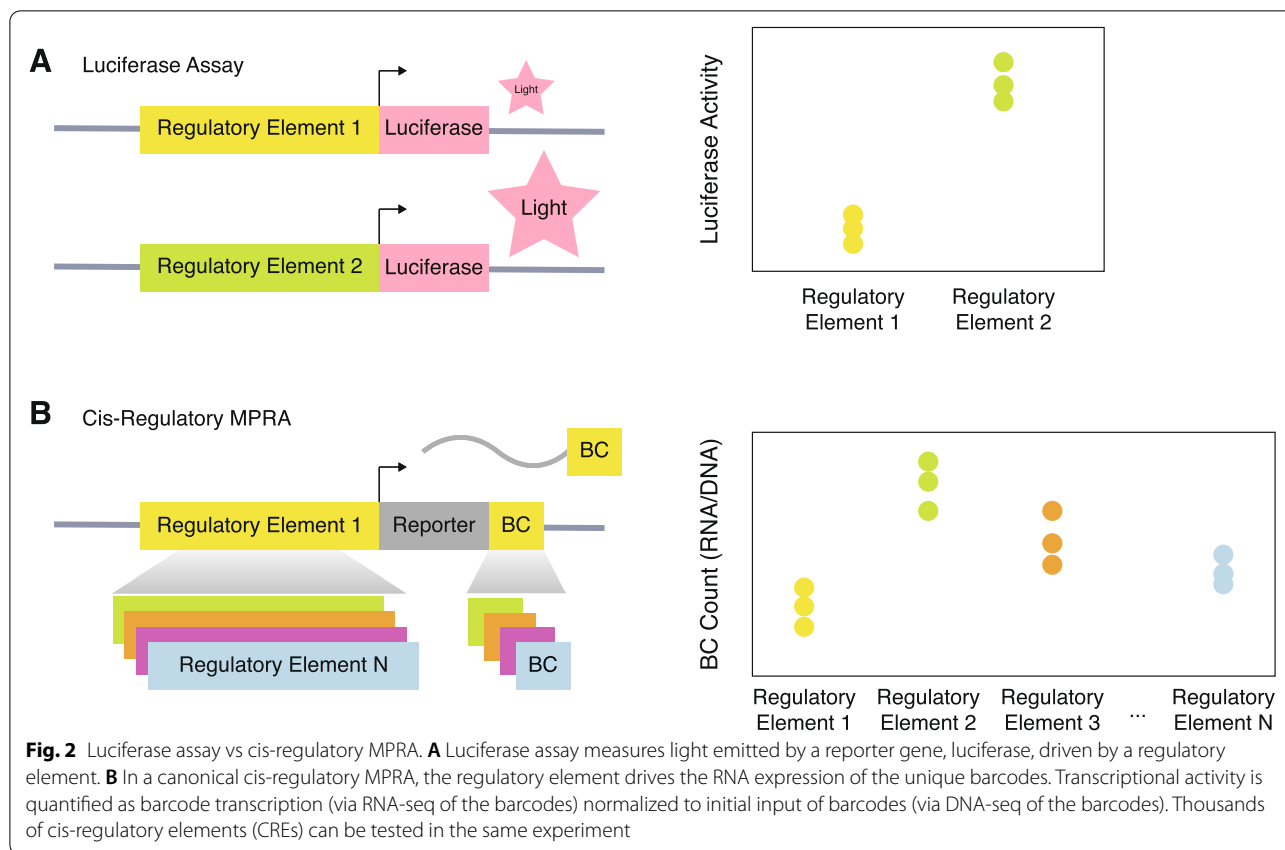
Several experimental and computational designs have been used to predict the causal variant at a given locus. Fine-mapping tools computationally predict potential causal variants based on association statistics and LD patterns [20–23], but different algorithms can yield conflicting results, and prioritized variants still require experimental validation [24]. Allele-specific chromatin accessibility (ASCA) can be used to determine if inherent genetic variation in a population of individuals affects chromatin accessibility, a proxy for gene regulatory activity, in relevant cell types [25, 26]. Genetic variants within a noncoding regulatory element that affects the function of that element are highly likely to be causal gene regulatory variants. Allele-specific chromatin accessibility when colocalized with GWAS suggests that the genetic variants are also causally associated with the trait or disease. However, ASCA experiments require large sample sizes of genetically diverse donors with both genotype and chromatin accessibility data, and they cannot independently test the effect of

multiple variants in high LD within a regulatory element [27, 28].

Functional validation assays fill the existing gap by experimentally demonstrating how genetic differences lead to phenotypic effects [29]. Gene regulatory activity of noncoding elements has historically been functionally validated via luciferase assays (Fig. 2A). A luciferase assay places a regulatory sequence of interest (sometimes containing a SNP) upstream of a luciferase reporter gene and quantifies the regulatory effect on expression via luminescence of luciferase [27]. However, luciferase assays lack the throughput to validate thousands of regulatory sequences at once because each regulatory element must be measured independently.

Massively parallel reporter assays (MPRA) have advanced the throughput of luciferase assays by enabling the simultaneous functional validation of regulatory activity of thousands of variants on a massive scale (Fig. 2B), often vastly narrowing down thousands of variants found by GWAS or quantitative trait loci (QTLs) in a single assay (Fig. 1). Rather than quantifying the luminescence of luciferase, MPRA measures the barcoded reporter gene expression via next-generation sequencing. Once the MPRA construct is introduced into cells of interest, the synthesized regulatory element drives the expression of its unique barcode, a random oligo sequence that uniquely tags the matching regulatory element. The initial input of the construct is quantified by the DNA counts of the barcodes, which is compared to the RNA counts to evaluate effects on expression (Fig. 2B).

MPRA has incredible potential for studying the noncoding genetic variants associated with



neurodevelopmental disorders. Whereas the majority of efforts have been made to characterize common variants identified by GWAS, wide application of WGS would further expand the utility of MPRA in characterizing various classes of variants located in the noncoding genome.

In this review, we will discuss the broad application of MPRA to functionally validate variants within the various regulatory contexts that encompass transcriptional and posttranscriptional regulation. We will then add important considerations for conducting MPRA, including limitations of MPRA experiments. We conclude by providing future directions of MPRA.

### MPRA for studying cis-regulatory elements

#### Canonical MPRA

The canonical MPRA design includes a CRE, a generic promoter, a reporter gene, and a unique barcode assigned to each regulatory element (Fig. 2B). Generally, CRE libraries of interest are made with mass oligonucleotide synthesis. To interrogate variant effects on gene regulation, the CRE can be modified to harbor a variant within its sequence. Additionally, every possible single-nucleotide mutation can be added to the CRE, called saturation mutagenesis. The impact of the variant on regulatory activity is measured through barcodes matched to each

unique variant. Because the barcode itself can have an influence on levels of expression, many barcodes are usually tested for each variant. Transcriptional activity is quantified as barcode transcription (via RNA-seq of the barcodes) normalized to initial input of barcodes (via DNA-seq of the barcodes). This allows systematic investigation of variant function within a noncoding region by comparing the gene regulatory activity between protective and risk alleles of a given variant. A growing body of research employs this strategy to identify functional regulatory variants within GWAS loci [30–34]. Key consideration in designing MPRA involves the use of proper controls [35]. For example, scrambled sequences of DNA in the relevant cell type can be used as negative controls to experimentally validate enhancers [35]. Likewise, a strong promoter or a known highly expressed sequence in the relevant cell type can be used as positive controls [35]. The canonical MPRA design has been adapted to fit the needs of differing types of CREs being tested [36–39].

#### Promoter

Mutations and variants in promoter regions can have a profound impact on gene expression. MPRA has been used to test the impact of variants in promoters on a massive scale. In comparison with the canonical MPRA

design, promoter MPRA lacks a CRE and alters the DNA sequence of the promoter region. Patwardhan et al. utilized promoter MPRA with saturation mutagenesis to screen the activity of mutated promoter sequences with attached barcodes (Fig. 3A) [40]. Barcode counts quantified via short-read sequencing provided a scalable read-out of promoter activity, which led to the identification of critical regions of a promoter that govern transcriptional efficiency [40].

In addition to introducing variation in the promoter sequences, a similar approach can be used to characterize promoter activity of any given sequence. Boer et al. developed a gigantic parallel reporter assay (GPRA) that measured the promoter activity of over 100 million randomly synthesized sequences [46]. The complexity of synthetic promoters surpasses the complexity of the human genome, allowing them to build a predictive model of how genetic sequence affects transcriptional regulation.

While many neurodevelopmental disorder-associated variants have been shown to be enriched in promoter regions [24, 47, 48], MPRA has yet to be adopted to systematically examine the regulatory function of these promoter variants. We expect MPRA will provide a useful avenue to elucidate the function of promoter variants associated with neurodevelopmental disorders.

### Enhancers

Enhancers are CREs that TFs bind to and activate gene expression [49]. Disease-associated risk variants are enriched in enhancers [50]. Despite their important roles in gene regulation and disease associations, the sequence logic of enhancers is not well understood. Therefore, MPRA has been widely adapted to experimentally test

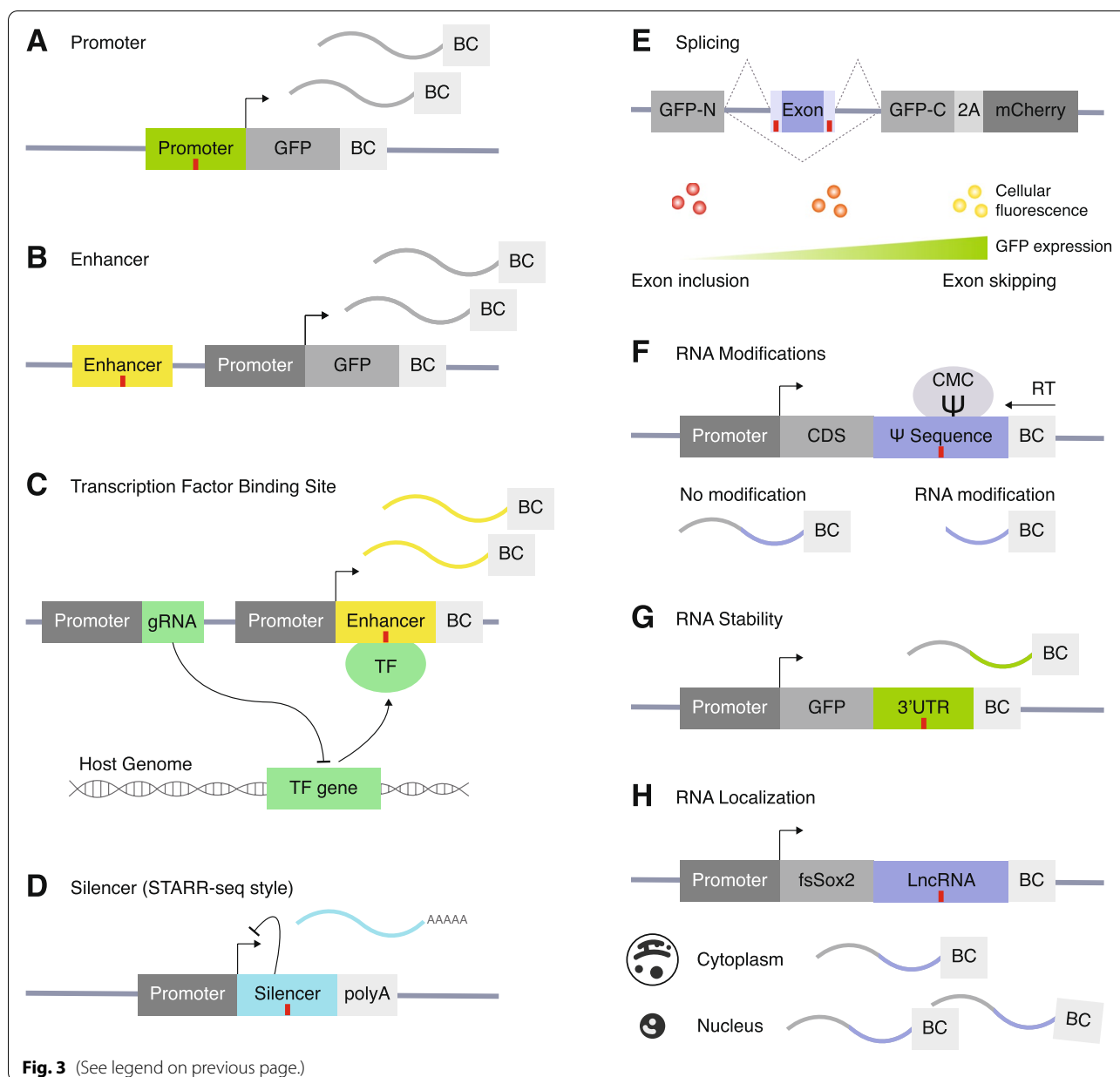
the function of enhancers and variants within enhancers [19, 51]. While MPRA can take on many forms to examine enhancer functions [52], generally, a putative enhancer element is coupled with a weak promoter (e.g., minimal promoter) that is followed by a reporter gene and a unique barcode (Fig. 3B).

Myint et al. used enhancer MPRA to screen 1049 schizophrenia- and 30 Alzheimer's disease-associated variants for differences in driving reporter gene expression [34]. They used two cell lines and identified 192 SNPs with significant differences in driving reporter gene expression [34]. Among the 192 variants, 148 showed allelic differences in K562 cells, 53 in SK-SY5Y cells, and only 9 showed allelic differences in both cell lines, demonstrating that genetic variants often exert their regulatory effects only within specific cell types [34]. As an additional example, Matoba et al. used MPRA to fine-map one novel ASD GWAS locus in HEK293T cells (Fig. 1) [19]. Of 98 variants tested, two were found to have significant differential allelic activity, with one variant (rs7001340) exhibiting strong effects. By integrating MPRA results with expression quantitative trait loci (eQTLs), they showed that an ASD-associated risk allele decreased the expression of *DDHD2* [19]. These examples highlight MPRA's ability to map disease-associated variants within putative enhancer regions.

Transcription factors (TF) recognize and bind to specific sequences within an enhancer, called TF binding motifs, to regulate gene expression. Variants within motifs can disrupt TF binding or create new motifs, altering regulatory activity. Though enhancer MPRA can identify if a variant affects enhancer activity, it does not experimentally validate which TF contributes to the altered regulation. TF-DNA interactions can be

(See figure on next page.)

**Fig. 3** MPRA designs for studying gene regulation. MPRA modifies the design of canonical cis-regulatory MPRA (described in Fig. 2B) that contains a cis-regulatory element (CRE), a promoter, a reporter gene, and a unique barcode (BC). Elements of this construct can be replaced or rearranged to test different types of CREs. The red vertical line indicates where a variant can be located. **A** Promoter MPRA contains a promoter harboring a variant, a reporter gene (e.g., GFP), and a unique BC. Image adapted from [40]. **B** Enhancer MPRA contains a regulatory element harboring a variant, a (minimal) promoter, a reporter gene, and a unique BC. **C** Transcription factor binding MPRA (TransMPRA) can be broken down into two components: (1) a promoter with a guide RNA (gRNA) that targets a transcription factor (TF) of interest and (2) a promoter, a test enhancer sequence harboring a variant, and a unique BC. The gRNA brings catalytically dead Cas9 protein with an attached Krüppel-associated box (dCas9-KRAB) which silences the expression of the TF gene. If the silenced TF interacts with the test enhancer, the downstream barcode expression is decreased. Image adapted from preprint [41]. **D** Silencer MPRA (in a STARR-seq style) contains a (strong) promoter and a test silencer harboring a variant. The silencer sequence can prevent self-transcription by silencing the promoter. Image adapted from [42]. **E** Splicing MPRA has minigene constructs that are inserted between a split-GFP reporter (GFP-N terminus and GFP-C terminus) and a peptide 2A (P2A) upstream of an mCherry reporter. Variants can be located in the variable intron sections on either side of the exon or within the exon. Inclusion of the middle exon disrupts GFP fluorescence, and cells can be FACS sorted into bins based on GFP:mCherry ratios. The GFP with or without the exon are quantified for exon inclusion or skipping via DNA-seq of the plasmid in each sorted bin. Image adapted from [43]. **F** RNA modification MPRA contains a promoter, an arbitrary coding sequence (CDS), a putative pseudouridine ( $\Psi$ ) sequence as 3' untranslated region (UTR), and a unique barcode. Once the library is introduced, cells are treated with N-cyclohexyl-N'- $\beta$ -(4-methylmorpholinium) ethylcarbodiimide (CMC) which binds to  $\Psi$  and prevents reverse transcription (RT). High-throughput sequencing of cDNA allows prediction of the exact base pair location of the  $\Psi$  RNA modification. Variants can be inserted anywhere in the CDS. Image adapted from [44]. **G** 3' UTR MPRA consists of a promoter, a reporter gene, a 3' UTR harboring a variant, and a BC. BC RNA counts reflect transcriptional stability modulated by 3' UTRs. **H** RNA localization MPRA consists of a promoter, a mutated *Sox2* gene that localizes in the cytoplasm (*fsSox2*), a lncRNA harboring a variant, and a unique barcode. Barcode expression from subcellular fractions is used to interrogate subcellular localization of lncRNA. Image adapted from [45]



measured using methods such as chromatin immunoprecipitation sequencing (ChIP-seq), or they can be inferred using CRISPR knockout screens that model the impact of TFs on gene regulatory programs [53–55]. A recently introduced technique (in preprint) called TransMPRA also addresses this question by combining MPRA with CRISPR interference and single-cell sequencing to measure the interaction between transacting factors and putative enhancers (Fig. 3C) [41]. In this system, a guide RNA (gRNA) for a known TF is packaged together with enhancer MPRA that are potentially directly targeted by the TF [41]. When introduced into cells expressing

dCas9-KRAB proteins, TF expression is inhibited, and enhancer activity is reduced only if the element is a downstream target of the TF [41]. Accordingly, TransMPRA provides an incredibly important tool to delineate potential transcriptional regulators for noncoding variants associated with neurodevelopmental disorders.

**Silencers**

MPRA has been adapted to test silencer elements, which are noncoding functional elements that lead to decreased expression of their target gene (Fig. 3D) [42]. Silencer MPRA differs from enhancer MPRA in two aspects.

First, enhancer MPRA uses a weak promoter (e.g., minimal promoter) to measure increases in gene expression elicited by the putative enhancer, while silencer MPRA uses a strong promoter (e.g., super core promoter, SCP1) that transcribes a high baseline level of the construct, so decreases in transcription can be detected. Second, silencer MPRA leverages the design of self-transcribing active regulatory region sequencing (STARR-seq), a sub-branch of MPRA (for more information about STARR-seq, please see the review [56]). STARR-seq places an uncharacterized CRE downstream of a strong core promoter followed by a polyA tail. This MPRA design does not require barcodes because the sequence of the transcribed putative silencer acts as the barcode [42]. While MPRA in the STARR-seq style has been widely adopted, it is important to consider that mRNA sequences could be affected by posttranscriptional effects such as mRNA degradation which would be indistinguishable from transcriptional effects [57].

Silencer MPRA has been applied to detect thousands of CREs acting as silencers, which were enriched for disease-associated SNPs [42], highlighting the need to decipher regulatory logic of transcriptional silencing in understanding disease etiology [42].

### MPRA for studying posttranscriptional regulation

#### *Splicing*

MPRA can be combined with methods that sequence populations of cells binned by fluorophore expression, called Sort-seq [58], to study posttranscriptional processes like alternative splicing. In splicing MPRA, a red fluorophore (mCherry) is constitutively expressed, and a three-exon, two-intron minigene construct is cloned into a plasmid in such a way that when the middle (tested) exon is skipped, a green fluorophore (GFP) is also expressed (Fig. 3E) [43]. Variants can be located in the variable intron sections on either side of the test exon or within the exon. Cells are sorted into bins using GFP:mCherry ratios by fluorescence-activated cell sorting (FACS), where a higher ratio indicates greater intron excision. Plasmid DNA is then sequenced within each bin to determine which variants affect splicing. In an experiment utilizing this assay, many of the variants that lead to differences in splicing were located outside of canonical splice sites in both exons and introns, demonstrating that novel types of genetic variation affect splicing [43].

Though splicing MPRA have not yet been used to validate neurodevelopmental disorder-associated variant function, alternative splicing is a critical process for neuronal fate specification during neurogenesis [59, 60], and differences in alternative splicing have been identified in postmortem brains from individuals with autism, schizophrenia, and bipolar disorder [61]. Rare

neurodevelopmental disorders can also be caused by alterations in alternative splicing. For example, familial dysautonomia, a degenerative sensory and autonomic nervous system disorder, is caused by a 5' splice site mutation in an intron of *IKBKAP* [62]. The mutation results in variable exclusion of exon 20 and reduced IKAP protein levels in neuronal tissue [63]. Identifying the mutation has allowed understanding of the disease mechanism [64] and testing of therapeutic treatments [65, 66]. Therefore, splicing MPRA have great potential to identify variants that contribute to abnormal splicing in neurodevelopmental and neuropsychiatric disorders.

#### *RNA editing*

RNA sequences can be modified posttranscriptionally in a process known as RNA editing, which can alter the function of a gene [67, 68]. Dysregulation of RNA editing has been shown in many nervous system disorders such as brain cancer, addiction, depression, Alzheimer's disease, amyotrophic lateral sclerosis (ALS), ASD, and intellectual disabilities (ID) [68, 69]. MPRA have been adapted to quantify RNA editing events such as uridine to pseudouridine ( $\Psi$ ) which changes RNA regulation and stability [44, 70]. RNA editing MPRA consists of a promoter, coding sequence (CDS), putative  $\Psi$  sequence containing variants within the 3' untranslated region (UTR), and a unique barcode (Fig. 3F) [44]. Once the MPRA library is introduced into cells, it is treated with a molecule that binds to the  $\Psi$  and prevents reverse transcription at the binding site [44]. Consequently, the exact base pair location of the  $\Psi$  alteration can be determined via high-throughput sequencing. This type of assay can show which underlying DNA sequences and variants lead to uridine to  $\Psi$  RNA editing, decoding much of the unknown regulatory code for RNA editing.

#### *RNA stability and translation*

In addition to transcriptional regulation, mRNA stability and translational control are another critical step that determines protein abundance. 5' and 3' UTRs are regulatory regions of the DNA that influence mRNA stability, localization, and translation [71, 72]. MPRA can be used to study the function of both 5' and 3' UTR sequences.

5' UTRs affect translational efficiency by altering ribosome loading [73]. In an example of 5' UTR MPRA, Sample et al. measured the impact of 5' UTR sequences on ribosome loading by having putative 5' UTR sequences inserted upstream to a GFP and a 3' UTR [73]. After introducing the constructs to the cells, 5' UTR sequences that are actively being translated in ribosomes are directly sequenced via polysome profiling [73]. This method identified 45 variants associated with disease that significantly affected ribosome loading [73]. 5' UTR MPRA is

an effective method to identify disease-associated variants that alter translational efficiency.

MPRA are also used to understand how variants within 3' UTR sequences affect stability of mRNA (Fig. 3G) [74, 75]. In the 3' UTR design, a promoter drives GFP expression which has a 3' UTR containing a variant, and a unique barcode matched to that variant. Barcoded expression of GFP and 3' UTR can be used to assess differences in RNA abundance (and hence stability). Lagunas et al. used this approach to assess the activity of > 500 de novo noncoding variants identified by WGS of ASD families, yielding 41 variants with differential stability effects in the brain [74]. This is in line with the previous findings that modifications in the 3' UTR are broadly linked to brain function and neurodevelopmental disorders [76–78].

### **RNA localization**

Location of RNA within the cell often closely aligns with its function. Massively parallel RNA assay (MPRNA) can test subcellular localization (e.g., nuclear vs. cytosolic) of long noncoding RNAs (lncRNAs) at scale [45]. The MPRNA construct consists of a cytosolic-localized *Sox2* variant (*fsSox2*), a DNA sequence that encodes a lncRNA, and a unique barcode (Fig. 3H) [45]. The *fsSox2* makes the baseline localization cytosolic; therefore, if the lncRNA creates a nuclear localization sequence, it will translocate to the nucleus. Once cells are transfected with the MPRNA construct, the cytoplasm and nucleus are isolated via subcellular fractionation [45]. The resulting barcode counts from RNA-seq inform lncRNA sequences that influence nuclear localization, which in turn affects the regulatory activity of the lncRNA [45]. Subcellular localization provides valuable insight into the function of lncRNAs, which is unknown compared to protein coding genes despite evidence found for their role in brain development and neurodevelopmental disorders such as ASD, Rett's syndrome, attention-deficit hyperactivity disorder (ADHD), and schizophrenia [79–92] and other neurological disorders such as Alzheimer's, Parkinson's, and Huntington's disease [82, 93].

### **Limitations of MPRA**

#### **MPRA cannot identify target genes**

Though MPRA is an incredibly useful tool to experimentally verify variant function, this assay is not without limitations. Enhancer, promoter, and silencer MPRA can effectively identify variants and elements with regulatory activity, but these assays cannot inform which gene(s) that the variants act on. Therefore, MPRA results need to be combined with other functional genomic data such as eQTLs and chromatin interaction profiles (Hi-C) to identify target genes [94]. While functional genomic

approaches can be a good starting point to discern variant function, acquiring functional genomic datasets that match the appropriate biological context can be challenging when a rare cell type or environmental perturbation is used.

A complementary approach to address this gap is Perturb-seq, which employs a pooled library of gRNAs associated with a unique barcode that modulates expression of the gRNAs' target genes. A Perturb-seq gRNA library could be designed for the MPRA-validated regulatory elements and introduced into a cell line expressing Cas9 protein. Introduced gRNAs can perturb the region of interest, and transcriptomic alterations can be profiled via scRNA-seq. The resulting data can explain the regulatory impact of perturbed elements within a given cell [95]. Moreover, because Cas9 perturbs host genomic DNA, it can further verify element function within the biologically relevant (epi)genetic context. Therefore, Perturb-seq can complement MPRA by shedding light on which genes and pathways MPRA-validated elements act on.

### **Genomic context**

Another disadvantage of MPRA is that it uses exogenous DNA constructs that do not model the endogenous regulatory environment at the variant location. Exogenous DNA constructs are either episomal or inserted into the DNA at random locations. An episomal construct is not subject to cis-regulatory effects such as chromatin accessibility and conformation. Instead, its activity is only modulated by trans-regulatory effects, like transcription factors. In contrast to episomal MPRA, MPRA delivered through a lentiviral vector (LentiMPRA) randomly integrates into the host genome, hence enabling functional characterization of regulatory elements and variants within the context of the host genome [96]. However, LentiMPRA constructs are randomly integrated into the genome, so the (epi)genomic context of the integrated site most likely differs from that of the host genome and may differ from one insert to another.

The effect of genomic context has been evaluated with a new technique called PatchMPRA [97]. This technique leverages a cell line that has multiple known landing pads, each labeled with a unique genomic barcode. Because chromatin architecture of each landing pad has been well characterized, (epi)genomic context can be accounted for when interpreting the MPRA results. Maricque et al. used PatchMPRA to test over 30,000 combinations of CREs and local chromatin architecture in K562 cells [97]. They found that the location of landing pads in the genome had significant effects on barcode expression [97]. In particular, the DNA sequence of CREs determines the intrinsic regulatory activity, which

is then fine-tuned by the chromatin environment [97]. While PatchMPRA enables systemic interrogation of interaction between regulatory elements and genomic contexts, it requires landing pads to be inserted into the host genome, which may not always be possible in some model systems.

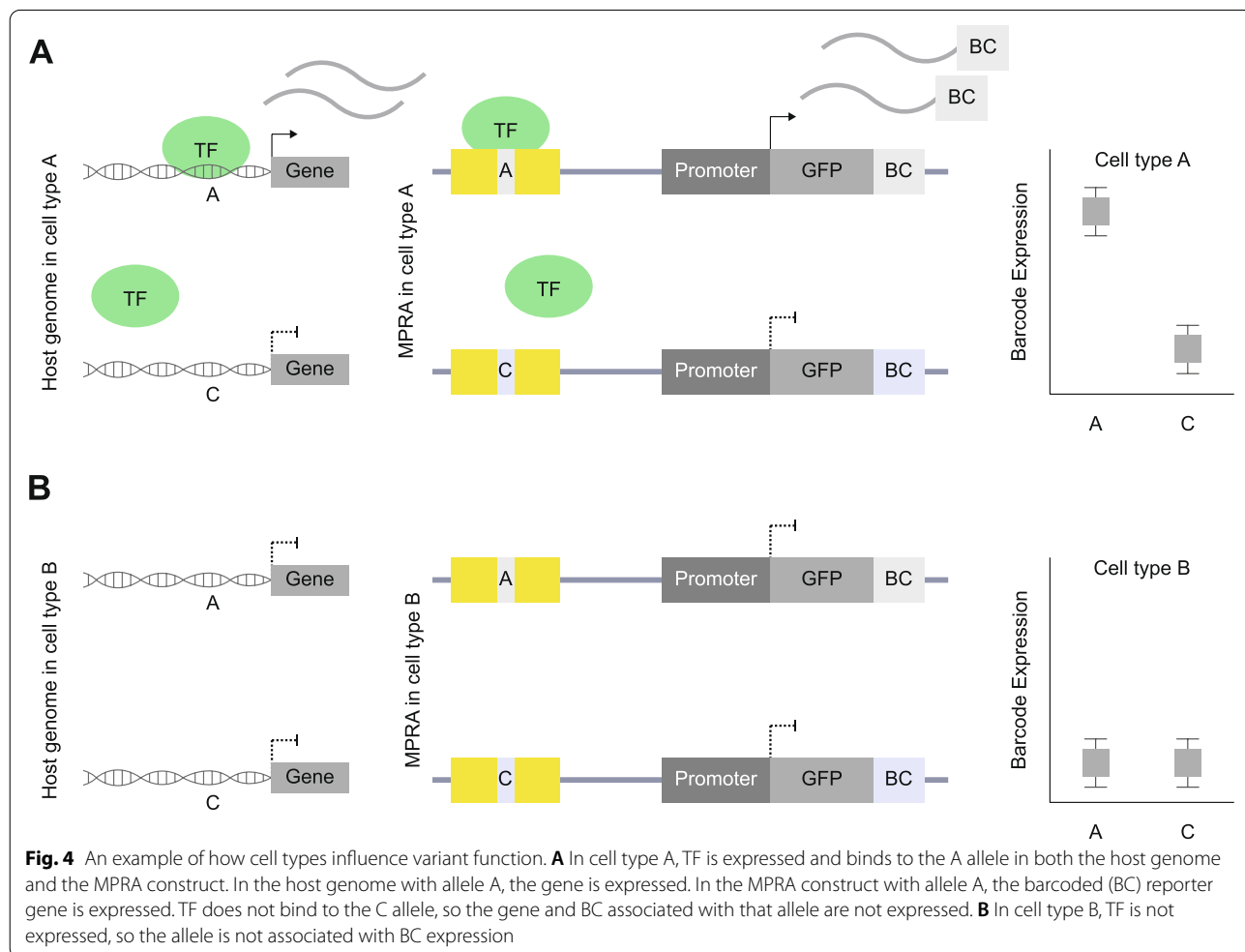
**Future directions**

**Choosing cell type and developmental time period**

Regulatory elements are often only functional within a given tissue or cell type [25, 98–101]. As regulatory elements display extensive tissue and cell type specificity [25, 98–101], it is important to choose the cell and/or tissue type carefully for an MPRA experiment. MPRA may give different results as to which CREs and variants have regulatory effects based on which cell type the MPRA is tested in (Fig. 4) [102]. This can be due to TFs being differentially expressed in different cell types [103] leading to cell type-specific regulatory element activity [25].

MPRA is most commonly performed, out of convenience, on cell types easily cultured and transfected in a lab (such as HEK293 cells [19]). Because TFs may only be expressed within specific cell types, MPRA results may change in other cell types (Fig. 4), so using a relevant brain cell type for neurodevelopmental disorders is a preferred experimental design. Genetic variation associated with multiple neurodevelopmental disorders is enriched in regulatory elements present in dorsal telencephalic neural progenitors and excitatory cortical neurons, making them optimal cell types to conduct MPRA experiments [25, 60, 98, 104–106] (Fig. 4).

The cell type specificity of MPRA has been previously demonstrated with an MPRA on random CREs tested in two cell lines: U87 glioblastoma cells and induced pluripotent stem cell (iPSC)-derived human neural progenitor cells (hNPCs) [102]. This study found a significant inverse correlation ( $R = -0.326$ ) between regulatory activity of CRE barcode expression in these two cell types [102]. The difference in regulatory activity was attributed to the





difference in TFs expressed in those cells and the presence of their binding motifs within the MPRA library. An enrichment in binding sites for TFs involved in brain development, such as SOX2, DBX1, and FOXP2, was found in hNPCs compared to U87 cells, as these TFs are more highly expressed in hNPCs [102]. These results underline the importance of choosing a relevant model system for studying brain-specific variant function.

To delineate variant effects on gene regulation in a cell type-specific manner *in vivo*, Cre recombinase-dependent MPRA packaged within an adeno-associated virus (AAV) has been developed [74]. In a preprint describing this method, mouse lines expressing Cre recombinase from an endogenous cell type-specific promoter is used to restrict the expression of an MPRA construct to a given cell type and location. This design has been used to test MPRA within excitatory neurons by injecting the Cre-dependent AAV MPRA into the cortex of Vglut1-IRES2-Cre-D mice [44]. Controlling for cell type specificity within an *in vivo* system is especially important as brain tissue is composed of heterogeneous cell types, and MPRA results from bulk tissue may mask the effects of variants functional in relatively less abundant cell types.

Just as cell and tissue types are important to MPRA, the development stage is a critical factor to consider when performing MPRA. This is especially important in studying genetic etiology of neurodevelopmental and psychiatric disorders which, by definition, have neurodevelopmental origin [60, 105, 107]. As such, investigating variant function during prenatal time periods when processes such as neurogenesis, gliogenesis, synaptogenesis, and pruning are occurring increases the likelihood of gaining neurodevelopmental relevant information via MPRA [108].

#### **Response MPRA: gene-environment interactions explored**

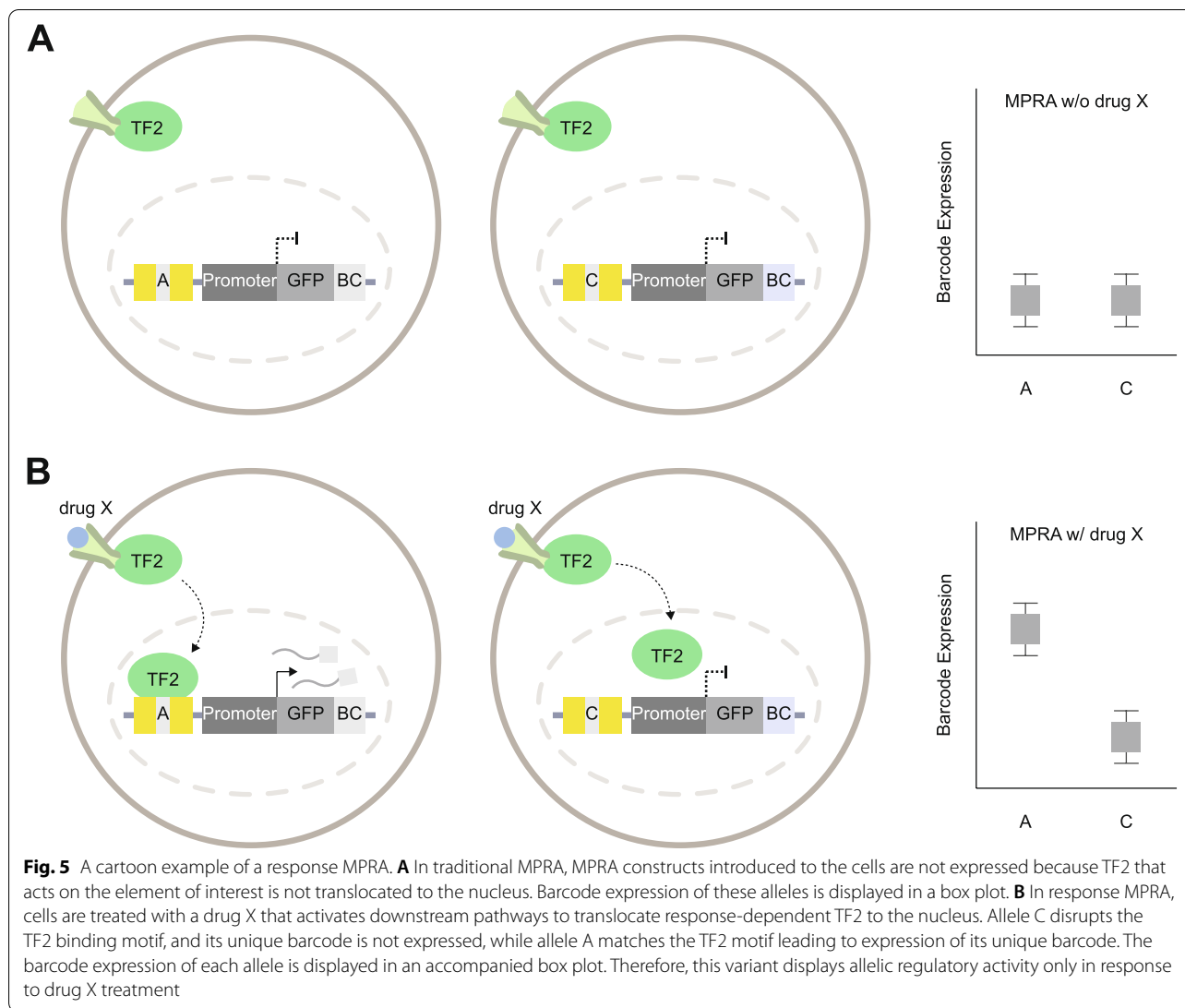
One critical, unanswered question in the field is the extent to which variant function is influenced by gene-environment interactions. External stimuli can alter a cellular pathway that has downstream effects on TF abundance and binding properties. Applying MPRA in this context can uncover a new class of variants that gain (or lose) regulatory effects upon exposure to external stimuli. Here, we propose a term “response MPRA” to describe MPRA performed in response to an external stimulus, such as exposure to hormones, drugs, or other biomolecules, as well as measuring gene regulatory effects in the context of a particular cell state (Fig. 5).

In an example of a response MPRA, Mulvey et al. [32] interrogated allelic regulatory activity of major depressive disorder (MDD) risk variants in response to *all-trans* retinoic acid (ATRA). ATRA, an acid derivative of retinol, activates the retinoic pathway, which has been

implicated with the risk for MDD [32]. ATRA administration to N2A neuroblastoma cells not only increased the magnitude of allelic regulatory activity of a subset of variants with retinoic receptor motifs but also unmasked allele-specific activity not otherwise detected in the traditional MPRA [32]. As retinoic acid is a potent driver of neuronal differentiation, further investigation is required to distinguish ATRA-dependent regulatory effects from cell type-specific regulatory effects. Still, the evidence suggests that a subset of variants may function only upon activation of the retinoic pathway, providing a biological context for MDD genetic risk factors.

An additional application of the response MPRA follows the paradigm of measuring gene regulatory activity in the context of a particular biological process. In neurons, a classic example of this involves measuring CRE response upon stimulation with known modulators of neuronal activity. In an early adaptation of this approach, Nguyen et al. compared promoter and enhancer activities in neurons treated with potassium chloride with their unstimulated controls [109]. While similar sequences were found to be associated with increased neuronal activity between promoters and enhancers, the authors identified specific TF binding sites enriched in promoters that led to greater overall activity in response to neuronal activity. Another important cellular process relevant to neurodevelopment is the proliferation of neural progenitor cells. Dysregulation of the cell cycle and increased proliferation have been associated with brain overgrowth phenotypes present in individuals with ASD [110, 111]. By isolating proliferating cells, marked by incorporation of thymidine analogues such as BrdU, one can identify elements specifically active during phases of the cell cycle. These examples demonstrate how MPRA can be used to assess the importance of cell state in mediating gene regulatory activity.

Given the limited therapeutic options available for neurodevelopmental and psychiatric disorders, response MPRA can provide a high-throughput framework to investigate the impact of a drug candidate on variant function, potentially leading to high-throughput testing for pharmacogenomics and personalized therapies based on genetic background. An example of such an approach could be testing ADHD-associated variants with methylphenidate (a common ADHD medication [112]) to determine which variants have altered functionality upon drug exposure. Identifying the alleles that are responsive to methylphenidate exposure may eventually lead to more informed clinical decisions based on patient genotype during ADHD treatment. Choosing an appropriate model system with disease relevance is critical to accurately assess drug response. In this regard, patient-derived iPSCs are ideal tools for drug discovery



efforts as iPSCs can differentiate into many cell types and can be scaled to meet coverage requirements for MPRA screening.

Finally, response MPRA can provide a useful tool to study gene-environment interactions by interrogating the variant function upon exposure to environmental factors associated with disease risk. Maternal exposure during pregnancy to valproic acid (VPA), a commonly prescribed antiepileptic medication, has been associated with risk for ASD [113], as well as several other neurodevelopmental disorders [114–116]. Studies have shown that VPA exposure can alter the proliferation and neurogenic capacities of neural progenitor cells during brain development [117], which can result in downstream deficits in brain structure [118] and cytoarchitecture [119]. Performing VPA-response MPRA within a progenitor cell type using ASD-associated variants could shed light

on which alleles have altered function upon VPA exposure. However, VPA is only one of many environmental factors that have been associated with ASD risk [120]. Overall, various classes of external stimuli, the number of variants whose function is altered upon stimulation, and the magnitude of these effects on gene regulation have yet to be explored and can provide novel insights into disease mechanisms.

### Conclusions

Many variants in noncoding regions have, and will continue to be, identified by large-scale studies such as GWAS, WGS, and QTL. Understanding the function of those variants, and which variants within a haplotype block are causal, is the next key step in moving from association to biological understanding. In this review, we outlined how MPRA can validate variant function

in a wide range of regulatory elements such as enhancers, promoters, silencers, and TF binding sites. We also described how MPRA can be used to garner mechanistic understanding of posttranscriptional regulation such as splicing, RNA modification, RNA stability, translation, and RNA localization. MPRA results can change based on cell type, stimulus state, and developmental time period, so these parameters must be carefully considered when designing an MPRA experiment. MPRA has limitations in associating regulatory effects to a target gene and lacks epigenetic context. Complementary approaches that range from other functional genomic resources (e.g., eQTLs, ASCA, and Hi-C) to other screening platforms (e.g., Perturb-seq) will extend knowledge gained from MPRA to provide a greater understanding of the mechanisms by which genetic variants affect brain structure, function, and development.

#### Abbreviations

AAV: Adeno-associated virus; ADHD: Attention-deficit hyperactivity disorder; AF: Allele frequency; ALS: Amyotrophic lateral sclerosis; ASCA: Allele-specific chromatin accessibility; ASD: Autism spectrum disorder; ATR: All-trans retinoic acid; BC: Barcode; CDS: Gene coding region; ChiP-Seq: Chromatin immunoprecipitation sequencing; CMC-N: Cyclohexyl-N'-(β-[N-methylmorpholino]ethyl) carbodiimide p-toluenesulfonate salt; CRE: Cis-regulatory element; CRISPR: Clustered regularly interspaced short palindromic repeats; dCas9: Catalytically dead Cas9 endonuclease; dCas9-KRAB: Dead Cas9 protein with an attached Krüppel-associated box; DDHD2: DDHD domain containing 2; DNA: Deoxyribonucleic acid; DNA-seq: Deoxyribonucleic acid sequencing; FACS: Fluorescence-activated cell sorting; GFP: Green fluorescent protein; GPRA: Gigantic parallel reporter assay; gRNA: Guide RNA; GWAS: Genome-wide association study; hNPCs: Human neural progenitor cells; ID: Intellectual disability; iPSCs: Induced pluripotent stem cells; LD: Linkage disequilibrium; lncRNA: Long noncoding RNA; MDD: Major depressive disorder; MPRA: Massively parallel reporter assay; MPRNA: Massively parallel RNA assay; mRNA: Messenger RNA; P2A: Peptide 2A; QTL: Quantitative trait loci; RNA: Ribonucleic acid; RNA-seq: RNA sequencing; RT: Reverse transcription; RT-PCR: Reverse transcription-polymerase chain reaction; SCP1: Super core promoter 1; SNP: Single-nucleotide polymorphism; STARR-seq: Self-transcribing active regulatory region sequencing; TF: Transcription factor; TFBS: Transcription factor binding site; TransMPRA: Transcription factor binding MPRA; UTR: Untranslated region; VPA: Valproic acid; WES: Whole exome sequencing; WGS: Whole genome sequencing; YFP: Yellow fluorescent protein; Ψ: Pseudouridine.

#### Acknowledgements

The authors thank the Stein and Won labs for their ongoing moral and scientific support. They would also like to acknowledge the existence of cats for their continual interruption and entertainment while writing this review.

#### Authors' contributions

JCM and JLB, outlined the layout of this review. JCM and JLB, co-wrote the entire draft with the supervision of JLS and HW. JCM and HW, designed and generated figures. OK, contributed to the "Response MPRA: gene-environment interactions explored" section. NM contributed to the "Choosing cell type and developmental time period" section. JCM, JLB, OK, JLS, and HW reviewed and edited the paper. The authors read and approved the final manuscript.

#### Funding

This research was supported by the National Institute of Mental Health (R00MH113823, DP2MH122403, H. W., R01MH118349, J. L. S., U01MH122509, J. L. S. and H. W.), National Human Genome Research Institute (UM1HG012003, H. W.), and a SPARK grant from the Simons Foundation Autism Research Initiative (H. W. and J. L. S.).

#### Availability of data and materials

Not applicable

#### Declarations

#### Ethical approval and consent to participate

Not applicable

#### Consent for publication

Not applicable

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. <sup>2</sup>UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

Received: 19 October 2021 Accepted: 1 September 2022

Published online: 09 September 2022

#### References

- Xiao X, Chang H, Li M. Molecular mechanisms underlying noncoding risk variations in psychiatric genetic studies. *Mol Psychiatry*. 2017;22(4):497–511.
- Roussos P, Mitchell AC, Voloudakis G, Fullard JF, Pothula VM, Tsang J, et al. A role for noncoding variation in schizophrenia. *Cell Rep*. 2014;9(4):1417–29.
- Sullivan PF, Geschwind DH. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell*. 2019;177(1):162–83.
- Turner TN, Eichler EE. The role of de novo noncoding regulatory mutations in neurodevelopmental disorders. *Trends Neurosci*. 2019;42(2):115–27.
- Tuncay IO, Parmalee NL, Khalil R, Kaur K, Kumar A, Jimale M, et al. Analysis of recent shared ancestry in a familial cohort identifies coding and noncoding autism spectrum disorder variants. *NPJ Genom Med*. 2022;7(1):13.
- Rodin RE, Dou Y, Kwon M, Sherman MA, D'Gama AM, Doan RN, et al. The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. *Nat Neurosci*. 2021;24(2):176–85.
- Callaghan DB, Rogic S, Tan PPC, Calli K, Qiao Y, Baldwin R, et al. Whole genome sequencing and variant discovery in the ASPIRE autism spectrum disorder cohort. *Clin Genet*. 2019;96(3):199–206.
- Liu Y, Chang X, Qu H, Glessner J, Tian L, Li D, et al. Non-coding structural variation differentially impacts attention-deficit hyperactivity disorder (ADHD) gene networks in African American vs Caucasian children. *Sci Rep*. 2020;10(1):15252.
- Ruzzo EK, Pérez-Cano L, Jung J-Y, Wang L-K, Kashef-Haghighi D, Hartl C, et al. Inherited and de novo genetic risk for autism impacts shared networks. *Cell*. 2019;178(4):850–66.e26.
- Werling DM, Brand H, An J-Y, Stone MR, Zhu L, Glessner JT, et al. An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat Genet*. 2018;50(5):727–36.
- Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC, et al. Genomic patterns of de novo mutation in simplex autism. *Cell*. 2017;171(3):710–22.e12.
- di Julio J, Bartha I, Wong EHM, Yu H-C, Lavrenko V, Yang D, et al. The human noncoding genome defined by genetic diversity. *Nat Genet*. 2018;50(3):333–7.
- Nord AS, West AE. Neurobiological functions of transcriptional enhancers. *Nat Neurosci*. 2019;23(1):5–14.
- Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*. 2015;16(4):197–212.

15. Li MJ, Yan B, Sham PC, Wang J. Exploring the function of genetic variants in the non-coding genomic regions: approaches for identifying human regulatory variants affecting gene expression. *Brief Bioinform.* 2015;16(3):393–412.
16. Abell NS, DeGorter MK, Gloudehans M, Greenwald E, Smith KS, He Z, et al. Multiple causal variants underlie genetic associations in humans. *Science.* 2022;375(6586):1247–54.
17. Gillmore JD, Gane E, Taubel J, Kao J, Fontana M, Maitland ML, et al. CRISPR-Cas9 In vivo gene editing for transthyretin amyloidosis. *N Engl J Med.* 2021;385(6):493–502.
18. Wolter JM, Mao H, Fragola G, Simon JM, Krantz JL, Bazick HO, et al. Cas9 gene therapy for Angelman syndrome traps Ube3a-ATS long non-coding RNA. *Nature.* 2020;587(7833):281–4.
19. Matoba N, Liang D, Sun H, Aygün N, McAfee JC, Davis JE, et al. Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Transl Psychiatry.* 2020;10(1):265.
20. Hormozdiari F, Kostem E, Kang EY, Pasianic B, Eskin E. Identifying causal variants at loci with multiple signals of association. *Genetics.* 2014;198(2):497–508.
21. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FIN-EMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics.* 2016;32(10):1493–501.
22. Pasianic B, Price AL. Dissecting the genetics of complex traits using summary association statistics. *Nat Rev Genet.* 2017;18(2):117–27.
23. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet.* 2018;19(8):491–504.
24. Mah W, Won H. The three-dimensional landscape of the genome in human brain tissue unveils regulatory mechanisms leading to schizophrenia risk. *Schizophr Res.* 2020;217:17–25.
25. Liang D, Elwell AL, Aygün N, Krupa O, Wolter JM, Kyere FA, et al. Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Nat Neurosci.* 2021;24(7):941–53.
26. Zhang S, Zhang H, Zhou Y, Qiao M, Zhao S, Kozlova A, et al. Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science.* 2020;369(6503):561–5.
27. Carter M, Shieh J. Chapter 15 - Biochemical assays and intracellular signaling. In: Carter M, Shieh J, editors. *Guide to Research Techniques in Neuroscience.* 2nd ed. San Diego: Academic Press; 2015. p. 311–43.
28. Cavalli M, Baltzer N, Umer HM, Grau J, Lemnian I, Pan G, et al. Allele specific chromatin signals, 3D interactions, and motif predictions for immune and B cell related diseases. *Sci Rep.* 2019;9(1):2695.
29. Rohde PD, Østergaard S, Kristensen TN, Sørensen P, Loeschcke V, Mackay TFC, et al. Functional validation of candidate genes detected by genomic feature models. *G3.* 2018;8(5):1659–68.
30. 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(1):2718.
31. Castaldi PJ, Guo F, Qiao D, Du F, Naing ZCC, Li Y, et al. Identification of functional variants in the FAM13A chronic obstructive pulmonary disease genome-wide association study locus by massively parallel reporter assays. *Am J Respir Crit Care Med.* 2019;199(1):52–61.
32. Mulvey B, Dougherty JD. Transcriptional-regulatory convergence across functional MDD risk variants identified by massively parallel reporter assays. *Transl Psychiatry.* 2021;11(1):403.
33. Movva R, Greenside P, Marinov GK, Nair S, Shrikumar A, Kundaje A. Deciphering regulatory DNA sequences and noncoding genetic variants using neural network models of massively parallel reporter assays. *PLoS One.* 2019;14(6):e0218073.
34. 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 B Neuropsychiatr Genet.* 2020;183(1):61–73.
35. Ashuach T, Fischer DS, Kreimer A, Ahituv N, Theis FJ, Yosef N. MPRAnalyze: statistical framework for massively parallel reporter assays. *Genome Biol.* 2019;20(1):183.
36. Klein JC, Agarwal V, Inoue F, Keith A, Martin B, Kircher M, et al. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat Methods.* 2020;17(11):1083–91.
37. Neumayr C, Pagani M, Stark A, Arnold CD. STARR-seq and UMI-STARR-seq: assessing enhancer activities for genome-wide-, high-, and low-complexity candidate libraries. *Curr Protoc Mol Biol.* 2019;128(1):e105.
38. Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houlston S, Lemak A, et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science.* 2017;357(6347):168–75.
39. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, et al. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol.* 2014;32(6):562–8.
40. 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(12):1173–5.
41. Calderon D, Ellis A, Daza RM, Martin B, Tome JM, Chen W, et al. TransM-PRA: a framework for assaying the role of many trans-acting factors at many enhancers bioRxiv. 2020;p.2020.09.30.321323.
42. Doni Jayavelu N, Jajodia A, Mishra A, Hawkins RD. Candidate silencer elements for the human and mouse genomes. *Nat Commun.* 2020;11(1):1061.
43. Cheung R, Insigne KD, Yao D, Burghard CP, Jones EM, Goodman DB, et al. A multiplexed assay for exon recognition reveals that an unappreciated fraction of rare genetic variants cause large-effect disruptions to splicing. *Mol Cell.* 2019;73(1):183–194.e8.
44. Safra M, Nir R, Farouq D, Vainberg Slutskiy I, Schwartz S. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res.* 2017;27(3):393–406.
45. Shukla CJ, McCorkindale AL, Gerhardinger C, Korthauer KD, Cabili MN, Shechner DM, et al. High-throughput identification of RNA nuclear enrichment sequences. *EMBO J.* 2018;37(6):e98452.
46. de Boer CG, Vaishnav ED, Sadeh R, Abeyta EL, Friedman N, Regev A. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat Biotechnol.* 2020;38(1):56–65.
47. Vogel Ciernia A, Laufer BI, Hwang H, Dunaway KW, Mordaunt CE, Coulson RL, et al. Epigenomic convergence of neural-immune risk factors in neurodevelopmental disorder cortex. *Cereb Cortex.* 2020;30(2):640–55.
48. An JY, Lin K, Zhu L, Werling DM, Dong S, Brand H, et al. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science.* 2018;362(6420):eaat6576.
49. Gasperini M, Tome JM, Shendure J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet.* 2020;21(5):292–310.
50. D'haene E, Vergult S. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genet Med.* 2021;23(1):34–46.
51. Inoue F, Kreimer A, Ashuach T, Ahituv N, Yosef N. Identification and massively parallel characterization of regulatory elements driving neural induction. *Cell Stem Cell.* 2019;25(5):713–27.e10.
52. Inoue F, Ahituv N. Decoding enhancers using massively parallel reporter assays. *Genomics.* 2015;106(3):159–64.
53. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet.* 2009;10(10):669–80.
54. Joung J, Konermann S, Gootenberg JS, Abudayyeh OO, Platt RJ, Brigham MD, et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc.* 2017;12(4):828–63.
55. Zhang H, Zhang Y, Zhou X, Wright S, Hyle J, Zhao L, et al. Functional interrogation of HOXA9 regulome in MLLr leukemia via reporter-based CRISPR/Cas9 screen. *Elife.* 2020;9:e57858.
56. Muerdter F, Boryń LM, Arnold CD. STARR-seq - principles and applications. *Genomics.* 2015;106(3):145–50.
57. Lee D, Shi M, Moran J, Wall M, Zhang J, Liu J, et al. STARRPeaker: uniform processing and accurate identification of STARR-seq active regions. *Genome Biol.* 2020;21(1):298.
58. Peterman N, Levine E. Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations. *BMC Genomics.* 2016;9(17):206.
59. Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell.* 2016;166(5):1147–62.e15.
60. Sey NYA, 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(4):583–93.

61. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362(6420):eaat8127.
62. Anderson SL, Coli R, Daly IW, Kichula EA, Rork MJ, Volpi SA, et al. Familial dysautonomia is caused by mutations of the IKAP gene. *Am J Hum Genet*. 2001;68(3):753–8.
63. Slaugenhaupt SA, Blumenfeld A, Gill SP, Leyne M, Mull J, Cuajungco MP, et al. Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am J Hum Genet*. 2001;68(3):598–605.
64. Dietrich P, Dragatsis I. Familial dysautonomia: mechanisms and models. *Genet Mol Biol*. 2016;39(4):497–514.
65. Lee G, Papapetrou EP, Kim H, Chambers SM, Tomishima MJ, Fasano CA, et al. Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature*. 2009;461(7262):402–6.
66. Lee G, Ramirez CN, Kim H, Zeltner N, Liu B, Radu C, et al. Large-scale screening using familial dysautonomia induced pluripotent stem cells identifies compounds that rescue IKBKAP expression. *Nat Biotechnol*. 2012;30(12):1244–8.
67. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem*. 2010;79:321–49.
68. Jung Y, Goldman D. Role of RNA modifications in brain and behavior. *Genes Brain Behav*. 2018;17(3):e12444.
69. Tran SS, Jun H-H, Bahn JH, Azghadi A, Ramaswami G, Van Nostrand EL, et al. Widespread RNA editing dysregulation in brains from autistic individuals. *Nat Neurosci*. 2019;22(1):25–36.
70. Charette M, Gray MW. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life*. 2000;49(5):341–51.
71. Mayr C. What are 3' UTRs doing? *Cold Spring Harb Perspect Biol*. 2019;11(10):a034728.
72. Araujo PR, Yoon K, Ko D, Smith AD, Qiao M, Suresh U, et al. Before it gets started: regulating translation at the 5' UTR. *Comp Funct Genomics*. 2012;28(2012):475731.
73. Sample PJ, Wang B, Reid DW, Presnyak V, McFadyen JJ, Morris DR, et al. Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat Biotechnol*. 2019;37(7):803–9.
74. Lagunas T, Plassmeyer SP, Friedman RZ, Rieger MA, Fischer AD, Aguilar Lucero AF, et al. A Cre-dependent massively parallel reporter assay allows for cell-type specific assessment of the functional effects of genetic variants in vivo. *bioRxiv*. 2021;p.2021.05.17.444514.
75. Griesemer D, Xue JR, Reilly SK, Ullrich JC, Kukreja K, Davis JR, et al. Genome-wide functional screen of 3'UTR variants uncovers causal variants for human disease and evolution. *Cell*. 2021;184(20):5247–5260. e19.
76. Wanke KA, Devanna P, Vernes SC. Understanding neurodevelopmental disorders: the promise of regulatory variation in the 3'UTR. *Biol Psychiatry*. 2018;83(7):548–57.
77. Göpferich M, George NO, Muelas AD, Bizyn A. Single cell 3'UTR analysis identifies changes in alternative polyadenylation throughout neuronal differentiation and in autism. *bioRxiv*. 2020;p.2020.08.12.247627.
78. Vaishnavi V, Manikandan M, Munirajan AK. Mining the 3' UTR of autism-implicated genes for SNPs perturbing microRNA regulation. *Genomics Proteomics Bioinformatics*. 2014;12(2):92–104.
79. Zhang S-F, Gao J, Liu C-M. The role of non-coding RNAs in neurodevelopmental disorders. *Front Genet*. 2019;20(10):1033.
80. Van De Vondervoort I, Gordebeke P, Khoshab N, Tiesinga P, Buitelaar J, Kozicz T, et al. Long non-coding RNAs in neurodevelopmental disorders. *Front Mol Neurosci*. 2013;6:53.
81. Ang CE, Ma Q, Wapinski OL, Fan S, Flynn RA, Lee QY, et al. The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *Elife*. 2019;8:e41770.
82. Li L, Zhuang Y, Zhao X, Li X. Long non-coding RNA in neuronal development and neurological disorders. *Front Genet*. 2018;9:744.
83. Wang Y, Zhao X, Ju W, Flory M, Zhong J, Jiang S, et al. Genome-wide differential expression of synaptic long noncoding RNAs in autism spectrum disorder. *Transl Psychiatry*. 2015;20(5):e660.
84. Wilkinson B, Campbell DB. Contribution of long noncoding RNAs to autism spectrum disorder risk. *Int Rev Neurobiol*. 2013;113:35–59.
85. Liu Y, Chang X, Hahn C-G, Gur RE, Sleiman PAM, Hakonarson H. Non-coding RNA dysregulation in the amygdala region of schizophrenia patients contributes to the pathogenesis of the disease. *Transl Psychiatry*. 2018;8(1):44.
86. Petazzi P, Sandoval J, Szczesna K, Jorge OC, Roa L, Sayols S, et al. Dysregulation of the long non-coding RNA transcriptome in a Rett syndrome mouse model. *RNA Biol*. 2013;10(7):1197–203.
87. Vieira AS, Dogini DB, Lopes-Cendes I. Role of non-coding RNAs in non-aging-related neurological disorders. *Braz J Med Biol Res*. 2018;51(8):e7566.
88. Ziats MN, Rennert OM. Aberrant expression of long noncoding RNAs in autistic brain. *J Mol Neurosci*. 2013;49(3):589–93.
89. Meng Q, Wang K, Brunetti T, Xia Y, Jiao C, Dai R, et al. The DGCR5 long noncoding RNA may regulate expression of several schizophrenia-related genes. *Sci Transl Med*. 2018;10(472):eaat6912.
90. Gudenan BL, Srivastava AK, Wang L. Integrative genomic analyses for identification and prioritization of long non-coding RNAs associated with autism. *PLoS One*. 2017;12(5):e0178532.
91. Zhang S, You L, Xu Q, Ou J, Wu D, Yuan X, et al. Distinct long non-coding RNA and mRNA expression profiles in the hippocampus of an attention deficit hyperactivity disorder model in spontaneously hypertensive rats and control Wistar Kyoto rats. *Brain Res Bull*. 2020;161:177–96.
92. Kahaei MS, Ghafouri-Fard S, Namvar A, Omrani MD, Sayad A, Taheri M. Association study of a single nucleotide polymorphism in brain cytoplasmic 200 long-noncoding RNA and psychiatric disorders. *Metab Brain Dis*. 2020;35(7):1095–100.
93. Bhattacharyya N, Pandey V, Bhattacharyya M, Dey A. Regulatory role of long non coding RNAs (lncRNAs) in neurological disorders: from novel biomarkers to promising therapeutic strategies. *Asian J Pharm Sci*. 2021;16(5):533–50.
94. Pratt BM, Won H. Advances in profiling chromatin architecture shed light on the regulatory dynamics underlying brain disorders. *Semin Cell Dev Biol*. 2022;121:153–60.
95. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*. 2016;167(7):1853–66.e17.
96. Gordon MG, Inoue F, Martin B, Schubach M, Agarwal V, Whalen S, et al. lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. *Nat Protoc*. 2020;15(8):2387–412.
97. Maricque BB, Chaudhari HG, Cohen BA. A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. *Nat Biotechnol*. 2018;10:1038/nbt.4285.
98. Hu B, Won H, Mah W, Park RB, Kassim B, Spiess K, et al. Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat Commun*. 2021;12(1):3968.
99. Consortium, Epigenomics R, Kundaje A, Meuleman W, Ernst J, Bilenky M, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317–30.
100. Jerber J, Seaton DD, Cuomo ASE, Kumasaka N, Haldane J, Steer J, et al. Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat Genet*. 2021;53(3):304–12.
101. Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science*. 2019;366(6469):1134–9.
102. Maricque BB, Dougherty JD, Cohen BA. A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of cis-regulatory activity in neural cells. *Nucleic Acids Res*. 2017;45(4):e16.
103. Keilwagen J, Posch S, Grau J. Accurate prediction of cell type-specific transcription factor binding. *Genome Biol*. 2019;20(1):9.
104. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*. 2018;362(6420):eaat7615.
105. de la Torre-Ubieta L, Stein JLL, Won H, Opland CKK, Liang D, Lu D, et al. The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell*. 2018;172(1–2):289–304.
106. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HA, et al. Genetic identification of brain cell types underlying schizophrenia. *Nat Genet*. 2018;50(6):825–33.
107. Spiess K, Won H. Regulatory landscape in brain development and disease. *Curr Opin Genet Dev*. 2020;18(65):53–60.

108. Silbereis JC, Pochareddy S, Zhu Y, Li M, Sestan N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron*. 2016;89(2):248–68.
109. Nguyen TA, Jones RD, Snavelly AR, Pfenning AR, Kirchner R, Hemberg M, et al. High-throughput functional comparison of promoter and enhancer activities. *Genome Res*. 2016;26(8):1023–33.
110. Courchesne E, Pramparo T, Gazestani VH, Lombardo MV, Pierce K, Lewis NE. The ASD living biology: from cell proliferation to clinical phenotype. *Mol Psychiatry*. 2018;24(1):88–107.
111. Marchetto MC, Belinson H, Tian Y, Freitas BC, Fu C, Vadodaria K, et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol Psychiatry*. 2017;22(6):820–35.
112. Zoëga H, Furu K, Halldórsson M, Thomsen PH, Sourander A, Martikainen JE. Use of ADHD drugs in the Nordic countries: a population-based comparison study. *Acta Psychiatr Scand*. 2011;123(5):360–7.
113. Christensen J, Grønberg TK, Sørensen MJ, Schendel D, Parner ET, Pedersen LH, et al. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA*. 2013;309(16):1696–703.
114. Roulet FI, Lai JKY, Foster JA. In utero exposure to valproic acid and autism—a current review of clinical and animal studies. *Neurotoxicol Teratol*. 2013;36:47–56.
115. Cohen MJ, Meador KJ, Browning N, May R, Baker GA, Clayton-Smith J, et al. Fetal antiepileptic drug exposure: adaptive and emotional/behavioral functioning at age 6 years. *Epilepsy Behav*. 2013;29(2):308–15.
116. Jentink J, Loane MA, Dolk H, Barisic I, Garne E, Morris JK, et al. Valproic acid monotherapy in pregnancy and major congenital malformations. *N Engl J Med*. 2010;362(23):2185–93.
117. Fujimura K, Mitsuhashi T, Shibata S, Shimoizato S, Takahashi T. In utero exposure to valproic acid induces neocortical dysgenesis via dysregulation of neural progenitor cell proliferation/differentiation. *J Neurosci*. 2016;36(42):10908–19.
118. Sawada K, Kamiya S, Aoki I. Neonatal valproic acid exposure produces altered gyrification related to increased parvalbumin-immunopositive neuron density with thickened sulcal floors. *PLoS One*. 2021;16(4):e0250262.
119. Zhao H, Wang Q, Yan T, Zhang Y, Xu H-J, Yu H-P, et al. Maternal valproic acid exposure leads to neurogenesis defects and autism-like behaviors in non-human primates. *Transl Psychiatry*. 2019;9(1):267.
120. Modabbernia A, Velthorst E, Reichenberg A. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. *Mol Autism*. 2017;17(8):13.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

