bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CRISPR tiling deletion screens reveal functional enhancers of neuropsychiatric risk genes and allelic compensation effects (ACE) on transcription

Authors: Xingjie Ren[1], Lina Zheng[2], Lenka Maliskova[1], Tsz Wai Tam[1], Yifan Sun[1], Hongjiang Liu[1], Jerry Lee[1], Maya Asami Takagi[1], Bin Li[3], Bing Ren[3,4,5], Wei Wang[2,3,6], Yin Shen[1,7,8,#]

Affiliations

[1]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA.

[2]Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA.

[3]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA.

[4]Center for Epigenomics, University of California, San Diego, La Jolla, CA, USA.

[5]Moores Cancer Center, University of California, San Diego, La Jolla, CA, USA.

[6]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA.

[7]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA

[8]Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA

[#]Corresponding author: Yin Shen, yin.shen@ucsf.edu

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Precise transcriptional regulation is critical for cellular function and development, yet the mechanism of this process remains poorly understood for many genes. To gain a deeper understanding of the regulation of neuropsychiatric disease risk genes, we identified a total of 39 functional enhancers for four dosage-sensitive genes, *APP*, *FMR1*, *MECP2*, and *SIN3A*, using CRISPR tiling deletion screening in human induced pluripotent stem cell (iPSC)-induced excitatory neurons. We found that enhancer annotation provides potential pathological insights into disease-associated copy number variants. More importantly, we discovered that allelic enhancer deletions at *SIN3A* could be compensated by increased transcriptional activities from the other intact allele. Such allelic compensation effects (ACE) on transcription is stably maintained during differentiation and, once established, cannot be reversed by ectopic *SIN3A* expression. Further, ACE at *SIN3A* occurs through dosage sensing by the promoter. Together, our findings unravel a regulatory compensation mechanism that ensures stable and precise transcriptional output for *SIN3A*, and potentially other dosage-sensitive genes.

Main

Optimal spatial-temporal gene regulation is pivotal to normal development. Mutations in *cis*-regulatory elements (CREs), such as enhancers, cause target gene misregulation and contribute to diseases[1,2]. To date, over one million candidate CREs (cCREs) have been mapped in the human genome based on biochemical signatures, including chromatin accessibility, histone modifications, and transcription factor (TF) binding sites[3,4]. cCREs are also enriched for variants identified by genome-wide association studies (GWAS) for complex diseases, signifying their potential contribution to human diseases through gene regulatory mechanisms[3]. However, how cCREs regulate target gene expression remains mostly uncharacterized.

Genetic analyses have identified numerous neuropsychiatric risk genes, many of which are dosage-sensitive genes[5], suggesting that precise regulation of gene expression is critical for maintaining normal neuronal function and preventing disease. For example, mutations and duplication in *APP*, a precursor protein of β-amyloid peptide[6] are causal factors in Alzheimer's disease[7]. Elevated *FMR1* transcription of *FMR1* premutations (55-200 CGG repeats at the 5' untranslated region) increases the risk of developing fragile X-associated tremor/ataxia syndrome (FXTAS), fragile X-associated primary ovarian insufficiency (FXPOI), and fragile X-associated neuropsychiatric disorders (FXAND), while full mutations of *FMR1* (>200 CGG repeats) completely inhibit *FMR1* transcription resulting in fragile X syndrome[8]. In another example of MeCP2, a methyl-CpG-binding protein[9], loss-of-function mutations in *MECP2* lead to Rett syndrome[10], and duplication of *MECP2* causes a neurodevelopmental disorder, *MECP2* duplication syndrome[11]. Finally, heterozygous loss-of-function variants in *SIN3A,* a transcriptional repressor[12], cause *SIN3A* haploinsufficiency, giving rise to neurodevelopmental syndromes including Witteveen-Kolk

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

syndrome and Autism Spectrum Disorder[13,14]. These observations of disease conditions resulting from gene dosage alterations underscore the essential role of regulatory mechanisms in safeguarding the genome against deleterious mutations, thereby preventing pathological shifts in gene expression.

To better understand the gene regulatory program for those dosage-sensitive genes, we performed unbiased CRISPR tilling deletion screening of enhancers for *APP*, *FMR1*, *MECP2,* and *SIN3A* using CREST-seq (for *cis*-regulatory element scan by tiling-deletion and sequencing)[15] during the differentiation of human induced pluripotent stem cell (iPSC) into excitatory neurons. Through extensive validation, we uncovered an unexpected transcriptional compensation mechanism that maintains the stable transcriptional output of *SIN3A* upon allelic enhancer deletions.

Results

Allelic tiling deletion CRISPR screens identify enhancers for neuropsychiatric risk genes

To identify functional enhancers for *APP*, *FMR1*, *MECP2,* and *SIN3A* genes in neurons (**Extended Data Fig. 1a,b**), we performed CREST-seq[15] for unbiased tiling deletion CRISPR screening of genomic sequences surrounding the gene of choice. These genes are strategically chosen due to their importance in both developmental and disease perspectives, as well as their involvement in pathogenesis linked to gene dosage alterations. Specifically, we generated allelically tagged EGFP or mCherry reporters in the WTC11 i^{3}N iPSC line[16] to monitor allelic gene expression during the CRISPR screens using fluorescence-activated cell sorting (FACS) (**Fig. 1a**). The WTC11 i^{3}N iPSC line contains the integrated doxycycline-inducible *Ngn2* at the *AAVS1* locus, which allows us to generate a large quantity of homogeneous excitatory neurons[16] (**Fig. 1a and Extended Data Fig. 1c**). For *APP* and *SIN3A*, EGFP and mCherry are tagged on each allele, and for X-linked *FMR1* and *MECP2*, we tagged them with either a mCherry or an EGFP reporter, respectively (**Fig. 1a, Extended Data Fig. 1a**). We designed approximately 11,000 to 17,000 paired-guide RNAs (pgRNAs) targeting 2-4 Mbp around each gene. pgRNAs mediated deletions had an average size of 2,000 to 3,500 bp and 15x or 20x coverage for each nucleotide (**Extended Data Fig. 2a-d**). We infected each iPSC reporter line with the corresponding lentivirus library expressing SpCas9 protein and pgRNAs, selected infected cells with puromycin for one week, and then differentiated iPSCs into excitatory neurons (**Fig. 1a**). 2 weeks after differentiation we sorted out neurons with reduced reporter expression using FACS (**Extended Data Fig. 3a**). To assess the screening strategy, we quantified the frequency of pgRNAs in each sample and calculated the fold change in pgRNA counts between FACS-sorted cells and control cells. As expected, positive control pgRNAs targeting EGFP and mCherry were significantly enriched in FACS-sorted populations with reduced reporter expression, whereas non-targeting negative control pgRNAs showed no enrichment, validating our screening strategy (**Fig. 1b**).

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We identified 39 enhancers for 4 genes using RELICS[17] (**Extended Data Fig. 3b and Supplementary Table 1**). On average, these functional enhancers are 315.3 kb away from the transcriptional start sites (TSSs) of their target genes, with 16 enhancers located within their target gene bodies (**Fig. 1c**). As anticipated, none of the identified enhancers overlap with the repressive chromatin marker H3K9me3 (**Fig. 1d**). 71.8% (28/39) enhancers overlap with active chromatin signatures profiled in WTC11 i^{3}N iPSC-derived excitatory neurons, including chromatin accessibility[18], H3K4me1, H3K4me3, H3K27ac, H3K36me3, and the binding of CTCF and RNA polymerase II[19], or cCREs annotated in excitatory neurons from the human brain samples[20] (**Fig. 1d**). Notably, 28.2% (11/39) of enhancers are not associated with the chromatin signatures of enhancers we examined. This is consistent with reports of the existence of hidden enhancers that do not have conventional chromatin marks for cCRE[21–23]. Interestingly, only 41.0% (16/39) of enhancers participate in H3K4me3 associated chromatin interactions[18] (**Fig. 1e)**, confirming the notion that while chromatin interactions are valuable for delineating enhancer-promoter relationships, other mechanisms also play a role in enhancer-mediated transcriptional regulation[24,25].

Functional validation of CREST-seq identified enhancers

We focused on validating enhancers located in gene bodies by examining their effects on target gene expression through CRISPR deletion followed by flow cytometry analysis (**Extended Data Fig. 4a**). For *FMR1*, deleting one enhancer (FMR1-E1, located in the first intron of *FMR1*) reduced expression of *FMR1-mCherry* in both iPSCs and excitatory neurons (**Fig. 2a,b and Extended Data Fig. 4b-d**). For MECP2, deleting three MECP2 gene body enhancers, MECP2-E3, MECP2-E8, and MECP2-E10, led to the downregulation of *MECP2-EGFP* in both iPSCs and excitatory neurons, while deleting MECP2-E6 caused downregulation of *MECP2-EGFP* only in excitatory neurons (**Fig. 2c and Extended Data Fig. 5a-c**), suggesting MECP2-E6 is a neuron-specific enhancer. The dependence of *MECP2* for the three shared enhancers was further confirmed with independent enhancer deletion clones (**Fig. 2d**). The reduction of *MECP2* transcription was more profound in clones with deletions of MECP2-promoter, MECP2-E8, and MECP2-E10 compared to MECP2-E6 (**Fig. 2e**), suggesting varied effects of enhancers on *MECP2* expression. Deleting APP-E3, located in the last intron of *APP*, led to a similar downregulation of *APP* as deleting the *APP* promoter in both iPSCs and excitatory neurons (**Fig. 2f and Extended Data Fig. 6a,b**).

In addition to gene body enhancers, we validated a distal enhancer, SIN3A-E4, for *SIN3A*. After Cas9 and pgRNA delivery, a subpopulation of cells exhibited significant downregulation of *SIN3A-EGFP* or *SIN3A-mCherry* in both iPSCs and 2-week excitatory neurons, confirming that SIN3A-E4 is a functional enhancer of *SIN3A* (**Fig. 2g**). As expected, we only observed the deletion of SIN3A-E4 on one of the two alleles consistent with the fact that *SIN3A* is a haploinsufficient gene[26] and an essential gene in neurons[27] (**Extended Data Fig. 7a-d**). Cells with further perturbation of the 19bp CTCF motif in SIN3A-E4 exhibited

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reduced *SIN3A-EGFP* or *SIN3A-mCherry* expression (**Fig. 2h,i**). Genotyping of cells with reduced *SIN3A-EGFP* or *SIN3A-mCherry* expression revealed various deletions, insertions, and substitutions at the CTCF motif (**Fig. 2j and Extended Data Fig. 8a**), confirming the importance of the CTCF binding motif in the SIN3A-E4 enhancer.

Enhancer annotation offers functional evidence for clinical copy number variants

21,217 clinical variants in ClinVar are copy number variants (CNVs) with only a few CNVs having experimental-based evidence of functional consequences (**Extended Data Fig. 9a**). Nearly a third of CNVs lack functional annotation and are classified as Variants of Uncertain Significance (VUS). Interestingly, VUS or other classifications are enriched for cCREs compared to pathogenic/likely pathogenic CNVs (**Extended Data Fig. 9b,c**) suggesting that VUS may contribute to human diseases by disrupting gene regulation. Indeed, we observed that several CNVs overlap with *SIN3A* and *MECP2* enhancers. This observation offers a potential functional interpretation for disease-associated CNVs, highlighting their role in regulating gene dosage (**Extended Data Fig. 9d,e**). To explore the potential regulatory function of CNVs, we used a hypergeometric test to assess the enrichment of 4,014 CNVs with lengths of 50bp to 5kb in distal cCREs identified in 222 distinct human cell types[28] and found cell type-selective significant enrichment of CNVs associated with 355 human diseases at cCREs of 218 cell types ($P < 0.05$) (**Fig. 2k and Supplementary Table 2**). For example, cCREs of melanocyte, oligo precursor, oligodendrocyte, and Schwann cells are enriched for CNVs in Rett syndrome patients, while cCREs of ventricular cardiomyocytes are enriched for CNVs in patients with hypertrophic cardiomyopathy. These findings suggest the involvement of the regulatory function of disease-associated CNVs in human diseases.

Allelic deletion of *SIN3A* enhancer triggers allelic compensation effects (ACE)

The dual reporter tagging of *SIN3A* enabled us to monitor the allelic *SIN3A* transcription followed by enhancer deletions. Remarkably, cells with reduced expression of *SIN3A-EGFP* exhibited increased expression of *SIN3A-mCherry*, and vice versa upon deleting the SIN3A-E4 enhancer (**Fig. 2g**), suggesting that enhancer deletion on one allele induced allelic compensation effects (ACE) from the other allele. As *SIN3A* is a haploinsufficient gene, we hypothesize that allelic enhancer perturbation may trigger ACE to maintain a steady level of transcriptional output, which may serve as a crucial genome defense mechanism against deleterious non-coding mutations affecting *SIN3A* expression. To examine whether other enhancer deletions could similarly trigger ACE, we deleted another three *SIN3A* enhancers located in various genomic regions, SIN3A-E2 (*CYP1A1* intron), SIN3A-E3 (*CYP1A2* exons, *CYP1A2* is not expressed in neurons with RPKM = 0), and SIN3A-E5 (non-coding intergenic regions) (**Fig. 3a**). After the delivery of Cas9 and pgRNAs for deleting these enhancers, cells exhibited significant downregulation of either *SIN3A-EGFP* or *SIN3A-mCherry* expression, but elevated reporter expression on the other allele

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in both iPSCs and 2-week excitatory neurons (**Fig. 3b,c and Extended Data Fig. 7a-d**), confirming that ACE is a general mechanism of *SIN3A* transcriptional regulation.

Bona fide enhancers only affect transcription in *cis*. To ensure observed allelic gene expression changes are due to enhancer deletion in *cis*, we picked two phased SNPs in the WTC11 genome. The first SNP is located in the last intron of *SIN3A* (chr15: 75374632, C/T, hg38), which was used for resolving the allelic information of tagged EGFP and mCherry reporters. The second SNP is located adjacent to SIN3A-E2 (chr15: 74721849, T/G, hg38), which was used for the identification of the allele with the enhancer deletion. Our results showed that cells with allelic enhancer deletions have reduced SIN3A-EGFP or SIN3A-mCherry expression from the same allele (**Extended Data Fig. 10a**). Therefore, we demonstrate that the ACE arises from the opposite allele, compensating for reduced *SIN3A* transcription caused by the enhancer deletion in *cis*.

Enhancer deletion-induced ACE can also be further confirmed with allelic gene expression analysis leveraging an SNP in the *SIN3A* intron in the WTC11 iPSC genome (**Fig. 3d,** chr15: 75374632, C/T, hg38). In wild-type clones (G+M+), we observed a near 1:1 expression ratio from both alleles. However, clones with allelic enhancer deletion with either reduced EGFP expression (G-M+) or reduced mCherry expression (G+M-) exhibit dominant expression from either the C allele or the T allele, respectively, in both iPSCs and 2-week excitatory neurons (**Fig. 3e**). More importantly, the total *SIN3A* mRNA level remains no changes across all the clones (**Fig. 3f**), suggesting that ACE is used to achieve the precise transcriptional output of *SIN3A*.

To explore the mechanism of ACE in response to enhancer deletions, we performed a time course analysis of allelic expression changes upon deleting one enhancer (SIN3A-E4) and compared that to deleting *SIN3A* promoter in iPSCs. Cells with the reduced SIN3A-EGFP or SIN3A-mCherry signals appeared two days after the delivery of Cas9 and pgRNAs (**Fig. 4a**). To track the ACE, we quantified the SIN3A-EGFP and SIN3A-mCherry signals in cells with either the enhancer or the promoter deletion. In cells with the SIN3A-E4 enhancer deletion, the downregulation of either *SIN3A-EGFP* or *SIN3A-mCherry* allele is positively correlated with the upregulation of the other allele over time (**Fig. 4b,c**, $R^2 = 0.92$ for the EGFP allele, $R^2 = 0.92$ for the mCherry allele). In contrast, we only observed the downregulation of either *SIN3A-EGFP* or *SIN3A-mCherry* allele in cells with the promoter deletion without apparent ACE from the opposite allele (**Fig. 4b,c**, $R^2 = 0.027$ for the EGFP allele, $R^2 = 0.08$ for the mCherry allele). To check the kinetics of ACE from the enhancer deletion, we calculated the slope between each pair of adjacent time points. The absolute slope value exceeded one after day 5, reached the summit at day 10, and dropped quickly at the end (**Fig. 4d**). The observed dynamic rate of ACE after enhancer deletion suggests that ACE is more potent as *SIN3A* expression approaches the level that triggers

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

haploinsufficiency after day 5. In addition, the ACE rate decreases as the total *SIN3A* expression level approaches the wild-type level. In contrast, promoter deletion-induced SIN3A downregulation remained constant after day 5 (**Fig. 4b,c**). Long-read RNA-seq data revealed *SIN3A* transcription from two TSSs[29], and we only deleted the promoter of the major *SIN3A* transcript (**Extended Data Fig. 11c,d**). Thus, the partial reduction of *SIN3A* expression from the promoter deletion allele may not be sufficient to induce ACE. These results demonstrate that ACE is a dynamic process initiated from significantly reduced expression of *SIN3A* from one allele.

To explore whether the established ACE persists during neuronal differentiation, we isolated single clones either with no enhancer deletion (SIN3A-EGFP+/SIN3A-mCherry+: G+M+), or with allelic enhancer deletions (SIN3A-EGFP-/SIN3A-mCherry+: G-M+; SIN3A-EGFP+/SIN3A-mCherry-: G+M-). We observed that transcriptional compensation remains unchanged after differentiating iPSCs into excitatory neurons (**Fig. 4e**), suggesting that the ACE of *SIN3A*, once established, can be heritably maintained throughout the differentiation.

Since ACE is triggered by allelic enhancer deletion-induced *SIN3A* downregulation, we wondered whether it can be reversed by elevating *SIN3A* expression. To test this, we ectopically expressed a *SIN3A* transgene driven by the *SIN3A* promoter, which resulted in about 1.7-fold expression of *SIN3A* compared to the endogenous expression level (**Extended Data Fig. 11a,b**). However, *SIN3A* overexpression is not sufficient for disrupting endogenous transcriptional compensation (**Fig. 4f)**. These results suggest ACE, once established, can not be reversed by increasing *SIN3A* expression.

The *SIN3A* promoter mediates allelic enhancer deletion-induced ACE

Next, we investigate how cells can sense reduced *SIN3A* expression upon enhancer deletion and initiate the process of ACE on transcription. As a transcriptional factor, SIN3A binds to its own promoter[30], suggesting autoregulatory feedback (**Extended Data Fig. 11c**). This prompted us to consider that the *SIN3A* promoter could mediate SIN3A dosage sensing to achieve an optimal transcriptional level of the *SIN3A* gene. To test whether the promoter is responsible for initiating ACE, we tested the activities of two *SIN3A* promoter reporters (P1, P1+P2) with and without shRNA-mediated downregulation of endogenous *SIN3A* expression. First, we showed both P1 and P1+P2 promoter reporters exhibit strong EGFP expression, confirming that they are active promoters (**Fig. 5a and Extended Data Fig. 11c**). Both P1 and P1+P2 promoter reporters exhibited a significant increase of promoter activity when endogenous *SIN3A* expression is reduced by *SIN3A* shRNA (**Fig. 5b-d**). Thus, the *SIN3A* promoter can counteract allelic enhancer deletion-induced downregulation by increasing its transcriptional activity. These results suggest allelic enhancer deletion leads to near complete loss of *SIN3A* in *cis*, resulting in less *SIN3A* binding at its own promoters, which triggers ACE via the upregulating of SIN3A from the trans allele. In

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contrast, allelic partial deletion of promoter retained partial *SIN3A* expression in *cis* (**Fig. 4c**), which is not sufficient to trigger ACE (**Fig. 5e**). Our ACE model can also explain the haploinsufficiency of *SIN3A* for the Witteveen-Kolk syndrome (WITKOS) patients with large deletions of the entire *SIN3A* locus including the *SIN3A* promoter[31–33] (**Extended Data Fig. 9d)**, while copy number loss variants overlapped with *SIN3A* enhancers identified from clinical samples are likely benign. In WITKOS patients, the loss of one copy of the promoter disrupts promoter-mediated SIN3A dosage sensing, resulting in only half of the normal expression of *SIN3A* from the intact wild-type allele. This reduced level of SIN3A is insufficient to support normal cellular function, leading to haploinsufficiency in WITKOS patients.

Leveraging the feature of protein binding to their own gene promoter, we matched protein-coding promoter sequences with the known TF binding motifs database[34,35] to identify promoters that could be bound by the TFs expressed from the same promoter. In total, we identified 530 human and 321 mouse TF genes with their promoters harboring their own binding motifs (**Extended Data Fig. 12a**). Gene ontology enrichment analysis for those genes yielded terms associated with transcriptional regulation, cis-regulatory region DNA binding, and nucleus localization, consistent with their roles as TFs (**Extended Data Fig. 12b**). Considering SIN3A is a transcriptional repressor, 279 human and 180 mouse repressive TFs could be subjected to enhancer deletion-induced ACE (**Extended Data Fig. 12a**). Leveraging RNA-seq data from human tissues in GTEx[36], we found that those 279 human genes are widely expressed across human tissues (**Extended Data Fig. 12c**). Since ACE is used to maintain the steady expression of associated genes, we further checked their dosage sensitivity using the ClinGen database with dosage-sensitive information for 1,545 genes[37] and a machine learning predicted genome wide gene dosage sensitivity map[38]. Among 279 genes, 45 were found in the ClinGen database and 270 were found in the dosage sensitivity map. In both analyses, there is a significant enrichment of human candidate genes in haploinsufficiency, instead of triplosensitivity (84.4% vs. 4.4% in ClinGen, 47.7% vs. 25.2% in gene dosage sensitivity map) (**Extended Data Fig. 12d**). These candidate genes suggest that ACE is a widespread gene regulatory mechanism for dosage-sensitive genes. The genes from our prediction are TFs, which drive precise transcription patterns[39]**,** and are known to be enriched for haploinsufficient genes[40] with genetic studies highlighting the significance of their dosage for normal development[41,42].

Discussion

In this study, we identified functional enhancers for four neuropsychiatric risk genes in iPSC-derived excitatory neurons using CREST-seq. Since *APP*, *FMR1*, *MECP2,* and *SIN3A* are dosage-sensitive genes associated with neuropsychiatric diseases, discovering their enhancers in neurons may offer new genomic loci for developing therapeutic interventions aimed at correcting their transcriptional output.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Functional enhancers are located in both gene-body and distal regions, with 28.2% of them lacking active chromatin markers commonly used for annotating candidate enhancers. Similar findings were reported for enhancers identified from CRISPR screens in mouse embryonic stem cells (mESCs)[23] and H1 human embryonic stem cells (ESCs)[43], and transgenic mouse reporter assays[21]. These findings reinforce the concept of the existence of hidden enhancers that do not have typical epigenetic features and emphasize the importance of characterizing regulatory elements in an unbiased manner with functional assays. We observed that only 41% of CREST-seq identified enhancers physically interacting with their target gene promoters in neurons, which could be attributed to two possibilities. One is that mechanisms other than chromatin interactions, including RNA polymerase tracking, TFs linking, and enhancer relocation, are used for transcriptional regulation[44]. Another possibility is that our study can identify enhancers that contribute to gene expression during the differentiation process, and their interaction with the promoter occurs in cells before they differentiate into neurons.

Enhancers outnumber protein-coding promoters, highlighting the complexity of the gene regulatory program, which remains inadequately comprehended. The “enhancer" terminology encompasses a variety of different classes of enhancers with distinct functional consequences on gene regulation. For example, some enhancers are redundant and may only cause transient transcriptional disruption when deleted[45,46]. The redundancy within the enhancer program is advantageous for achieving precise and resilient gene expression. Other enhancers, such as shadow enhancers[47], exert an additive function in transcriptional output, whereby multiple enhancers collectively contribute to the desired transcriptional level of target genes[48,49]. Our study unveils an additional layer of complexity to the gene regulation program by uncovering ACE upon allelic enhancer deletion for dosage-sensitive genes.

It is crucial for diploid organisms to maintain finely tuned expression levels for dosage-sensitive genes, including TFs and haploinsufficient genes. These genes play pivotal roles in many fundamental biological processes, and any change in the transcription level of these genes or a loss-of-function mutation on one of the alleles will render them insufficient for their function. Therefore, a precision and robust transcriptional control mechanism must be established to guarantee optimal transcriptional output. Typically, multiple enhancers participate in regulating a target gene, increasing the vulnerable genomic space subjected to deleterious mutations that could adversely affect transcriptional control. We suggest that ACE is one type of genetic compensation mechanism[50,51], which serves as a defense mechanism for overcoming adverse effects caused by enhancer mutations and accounting for widespread sensitivity to TF dosages during development.

Our results demonstrate that promoter sequences play a critical role in detecting reduced gene dosage and initiating transcriptional compensation through the binding of their own protein products. Compared

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to allelic enhancer deletion, we didn't observe transcriptional compensation in the allelic deletion of the *SIN3A* promoter (**Fig. 2g, 4b, 4c**). This is possibly due to the cells with allelic promoter deletion still having sufficient SIN3A binds to the promoter on the other allele. However, the reduced SIN3A level can be directly detected in the cells with allelic enhancer deletion, as they possess two copies of the *SIN3A* promoter. Long read RNA-seq study showed *SIN3A* transcription from two TSSs[52]. The promoter region we deleted only covered the TSS with stronger transcriptional activity (**Extended Data Fig. 11c,d**). Thus, another possibility is that *SIN3A* expression level from the allele with the promoter deletion is higher than the level from the allele with enhancer deletion, possibly due to compensation of the other intact promoter, which did not reach the threshold needed for initiating transcriptional compensation. Our transcriptional compensation model offers one explanation of why disturbing enhancers for dosage-sensitive genes don't seem to affect the cellular and developmental processes. To validate the transcriptional compensation of additional candidate genes, identifying their enhancers and performing allelic gene expression analysis in cells with deletion or perturbation of one copy of enhancer is needed. Further testing of the enhancer-deletion-triggered transcriptional compensation mechanism *in vivo* will solidify our understanding of how dosage-sensitive genes achieve robust transcriptional output and normal development.

Figures:

Figure 1. Identification and analysis of enhancers of four neuropsychiatric risk genes.

Figure 2. Validating CREST-seq identified enhancers.

Figure 3. Allelic enhancer deletion induces transcriptional compensation of *SIN3A*.

Figure 4. Allelic enhancer deletion-induced allelic compensation effect (ACE) is a dynamic process.

Figure 5. The *SIN3A* promoter mediates allelic enhancer deletion-induced allelic compensation effect (ACE).

Extended Data Figure 1. Engineered reporter cell lines and gene expression.

Extended Data Figure 2. pgRNA libraries of *APP*, *FMR1*, *MECP2*, and *SIN3A*.

Extended Data Figure 3. CREST-seq screens and data analysis.

Extended Data Figure 4. Enhancer validation strategy and validation of *FMR1* enhancer.

Extended Data Figure 5. *MECP2* enhancer validations.

Extended Data Figure 6. *APP* enhancer validation.

Extended Data Figure 7. *SIN3A* enhancer validations.

Extended Data Figure 8. Editing outcomes of CTCF sgRNAs.

Extended Data Figure 9. The regulatory function of copy number variants.

Extended Data Figure 10. *cis*-regulation of *SIN3A* by the SIN3A-E2 enhancer.

Extended Data Figure 11. *SIN3A* ectopic expression and *SIN3A* promoter reporter assay.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Figure 12. Transcriptional compensation is associated with gene dosage sensitivity.

Supplementary Tables:

Supplementary Table 1. List of identified enhancers

Supplementary Table 2. Enhancer enrichment analysis of clinical copy number variants

Supplementary Table 3. DNA oligo sequences for donor cloning, RT-qPCR, genotyping and library preparation

Supplementary Table 4. sgRNA sequences for enhancer validation and generating reporter cell lines

Supplementary Table 5. shRNA sequences for *SIN3A* knockdown

Supplementary Table 6. Information of datasets used in this study

Supplementary Table 7. Candidate transcriptional compensation genes

Acknowledgments

This work was supported by the NIH grant UM1HG009402 (to Y.Shen. and B.R.). This work was made possible in part by the NIH grants P30DK063720 and S101S10OD021822-01 to the UCSF Parnassus Flow Cytometry Core. Sequencing was performed at the UCSF CAT, supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants.

Author contributions

X.R. and Y.Shen. designed the study. X.R. and B.L. designed the pgRNA libraries. X.R., L.M., T.W.T., Y.Sun, J.L., and M.A.T. performed the experiments. X.R., L.Z., Y.Sun, H.L., W.W., and Y.Shen. contributed to data analysis and interpretation. X.R. and Y.Shen. prepared the manuscript with input from all authors. Y.Shen., W.W., and B.R. supervised the work and obtained funding.

Competing interests

B.R. is a co-founder and consultant of Arima Genomics Inc. and co-founder of Epigenome Technologies. The other authors declare that they have no competing interests.

Data availability statement

The CRISPR screen datasets used in this study are available at the ENCODE portal (www.encodeproject.org) and accession numbers are ENCSR783CGW (APP pgRNA plasmid library), ENCSR364KFC (APP control), ENCSR678GDA (Low APP-EGFP), ENCSR952RDF (Low APP-mCherry), ENCSR493NRD (SIN3A pgRNA plasmid library), ENCSR284PQK (SIN3A control), ENCSR113CEG (Low SIN3A-mCherry), ENCSR750UIY (Low SIN3A-EGFP), ENCSR888FDQ (FMR1 pgRNA plasmid library), ENCSR466IBU (FMR1 control), ENCSR562YXE (Low FMR1-mCherry), ENCSR473BRJ (MECP2 pgRNA plasmid library), ENCSR072YHQ (MECP2 control), and

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ENCSR119JRG (Low MECP2-EGFP). Public datasets used in this study are listed in **Supplementary Table 6**.

Methods:

Generating the reporter iPSC lines

To monitor allelic gene expression, we generated C-terminal allelically tagged human iPSC lines for *APP* (*APP-EGFP/mCherry*), *SIN3A* (*SIN3A-EGFP/mCherry*), *FMR1* (*FMR1-mCherry*), *MECP2* (*MECP2-EGFP*) using CRISPR/Cas9-mediated homology-directed repair (HDR). The parental cell line we used was the WTC11 i^{3}N iPSC line, which has doxycycline-inducible *Ngn2* integrated at the AAVS1 safe harbor locus. For *SIN3A* and *APP*, we generated EGFP and mCherry donor vectors with identical homology arms. For *MECP2* and *FMR1*, we generated an EGFP donor for *MECP2* and a mCherry donor for *FMR1*. We designed sgRNAs with the targeting site within 100bp upstream or downstream of each stop codon to knock in the reporters at the C-terminus of the coding region of each gene. We amplified the genomic regions of 500 to 1000bp upstream and downstream of the stop codon for each target gene as homology arms and inserted the EGFP or mCherry sequences between homology arms. To prevent EGFP and mCherry from affecting target gene function, we added a GS linker and T2A sequence between the C-terminal of the target gene and the N-terminal of EGFP or mCherry. We also mutated sgRNA target sites or PAM sequences on donor vectors to prevent the CRISPR/Cas9 system from cutting donor vectors during the HDR, without altering the encoded amino acids. We cloned all donor vectors by Gibson assembly (NEB, E2621S) and verified them through Sanger sequencing.

We *in vitro* transcribed all sgRNA using the Precision gRNA Synthesis Kit (Invitrogen, A29377), and obtained Cas9-NLS protein from QB3 MacroLab at the University of California, Berkeley. We delivered the CRISPR/Cas9 machinery into iPSC in ribonucleoprotein (RNP) format and donor vectors in plasmid format. To assemble RNP complex, we incubated the *in vitro* transcribed sgRNAs with Cas9-NLS protein at 20-25°C for 15 min. We then mixed the assembled RNP complex with EGFP and/or mCherry donor vectors and delivered them into WTC11 i^{3}N iPSCs using nucleofection (Lonza, VPH-5012). After nucleofection, we seeded the cells into Matrigel-coated (Corning, 354277) wells for recovery. Three to four days later, we sorted the EGFP and mCherry double-positive cells (for *SIN3A* and *APP*), EGFP-positive cells (for *MECP2*), or mCherry-positive cells (for *FMR1*) into Matrigel-coated (Corning, 354277) 96-well plates with one cell per well using fluorescence-activated cell sorting (FACS) to generate clonal allelically tagged reporter cell lines. After about two weeks, we expanded the viable clones and analyzed them to establish reporter cell lines. We validated the individual clonal reporter cell lines with genotyping PCR followed by Sanger sequencing and flow cytometry analysis. The step-by-step protocol can be found

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

at STAR Protocols[53]. DNA sequences of oligos and sgRNAs are listed in **Supplementary Tables 3 and 4**.

Cell culture and neuronal differentiation

The WTC11 i^{3}N iPSCs were cultured on Matrigel-coated (Corning, 354277) plates and maintained in Essential 8 medium (Thermo Fisher Scientific, A1517001), and passaged with Accutase (STEMCELL Technologies, 07920) and 10-μM ROCK inhibitor Y-27632 (STEMCELL Technologies, 72302). The Human embryonic kidney (HEK) 293T cells were cultured in Dulbecco's Modified Eagle medium (Gibco, 11995065) with 10% fetal bovine serum (FBS) (HyClone, SH30396.03), and passaged with trypsin-EDTA (Gibco, 25200072). All the cells were grown with 5% CO_2 at 37°C and verified mycoplasma-free using the MycoAlert Mycoplasma Detection Kit (Lonza, LT07-218). The differentiation of WTC11 i^{3}N iPSCs into excitatory neurons was performed using a two-step differentiation protocol. Briefly, iPSCs were cultured on Matrigel-coated plates with pre-differentiation media containing doxycycline (2 μg/mL; Sigma-Aldrich, D9891) for three days, with daily media changes. After three days, the pre-differentiated cells were dissociated with Accutase (STEMCELL Technologies, 07920) and subplated on Poly-L-Ornithine-coated (15 μg/mL; Sigma-Aldrich, P3655) plates with maturation media containing doxycycline. The maturation media were changed weekly by removing half of the media from each well and adding an equal amount of fresh media without doxycycline. The detailed protocol is accessible at the ENCODE portal (https://www.encodeproject.org/documents/d74fb151-366c-4450-9fa0-31cc614035f9/).

sgRNA library design, cloning, packaging

To perform tiling deletion CRISPR screens, we designed paired sgRNA (pgRNA) library for each target locus, including *SIN3A* (chr15: 74,370,000-76,461,000, hg38), *APP* (chr21: 24,880,000-27,180,000, hg38), *FMR1* (chrX: 146,000,000-150,000,000, hg38), and *MECP2* (chrX: 153,000,000-155,100,000). We first selected all the available sgRNAs within each target region from the sgRNA database generated in CREST-seq[15] and added a G at the start of the sgRNAs that didn't start with G. Then, we removed sgRNAs containing any transcriptional termination sequences (AATAAA, TTTTT, TTTTTT) or BsmBI cut sites (CGTCTC, GAGACG). After filtering, we paired sgRNAs sequentially to generate pgRNA libraries. For each library, the average distance between each sgRNA pair is about 2000 to 3000bp, and the average coverage of sgRNA pairs across each nucleotide in the target region is 15 or 20. To design non-targeting negative control pgRNAs, we first identified unique 20bp long DNA sequences that weren't followed by the NGG PAM sequence and added a G at the start of the sequences that didn't start with G. We removed DNA sequences containing any sequence of TTT, TTNTT, TTTTTT, AATAAA, AAAAA, CGTCTC, or GAGACG. Next, we paired them into pairs with an average distance between two sequences about 1500bp to 2000bp. For positive control pgRNAs targeting EGFP and mCherry, we manually designed 10 sgRNAs targeting EGFP or mCherry sequence and named them with numbers 1

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to 10 according to their locations in EGFP or mCherry sequence from N terminal to C terminal. We further generated pgRNAs by pairing sgRNA1 to sgRNA6, sgRNA2 to sgRNA7, and so on. The oligo libraries of *APP*, *FMR1*, and *SIN3A* were made by following the template of CTTGGAGAAAAGCCTTGTTT{sgRNA1}GTTTAGAGACG{10nt_random_sequence}CGTCTCACACC{sgRNA2}GTTTTAGAGCTAGAAATAGCAAGTT, and the oligo library of *MECP2* was made by following the template of TGTGGAAAGGACGAAACACC{sgRNA1}GTTTAAGAGACG{10nt_random_sequence}CGTCTCTTGTTT{sgRNA2}GTTTTAGAGCTAGAAATAGCAAGTT. We synthesized the designed pgRNA libraries (Twist Bioscience) and cloned into lentiCRISPRv2 plasmid with mouse U6 promoter for *APP*, *FMR1* and *SIN3A*, and cloned into lentiCRISPRv2 plasmid with human U6 promoter for *MECP2*.

We used a two-step cloning strategy to clone these pgRNA libraries. First, we amplified the pgRNAs from the synthesized oligo pool with NEBNext High-Fidelity 2× PCR Master Mix (NEB, M0541S). For each 50μl PCR reaction, we used 0.5μl 20nM oligo pool as a template. The PCR reaction was performed as follows: 98°C 30s; 98°C 10s, 55°C 30s, 72°C 30s, for 15 cycles; 72°C 5min; 4°C hold. The amplified oligo pool was purified and inserted into BsmBI digested lentiCRISPRv2 plasmids via Gibson assembly (NEB, E2621L). The assembled products were transformed into NEB 5-α electrocompetent *Escherichia coli* cells (NEB, C2989K) by electroporation. Millions (1000× of pgRNA library size) of independent bacterial colonies were cultured, and pgRNA library plasmids from first-step cloning were extracted with the Qiagen EndoFree Plasmid Mega Kit (Qiagen, 12381). Second, we digested the pgRNA library plasmids from first-step cloning with BsmBI and purified the product with gel extraction (MACHEREY-NAGEL, 740609.250S). Then, a DNA fragment containing a sgRNA scaffold and another U6 promoter was ligated to the BsmBI digested pgRNA library plasmids using T4 ligase (NEB, M0202M). The ligated products were electroporated into NEB 5-α electrocompetent *Escherichia coli* cells (NEB, C2989K), and millions (1000× of pgRNA library size) of bacterial colonies were cultured for each library. The final plasmid libraries were extracted with the Qiagen EndoFree Plasmid Mega Kit (QIAGEN, 12381). To check the quality of each pgRNA plasmid library, we amplified the pgRNA cassette from the cloned plasmid library by three rounds of PCR with NEBNext High-Fidelity 2× PCR Master Mix (NEB, M0541S). The DNA sequences of oligos used for pgRNA library cloning are listed in **Supplementary Table 3.** The prepared libraries were sequenced with paired-end deep sequencing.

Four pgRNA libraries were packaged into lentivirus libraries individually in HEK293T cells. The titration of each lentivirus library was tested in their associated reporter iPSC lines. The detailed steps for lentivirus packaging and titration were the same as previously described[54].

CREST-seq screen

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To identify enhancers for *APP*, *FMR1*, *MECP2,* and *SIN3A*, we performed CREST-seq screens in excitatory neurons differentiated from each reporter cell line. For each screen, we seeded the reporter iPSCs in Matrigel-coated 6-well plates with one million cells per well, and the total cell number was about 2,000 times the total oligo number in each pgRNA library. 24 hours later, we transduced the lentiviral library into the iPSCs at a multiplicity of infection (MOI) of 0.5 with polybrane (8 μg/mL; Millipore, TR-1003-G) and spun at 1000 RCF at 37°C for 90 min. The next day, we passaged the infected cells with Accutase (STEMCELL Technologies, 07920) and treated them with puromycin (500 ng/mL; Sigma-Aldrich, P8833) at forty-eight hours after infection for 7 days to get rid of uninfected cells. Then, we differentiated the infected cells into excitatory neurons. Two weeks after differentiation, we treated the excitatory neurons with Papain (20U/mL; Sigma, P4762) and DNase Ⅰ (100U/mL; Sigma, DN25) for 30 min at 37°C to dissociate them into single cells. We collected the dissociated neurons with DMEM/F12 media (Gibco, 11330032) plus 10% FBS (HyClone, SH30396.03) and pelleted at 200 RCF and 25°C for 10 min. We resuspended the cell pellets in the HBSS buffer (Gibco, 14175095) with 0.5% FBS for FACS. We collected about 500,000 cells with reduced expression of EGFP or mCherry reporter for each screen. We extracted the genomic DNA from FACS-isolated cells and control cells without FACS via cell lysis and digestion (100 mM pH 8.5 Tris-HCl, 5 mM EDTA, 200 mM NaCl, 0.2% SDS, and 100 μg/mL proteinase K), phenol: chloroform (Thermo Fisher Scientific, 17908) extraction, and isopropanol (Fisher Scientific, BP2618500) precipitation. We amplified the pgRNA cassette from the genomic DNA by performing three rounds of PCR using 500 ng of genomic DNA for each reaction and NEBNext High-Fidelity 2× PCR Master Mix (NEB, M0541S). We deep sequenced the purified libraries with paired-end sequencing. Detailed information on screening is available at the ENCODE portal (https://www.encodeproject.org/documents/c1194c4c-ba28-4e37-a13f-3dde86d03241/). The DNA sequences of oligos used for pgRNA libraries preparation are listed in **Supplementary Table 3**.

Analysis of CREST-seq screens

To quantify the frequency of pgRNAs in each sample, we aligned the paired-end sequencing data to the sequences of designed pgRNAs using BWA[55] (bwa-0.7.17) with default parameters and only the paired reads that exactly matched the designed pgRNA were counted as the frequency of each pgRNA. To evaluate the performance of the FACS-based screening strategy we used for CREST-seq screens, we checked the fold change and *P* value of each pgRNA in each screen by comparing libraries made from sorted cells and control libraries made from unsorted cells. We performed analysis using CRISPY with default settings, and the total mapped reads normalized read counts of each screen were used as input for CRISPY. For *SIN3A* and *APP* screens, we analyzed the libraries for the EGFP allele and mCherry allele separately. Significant enrichment of positive control pgRNAs targeting EGFP and mCherry demonstrated the success of these screens. We further identified functional enhancers for each target

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gene using RELICS (v.2.0)[17]. RELICS splits the region of interest into segments and applies a Bayesian hierarchical model to identify functional sequences supported by the screening data. We prepared the input files to provide genomic coordinates and the total mapped reads normalized read counts of each pgRNA in the standard input format for RELICS. We labeled pgRNAs overlapping 5'TUR and exons of each target gene as known functional sequences and the designed negative controls as negative controls for RELICS. Then, RELICS identified the functional sequences for each screen using the default settings for RELICS v.2.0 (min_FS_nr:30, glmm_negativeTraining:negative_control, crisprSystem:dualCRISPR). We merged the identified adjacent functional sequences and calculated the median RELICS score for each merged DNA fragment using bedtools (v2.26.0). The merged fragments with a median RELICS score >0.2 and more than one functional sequence were considered enhancers.

Chromatin signature analysis of identified enhancers

We checked the overlap between chromatin signatures and identified enhancers using bedtools intersect (v2.26.0). For the marks including H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K9me3, CTCF, and RNA polymerase II, we downloaded the original sequencing files from Gene Expression Omnibus database under accession number GSE167259. We aligned them to the GRCh38/hg38 reference genome using ENCODE chip-seq-pipeline2 (v2.1.6) with the standard setting. We used the overlap optimal peaks for chromatin signature analysis. For cCREs in excitatory neurons from the human brain samples, we downloaded the bed files containing identified cCREs from 38 excitatory neuron subtypes (http://catlas.org/catlas_downloads/humanbrain/cCREs/) and merged them together using bedtools (v2.26.0). We used the merged bed file containing all the cCREs in excitatory neurons for chromatin signature analysis. For accessible genomic regions, we used the ATAC-seq peaks identified in WTC11 i^3N iPSC-derived excitatory neurons[18].

Validation of identified enhancers

We performed the validation experiments for enhancers and promoters by using paired sgRNA-mediated CRISPR deletion. For each region, we designed two sgRNAs to delete the target region (sgRNA sequences are listed in **Supplementary Table 4**). To clone the two sgRNAs into lentiCRISPRv2 vector (Addgene, #52961), we amplified the sgRNA scaffold and mouse U6 promoter using two oligos containing the designed sgRNA sequences, and inserted the amplified DNA fragments into the lentiCRISPR v2 vector (Addgene, #52961) using Gibson assembly (NEB, E2621L). The resulting plasmid contains two sgRNAs with the pattern of hU6-sgRNA1-mU6-sgRNA2. After validating the sgRNA sequences via Sanger sequencing, we individually packaged each plasmid into lentivirus using the same procedure as previously described[56]. We performed validation experiments individually by infecting the reporter cell lines with the associated lentivirus. About 200,000 reporter iPSCs were seeded into a Matrigel-coated cell in a 24-well plate, and the cells were infected with lentivirus 24 hours after seeding

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

using the spin infection method we used for the CREST-seq screen. Forty-eight hours after infection, we treated the cells with puromycin (500 ng/mL; Sigma-Aldrich, P8833) for 7 days. For the validation in iPSCs, we cultured the infected iPSCs for 7 days without puromycin treatment and performed flow cytometry analysis. For the validation in excitatory neurons, we differentiated the infected iPSCs into excitatory neurons and analyzed the neurons with flow cytometry at 14 days after differentiation. For *SIN3A* and *MECP2* validations, we established single-cell clones using FACS-mediated single-cell sorting.

Flow cytometry analysis and fluorescence-activated cell sorting

The cells for flow cytometry analysis and FACS were dissociated into single cells using Accutase (STEMCELL Technologies, 07920) for iPSCs and Papain (Sigma, P4762) for excitatory neurons. The iPSCs were resuspended with FACS buffer (1× DPBS, 2mM EDTA, 25mM HEPES pH7.0, and 1% FBS), and neurons were resuspended with HBSS buffer (Gibco, 14175095) with 0.5% FBS. We used the same gate setting for both flow cytometry analysis and FACS. First, cells were separated from the debris based on the forward scatter area (FSC-A) and side scatter area (SSC-A). Then, single cells were separated using a single cell gate based on the width and area metrics of the forward scatter (FSC-W versus FSC-A) and side scatter (SSC-W versus SSC-A). Further, the gates for EGFP and mCherry signal baselines were set using cells without EGFP and mCherry signals. Flow cytometry analyses were performed on BD LSR II and BD LSRFortessa Flow Cytometers. FACS experiments were conducted on a BD FACSAria II instrument using a 100-μm nozzle. All the plots associated with flow cytometry analysis and FACS were made by using FlowJo (v10.7.2).

Time-course analysis of *SIN3A* transcriptional compensation

To monitor the transcriptional compensation of *SIN3A*, we seeded *SIN3A-EGFP/mCherry* iPSCs in a Matrigel-coated 12-well plate with 200,000 cells per well. 24 hours later, we infected the cells with lentivirus expressing Cas9 and pgRNAs targeting the *SIN3A* promoter and an enhancer. After infection, we dissociated the cells with Accutase at each time point. We used one-third of the cells for flow cytometry analysis and maintained two-thirds for analysis at the next time point. We analyzed the cells using BD LSRFortessa Flow Cytometers and analyzed the data using FlowJo (v10.7.2).

RT-qPCR and allelic gene expression

We extracted total RNA from each sample using QIAGEN plus mini RNA kit (Qiagen, 74134), and 1μg total RNA was used to make cDNA with iScript cDNA synthesis kit (Bio-Rad, 1708891). To check the allelic expression of *SIN3A*, we used one SNP located in the *SIN3A* intron. We amplified the SNP region from each cDNA sample and added deep sequencing adaptors via PCR to prepare a sequencing library for each sample. The amplicons in each purified library were analyzed by deep sequencing (DNA oligos

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are listed in **Supplementary Table 3**). The copy number of each allele of *SIN3A* in each sample was counted using a 21bp window with the SNP in the middle. The total expression levels of *SIN3A* and *MECP2* were analyzed on a Roche LightCycler 96 instrument using Luminaris HiGreen qPCR Master Mix (Thermo Scientific, K0992) (DNA oligos are listed in **Supplementary Table 3**). Data were normalized to *GAPDH*.

CTCF motif deletion

We scanned transcription factor motifs in SIN3A-E4 using FIMO (v5.4.1)[57] with human motif database HOCOMOCO (HOCOMOCOv11 full annotation)[34] and default settings. We focused on a CTCF motif and designed two sgRNAs with spacer sequence overlapping CTCF motifs. We cloned the two sgRNAs into the lentiCRISPRv2 vector (Addgene, #52961) individually and packaged them into lentivirus. We infected the SIN3A-EGFP/mCherry iPSCs with each lentivirus separately and treated the cells with puromycin for 7 days. After puromycin treatment, we cultured the cells for an additional 7 days. We then isolated cells with reduced expression levels of EGFP or mCherry reporters from each cell pool using FACS and extracted the genomic DNA from these isolated cells. To check the DNA sequences in the sgRNA targeting sites, we amplified the sgRNA target sites with PCR and deep sequenced the amplicons (DNA oligos are listed in **Supplementary Table 3**). The deep sequencing data of each sample was analyzed using CRISPRssor2[58].

ClinVar variants enrichment analysis

We downloaded the clinical variants found in patient samples from the ClinVar database[59] (https://www.ncbi.nlm.nih.gov/clinvar/, version 2023-08). Copy number variants (CNVs) are variants equal to or larger than 50 bp. For genomic localization enrichment analysis, we checked the overlap between CNVs and protein coding regions, promoter regions, and distal cCREs, and performed a two-sided Fisher's exact test to determine the significance of enrichment. Protein coding regions were obtained from the GENCODE GTF file (GENCODE v44 annotation) using the features CDS, start_codon, or stop_codon. Promoter regions and distal cCREs were obtained from a comprehensive list of cCREs identified in 222 distinct human cell types[60]. We classified Promoter and Promoter Proximal regions as promoter regions. For enrichment analysis of CNVs in distal cCREs, we filtered the CNVs with the variant length within 50 to 5000 bp and separated them into different groups based on associated diseases, and used the distal cCREs identified from 222 distinct human cell types[3]. For each cell type and disease-associated CNVs group combination, we computed the number of intersections between disease-associated CNVs and cell-type-associated distal cCREs. We compared the cell type specific intersection number with the number of intersections between disease-associated CNVs and the entire set of distal cCREs from all cell types, using a hypergeometric test to evaluate the statistical significance of cell type specific enrichment. We used $P > 0.05$ as the cutoff for significant enrichment.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The overexpression of *SIN3A*

To overexpress *SIN3A*, we constructed the *SIN3A* promoter P1 controlled *SIN3A* expression plasmid (SIN3A-Pr > SIN3A-P2A-BFP). We amplified the *SIN3A* promoter P1 region (chr15: 75451566 - 75452299, hg38) from the genomic DNA of WTC11 i^{3}N iPSCs, SIN3A coding region from cDNA made from total mRNA of WTC11 i^{3}N iPSCs, and BFP from a plasmid (Addgene, #102244) using NEBNext High-Fidelity 2× PCR Master Mix (NEB, M0541S). We inserted these three fragments into the pLS-SceI plasmid (Addgene, #137725) and replaced the minimal promoter and EGFP sequences using Gibson assembly (NEB, E2621L) to construct the *SIN3A* expression plasmid. We packaged the *SIN3A* expression plasmid into lentivirus and delivered it into iPSCs via spin infection. To check the expression level of *SIN3A* in the infected cells, we isolated BFP-positive cells using FACS and extracted the total mRNA from BFP-positive cells using QIAGEN plus mini RNA kit (Qiagen, 74134), and 1µg total RNA was used to make cDNA with iScript cDNA synthesis kit (Bio-Rad, 1708891). The total *SIN3A* expression levels were analyzed on a Roche LightCycler 96 instrument using Luminaris HiGreen qPCR Master Mix (Thermo Scientific, K0992). The DNA sequences of oligos are listed in **Supplementary Table 3**. Data were normalized to *GAPDH*.

SIN3A promoter reporter assay

We used *SIN3A* promoter reporter to test the transcriptional activity of the *SIN3A* promoter under wild type and SIN3A knockdown conditions. To construct the *SIN3A* promoter reporter plasmids, we modified one lentivirus EGFP reporter plasmid (Addgene, #137725) by replacing the scaffold-attached region (SAR)[61,62] with human anti-repressor element 40[63], and used the modified plasmid as a backbone for *SIN3A* promoter reporter plasmid cloning.

We picked two regions (P1, Chr15: 75451566-75452299; P2, Chr15: 75453777-75454850) as *SIN3A* promoter based on the ATAC-seq and SIN3A ChIP-seq data. We amplified these two regions from the genomic DNA of WTC11 i^{3}N iPSCs and inserted them before the start of the EGFP sequence in the modified EGFP plasmid via Gibson assembly (NEB, E2621L), and constructed P1-EGFP and P1+P2-EGFP reporter plasmids. To knockdown *SIN3A* expression, we used shRNA-mediated knockdown. We designed a shRNA targeting *SIN3A* mRNA using the DSIR tool (http://biodev.extra.cea.fr/DSIR/DSIR.html) and used a human control shRNA from a previous study[64]. To clone shRNA expression plasmids, we replaced the sgRNA scaffold and Cas9 expression cassette in lentiCRISPRv2 (Addgene, #52961) vector with EF1α-HygR-BFP. shRNAs were cloned into the modified lentiCRISPRv2 vector under the control of a human U6 promoter and packaged into lentivirus for cell transduction. The cloned plasmids were verified using Sanger sequencing and packaged into lentivirus. To test the knockdown efficiency of *SIN3A* shRNA, we infected SIN3A-EGFP/mCherry reporter cell line

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with lentivirus containing *SIN3A* shRNA or control shRNA and checked the SIN3A-EGFP and SIN3A-mCherry signals with flow cytometry six days after infection. We used the EGFP and mCherry signals from WTC11 i^{3}N cells as baselines and calculated the knockdown efficiency of *SIN3A* shRNA relative to control shRNA. The average knockdown efficiency from SIN3A-EGFP and SIN3A-mChery alleles was used as knockdown efficiency of *SIN3A* shRNA. To test the *SIN3A* promoter reporter, we infected WTC11 i^{3}N iPSCs with P1-EGFP and P1+P2-EGFP lentivirus individually with MOI<0.1, and isolated the EGFP positive cells using FACS. Then, we infected the FACS-isolated P1-EGFP and P1+P2-EGFP cells with lentivirus containing control shRNA and *SIN3A* shRNA individually and checked the EGFP signal with flow cytometry six days after infection. We performed all the experiments in three biological replicates and analyzed them with BD LSRFortessa Flow Cytometer and FlowJo (v10.7.2). The sequences of shRNAs are listed in **Supplementary Table 5**.

Identification and analysis of candidate transcriptional compensation genes

To identify candidate transcriptional compensation genes, we extracted the promoter sequences (+/- 1kb of TSS) for each protein-coding gene in the human (GENCODE v44 annotation) and mouse (GENCODE vm33 annotation) genomes. Then, we searched for transcription factor (TF) binding motifs in these promoter sequences using FIMO[57] (v5.5.4) (P<0.0001) and TF motifs from HOCOMOCO (HOCOMOCOv11 full annotation)[34] and JASPER (JASPAR2022 CORE vertebrates)[65] databases. GO term analysis was performed using Enrichr[66]. The identity of each TF was annotated using UniProtKB (activator or repressor) and Gene Ontology (AmiGO 2 with the terms “DNA-binding transcription activator activity” or “DNA-binding transcription repressor activity”). The expression of the identified candidate transcriptional compensation genes was checked using bulk tissue RNA-seq data from GTEx (RNASeQCv1.1.9)[67].

Allelic analysis of *SIN3A* enhancer-mediated *cis*-regulation

We identified phased SNPs using WTC11 whole genome sequence data (https://www.allencell.org/genomics.html). To perform allelic analysis of *SIN3A* enhancer-mediated cis-regulation, we selected one phased SNP located in the last intron of *SIN3A* (chr15: 75374632, C/T, hg38) and another phased SNP near SIN3A-E2 (chr15: 74721849, T/G, hg38). To link the *SIN3A* alleles to the tagged EGFP and mCherry reporters, we amplified the genomic region covering the *SIN3A* intron SNP and reporters using TaKaRa LA Taq DNA Polymerase (TaKaRa, RR042A), genomic DNA from SIN3A-EGFP/mCherry iPSCs, and reporter specific primers (GFP-Rs1, mCherry-Rs1, SIN3A_intron_SNP1-R). Then, we sequenced the PCR product using Sanger sequencing to confirm the relationship between *SIN3A* intron SNP and reporters. To check the enhancer deletion allele, we infected the SIN3A-EGFP/mCherry iPSCs with lentivirus expressing Cas9 and pgRNAs targeting SIN3A-E2 followed by puromycin treatment for seven days. Then, we isolated the cells with reduced expression levels of *SIN3A*-

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

EGFP or *SIN3A-mCherry* using FACS. We extracted the genomic DNA from FACS-isolated cells using QuickExtract DNA Extraction Solution (Biosearch Technologies, QE0905T). We amplified the allele with enhancer deletion from each genomic DNA using TaKaRa LA Taq DNA Polymerase (TaKaRa, RR042A) and primers targeting the SIN3A-E2 region (SIN3A_En_SNP-F, SIN3A_En_SNP-R). We then performed TOPO cloning (Invitrogen, 450071) and sequenced 6 colonies from each sample using Sanger sequencing to verify the sequences. The DNA sequences of oligos used in this experiment are listed in **Supplementary Table 3**.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures:

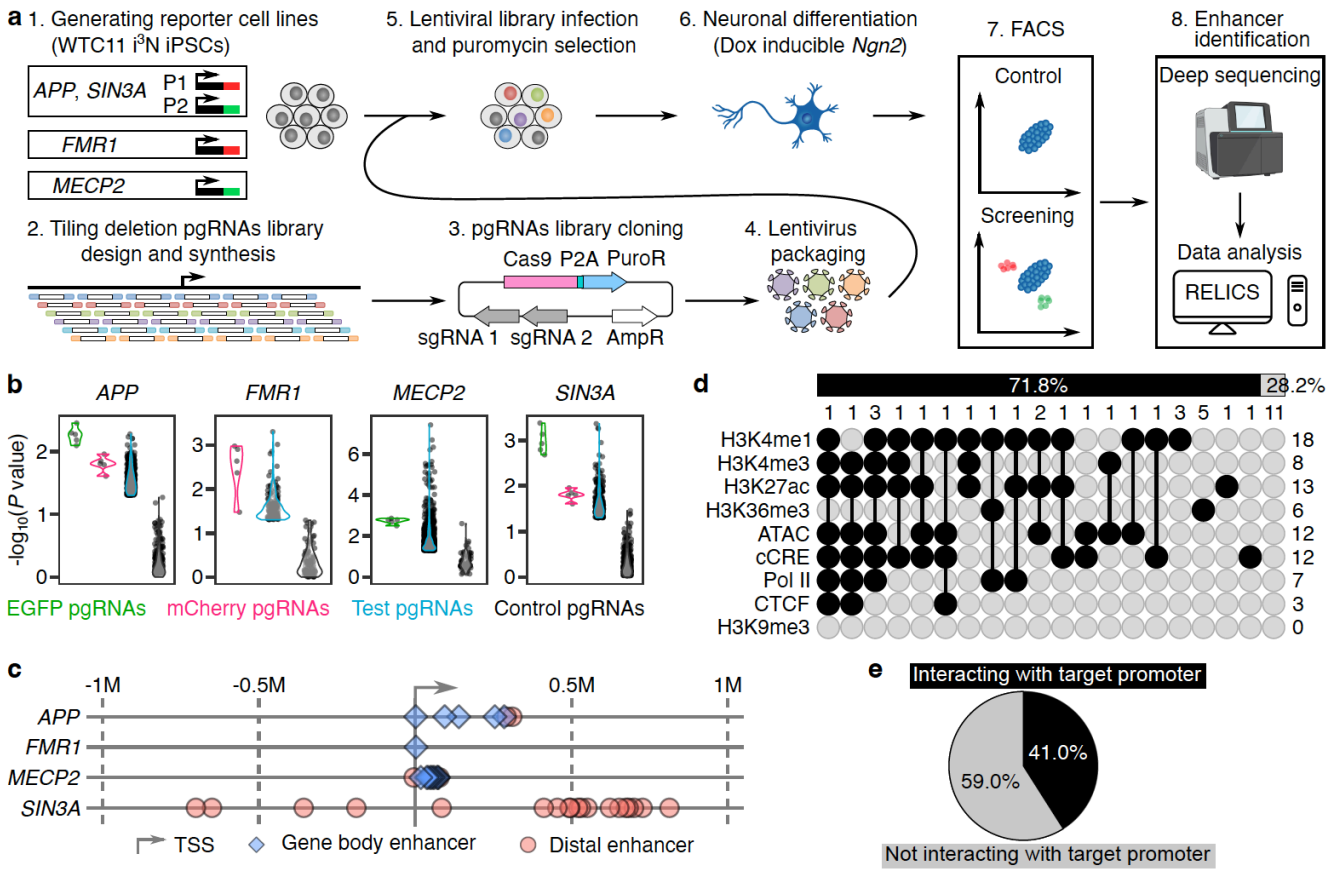
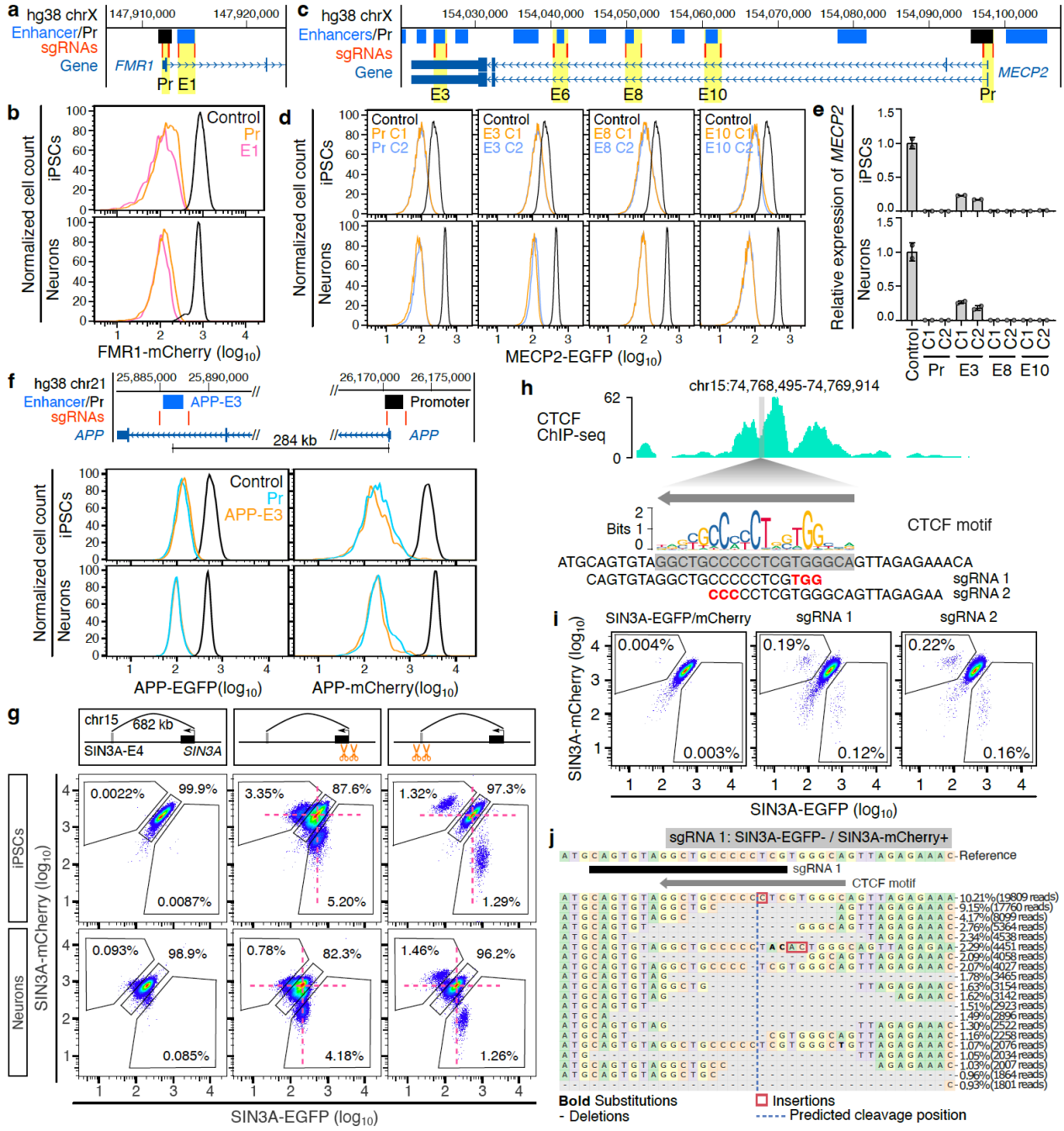


Figure 1. Identification and analysis of enhancers of four neuropsychiatric risk genes. a, The workflow of identifying enhancers of *APP*, *FMR1, MECP2*, and *SIN3A* in iPSC-induced excitatory neurons using CRISPR tilling deletion screening. **b,** The *P* value distribution of enriched pgRNAs ($log_2FC>0$) in each screen. The positive control pgRNAs targeting EGFP and mCherry and some of the test pgRNAs are significantly enriched in each screen. The negative control pgRNAs are not significantly enriched. **c,** The distribution of identified enhancers of *APP*, *FMR1, MECP2*, and *SIN3A*, relative to TSS of each target gene. **d,** Upset plot showing the overlap between identified enhancers and each chromatin feature. The numbers in each row and column indicate the total number of enhancers in each category. **e,** The percentage of enhancers interacting and not interacting with target promoters based on H3K4me3 PLAC-seq data.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2. Validating CREST-seq identified enhancers. a, Genome browser screenshot showing gene body enhancer of *FMR1* and sgRNAs targeting *FMR1* promoter and enhancer. **b,** Flow cytometry plots showing the significant downregulation of FMR1-mCherry expression after deleting *FMR1* promoter and FMR1-E1 enhancer in both iPSCs and excitatory neurons. Positive controls (black line) are the FMR1-mCherry reporter cells. **c,** Genome browser screenshot showing identified enhancers of *MECP2* and sgRNAs targeting *MECP2* promoter and enhancers. **d,** Single clones of *MECP2* promoter or enhancers deletion showing significant downregulation of MECP2-EGFP in both iPSCs and excitatory neurons. Positive controls (black line) are the MECP2-EGFP cells. C1 and C2 indicate two independent clones. **e,** RT-qPCR results showing the significant downregulation of *MECP2* expression in each clone ($P < 0.05$ for all the clones, two-tailed two-sample *t*-test; n = 2). Data are mean ± SEM. **f,** Flow cytometry plots showing the significant downregulation of APP-EGFP or APP-mCherry in *APP* promoter and APP-E3 deletion cells. Positive controls (black line) are the APP-EGFP/mCherry reporter cells. **g,** Flow cytometry plots showing the downregulation of SIN3A-EGFP or SIN3A-mCherry in *SIN3A* promoter and SIN3A-E4 deletion cells. Red dashed lines indicate the position of SIN3A-EGFP/mCherry double positive cells. **h,** The genome browser screenshot showing the CTCF ChIP-seq signal in SIN3A-E4 enhancer region in WTC11 iPSCs. The CTCF motif was obtained from JASPAR. Two sgRNAs were designed to target the CTCF motif. PAM sequences were in red. **i,** Flow cytometry plots showing the downregulation of SIN3A-EGFP or SIN3A-mCherry in sgRNA1 and sgRNA2 infected cells. **j,** The editing outcomes of sgRNA 1 in the cells of SIN3A-EGFP-/SIN3A-mCherry+. **k,** The enrichment of disease-associated CNVs in distal cCREs identified in diverse cell types in the human body. Heatmap shows the data from diseases with at least 10 CNVs and *P* value less than 1×10^{-5} in at least one cell type.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

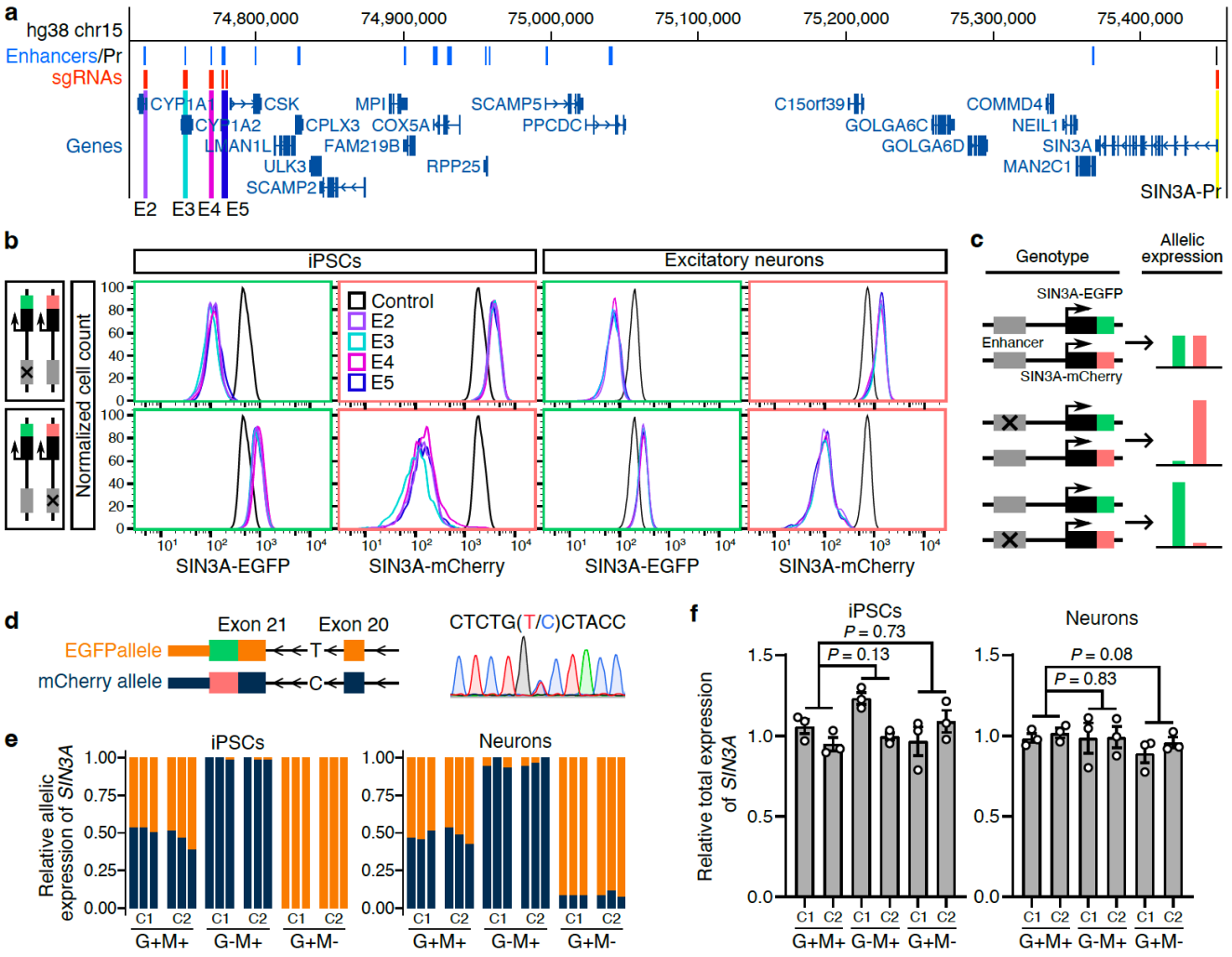


Figure 3. Allelic enhancer deletion induces transcriptional compensation of *SIN3A*. a, Genome browser screenshot showing enhancers of *SIN3A* and sgRNAs targeting *SIN3A* promoter and enhancers. **b,** Flow cytometry plots showing the significant downregulation of SIN3A-EGFP and SIN3A-mCherry expression after deleting *SIN3A* enhancers. Positive controls (black lines) are SIN3A-EGFP/mCherry reporter cells. **c,** The model of the allelic expression pattern of *SIN3A* and the associated genotype. **d,** Sanger sequencing shows the SNP in *SIN3A* intron. **e,** Allelic gene expression analysis using the SNP located in *SIN3A* intron shows dominant expression from one allele in G-M+ (SIN3A-EGFP-/SIN3A-mCherry+) and G+M- (SIN3A-EGFP+/SIN3A-mCherry-) clones in both iPSCs and 2-week excitatory neurons. C1 and C2 indicate two independent clones, and each clone has three biological replicates. Dark blue color indicates the C allele, and orange color indicates the T allele. **f,** RT-qPCR results showing the total *SIN3A* expression in each clone relative to *GAPDH*. Each clone has three biological replicates. *P* values were determined using the two-tailed two-sample *t*-test.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

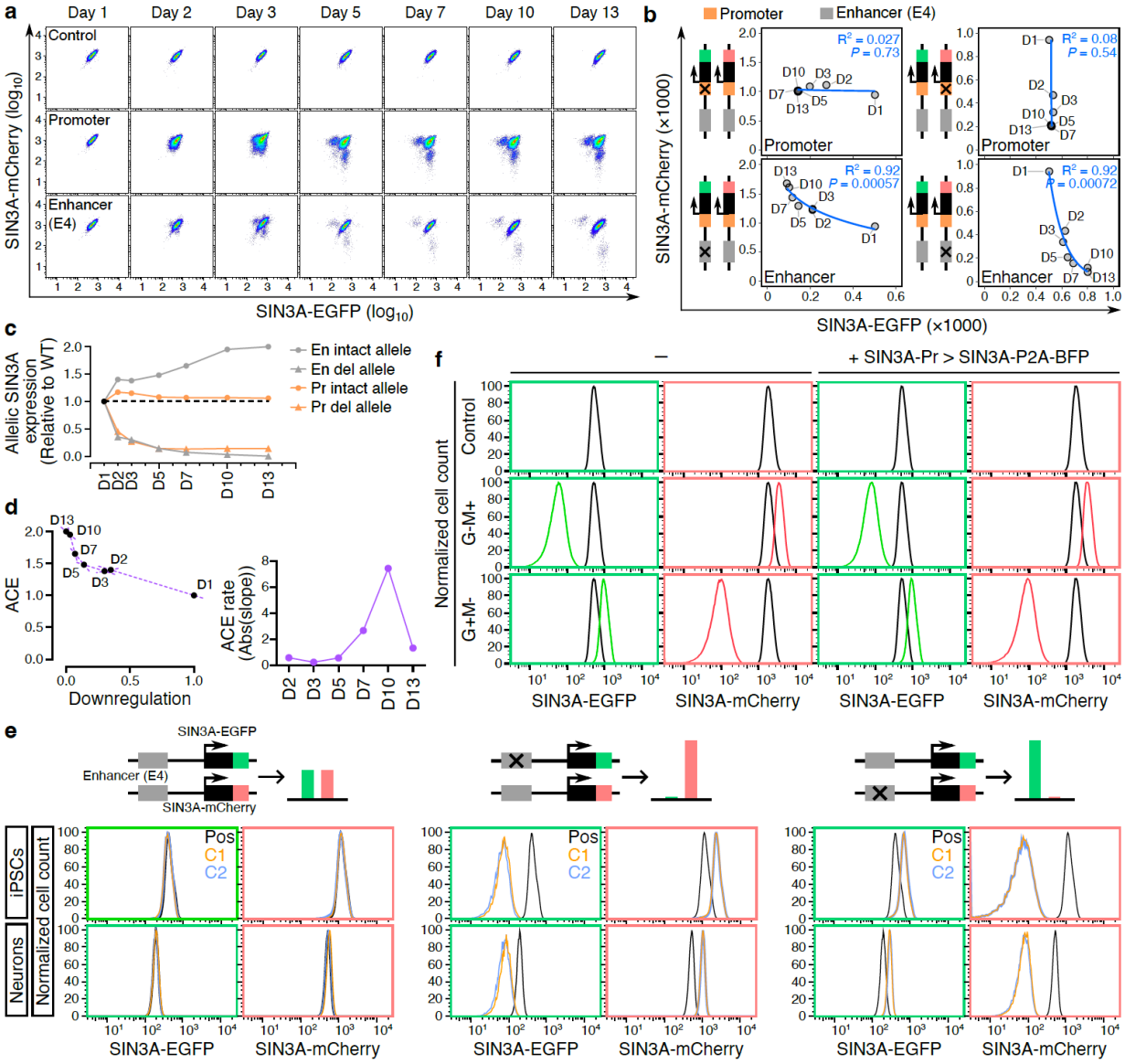


Figure 4. Allelic enhancer deletion-induced allelic compensation effect (ACE) is a dynamic process. a, Flow cytometry plots showing the expression of SIN3A-EGFP and SIN3A-mCherry in control cells (SIN3A-EGFP/mCherry reporter cells) and cells infected with pgRNAs targeting *SIN3A* promoter and SIN3A-E4 enhancer. The dates refer to the days following the lentivirus infection. **b,** Dot plots showing the expression trend of SIN3A-EGFP and SIN3A-mCherry signals in the cells with reduced expression level of SIN3A-EGFP or SIN3A-mCherry in panel a. Trendlines are based on logarithmic model. **c,** Allelic promoter and enhancer deletion-induced downregulation of *SIN3A*. Dots indicate the levels of SIN3A-EGFP or SIN3A-mCherry in cells with allelic promoter or enhancer deletions. The black dashed line indicates allelic expression levels from wild-type cells. **d,** The ACE rate of *SIN3A* enhancer E4 deletion. The average downregulation and transcriptional compensation resulting from enhancer deletion on the EGFP and mCherry alleles were used to calculate the slope between each pair of adjacent time points. **e,** Flow cytometry plots showing the SIN3A-EGFP and SIN3A-mCherry signals from each clone in iPSCs and neurons. Positive control is SIN3A-EGFP/mCherry reporter cells. C1 and C2 indicate two independent clones of each genotype. **f,** Flow cytometry plots showing the SIN3A-EGFP and SIN3A-mCherry signals in the cells with and without ectopic *SIN3A* expression. SIN3A-EGFP/mCherry reporter cells were used as control.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

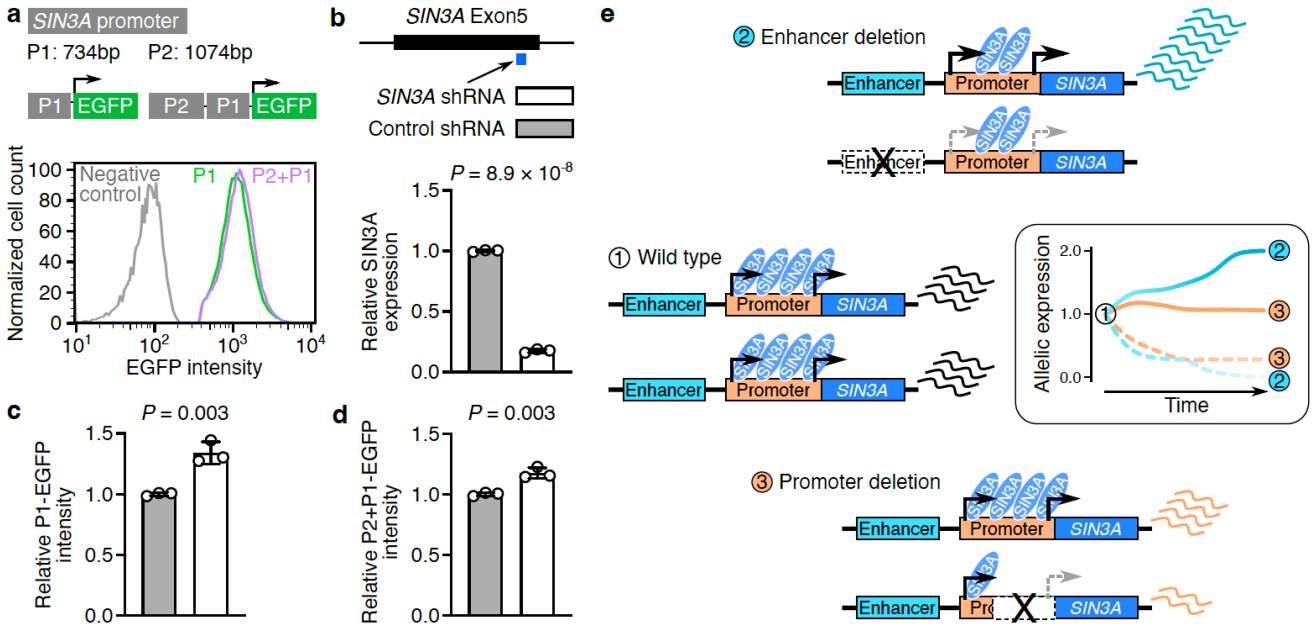
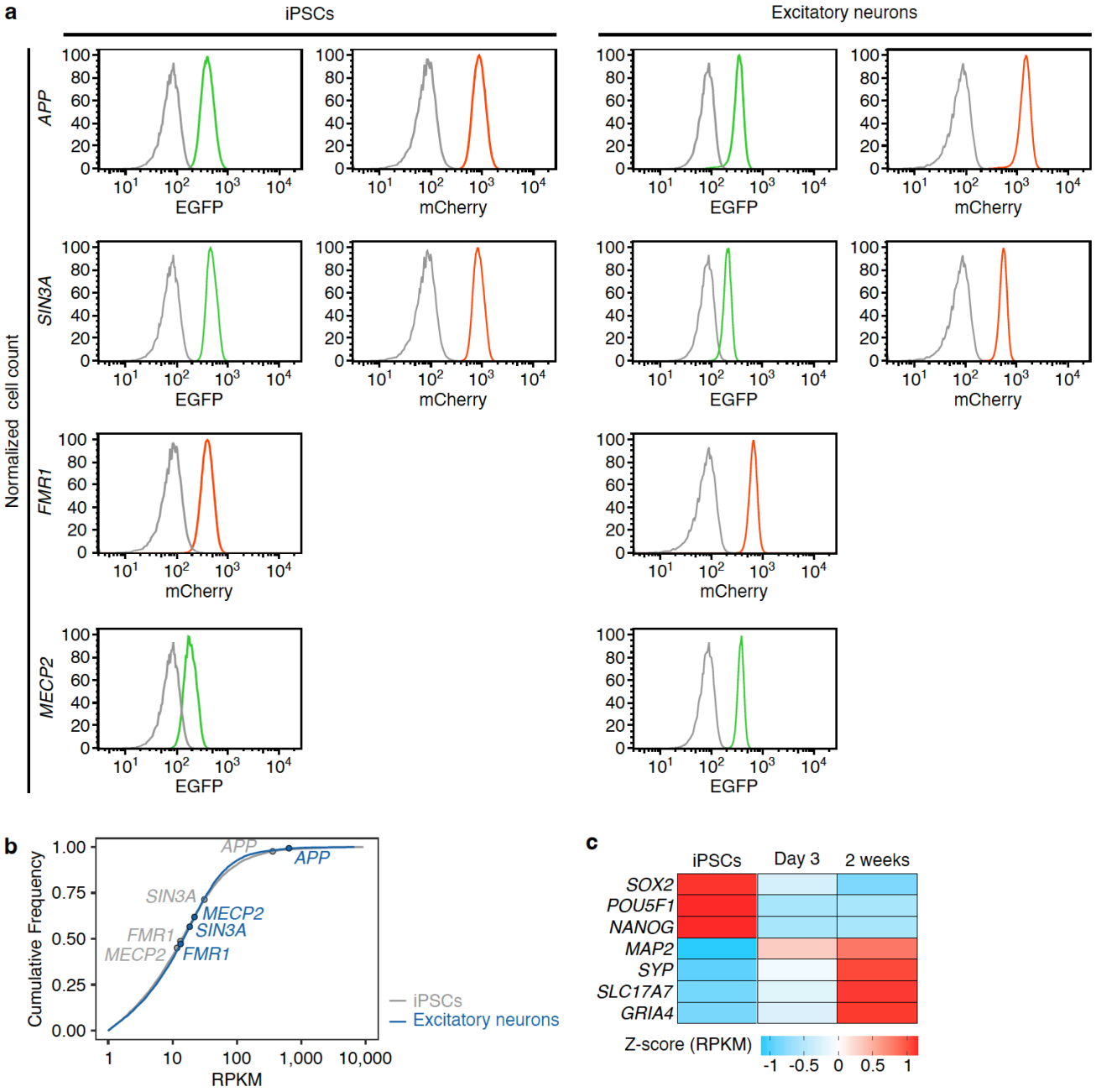


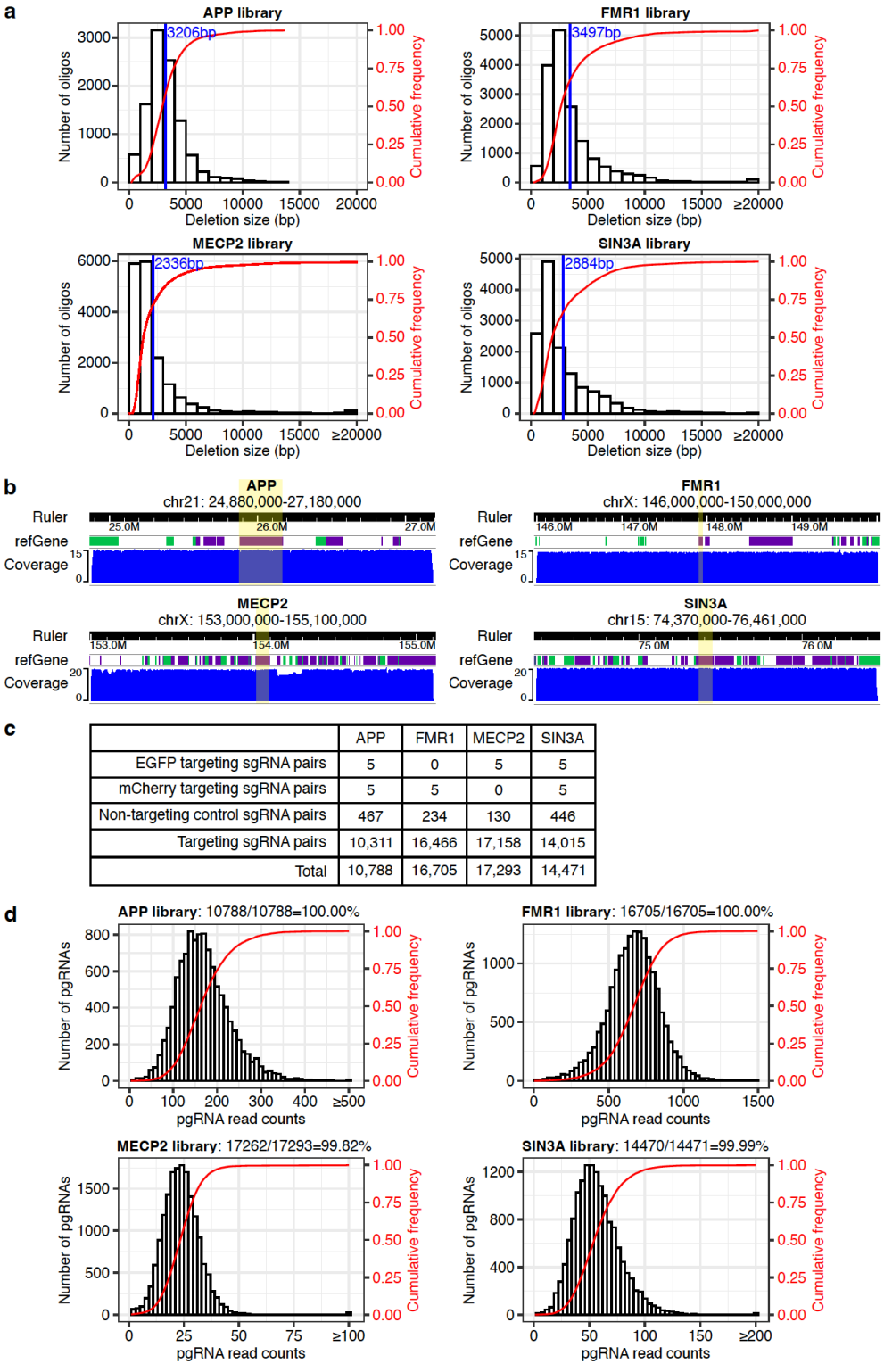
Figure 5. The *SIN3A* promoter mediates allelic enhancer deletion-induced allelic compensation effect (ACE). a, Flow cytometry plots showing the EGFP expression from *SIN3A* promoter reporters. **b,** shRNA-mediated downregulation of *SIN3A*. **c,d,** *SIN3A* promoter reporters show significantly higher EGFP intensity in cells with *SIN3A* shRNA, compared to cells with control shRNA. *P* values in panels **b-d** were determined using the two-tailed two-sample *t*-test. **e,** The working model of allelic enhancer deletion-induced ACE. *SIN3A* is evenly expressed from two alleles in wild-type cells. Allelic enhancer deletion causes downregulation of *SIN3A* from the enhancer deletion allele (sky blue dashed line), which triggers ACE from the intact allele (sky blue solid line). Allelic partial promoter deletion causes partial downregulation of *SIN3A* (orange dashed line) without ACE (orange solid line).

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Extended Data Figure 1. **Engineered reporter cell lines and gene expression. a,** Flow cytometry plots showing the expression of EGFP and mCherry reporters in APP-EGFP/mCherry, SIN3A-EGFP/mCherry, FMR1-mCherry, and MECP2-EGFP reporter cell lines. The expression of reporters was checked in both iPSCs and excitatory neurons. Gray lines are signals from negative control cells, WTC11 i^{3}N. **b,** RNA-seq data shows the expression of *APP*, *FMR1, MECP2,* and *SIN3A* in iPSCs and 2-week excitatory neurons. The genes were ranked on RPKM. **c,** The expression of cell type marker genes in iPSCs and excitatory neurons.

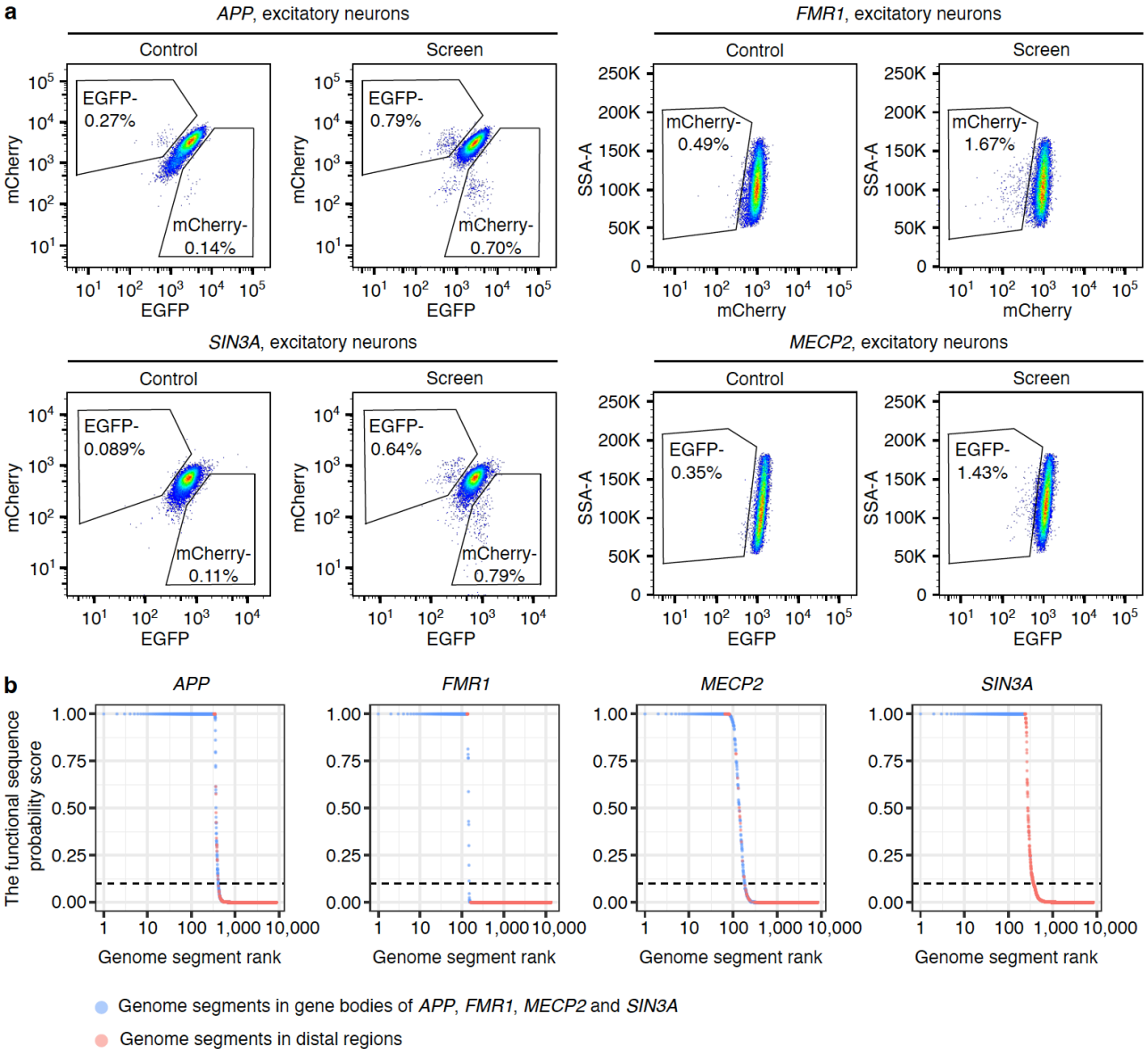
bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



	APP	FMR1	MECP2	SIN3A
EGFP targeting sgRNA pairs	5	0	5	5
mCherry targeting sgRNA pairs	5	5	0	5
Non-targeting control sgRNA pairs	467	234	130	446
Targeting sgRNA pairs	10,311	16,466	17,158	14,015
Total	10,788	16,705	17,293	14,471

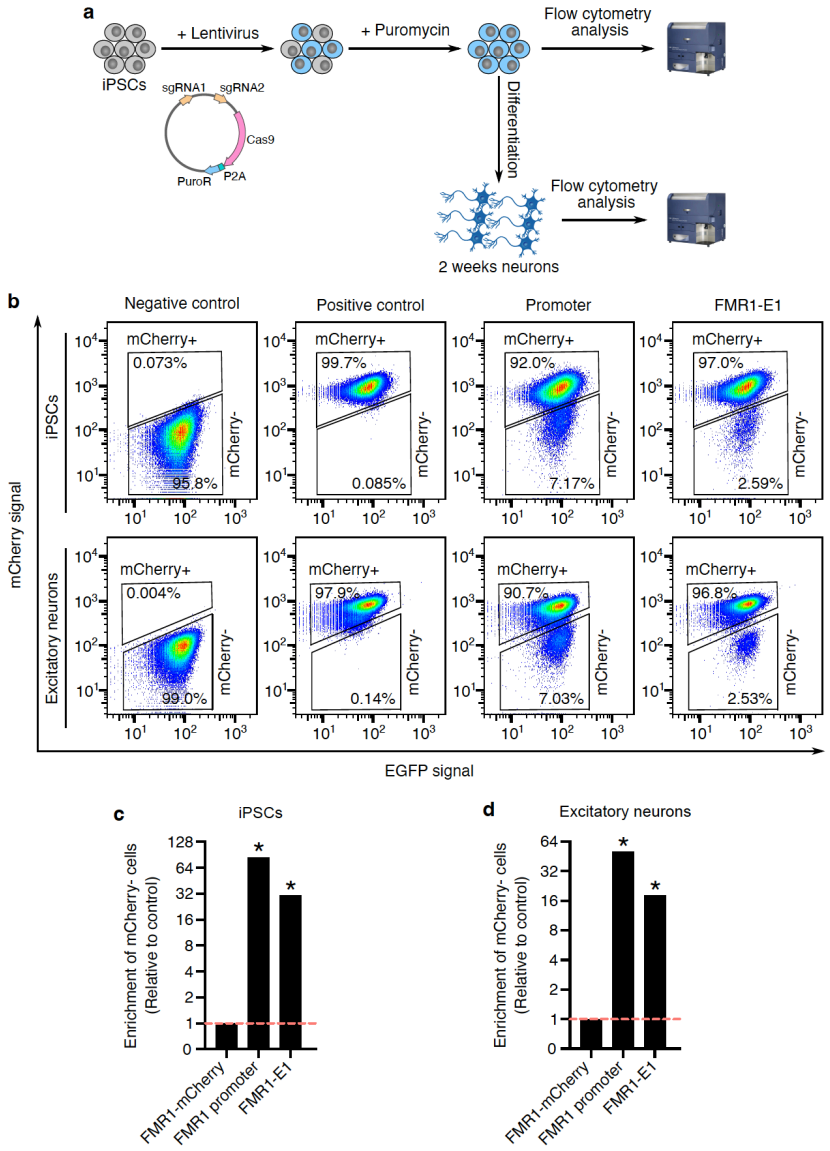
Extended Data Figure 2. **pgRNA libraries of *APP*, *FMR1*, *MECP2*, and *SIN3A*. a,** The distribution of deletion size of pgRNA libraries. Blue lines indicate the average deletion size of each pgRNA library. **b,** The coverage of pgRNA libraries. The gene body regions of each gene were labeled with yellow. **c,** The composition of pgRNA libraries. **d,** The distribution of pgRNA read counts and cumulative frequency in cloned plasmid libraries. More than 99% of designed pgRNA were recovered in each plasmid library.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



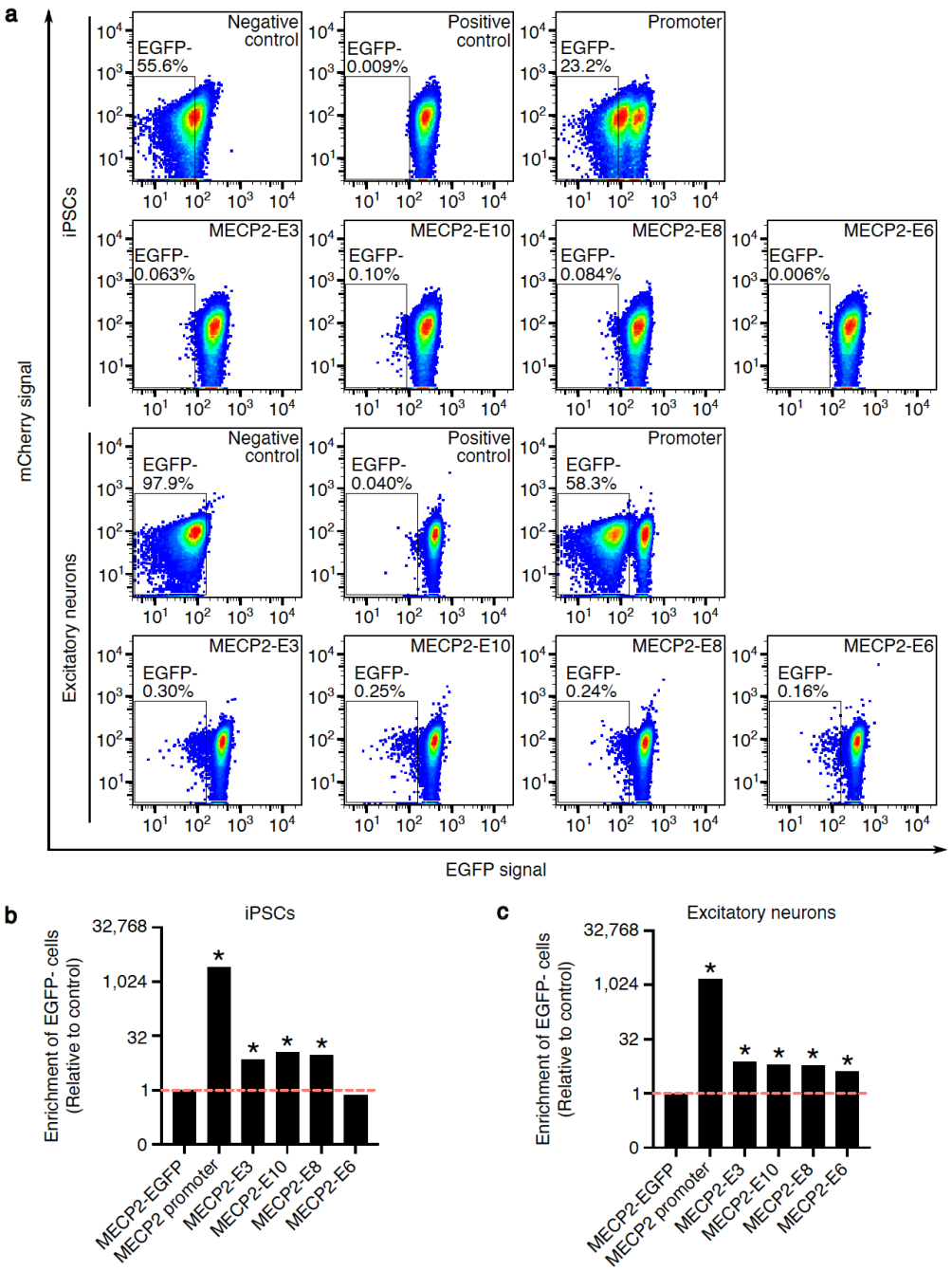
Extended Data Figure 3. CREST-seq screens and data analysis. a, The representative FACS plots showing the sorting strategies used for CREST-seq screens. The reporter cells without pgRNA library infection were used as the control for each screen. **b,** The functional sequence probability score of genome segments in RELICS analysis for each screen. The black dashed lines indicate the default cutoff of the functional sequence probability score (score = 0.1) in RELICS. The genome segments with a score >0.1 were identified as functional sequences in RELICS analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



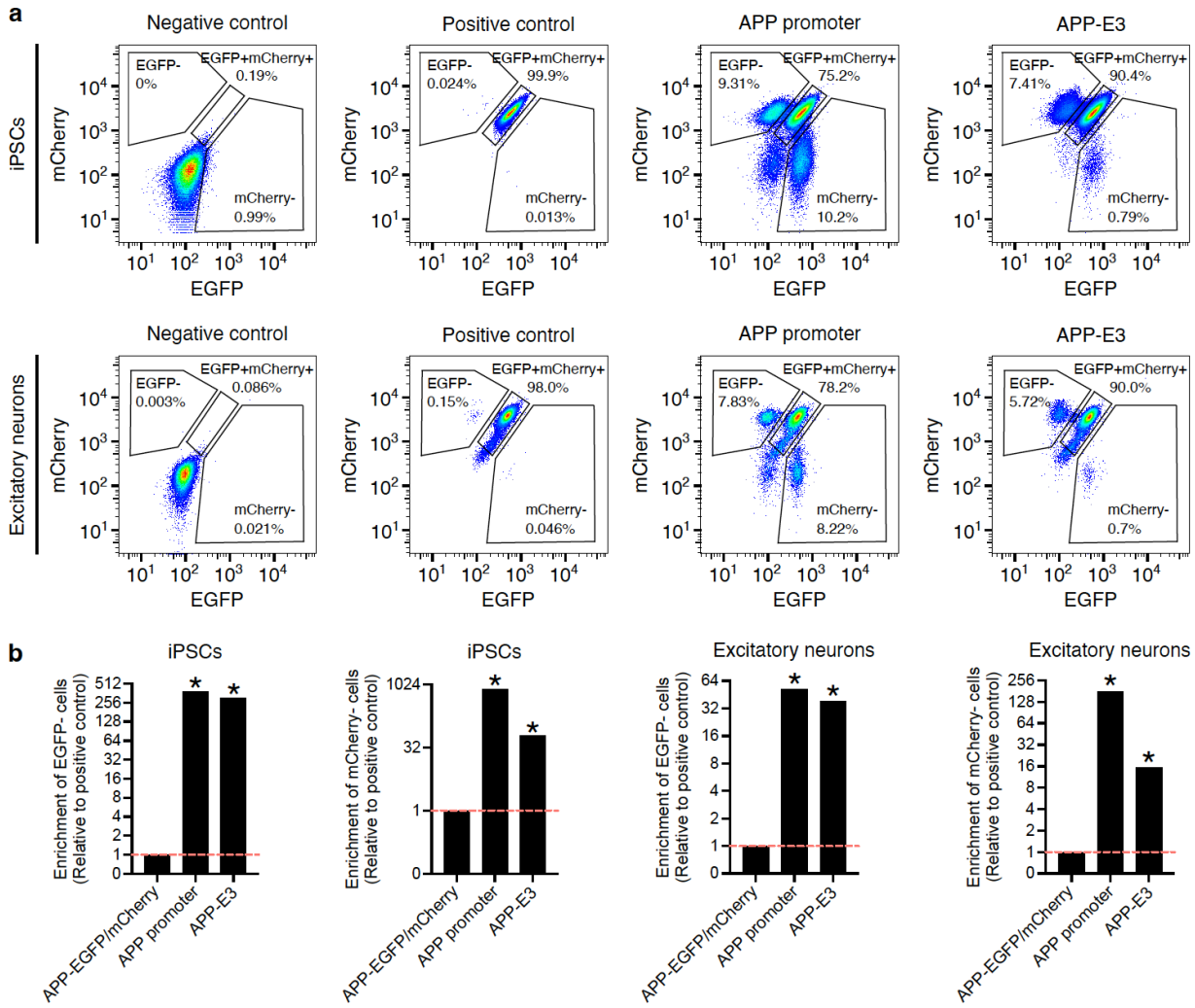
Extended Data Figure 4. Enhancer validation strategy and validation of *FMR1* enhancer. a, The flow cytometry based strategy for enhancer validation. **b,** Flow cytometry plots showing the percentage of cells with reduced FMR1-mCherry expression in each condition. The negative control is the WTC11 i^{3}N cells. The positive control is the FMR1-mCherry reporter cells. **c,** Bar graphs showing the significance of the relative enrichment of cells with reduced expression of FMR1-mCherry compared to positive control cells. *P* values were determined using the two-sided Fisher's exact test. * $P < 0.0001$.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



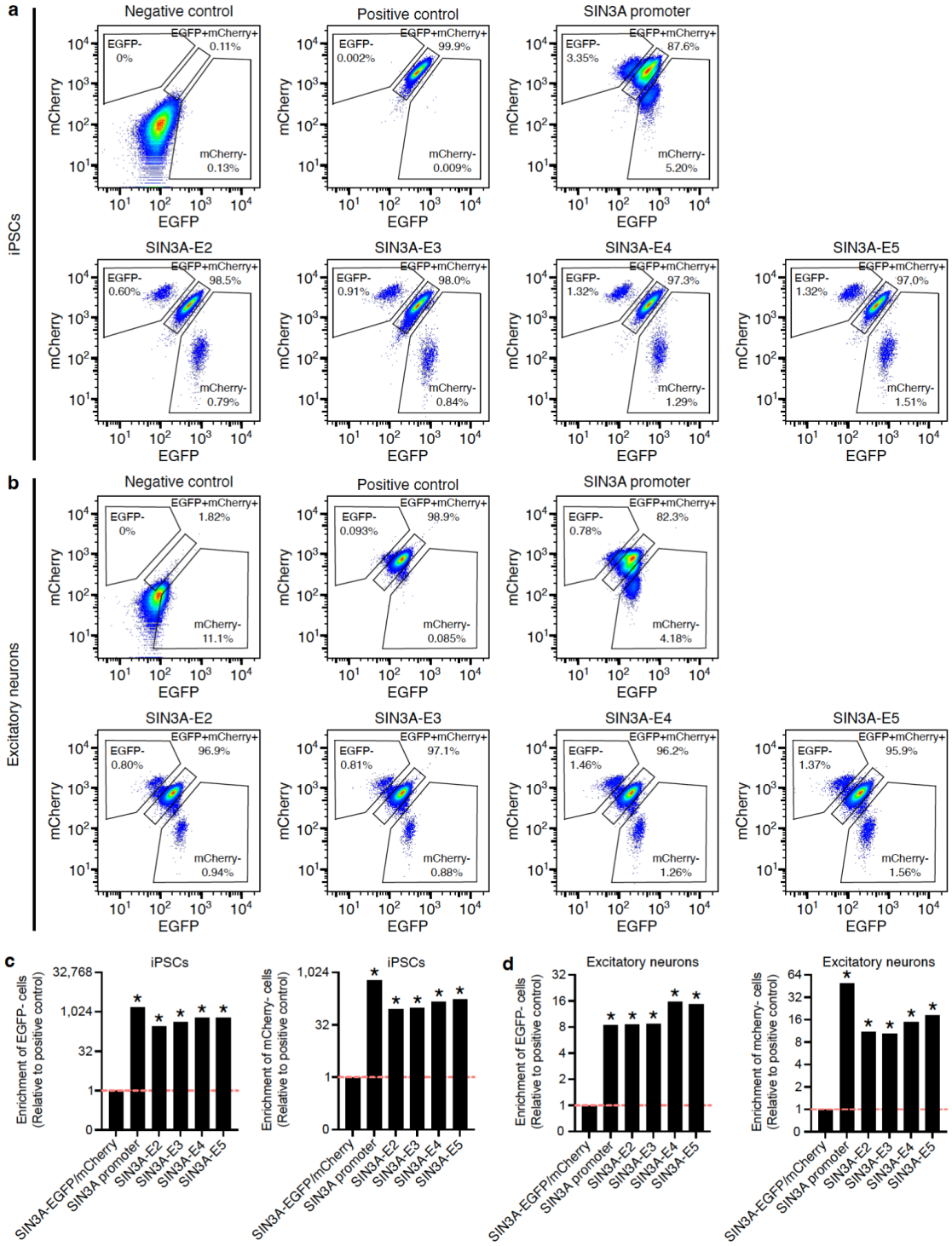
Extended Data Figure 5. *MECP2* enhancer validations. a, Flow cytometry plots showing the percentage of cells with reduced MECP2-EGFP expression in each condition. The negative control is the WTC11 i^{3}N cells. The positive control is the MECP2-EGFP reporter cells. **b,c,** Bar graphs showing the significance of the relative enrichment of cells with reduced expression of MECP2-EGFP compared to positive control cells. *P* values were determined using the two-sided Fisher's exact test. * $P < 0.0001$.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



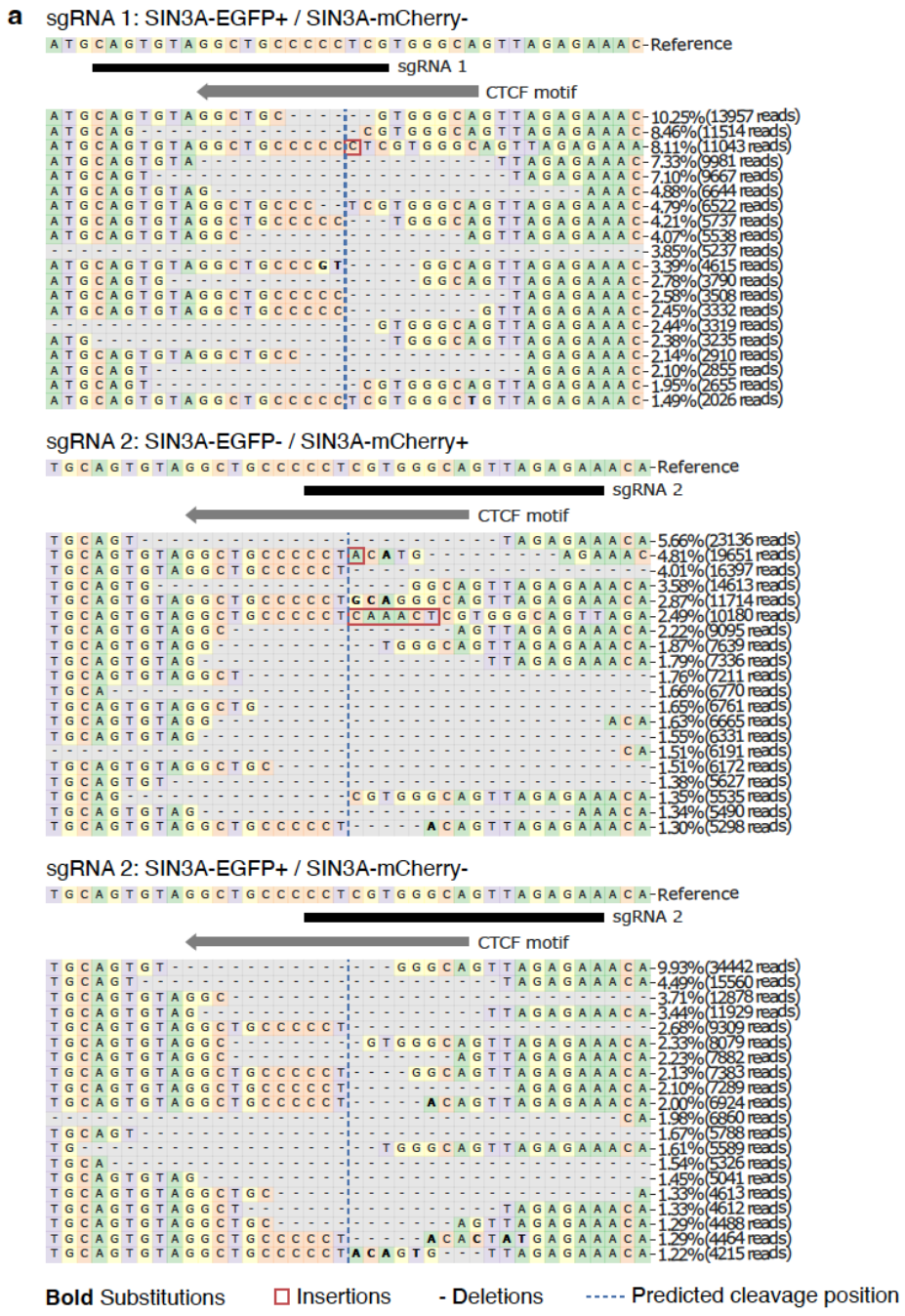
Extended Data Figure 6. *APP* enhancer validation. a, Flow cytometry plots showing the percentage of cells with reduced expression of APP-EGFP or APP-mCherry signals in each condition. The negative control is the WTC11 i^{3}N cells. The positive control is the APP-EGFP/mCherry reporter cells. **b,** Bar graphs showing the significance of the relative enrichment of cells with reduced expression of APP-EGFP or APP-mCherry compared to positive control cells. *P* values were determined using the two-sided Fisher's exact test. * $P < 0.0001$.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



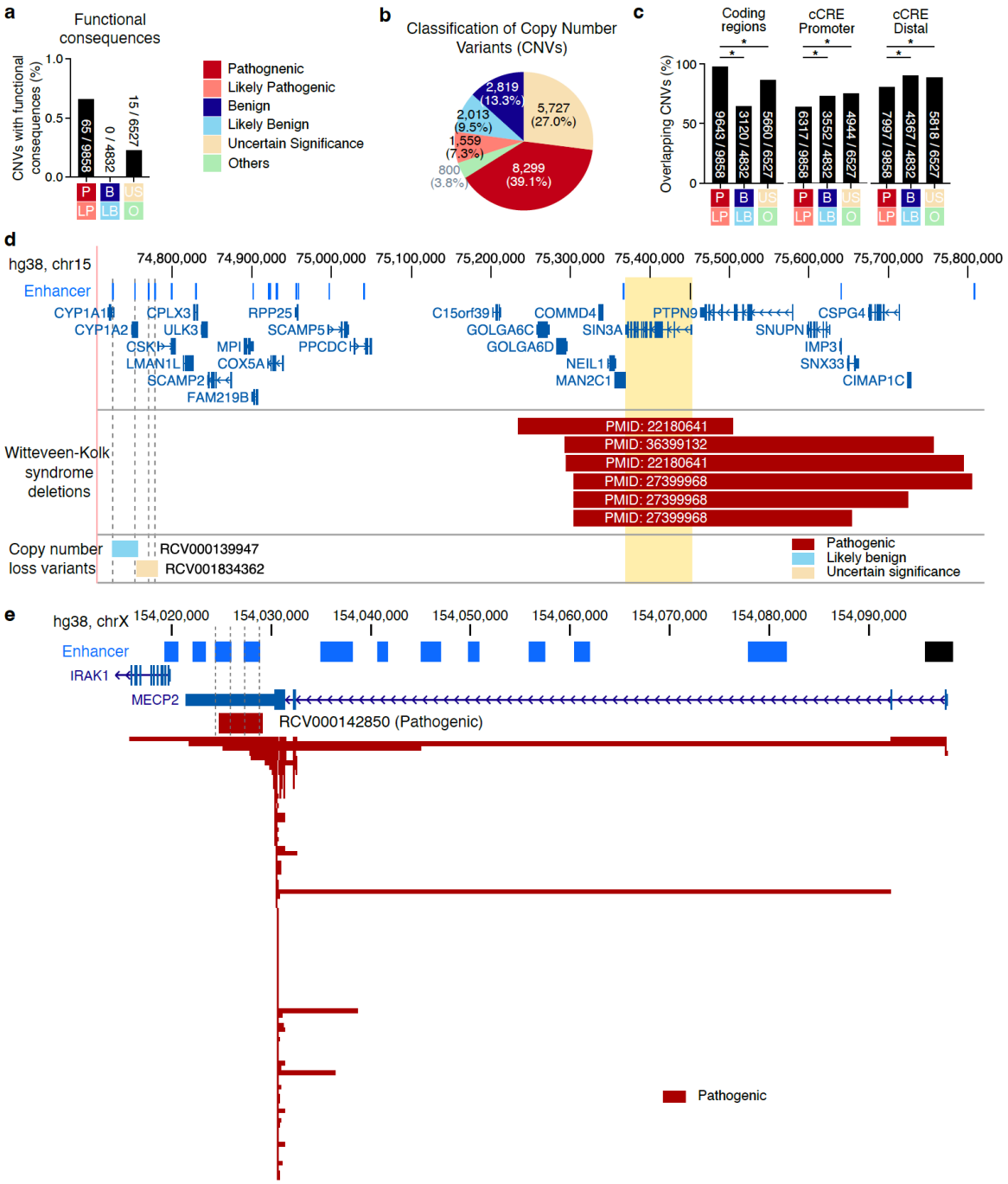
Extended Data Figure 7. *SIN3A* enhancer validations. a,b, Flow cytometry plots showing the percentage of cells with reduced expression of SIN3A-EGFP or SIN3A-mCherry in each condition. The negative control is the WTC11 i^{3}N cells. The positive control is the SIN3A-EGFP/mCherry reporter cells. **c,d,** Bar graphs showing the significance of the relative enrichment of cells with reduced expression of SIN3A-EGFP or SIN3A-mCherry compared to positive control cells. *P* values were determined using the two-sided Fisher's exact test. * $P < 0.0001$.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



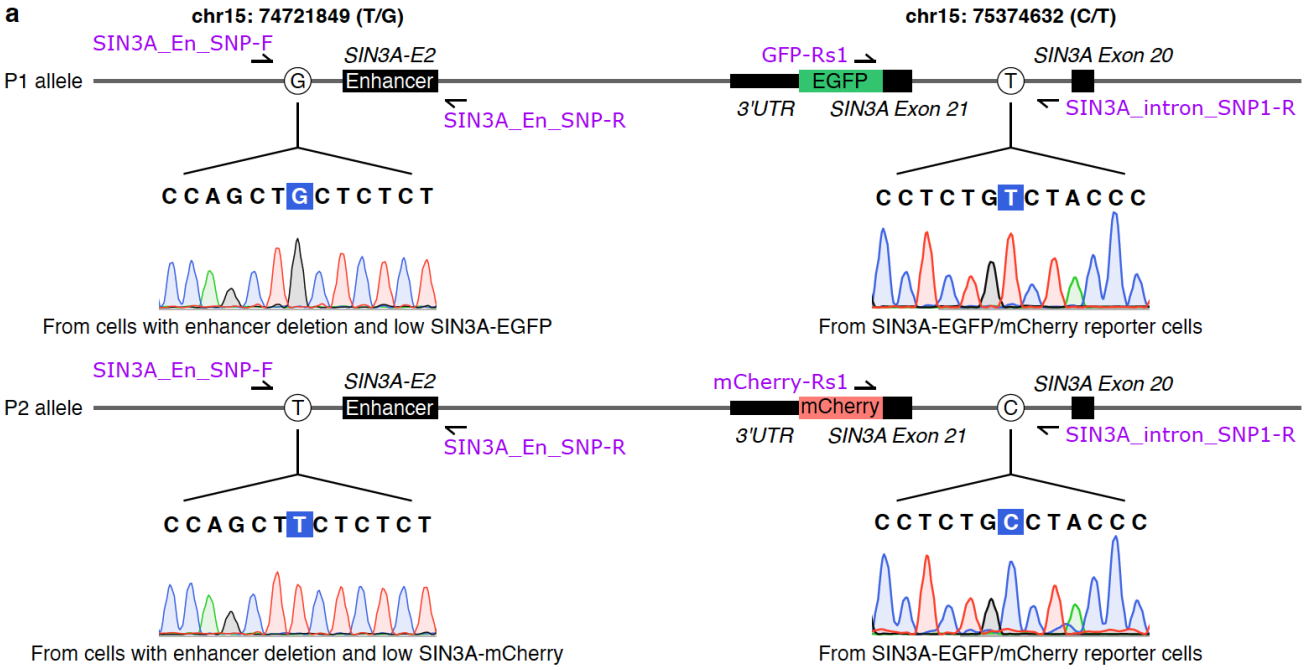
Extended Data Figure 8. Editing outcomes of CTCF sgRNAs. a, CRISPResso2 analysis of the targeted sequencing data shows the genome editing outcomes at the CTCF motif in the cells with reduced expression of SIN3A-EGFP or SIN3A-mCherry.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



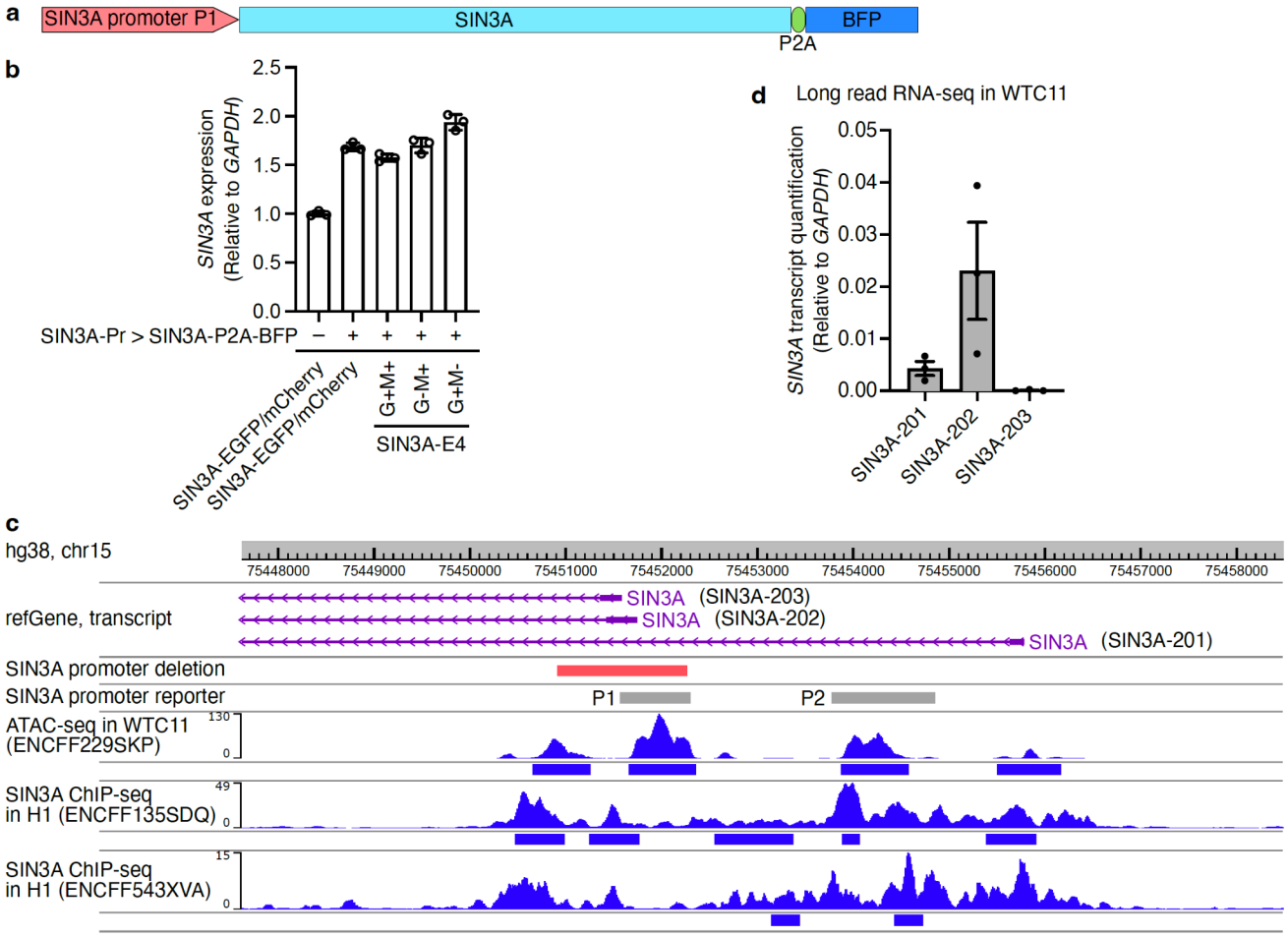
Extended Data Figure 9. The regulatory function of copy number variants. a, The percentage of copy number variants (CNVs) with experimental evidence-based functional consequences. Numbers are displayed in the format of CNVs with functional consequences / total CNVs in each category. **b**, The classification of copy number variants (size ≥ 50bp) in ClinVar. **c**, The overlap between CNVs and coding regions, promoter regions, and distal cCREs. Numbers are displayed in the format of overlapping CNVs / total CNVs in each category. *P* values determined by two-sided Fisher's exact test. * $P < 1\times10^{-15}$. **d**, The overlap between SIN3A enhancers, SIN3A gene, and genetic variants including heterozygous deletions from Witteveen-Kolk syndrome patients and two copy number loss variants in ClinVar. **e**, The overlap between *MECP2* enhancer and copy number variants in *MECP2* locus. In total, 155 clinical deletion/copy number loss variants overlapping with *MECP2* coding regions were interpreted as pathogenic variants and associated with Rett syndrome. RCV000142850 is a 4.3kb copy number loss variant located in the 3'UTR of *MECP2*, and it was interpreted as a pathogenic variant.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



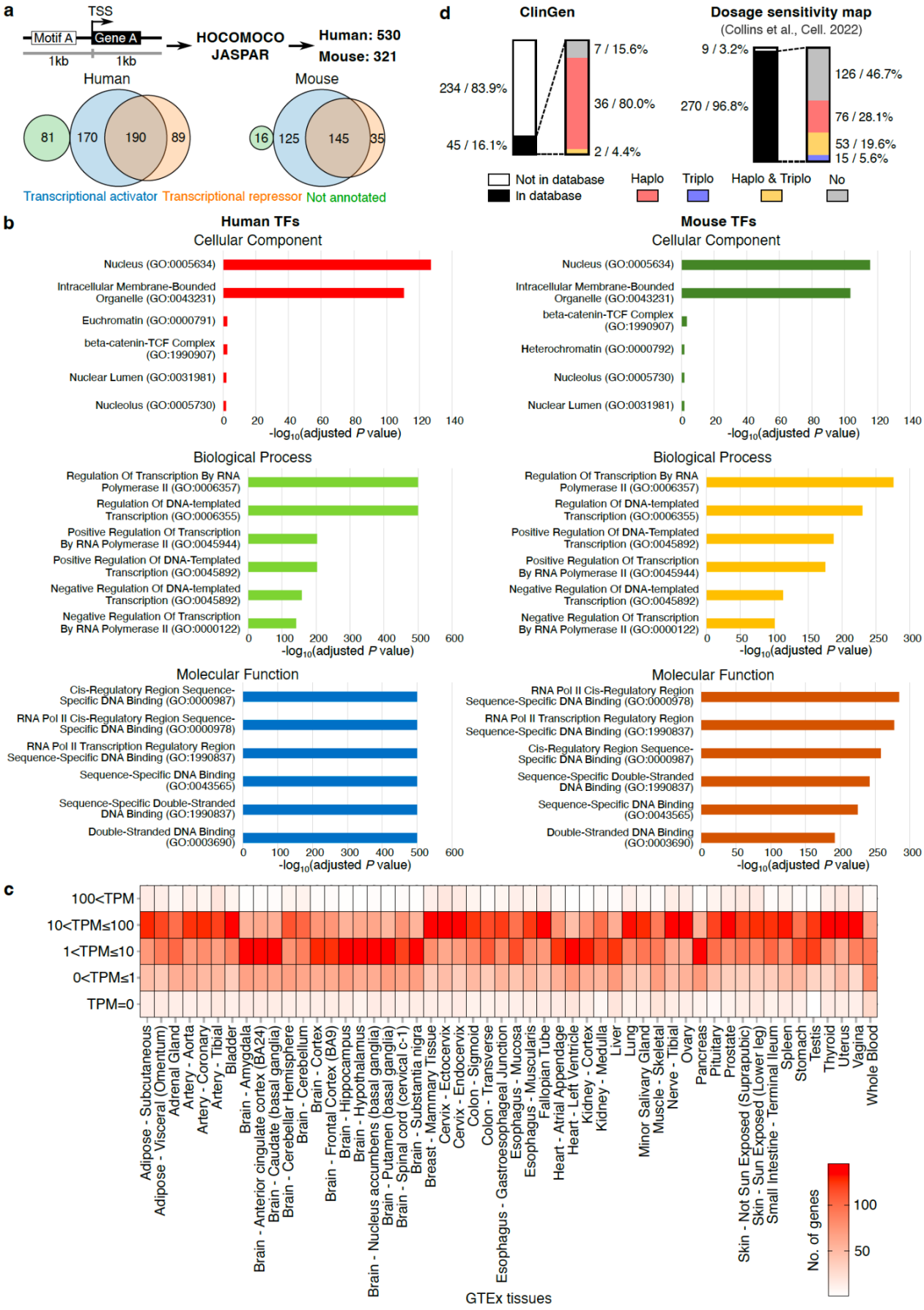
Extended Data Figure 10. *cis*-regulation of *SIN3A* by the SIN3A-E2 enhancer. a, Sanger sequencing data showing the genotype of each allele of *SIN3A* enhancer and *SIN3A*. P1 and P2 alleles are identified using the phased variants in WTC11 genome. Both *SIN3A* enhancer region and *SIN3A* region are amplified using genomic DNA from indicated cells, and the phased variants in amplified regions are confirmed using Sanger sequencing.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Extended Data Figure 11. *SIN3A* ectopic expression and *SIN3A* promoter reporter assay. a, The *SIN3A* promoter P1 controlled SIN3A-P2A-BFP expression cassette. **b,** RT-qPCR results show the expression levels of *SIN3A* in control condition and overexpression conditions. Data are mean ± SD from three technical replicates. **c,** WashU Epigenome Browser snapshot showing *SIN3A* transcripts from refGene, *SIN3A* promoter deletion region in validation experiments, two promoter regions used for *SIN3A* promoter reporter assay, ATAC-seq signal in WTC11 iPSCs, and *SIN3A* ChIP-seq signals in H1 cells. **d**, The expression of *SIN3A* transcripts from long read RNA-seq data in WTC11 cells. Data are mean ± SEM from three biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Figure 12. Transcriptional compensation is associated with gene dosage sensitivity. a, The strategy used for identifying candidate genes with transcriptional compensation. Venn diagrams show the distribution of transcriptional activators and transcriptional repressors in 530 human transcription factors (TFs) and 321 mouse TFs. **b,** The significant enrichment of human and mouse TFs in cellular component, biological process, and molecular function. **c,** The expression of the identified candidate transcriptional compensation genes (transcriptional repressor) in human tissues. The expression data were obtained from GTEx. **d,** The distribution of identified candidate transcriptional compensation genes in ClinGen and Dosage sensitivity map. Haplo: haploinsufficiency. Triplo: triplosensitivity.

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reference:

1. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725–1735 (2003).
2. Long, H. K. *et al.* Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell* **27**, 765-783.e14 (2020).
3. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001.e19 (2021).
4. ENCODE Project Consortium *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
5. Collins, R. L. *et al.* A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**, 3041-3055.e25 (2022).
6. Thinakaran, G. & Koo, E. H. Amyloid precursor protein trafficking, processing, and function. *J Biol Chem* **283**, 29615–29619 (2008).
7. Guerreiro, R., Brás, J. & Hardy, J. SnapShot: genetics of Alzheimer's disease. *Cell* **155**, 968-968.e1 (2013).
8. Sodhi, D. K. & Hagerman, R. Fragile X Premutation: Medications, Therapy and Lifestyle Advice. *Pharmgenomics Pers Med* **14**, 1689–1699 (2021).
9. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905–914 (1992).
10. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**, 185–188 (1999).
11. Pejhan, S. & Rastegar, M. Role of DNA Methyl-CpG-Binding Protein MeCP2 in Rett Syndrome Pathobiology and Mechanism of Disease. *Biomolecules* **11**, 75 (2021).
12. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).
13. Balasubramanian, M. *et al.* Comprehensive study of 28 individuals with SIN3A-related disorder underscoring the associated mild cognitive and distinctive facial phenotype. *Eur J Hum Genet* **29**, 625–636 (2021).
14. Witteveen, J. S. *et al.* Haploinsufficiency of MeCP2-interacting transcriptional co-repressor SIN3A causes mild intellectual disability by affecting the development of cortical integrity. *Nat Genet* **48**, 877–887 (2016).
15. Diao, Y. *et al.* A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* **14**, 629–635 (2017).
16. Wang, C. *et al.* Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. *Stem Cell Reports* **9**, 1221–1233 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Fiaux, P. C., Chen, H. V., Chen, P. B., Chen, A. R. & McVicker, G. Discovering functional sequences with RELICS, an analysis method for CRISPR screens. *PLoS Comput Biol* **16**, e1008194 (2020).
18. Yang, X. *et al.* Functional characterization of gene regulatory elements and neuropsychiatric disease-associated risk loci in iPSCs and iPSC-derived neurons. Preprint at https://doi.org/10.1101/2023.08.30.555359 (2023).
19. Wu, W. *et al.* Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593**, 440–444 (2021).
20. Li, Y. E. *et al.* A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023).
21. Mannion, B. J. *et al.* Uncovering Hidden Enhancers Through Unbiased *In Vivo* Testing. Preprint at https://doi.org/10.1101/2022.05.29.493901 (2022).
22. Diao, Y. *et al.* A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* **14**, 629–635 (2017).
23. Rajagopal, N. *et al.* High-throughput mapping of regulatory DNA. *Nat Biotechnol* **34**, 167–174 (2016).
24. Popay, T. M. & Dixon, J. R. Coming full circle: On the origin and evolution of the looping model for enhancer-promoter communication. *J Biol Chem* **298**, 102117 (2022).
25. Panigrahi, A. & O'Malley, B. W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol* **22**, 108 (2021).
26. Witteveen, J. S. *et al.* Haploinsufficiency of MeCP2-interacting transcriptional co-repressor SIN3A causes mild intellectual disability by affecting the development of cortical integrity. *Nat Genet* **48**, 877–887 (2016).
27. Witteveen, J. S. *et al.* Haploinsufficiency of MeCP2-interacting transcriptional co-repressor SIN3A causes mild intellectual disability by affecting the development of cortical integrity. *Nat Genet* **48**, 877–887 (2016).
28. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001.e19 (2021).
29. Pardo-Palacios, F. J. *et al.* Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nat Methods* **21**, 1349–1363 (2024).
30. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813–1831 (2012).
31. Mefford, H. C. *et al.* Further clinical and molecular delineation of the 15q24 microdeletion syndrome. *J Med Genet* **49**, 110–118 (2012).
32. Coenen-van der Spek, J. *et al.* DNA methylation episignature for Witteveen-Kolk syndrome due to SIN3A haploinsufficiency. *Genet Med* **25**, 63–75 (2023).

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33. Witteveen, J. S. *et al.* Haploinsufficiency of MeCP2-interacting transcriptional co-repressor SIN3A causes mild intellectual disability by affecting the development of cortical integrity. *Nat Genet* **48**, 877–887 (2016).
34. Kulakovskiy, I. V. *et al.* HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* **46**, D252–D259 (2018).
35. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **52**, D174–D182 (2024).
36. Melé, M. *et al.* Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
37. Rehm, H. L. *et al.* ClinGen--the Clinical Genome Resource. *N Engl J Med* **372**, 2235–2242 (2015).
38. Collins, R. L. *et al.* A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**, 3041-3055.e25 (2022).
39. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613–626 (2012).
40. Seidman, J. G. & Seidman, C. Transcription factor haploinsufficiency: when half a loaf is not enough. *J Clin Invest* **109**, 451–455 (2002).
41. Naqvi, S. *et al.* Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nat Genet* **55**, 841–851 (2023).
42. Farley, E. K. *et al.* Suboptimization of developmental enhancers. *Science* **350**, 325–328 (2015).
43. Diao, Y. *et al.* A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* **14**, 629–635 (2017).
44. Grosveld, F., van Staalduinen, J. & Stadhouders, R. Transcriptional Regulation by (Super)Enhancers: From Discovery to Mechanisms. *Annu Rev Genomics Hum Genet* **22**, 127–146 (2021).
45. Diao, Y. *et al.* A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res* **26**, 397–405 (2016).
46. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).
47. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr Biol* **20**, 1562–1567 (2010).
48. Lin, X. *et al.* Nested epistasis enhancer networks for robust genome regulation. *Science* **377**, 1077–1085 (2022).
49. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol Cell* **66**, 285-299.e5 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

50. El-Brolosy, M. A. *et al.* Genetic compensation triggered by mutant mRNA degradation. *Nature* **568**, 193–197 (2019).
51. Vande Zande, P., Siddiq, M. A., Hodgins-Davis, A., Kim, L. & Wittkopp, P. J. Active compensation for changes in TDH3 expression mediated by direct regulators of TDH3 in Saccharomyces cerevisiae. *PLoS Genet* **19**, e1011078 (2023).
52. Pardo-Palacios, F. J. *et al.* Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nat Methods* **21**, 1349–1363 (2024).
53. Ren, X., Takagi, M. A. & Shen, Y. Efficient bi-allelic tagging in human induced pluripotent stem cells using CRISPR. *STAR Protoc* **4**, 102084 (2023).
54. Ren, X. *et al.* Parallel characterization of cis-regulatory elements for multiple genes using CRISPRpath. *Sci Adv* **7**, eabi4360 (2021).
55. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
56. Ren, X. *et al.* Parallel characterization of cis-regulatory elements for multiple genes using CRISPRpath. *Sci Adv* **7**, eabi4360 (2021).
57. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
58. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**, 224–226 (2019).
59. Landrum, M. J. *et al.* ClinVar: improvements to accessing data. *Nucleic Acids Res* **48**, D835–D844 (2020).
60. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001.e19 (2021).
61. Inoue, F. *et al.* A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* **27**, 38–52 (2017).
62. Klehr, D., Maass, K. & Bode, J. Scaffold-attached regions from the human interferon beta domain can be used to enhance the stable expression of genes under the control of various promoters. *Biochemistry* **30**, 1264–1270 (1991).
63. Kwaks, T. H. J. *et al.* Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. *Nat Biotechnol* **21**, 553–558 (2003).
64. Parra Bravo, C. *et al.* Human iPSC 4R tauopathy model uncovers modifiers of tau propagation. *Cell* **187**, 2446-2464.e22 (2024).
65. Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165–D173 (2022).
66. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2024.10.08.616922; this version posted October 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

67. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).