Rare Genetic Variants in Antisense Long Non-coding RNAs Contribute to Obsessive-Compulsive Disorder

Authors: Seulgi Jung[1,2], Madison Caballero[1,2], Shelby Smout[1,2], Behrang Mahjani[1-7]

Corresponding author: Behrang Mahjani, PhD, One Gustave L. Levy Place, Box 1230, New York, NY 10029, behrang.mahjani@mssm.edu

Affiliations:

[1]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[4]Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[5]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[6]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm

[7]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Word Count: 2000

Keywords: OCD, lncRNA, antisense lncRNAs, non-coding

Abstract

Obsessive-compulsive disorder (OCD) is a prevalent neuropsychiatric disorder with an incompletely characterized genetic basis, hindering targeted treatment approaches. Using whole-genome sequencing from the *All of Us* cohort (2,276 OCD cases, 13,517 controls), we assessed rare variants within antisense long non-coding RNAs (lncRNAs). OCD cases exhibited significant enrichment of these variants, particularly within evolutionarily constrained antisense transcripts (relative risk = 1.52, P = 0.005). Burden analyses identified two lncRNAs robustly associated with OCD: *A2ML1-AS1* (OR = 4.5, P = 0.001) and *NFIB-AS1* (OR = 3.6, P = 0.002). Rare variants in these lncRNAs localized to regulatory regions influencing adjacent neurodevelopmental genes and correlated with gene expression in brain regions implicated in OCD. Given their critical roles in gene regulation, tissue-specific expression, and responsiveness to cellular states, antisense lncRNAs represent promising biomarkers and therapeutic targets. Our findings expand the genetic architecture of OCD and underscore the potential of RNA-based precision therapies for personalized clinical interventions.

Introduction

Obsessive-compulsive disorder (OCD), characterized by intrusive thoughts and repetitive behaviors, affects 1–3% of the global population and typically manifests in childhood. Current therapeutic options offer limited efficacy, partly due to an incomplete understanding of the disorder's biological basis.[1–3] Highly conserved neurodevelopmental processes and cortico-striatal-thalamo-cortical circuits are strongly implicated in OCD pathophysiology.[4,5] However, while genome-wide association studies have identified several protein-coding risk loci, these explain only a fraction of OCD heritability, suggesting contributions from unexplored regulatory elements.[6]

Long non-coding RNAs (lncRNAs), transcripts longer than 200 nucleotides without protein-coding potential, have emerged as critical regulators of gene expression through mechanisms including chromatin remodeling, enhancer modulation, RNA processing, and nuclear architecture.[7,8] Approximately 40% of identified lncRNAs display tissue-specific expression in the mammalian nervous system, underscoring their roles in neurodevelopment and brain function.[9,10] Increasing evidence implicates dysregulated lncRNAs in various neuropsychiatric disorders, including autism spectrum disorder,[11] Rett syndrome,[12] schizophrenia,[13] and fragile X syndrome[14] highlighting the potential of these transcripts in shaping neurodevelopmental pathways.

Antisense lncRNAs originate from the opposite strand of protein-coding genes and exhibit varying degrees of overlap, making them particularly intriguing for their precise cis-regulatory effects on adjacent coding genes.[15,16] Antisense lncRNAs modulate the expression of proximal genes through a variety of mechanisms, including transcriptional interference, where antisense transcription impedes the progression of RNA polymerase on the sense strand, and epigenetic modifications, such as the recruitment of chromatin-modifying enzymes that alter histone modifications or DNA methylation patterns, thereby influencing gene expression.[17] The regulatory specificity of antisense lncRNAs positions them as attractive therapeutic targets, amenable to interventions such as antisense oligonucleotides,[18] small interfering RNAs,[19] or CRISPR-based RNA editing,[20] potentially offering enhanced specificity and reduced side effects compared to traditional pharmacotherapies.[3,21]

Despite growing appreciation for lncRNAs' roles in brain function,[22,23] their genetic underpinnings in childhood-onset psychiatric disorders remain largely unexplored.[24] Addressing this critical gap, we analyzed whole-genome sequencing data from the All of Us Research Program[25] to systematically identify rare genetic variants within antisense lncRNAs associated with OCD. Our analyses revealed significant enrichment of variants within specific antisense lncRNAs (*A2ML1-AS1* and *NFIB-AS1*), providing novel insights into OCD pathogenesis. By delineating how dysregulated antisense lncRNAs influence neurodevelopmental gene expression, our findings open new possibilities for targeted therapeutic strategies and precision medicine in childhood neuropsychiatric disorders.

Results

Our primary goal was to characterize the landscape of rare genetic variants (minor allele frequency <0.1%) within antisense lncRNAs and evaluate their association with OCD risk. We acquired whole-genome sequencing (WGS) data from 2,276 OCD-affected individuals and 13,517 controls from the All of Us Research Program (Supplementary Figure 1a-c, see Methods). The mean age of OCD onset was 12.4 years (SD = 4.7), and 54.3% of participants were assigned female at birth.

We identified rare variants within antisense lncRNAs from the Ensembl database or immediately 5kb up or downstream to capture *cis*-regulatory elements. After rigorous quality control, including filtering for sequencing depth (>10x), mapping quality (>20), and other variant quality metrics, we identified 2,692 rare variants within 100 total antisense lncRNAs. Among the total rare variants, 61.5% were intronic (lncRNAs are subject to splicing mechanisms),[26] 16.5% were downstream, 13.7% were upstream, and 8.3% were exonic (Figure 1).

We observed a significant enrichment of total rare variants in antisense lncRNA among individuals with OCD compared to matched controls (relative risk = 1.10, 95% CI: 1.01–1.21, P = 0.03, binomial test). Using the loss-of-function observed/expected upper bound fraction (LOEUF) scores for the lncRNA's cognate sense gene,[27] we identified a significant enrichment of genes in the first decile, representing the most evolutionarily constrained genes (relative risk = 1.52, 95% CI: 1.13–2.03, *P* = 0.005, binomial test; Figure 2a). Results were consistent when using AlphaMissense scores[28] to grade cognate sense gene conservation (relative risk = 1.40, 95% CI: 1.02–1.88, *P* = 0.03, binomial test; Figure 2b).

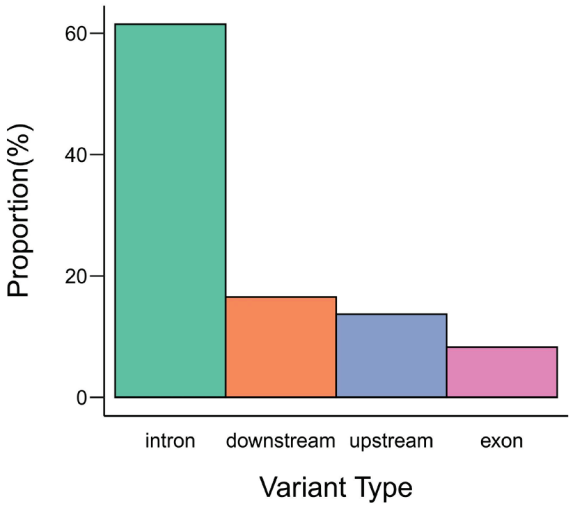

Figure 1. Classification of rare variants in antisense lncRNAs. Proportion of rare variants by genomic annotation, including 1,655 intronic (61.5%), 445 downstream (16.5%), 369 upstream (13.7%), and 223 exonic (8.3%) variants.

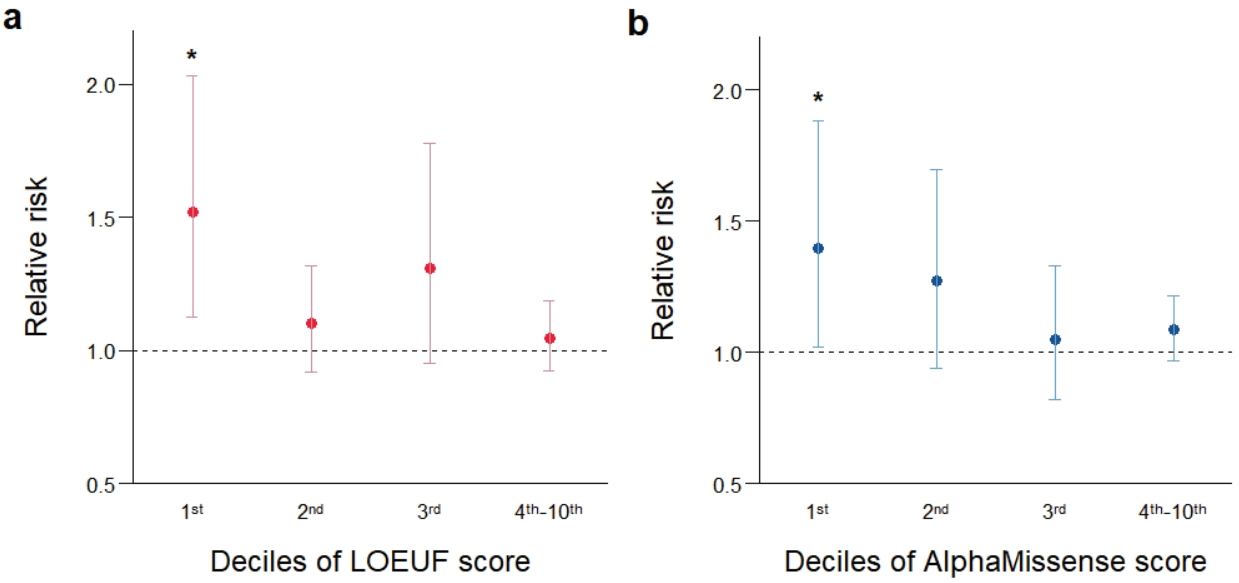

Figure 2. Enrichment of antisense lncRNA rare variants in OCD. Relative risk of rare variants in antisense lncRNAs across deciles of (a) LOEUF and (b) AlphaMissense scores. The first decile includes most constrained genes. Both panels show risk ratios with 95% confidence intervals. The dashed horizontal line represents relative risk = 1, indicating no effect. Points above this line suggest increased risk. $^{*}P<0.05$ using a binomial test between OCD cases and unaffected controls.

Antisense lncRNAs associated with the risk of OCD

To identify risk antisense lncRNAs for OCD, we performed a burden test comparing cases and controls based on cumulative rare variant counts (Supplementary Table 1; see Methods). Among antisense transcripts analyzed, two emerged with particularly robust association signals with false discovery rate < 0.1: *A2ML1-AS1* and *NFIB-AS1* (Figure 3 and Supplementary Figure 2a and b). Rare variants, clustering at the 3′ end of *A2ML1-AS1* (chr12:8,782,890–8,783,100 bp, GRCh38), were significantly enriched in OCD cases (OR = 4.5, P = 0.001). Notably, this region corresponds to a canonical 5′ UTR of the nearby cognate sense *RIMKLB* transcript and exhibits DNase I hypersensitivity in brain tissues, suggesting potential regulatory activity, such as enhancer, promoter, or transcription factor binding sites (Supplementary Figure 3a). Rare variants in *NFIB-AS1* similarly exhibited a significant burden in OCD cases compared to controls (OR = 3.6, P = 0.002). These variants, localized to chr9:14,398,500–14,399,000 bp, overlapped a rarely transcribed, brain-specific non-canonical 5′ UTR of the cognate sense *NFIB* gene and a region of open chromatin in motor neuron cells and brain tissues (Supplementary Figure 3b), suggesting a potential regulatory role in *NFIB* expression.

Due to direct overlap, rare variant burden in antisense lncRNAs may primarily reflect regulatory elements of protein-coding genes, such as 5′ UTRs. Therefore, to further implicate the antisense

lncRNAs, we extended our burden analysis to include *A2ML1*, *NFIB*, and *RIMKLB* transcripts with the same 5 kb extensions. This analysis revealed that rare variants in the extended regions of those three genes were not significantly enriched in OCD cases (Supplementary Table 2), suggesting that the rare variant association is likely driven by the antisense lncRNAs rather than broader regulatory regions of the protein-coding genes.

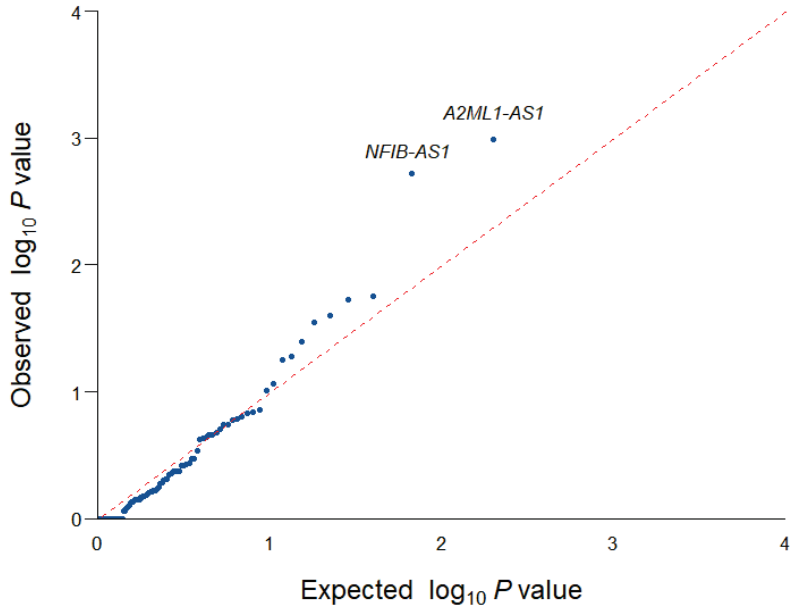

Figure 3. Quantile-quantile plot of *P* values from the burden test. This quantile-quantile plot displays the observed -$\log_{10}(P)$ values against the expected -$\log_{10}(P)$ values under the null hypothesis. The red dashed line represents the expected distribution under the assumption of no association. Dots with gene names showed a false discovery rate < 0.1 in the burden test.

Gene expression correlation analysis

To determine whether *A2ML1-AS1* and *NFIB-AS1* and their respective sense genes are expressed in brain tissues, we analyzed single-cell RNA sequencing data using the CellxGene platform (https://cellxgene.cziscience.com/gene-expression). Both antisense lncRNAs and their sense genes were expressed across brain tissues supporting biological relevance (Supplementary Figure 4).

To evaluate the correlation between the expression levels of two risk-associated antisense lncRNAs for OCD and their respective sense genes, we utilized gene expression profiling interactive analysis (GEPIA). Using gene expression data from 13 brain tissues in the GTEx database, we performed Pearson correlation tests for the following pairs: *A2ML1-AS1* and *A2ML1*, *A2ML1-AS1* and *RIMKLB*, and *NFIB-AS1* and *NFIB*. Notably, *A2ML1-AS1* expression exhibited a significant positive correlation with *RIMKLB* but not with *A2ML1* (Supplementary

Figure 5a,b). *NFIB-AS1* expression demonstrated a significant positive correlation with *NFIB* (Supplementary Figure 5c).

Discussion

This study identified antisense lncRNAs as novel genetic contributors to OCD. The burden of rare variants within antisense lncRNAs, particularly those with highly evolutionarily constrained cognate sense genes, suggest these non-coding elements modulate essential neurodevelopmental pathways underlying OCD pathophysiology. We further identified *A2ML1-AS1* and *NFIB-AS1* as significantly associated with OCD risk, positioning them as high-priority candidates for mechanistic investigation.

Intriguingly, our expression correlation analysis revealed that *A2ML1-AS1* positively correlates with *RIMKLB* rather than *A2ML1*, highlighting the nuanced regulatory activity of lncRNAs. In contrast, *NFIB-AS1* exhibits a strong positive correlation with its cognate sense gene *NFIB*, consistent with a classical *cis*-regulatory role. *NFIB* encodes a transcription factor essential for proper brain development and neuronal differentiation,[29] and its dysregulation has been implicated in multiple neuropsychiatric disorders.[30–34]

While previous genome-wide association studies have primarily identified common variants with modest effects on OCD risk,[6] our investigation highlights rare variants in non-coding regions that may exert stronger functional consequences. The substantial enrichment of these variants in highly constrained genes suggests that antisense lncRNAs associated with essential neurodevelopmental genes are particularly relevant to OCD etiology. Although lncRNAs have been implicated in several other neuropsychiatric conditions, the specific role of antisense lncRNAs in OCD has remained largely unexplored until now. Our study provides the first systematic evidence linking specific antisense lncRNA variants to OCD risk, distinguishing our work from previous investigations in this field.

Several important limitations of our study should be acknowledged. Replication in larger, more diverse cohorts will be essential to validate our findings and ensure broader generalizability across populations. Rigorous experimental validation will also be critical for definitively determining whether the identified variants functionally alter antisense lncRNA activity and downstream gene expression. Furthermore, longitudinal clinical studies will be required to delineate precisely how antisense lncRNA variants influence OCD onset, progression, symptom dimensions, and treatment response patterns. Given variant rarity, investigations employing patient-derived induced pluripotent stem cells and brain organoids are ideal systems to provide crucial mechanistic insights and facilitate screening of potential therapeutic interventions.

The recent FDA approval of antisense therapeutics for neurological disorders such as spinal muscular atrophy[35] and ongoing clinical trials for conditions including Duchenne muscular dystrophy[36] and amyotrophic lateral sclerosis[37] highlight the clinical feasibility of translating these approaches into effective treatments for OCD. Patients with treatment-resistant OCD or those

belonging to genetically defined subgroups could particularly benefit from such targeted interventions.

Methods

Study population and data acquisition

The All of Us Research Program, led by the National Institutes of Health (NIH), is a pioneering precision medicine initiative designed to advance health research and medical discoveries by collecting diverse data from at least one million participants across the United States.[38] In the current All of Us Research Program workspace (controlled tier, version 8), we identified samples diagnosed with OCD, of which 2,764 had available whole-genome sequencing (WGS) data. OCD diagnoses were identified using International Classification of Diseases, 10th Revision (ICD-10) codes (F42.0–F42.9). For case-control matching, we selected these 2,764 OCD cases and 271,481 individuals without mental disorders (ICD-10 F01-F99) from the WGS dataset. Using the R package optmatch, we performed 1:5 case-control matching based on sex and principal components (PCs) 1–5, which were derived from WGS data and provided by the All of Us Research Program.

Whole-genome sequencing and variant calling

From the whole-genome sequencing (WGS) dataset of 414,830 samples, we retained 2,764 cases and 13,820 matched controls after applying quality control (QC) filters. Individuals were excluded if they had fewer than 2.4 million or more than 5.0 million total variants, over 100,000 variants absent from gnomAD v3.1,[27] or a heterozygous-to-homozygous variant ratio exceeding 3.3.

We further filtered out low-quality variants based on the following criteria: genotype quality (GQ) ≤ 20, read depth (DP) ≤ 10, allele balance (AB) ≤ 0.2 for heterozygotes, ExcessHet < 54.69, and quality scores (QUAL) < 60 for single nucleotide variants (SNVs) and < 69 for short insertions and deletions (InDels). Variants located in low-complexity regions were also removed. Additional QC steps excluded samples with call rates more than three standard deviations below the mean, individuals with close genetic relatedness (kinship coefficient > 0.1) based on Kinship-based INference for Genome-wide association studies (KING),[39] and those with sex discrepancies.

After filtering, 2,276 cases and 13,517 controls remained. Further exclusions included homozygous reference calls with < 90% read depth supporting the reference allele or GQ < 25, heterozygous calls with AB < 30%, GQ < 25, or a binomial probability > 1×10^{-9} for the observed allele balance centered at 0.5, and homozygous alternate calls with < 90% read depth supporting the alternate allele or GQ < 25. Variants with a call rate ≤ 90% and Hardy-Weinberg equilibrium P < 1×10^{-12} were also removed.

Annotation of antisense lncRNAs and variant selection

We utilized the Variant Effect Predictor (VEP) to predict the potential functional effects of the variants.[40] To filter rare variants, we included those with an allele count of ≤ 5 in both our case-control cohort and the non-psychiatric subset of the gnomAD database v3.1.[27] Antisense long non-coding RNAs (lncRNAs) were annotated using the Ensembl database.[41] Transcripts were included if they met the following criteria: ≥200 nucleotides in length, lacking protein-coding potential, and transcribed from the antisense strand of annotated protein-coding genes.

Rare variants located within ±5 kb of antisense lncRNA transcripts were extracted for analysis. Variants were categorized as exonic, intronic, upstream, or downstream based on their position relative to annotated transcript structures. Variants in low-complexity regions or those with poor sequencing coverage were excluded from further analysis.

Statistical analysis

Rare variant burden analyses were conducted using Fisher's exact tests to compare the cumulative number of rare variants in each antisense lncRNA between OCD cases and controls. Relative risk ratios (RRs) with 95% confidence intervals (CIs) were calculated to quantify the enrichment of rare variants in OCD cases relative to controls using and binomial tests. False discovery rates (FDRs) were controlled using the Benjamini–Hochberg method, with an FDR threshold of <0.1 defining statistically significant associations. Odds ratios (ORs) and 95% CIs were also computed to further quantify associations.

For gene constraint analysis, cognate sense genes were stratified into deciles based on their loss-of-function observed/expected upper bound fraction (LOEUF) scores (range: 0–2, where lower scores indicate stronger selective constraint)[27] and AlphaMissense scores (range: 0–1, where higher scores reflect stronger selective pressure against missense variants).[28] Enrichment analyses were performed within each decile to identify differential burden patterns among OCD-associated antisense lncRNAs using binomial tests.

Transcriptomic analyses

To explore cell type-specific gene expression patterns in neural tissue, we leveraged single-cell RNA sequencing data accessed through the CELLxGENE platform (https://cellxgene.cziscience.com/gene-expression).[42] All 15 ADHD risk genes were profiled across diverse neural cell populations, including neurons and glial cells, to provide insight into their potential roles in ADHD pathophysiology.

To investigate cell type-specific gene expression patterns, we utilized the CELLxGENE platform (https://cellxgene.cziscience.com/gene-expression) to analyze publicly available single-cell RNA sequencing (scRNA-seq) data.[42] This platform aggregates and standardizes large-scale single-cell datasets, allowing for the exploration of gene expression across diverse tissues and cell types. We queried the expression of *NFIB*, *RIMKLB*, *A2ML1*, *RP11-120J1.1* (*NFIB-AS1*), and

A2ML1-AS1 across multiple tissues, including brain, cortex, blood, bone marrow, embryo, and head. The analysis included data from millions of single cells, and expression levels were assessed by gene-specific detection in each tissue. For visualization, gene expression levels were represented using dot plots, where the color intensity reflected normalized gene expression, and the dot size indicated the proportion of cells expressing the gene.

To evaluate the correlation between the expression levels of antisense lncRNAs and their respective sense genes, we utilized the gene expression profiling interactive analysis (GEPIA) platform.[43] GEPIA integrates transcriptomic data from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) project, providing normalized gene expression values for correlation analysis.[44–46]

For each gene pair, we extracted transcript per million (TPM)-normalized expression values from 13 brain tissues in the GTEx database and performed Pearson correlation analysis using GEPIA's Gene Correlation Analysis module. Log_2-transformed TPM values were used to generate scatter plots, where each dot represents an individual sample. The Pearson correlation coefficient (R) and corresponding *P* value were computed to assess statistical significance. Statistical significance was defined as $P < 0.05$.

Acknowledgements: This study was supported by a grant from Beatrice and Samuel A. Seaver Foundation (BM).

The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the All of Us Research Program would not be possible without the partnership of its participants.

Author contributions:

Study concept and design: SJ, BM

Acquisition, analysis, or interpretation of data: MC, SJ, BM

Drafting of the manuscript: All authors

Critical revision of the manuscript for important intellectual content: All authors

Statistical analysis: MC, SJ, BM

Obtained funding: BM
Study supervision: BM

Competing interests: The authors declare no competing interests.

Data availability: All of Us controlled tier data is available for registered institutions and cohort selection is detailed in the workbench titled "Detecting the prevalence of ultra-rare gene mutations." Due to stringent privacy protections and data sharing restrictions enforced by the All of Us Research Program, individual-level variant data cannot be shared publicly. Researchers must adhere to specific guidelines to protect participant confidentiality, which includes not reporting or disseminating any data that could allow the re-identification of participants, particularly those involving participant counts between 1 and 20 without employing approved data obscuration methods.

Code availability: Code and resources used in this study is available on GitHub at (https://github.com/MahjaniLab/OCD_WGS). For any further inquiries or requests for code not available in the repository, please contact the corresponding author.

References:

1. Ruscio, A. M., Stein, D. J., Chiu, W. T. & Kessler, R. C. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Mol. Psychiatry* **15**, 53–63 (2010).
2. Mahjani, B., Bey, K., Boberg, J. & Burton, C. Genetics of obsessive-compulsive disorder. *Psychol. Med.* **51**, 2247–2259 (2021), PCMID: PMC8477226.
3. Hirschtritt, M. E., Bloch, M. H. & Mathews, C. A. Obsessive-compulsive disorder: Advances in diagnosis and treatment. *JAMA* **317**, 1358–1367 (2017), PMID: 28384832.
4. Mataix-Cols, D., Cullen, K. & Lange, K. Neural substrates of symptoms of obsessive-compulsive disorder in children: a developmental perspective. *Br J Psychiatry* **223**, 12–22 (2023).
5. Harrison, B. J. *et al.* Altered corticostriatal functional connectivity in obsessive-compulsive disorder. *Arch. Gen. Psychiatry* **66**, 1189–1200 (2009), PMID: 19884607.
6. Strom, N. I. *et al.* Genome-wide association study identifies 30 obsessive-compulsive disorder associated loci. *Genetic and Genomic Medicine* (2024).
7. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407 (2018), PCMID: PMC5978744.

8. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021), PCMID: PMC7754182.

9. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012), PCMID: PMC3431493.

10. Aprea, J. & Calegari, F. Long non-coding RNAs in corticogenesis: deciphering the non-coding code of the brain. *EMBO J.* **34**, 2865–2884 (2015), PCMID: PMC4687686.

11. Delalle, I., Westphal, H. & Creuzet, S. E. Long noncoding RNAs in neurodevelopment and autism. *Prog Mol Biol Transl Sci* **174**, 181–209 (2020).

12. Petazzi, P. *et al.* Dysregulation of the long non-coding RNA transcriptome in a Rett syndrome mouse model. *RNA Biol.* **10**, 1197–1203 (2013), PCMID: PMC3849168.

13. Cai, C. *et al.* Is human blood a good surrogate for brain tissue in transcriptional studies? *BMC Genomics* **11**, 589 (2010), PCMID: PMC3091510.

14. Pastori, C. *et al.* Comprehensive analysis of the transcriptional landscape of the human FMR1 gene reveals two new long noncoding RNAs differentially expressed in Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome. *Hum. Genet.* **133**, 59–67 (2014), PCMID: PMC3898532.

15. Khorkova, O., Hsiao, J. & Wahlestedt, C. Basic biology and therapeutic implications of lncRNA. *Adv. Drug Deliv. Rev.* **87**, 15–24 (2015), PCMID: PMC4544752.

16. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.* **10**, 637–643 (2009), PCMID: PMC2850559.

17. Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription. *Nature Reviews Genetics* **14**, 880–893 (2013).

18. Watts, J. K. & Corey, D. R. Silencing disease genes in the laboratory and the clinic. *J. Pathol.* **226**, 365–379 (2012), PCMID: PMC3916955.

19. Gavrilov, K. & Saltzman, W. M. Therapeutic siRNA: principles, challenges, and strategies. *Yale J. Biol. Med.* **85**, 187–200 (2012), PCMID: PMC3375670.

20. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280–284 (2017), PCMID: PMC5706658.

21. Pittenger, C. The pharmacological treatment of obsessive-compulsive disorder. *Psychiatr. Clin. North Am.* **46**, 107–119 (2023), PMID: 36740347.

22. Lin, M. *et al.* RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One* **6**, e23356 (2011), PCMID: PMC3168439.

23. Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L. & Barry, G. Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. *Neuron* **88**, 861–877 (2015), PMID: 26637795.

24. Akula, N. *et al.* RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Mol. Psychiatry* **19**, 1179–1185 (2014), PCMID: PMC5560442.

25. All of Us Research Program Investigators. The 'All of Us' Research Program. *N Engl J Med* **381**, 668–676 (2019).

26. Basu, K., Dey, A. & Kiran, M. Inefficient splicing of long non-coding RNAs is associated with higher transcript complexity in human and mouse. *RNA Biol* **20**, 563–572 (2023), PCMID: PMC10405767.

27. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020), PCMID: PMC7334197.

28. Cheng, J. *et al.* Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023), PMID: 37733863.

29. Özkan, A. *In vitro and in silico investigation of NFIB-SUMO interactions*. (2024).

30. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address: douglas.ruderfer@vanderbilt.edu & Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715.e16 (2018), PCMID: PMC6432650.

31. Bigdeli, T. B. *et al.* Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans. *Schizophr Bull* **47**, 517–529 (2021), PCMID: PMC7965063.

32. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022), PCMID: PMC9392466.

33. Dang, X. *et al.* Gene-level analysis reveals the genetic aetiology and therapeutic targets of schizophrenia. *Nat Hum Behav* (2025) doi:10.1038/s41562-024-02091-4, PMID: 39753749.

34. Fu, J. M. *et al.* Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022), PCMID: PMC9653013.

35. Finkel, R. S. *et al.* Nusinersen versus sham control in infantile-onset spinal muscular atrophy. *N. Engl. J. Med.* **377**, 1723–1732 (2017).

36. Mcdonald, C., Camino, E., Escandon, R. & Finkel, R. S. Draft Guidance for Industry: Duchenne Muscular Dystrophy, Becker Muscular Dystrophy, and Related Dystrophinopathies - Developing Potential Treatments for the Future. *Journal of*

Neuromuscular Diseases (2024).

37. Vissing, J. Therapeutic advances in neuromuscular diseases in 2023. *Lancet Neurol.* **23**, 24–25 (2024), PMID: 38101891.
38. All of Us Research Program Investigators *et al.* The 'All of Us' Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019), PCMID: PMC8291101.
39. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010), PCMID: PMC3025716.
40. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016), PCMID: PMC4893825.
41. Harrison, P. W. *et al.* Ensembl 2024. *Nucleic Acids Res* **52**, D891–D899 (2024), PCMID: PMC10767893.
42. CZI Single-Cell Biology Program *et al.* CZ CELL×GENE Discover: A single-cell data platform for scalable exploration, analysis and modeling of aggregated data. (2023) doi:10.1101/2023.10.30.563174.
43. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* **47**, W556–W560 (2019), PCMID: PMC6602440.
44. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113–1120 (2013), PCMID: PMC3919969.
45. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015), PCMID: PMC4547484.
46. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585 (2013), PCMID: PMC4010069.