Title

Mechanisms of genetic variation underlying neurodevelopmental disorder phenotypes in the human cortex

ISBN

9798189299070

Author

Hennick, Kelsey

Publication Date

2026-09-03

Peer reviewed|Thesis/dissertation

Mechanisms of genetic variation underlying neurodevelopmental disorder phenotypes in the human cortex

by
Kelsey Hennick

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Barbara Panning
B18F20197C95417...

Barbara Panning
Chair

DocuSigned by:
Tomasz Nowakowski
DocuSigned by:
A4...

Tomasz Nowakowski

DocuSigned by:
John Rubenstein
DocuSigned by:
450...

John Rubenstein

DocuSigned by:
Helen Willsey
CAA5EAC85A9B44D...

Helen Willsey

Committee Members

To my family, given and chosen. To the mentors and mentees who made me the scientist I am today. To the colleagues and community I found along the way.

ACKNOWLEDGMENTS

The work presented here is the culmination of years of support and encouragement from more people than I can name, though I will give it a try. First, none of this work would have been possible without my advisor Dr. Tomasz Nowakowski. Tom saw something in me (what it was I'm still not sure) and emailed me about meeting at my program retreat my first month of graduate school. I knew immediately I would want to join the lab, and though it's been a journey of the good, the bad, and the ugly, Tom was always there to provide critical guidance and reassurance. I'm particularly grateful for his patience, as I attempted to manage a several collaborations and teach myself bioinformatics, and throughout it all he kept his pestering to a minimum. As I'm on my way out, I can admit that Tom was (almost) always right, and I appreciate the room to explore all paths before ending up on the right one per his suggestion months before. I've learned an indescribable amount about neurodevelopment, grant writing, career development, and much more from Tom's mentorship, and wouldn't be the scientist I am today without his guidance. I came into graduate school with the goal of pursuing an industry job, not thinking I would be competitive for the academic job market. The first time I received encouragement to pursue the academic path was from Tom, so thank you for believing in me and putting me on this path.

Beyond Tom's direct mentorship, I was surrounded by several other faculty who provided invaluable feedback and support throughout graduate school. To my thesis committee, Helen Willsey, John Rubenstein, and Barbara Panning, thank you for your commitment to my success. A special thanks to Helen for having an open door whenever I needed extra support.

Thank you to my qualifying exam committee, Mercedes Paredes, Jeremy Willsey, and Nadav Ahituv, for sparing me the trauma and allowing me to pass.

To the members of the Simons collaboration, Huda Zoghbi, Evan Eichler, Aravinda Chakravarti, and their trainees Yang Sui, Rebecca Meyer-Schuman, Fisher Cherry, Druha Karunakaran, and Hanna Berk-Rauch, thank you for being such an integral and inspiring part of my grad school journey, and for invaluable contributions to the work presented here.

To my other coauthors, Belinda Wang and Rasika Vartak for being incredible collaborators and scientists, and letting me lead part of such a fun and impactful project. Particularly to Belinda, who has been there to provide support with everything from linear mixed models to postdoc decisions, thank you for being such a great mentor.

Thank you to the members of the Nowakowski lab, past and present: Matt Keefe, Chang Kim, David Shin, Galina Popova, Sih-Rong Wu, Marilyn Steyert. I can't overstate how much I have learned from you all and appreciate all the advice you've given me over the years. Particularly Chang who taught me bioinformatics, and who also was always there for late night whiskey in lab. And a special thanks to Matt who was more invested in solving my problems than I was, and for being such a great and consistent part of my life outside of lab.

Thank you to the Willsey lab who adopted me as one of their own and contributed to some of my favorite moments from grad school, particularly Kate McCluskey who made commuting from the east bay much less painful (even fun?).

Thank you to my GRAD210 cohort and particularly D'Anne Duncan, who is more committed to student success than anyone, and who is an integral part of the UCSF community. Thank you for giving me the space to grow in your class and beyond.

To the Tetrad class of 2019, thanks for being there through Bioreg suffering, a global pandemic, and everything else thrown at us over the years. A special thanks to the incredible friends who I

couldn’t have done grad school without, Jess Mella, Anna Morrison, Jocelyne Fadiga, Liz Bond, Fredrick Leon, and Nathan Cho. An additional thank you to Tetrad student Jesslyn Park for being such a committed and dedicated scientist and friend.

Beyond science, friends and family supported me in a myriad of other ways during graduate school. To my consistent and ever-growing community at Bridges Rock Gym, thank you all for giving me such a needed break from academia every Saturday.

To Jayden Ross, Rex Serrano, and Rachel Haeseker, thank you for being the fun and chaotic juxtaposition to grad school I needed to make it through. Couldn’t have done it without you all.

To Hannah Kramer and Cayley Winston-Coren, thanks for being by my side through it all—undergrad, grad school applications, and the entirety of grad school. To say my life wouldn’t be what it is today without you both is an understatement, and I can’t emphasize enough how valuable your support and friendship has been.

To Dan Kramer, I could write a separate thesis on the ways in which you have supported and loved me during this chapter, but I’ll keep it brief and say thank you for all that you’ve done to get me here. I would not have made it to this point without you. And of course thank you to our stinky son Oakley for living up to his title of emotional support animal and creating an unimaginable amount of both chaos and joy in our lives.

Lastly, thank you to my family. To my parents for supporting me in more ways than I can describe, for encouraging me to follow my passions, to take breaks and enjoy the journey, for being such a constant source of emotional support, thank you. To my sister Lindsey, thank you for bringing comedic relief to every family event, for inspiring me to live my best life, for being a wonderful little sister, thank you. And to my grandparents, my Saba and Savta, thank you for showing me

what unwavering dedication and hard work looks like, for creating such a loving and supportive family in the US, for always pushing me to succeed, and for being my biggest supporters along the way. Thank you.

CONTRIBUTIONS

All work was conducted under the supervision of Dr. Tomasz Nowakowski. The work presented in Chapter 2 will be submitted to *Neuron* as a review article in collaboration with Urvashi Thopte and Dr. Cathryn Cadwell; publication is expected in 2027. The work presented in Chapter 3 was published in *Science* and was done in collaboration with Drs. Belinda Wang, Rasika Vartak, Yefim Zaltsman, Zun Zar Chi Naing, Helen Willsey, Ruth Hüttenhain, A. Jeremy Willsey, Kirsten Obernier, Matthew State, and Nevan Krogan. The work presented in Chapter 3 is currently available on bioRxiv and under revision at *Nature* and was done in collaboration with Drs. Yang Sui, Evan Eichler, Huda Zoghbi, and Aravinda Chakravarti; publication is expected in 2026.

Citation for Chapter 2 review:

Hennick K, Thopte U, Cadwell C, Nowakowski TJ. Emerging roles of retinoic acid signaling in human forebrain development and related disorders. (in preparation).

Citation for Chapter 3 manuscript:

Wang B*, Vartak R*, **Hennick KM*,** Zaltsman Y*, Naing ZZC*, *et al.* Autism mutations rewire protein interaction networks to drive neurodevelopmental pathology. *Science* in press (2026).

Citation for Chapter 4 manuscript:

Hennick K*, Sui Y*, *et al.* Sex differences in the developing human cortex intersect with genetic risk of neurodevelopmental disorders. bioRxiv [Preprint]. 2025 Sep 5:2025.09.04.674293. doi: 10.1101/2025.09.04.674293.

"Unfortunately, nature seems unaware of our intellectual need for convenience and unity, and very often takes delight in complication and diversity."

Santiago Ramón y Cajal, 1906

Mechanisms of genetic variation underlying neurodevelopmental disorder phenotypes in the human cortex

Kelsey Hennick

ABSTRACT

Development of the human cortex is a complex process that requires tightly controlled spatiotemporal regulation of gene expression. Genetic variation or altered regulatory dynamics can impact downstream protein function, which often leads to neurodevelopmental disorders (NDDs). These NDDs are widespread in genetic etiology and trait presentation, and often are sex biased in diagnosis. While the last few decades have identified hundreds of high confidence risk genes implicated in these NDDs, the convergent protein biological functions impacted by mutations in these genes, as well as the sex bias in risk, is still unclear. Here, I present two main chapters aimed at understanding the complex genetic mechanisms that give rise to NDDs. First, I show that genetic variation across risk genes gives rise to convergent rewiring of protein-protein interactions, using the FOXP1-FOXP4 interaction as an example. I highlight a novel, necessary role for FOXP4 in driving the developmental phenotypes associated with FOXP1 mutations. Then, I highlight sex differences in the midgestation cortex, identifying critical sex-biased gene expression and regulatory programs that underlie sex-specific mutational burden identified in NDD cohorts. This work is preceded by a literature review of the critical and novel roles of the morphogen retinoic acid during cortex development and its relation to NDD biology. Together, this work takes important steps forward in understanding how genetic variation leads to altered neurodevelopment, and therefore NDD vulnerability, at the molecular, cellular, and developmental scale.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASD	autism spectrum disorder
CP	cortical plate
CTX	cortex
CUT&Tag	cleavage under targets and tagmentation
EN	excitatory neuron
GW	gestational week
hcASD	high confidence autism spectrum disorder (gene)
hiPSC	human induced pluripotent stem cell
IN	inhibitory neuron
IPC	intermediate progenitor cell
NDD	neurodevelopmental disorder
OPC	oligodendrocyte precursor cell
PFC	prefrontal cortex
PPI	protein-protein interaction
RG	radial glia
scRNA-seq	single cell RNA sequencing

snMultiome single nucleus multiome

VZ ventricular zone

CHAPTER 1: INTRODUCTION

The cerebral cortex

The cortex is an incredibly complex and diverse structure that originates from the neuroepithelium lining the lateral ventricles. The tightly controlled expansion and differentiation of these neuroepithelium stem cells eventually gives rise to the human cortex comprised of tens of billions of neurons across several cell types defined by their gene expression, chromatin accessibility, morphology, circuits, and location.[1] The cortex is comprised of 6 layers with distinct cytoarchitecture and cell types. Aside from the outermost layer 1 of the cortex, neurons are generated in an inside out fashion, first populating layer 6, then 5, and so on.[2] Morphogens and growth factors from the ventricles and meninges govern the trajectory of cortical neuron differentiation, and additionally contribute to forebrain patterning during development.[3–7] These extrinsic cues initiate transcriptional cascades within dividing progenitors that signal when to differentiate and what neuronal subtypes to produce.[8–10] Together, the coordination of these complex events builds the human cortex.

The neurons produced by cortical progenitors are almost exclusively excitatory neurons,[2,11–13] while inhibitory neurons are born in the ventral forebrain and migrate into the cortex.[14–18] The balance not only of excitatory and inhibitory neurons, but also of neuronal subtypes across the cortex, is critical for proper brain development and function.[19,20] These excitatory and inhibitory neurons create circuits with each other,[1] as well as with neurons of other brain regions, including the thalamus.[21] Altering the balance or function of both classes of neurons can therefore have devastating effects on the brain.[22]

The cortex itself is incredibly complex, but this structure, specifically the prefrontal cortex, is expanded and evolved in humans[23,24] and contributes substantially to larger brain volume and high

order cognitive functioning, such speech and executive function.[25–27] This massive expansion of the human prefrontal cortex compared to other mammals is due in part to a larger progenitor pool, protracted neurogenesis, increased cell type diversity, and expanded morphogen signaling.[28–31] While this additional complexity has led to human-specific traits, it has also led to human-specific disorders, such autism. The remarkable expansion of cellular diversity and function has been studied for decades,[32–35] and additional genome wide association studies has allowed us to link cell type specific gene expression to genes implicated in neurodevelopmental conditions.[36–39] To this end, understanding how gene expression is regulated throughout development provides a window into understanding the origins of neurodevelopmental disorders (NDDs).

Gene expression and regulation

The subtypes of neurons are specified by the unique suite of genes expressed that control cellular function and behavior.[32,40] While all cells have the ability to express all genes, additional layers of gene regulation ensure that only those that are necessary are expressed in a given cell type at a given time.[41,42] These layers include chromatin accessibility, transcription factor and cofactor expression, and exposure to morphogens and growth factors.

In addition to mapping cell type specific gene expression, more recent work has emphasized the importance of similarly mapping cell type specific chromatin accessibility,[43–45] providing a regulatory map unique to the myriad of cell types in the brain. Within these regulatory regions are motifs specific to transcription factors critical for modulating gene expression. Integrated analysis of gene expression, chromatin accessibility, and motif enrichment therefore provides a framework with which gene regulatory networks can be inferred.[41]

This genomic characterization is critical as we gain additional understanding of NDD pathobiology and the importance of these regulatory networks during development in this regard.[46,47] The expansion of sequencing technologies to include the entire genome has led to the discovery of hundreds of thousands of noncoding variants associated with NDDs.[48,49] Integrating the regulatory map with this reference of noncoding variation allows us to dissect the contribution of these variants on gene regulation, and ultimately neurodevelopment as it relates to NDDs.

Neurodevelopmental disorders

This spatiotemporal control of cortical neuron differentiation is critical for development, and disruptions along this trajectory can lead to NDDs. As discussed, these disruptions effect gene expression and regulation, which in turn disrupts neurogenesis and cell function.

Although many brain regions have been implicated in NDDs, the cortex is among the most highly replicated across studies.[50] Specifically, the early born deep layer cortical neurons that arise during midgestation confer significant risk to NDD pathobiology, as identified by NDD risk gene co-expression analysis.[36] Postmortem studies of autistic brains have additionally identified supernumerary neurons in the deep cortical layers and subplate,[51] and autistic individuals have reported dysregulation in neuronal excitability leading to epilepsy.[52,53]

Leveraging the high confidence association of tens to hundreds of genes with autism and other neurodevelopmental conditions provides a window into their importance during cortex development. Studying the function of these genes independently allows us to study their function during critical developmental processes. Since there is no causal gene (or even handful of genes) for autism, expanding our view to highlight convergence of gene function is critical to understanding the etiology of these disorders, as many recent studies have done.[39,54,55]

Sex differences in development and disorder prevalence

Beyond the complex genetic presentation of autism and related disorders, there is also an observed sex bias in diagnosis of these individuals (or probands), where males are around 3-4 times more likely to be diagnosed with autism than females.[56] Decades of work have been dedicated to explaining this phenomenon, studying the contribution of hormones such as testosterone and estrogen to autism risk and resilience during development,[57–59] as well as the overall brain volume and activity between males and females.[60,61] More recently, studies begun to identify the differences in gene expression between males and females across development and postnatal periods.[62] Brain regions that exhibit high sex dimorphism, such as the bed nucleus of the stria terminalis and amygdala, have been of interest and have been molecularly characterized, but the extent to which sex differences in the cortex exist is unclear.

Focus of this dissertation

The work described in this dissertation approaches the question of what genetic, molecular, and functional mechanisms contribute to NDD pathobiology from two angles: first, how does autism associated variation in high confidence genes affect this gene's function, independently and as a protein network, as the cortex develops; and second, what sex specific differences exist in the mid gestation cortex that underlie sex biased penetrance of NDDs. Chapter 2 prefaces this work with a literature review of a critical morphogen for forebrain development, retinoic acid, and its role in cortex function and arealization. Chapter 3 then maps the protein interaction network of high confidence autism risk genes, identifying a critical interaction between transcription factors FOXP1 and FOXP4 required for proper differentiation of cortical neurons. Chapter 4 describes sex-specific biology in the midgestation cortex, highlighting a novel gene regulatory network

driven by MEF2C that is biased in females. This work also describes the X chromosome as a driver of sex specific vulnerability that is dependent on the nature of the mutation (*de novo* or inherited). Lastly, the impact of this work is described in the conclusion, including outlook and potential limitations of these studies.

A note on neurodevelopmental disorders and sex bias

Neurodevelopmental disorders (or conditions) such as autism exist on a spectrum. Presentation, traits, and required support vary, and this work is not done from the lens of "curing" these conditions. Rather, if we know these genes to be important for neurodevelopment as they are highly associated with diagnosis of these conditions, studying gene function and how it differs with disorder-related variation allows us to better understand how the brain develops.

The sex bias we see in autism diagnosis may be due to underlying biology, but it also could be due to biases in diagnostic criteria and societal pressure to "mask" that is perceived differently between males and females. In recognition and respect for trans, intersex, and gender nonconforming individuals, throughout this dissertation "male" refers strictly to individuals with XY chromosomes, and "female" refers to individuals with XX and this work does not reflect gender expression in any way.[63–65]

References

1. Cadwell, C. R., Bhaduri, A., Mostajo-Radji, M. A., Keefe, M. G. & Nowakowski, T. J. Development and Arealization of the Cerebral Cortex. *Neuron* **103**, 980–1004 (2019).

2. Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nat. Rev. Neurosci.* **9**, 110–122 (2008).

3. Ragsdale, C. W. & Grove, E. A. Patterning the mammalian cerebral cortex. *Curr. Opin. Neurobiol.* **11**, 50–58 (2001).

4. Siegenthaler, J. A. *et al.* Retinoic Acid from the Meninges Regulates Cortical Neuron Generation. *Cell* **139**, 597–609 (2009).

5. Wilson, L., Gale, E., Chambers, D. & Maden, M. Retinoic acid and the control of dorsoventral patterning in the avian spinal cord. *Dev. Biol.* **269**, 433–446 (2004).

6. Kudoh, T., Wilson, S. W. & Dawid, I. B. Distinct roles for Fgf, Wnt and retinoic acid in posteriorizing the neural ectoderm. *Development* **129**, 4335–4346 (2002).

7. Shimamura, K. & Rubenstein, J. L. R. Inductive interactions direct early regionalization of the mouse forebrain. *Development* **124**, 2709–2718 (1997).

8. Telley, L. *et al.* Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science* **351**, 1443–1446 (2016).

9. Ypsilanti, A. R. & Rubenstein, J. L. R. Transcriptional and epigenetic mechanisms of early cortical development: An examination of how Pax6 coordinates cortical development. *J. Comp. Neurol.* **524**, 609–629 (2016).

10. Shirasaki, R. & Pfaff, S. L. Transcriptional Codes and the Control of Neuronal Identity. *Annu. Rev. Neurosci.* **25**, 251–281 (2002).

11. Delgado, R. N. *et al.* Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* **601**, 397–403 (2022).

12. Keefe, M. G., Steyert, M. R. & Nowakowski, T. J. Lineage-resolved atlas of the developing human cortex. *Nature* **647**, 194–202 (2025).

13. Gorski, J. A. *et al.* Cortical Excitatory Neurons and Glia, But Not GABAergic Neurons, Are Produced in the Emx1-Expressing Lineage. *J. Neurosci.* **22**, 6309–6314 (2002).

14. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. R. Interneuron Migration from Basal Forebrain to Neocortex: Dependence on *Dlx* Genes. *Science* **278**, 474–476 (1997).

15. Anderson, S. A., Marín, O., Horn, C., Jennings, K. & Rubenstein, J. L. R. Distinct cortical migrations from the medial and lateral ganglionic eminences. *Development* **128**, 353–363 (2001).

16. Xu, Q., Tam, M. & Anderson, S. A. Fate mapping Nkx2.1-lineage cells in the mouse telencephalon. *J. Comp. Neurol.* **506**, 16–29 (2008).

17. Letinic, K., Zoncu, R. & Rakic, P. Origin of GABAergic neurons in the human neocortex. *Nature* **417**, 645–649 (2002).

18. Hansen, D. V. *et al.* Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat. Neurosci.* **16**, 1576–1587 (2013).

19. Rubenstein, J. L. R. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).

20. Dehghani, N. *et al.* Dynamic Balance of Excitation and Inhibition in Human and Monkey Neocortex. *Sci. Rep.* **6**, 23176 (2016).

21. Vanderhaeghen, P. & Polleux, F. Developmental mechanisms patterning thalamocortical projections: intrinsic, extrinsic and in between. *Trends Neurosci.* **27**, 384–391 (2004).

22. Nelson, S. B. & Valakh, V. Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. *Neuron* **87**, 684–698 (2015).

23. Petrides, M., Tomaiuolo, F., Yeterian, E. H. & Pandya, D. N. The prefrontal cortex: Comparative architectonic organization in the human and the macaque monkey brains. *Cortex* **48**, 46–57 (2012).

24. Preuss, T. M. & Wise, S. P. Evolution of prefrontal cortex. *Neuropsychopharmacology* **47**, 3–19 (2022).

25. Gabrieli, J. D. E., Poldrack, R. A. & Desmond, J. E. The role of left prefrontal cortex in language and memory. *Proc. Natl. Acad. Sci.* **95**, 906–913 (1998).

26. Friedman, N. P. & Robbins, T. W. The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology* **47**, 72–89 (2022).

27. Kerns, J. G., Cohen, J. D., Stenger, V. A. & Carter, C. S. Prefrontal Cortex Guides Context-Appropriate Responding during Language Production. *Neuron* **43**, 283–291 (2024).

28. Hansen, D. V., Lui, J. H., Parker, P. R. L. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).

29. Shibata, M. *et al.* Regulation of prefrontal patterning and connectivity by retinoic acid. *Nature* **598**, 483–488 (2021).

30. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).

31. Vanderhaeghen, P. & Polleux, F. Developmental mechanisms underlying the evolution of human cortical circuits. *Nat. Rev. Neurosci.* **24**, 213–232 (2023).

32. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

33. Miller, D. J., Bhaduri, A., Sestan, N. & Kriegstein, A. Shared and derived features of cellular diversity in the human cerebral cortex. *Curr. Opin. Neurobiol.* **56**, 117–124 (2019).

34. Siletti, K. *et al.* Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).

35. Bhaduri, A. *et al.* An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).

36. Willsey, A. J. *et al.* Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism. *Cell* **155**, 997–1007 (2013).

37. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e23 (2020).

38. 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).

39. Wamsley, B. *et al.* Molecular cascades and cell type–specific signatures in ASD revealed by single-cell genomics. *Science* **384**, eadh2602 (2024).

40. Wang, L. *et al.* Molecular and cellular dynamics of the developing human neocortex. *Nature* https://doi.org/10.1038/s41586-024-08351-7 (2025) doi:10.1038/s41586-024-08351-7.

41. Ding, J. W. *et al.* Dissecting gene regulatory networks governing human cortical cell fate. *Nature* **651**, 732–742 (2026).

42. Deng, C. *et al.* Massively parallel characterization of regulatory elements in the developing human cortex. *Science* **384**, eadh0559 (2024).

43. Mannens, C. C. A. *et al.* Chromatin accessibility during human first-trimester neurodevelopment. *Nature* https://doi.org/10.1038/s41586-024-07234-1 (2024) doi:10.1038/s41586-024-07234-1.

44. Ziffra, R. S. *et al.* Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).

45. Braun, E. *et al.* Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, (2023).

46. An, J.-Y. *et al.* Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* **362**, (2018).

47. Shin, T. *et al.* Rare variation in non-coding regions with evolutionary signatures contributes to autism spectrum disorder risk. *Cell Genomics* **4**, 100609 (2024).

48. Wang, T. *et al.* Integrated gene analyses of de novo variants from 46,612 trios with autism and developmental disorders. *Proc. Natl. Acad. Sci.* **119**, (2022).

49. Sui, Y. *et al.* Using the linear references from the pangenome to discover missing autism variants. *Nat. Commun.* **17**, 1681 (2026).

50. Willsey, H. R., Willsey, A. J., Wang, B. & State, M. W. Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. *Nat. Rev. Neurosci.* **23**, 323–341 (2022).

51. Avino, T. & Hutsler, J. J. Supernumerary neurons within the cerebral cortical subplate in autism spectrum disorders. *Brain Res.* **1760**, 147350 (2021).

52. Bolton, P. F. *et al.* Epilepsy in autism: features and correlates. *Br. J. Psychiatry* **198**, 289–294 (2011).

53. Tuchman, R. & Cuccaro, M. Epilepsy and Autism: Neurodevelopmental Perspective. *Curr. Neurol. Neurosci. Rep.* **11**, 428–434 (2011).

54. Sun, N. *et al.* Autism genes converge on microtubule biology and RNA-binding proteins during excitatory neurogenesis. Preprint at https://doi.org/10.1101/2023.12.22.573108 (2023).

55. Paulsen, B. *et al.* Autism genes converge on asynchronous development of shared neuron classes. *Nature* **602**, 268–273 (2022).

56. Werling, D. M. & Geschwind, D. H. Understanding sex bias in autism spectrum disorder. *Proc. Natl. Acad. Sci.* **110**, 4868–4869 (2013).

57. Willsey, H. R. *et al.* Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. *Neuron* **109**, 788-804.e8 (2021).

58. Knickmeyer, R. C. & Baron-Cohen, S. Topical Review: Fetal Testosterone and Sex Differences in Typical Social Development and in Autism. *J. Child Neurol.* **21**, 825–845 (2006).

59. Mishra, J. S. *et al.* Elevated maternal testosterone induces sex-specific neurodevelopmental changes and ASD-related behavioral phenotypes in rat offspring. *Pediatr. Res.* **99**, 1954–1963 (2026).

60. Lombardo, M. V. *et al.* Fetal Testosterone Influences Sexually Dimorphic Gray Matter in the Human Brain. *J. Neurosci.* **32**, 674–680 (2012).

61. Auyeung, B., Lombardo, M. V. & Baron-Cohen, S. Prenatal and postnatal hormone effects on the human brain and cognition. *Pflüg. Arch. - Eur. J. Physiol.* **465**, 557–571 (2013).

62. Velmeshev, D. *et al.* Single-cell analysis of prenatal and postnatal human cortical development. *Science* **382**, (2023).

63. Miyagi, M., Guthman, E. M. & Sun, S. D.-K. Transgender rights rely on inclusive language. *Science* **374**, 1568–1569 (2021).

64. Massa, M. G., Aghi, K. & Hill, M. J. Deconstructing sex: Strategies for undoing binary thinking in neuroendocrinology and behavior. *Horm. Behav.* **156**, 105441 (2023).

65. Garcia-Sifuentes, Y. & Maney, D. L. Reporting and misreporting of sex differences in the biological sciences. *eLife* **10**, e70817 (2021).

CHAPTER 2: EMERGING ROLES OF RETINOIC ACID SIGNALING IN HUMAN-SPECIFIC FOREBRAIN DEVELOPMENT AND RELATED DISORDERS

Introduction

Retinoic acid (RA), a metabolite of Vitamin A, is a well-established morphogen that plays a central role in a wide range of early developmental processes. RA is a critical signaling molecule that drives induction, patterning and morphogenesis during embryonic development and has been studied in animal models for decades.[1–8] RA signaling is critical for central nervous system development, contributing to development of ventral telencephalic structures such as the striatum,[9–11] as well as caudal structures including the hindbrain and spinal cord.[12,13] However, recent work using primary human and non-human primate models, as well as human induced pluripotent stem cell (iPSC)-derived cerebral organoid models, has identified novel roles of RA in the developing cerebral cortex and dorsal forebrain.[14–16] These studies demonstrate that RA can exert distinct and even opposing effects depending on where and when it acts. This suggests that RA is pleiotropic, where its regulatory effects depend on spatial context, developmental timing, receptor composition, and ligand availability. To this end, the complex and nuanced developmental effects of RA signaling are not fully understood.

Here, we review the role of RA signaling during neurodevelopment, particularly in humans, with a focus on the newer emerging roles of RA and its receptors in shaping the patterning and organization of the forebrain and neuronal identity during development. By integrating molecular, genetic and functional insights from recent studies, we aim to distinguish the contextual factors underlying the versatile and complex regulatory properties of RA signalling.

Mechanism of RA signaling

Synthesis of RA

All-trans retinoic acid (hereafter, RA) is derived from retinol (vitamin A) from a series of oxidation reactions involving first aldehyde dehydrogenases (ADHs), followed by retinaldehyde dehydrogenases (RALDH1, RALDH2, RALDH3). CYP26A1 eventually metabolizes and degrades RA. In the brain, the main source of RA is the meninges, where disruption of the meninges during development results in dysregulated neurogenesis.[17]

Once it is produced, RA is shuttled into the nucleus where it acts as a ligand for the Retinoic Acid Receptors (RARs, RARb, RARg). The RARs heterodimerize with Retinoid X Receptors (RXRa, RXRb, RXRg), and subsequently act as transcription factors that bind to regulatory regions of DNA (retinoic acid response elements, RAREs) and modulate expression of target genes. Some data suggests that the RAR:RXR heterodimer is constitutively bound to DNA, but only initiates transcription in the presence of RA. Other work argues ubiquitous binding of RXR, and RAR is only recruited once activated by RA. In either case, target genes are typically transcribed in the presence of RA. The synthesis of RA and subsequent transcriptional regulation of RA target genes is reviewed in Duester 2008[18] and Cunningham and Duester 2015.[19]

Genes regulated by RA

RA signaling initiates expression of genes required for a wide range of biological processes. RA signaling is involved in maintaining pluripotency with acute exposure,[20] promoting neuronal differentiation,[21–23] neural tube patterning and hindbrain specification,[24–28] and ventral forebrain patterning of the ganglionic eminences.[9–11] To this end, RA target genes that are widespread in

function and include HOX genes (*HOXA1-4*, *HOXB1-6*)[29,30] as well as RA metabolism and signaling genes (*CYP26A1, RARB, CRABP2)*. Most genes are only bound by RARs in the presence of RA, but a small subset of genes, including *Cyp26a1* and *Rarb* have been shown to be constitutively bound by RARs, and only transcribed in the presence of RA in the mouse. Activation of RXRs is only necessary for a subset of transcriptional events.[31] Recent work profiling genes expressed upon RA treatment in a cortical organoid model of human neurodevelopment has identified several genes differentially expressed between vitamin A and no vitamin A treatment, including *BCL11B, MEIS2,* and *SOX.5*[14] Additionally, a recent study profiling RA target genes specifically in the human midgestation dorsolateral prefrontal cortex identified 5 gene modules by weighted gene co-expression analysis integrating paired single nucleus RNA and ATAC sequencing together with RARA, RARB, and RXRG CUT&Tag, including genes involved in forebrain development and axonogenesis.[16] This study highlights *MEIS2* as a critical regulator of RA-induced gene expression, and is a direct target of RAR/RXR. For an extensive characterization of RA induced genes, see Balmer and Blomhoff.[29]

Genetic and epigenetic determinants of RA-induced gene expression

RAR and RXR recruitment of co-factors allows for tight regulation of target genes. Changes in conformation of receptors when bound to RA allow for exchange of co-factors and initiation of transcription. In the absence of RA, RARs bind to RAREs in a non-permissive conformation. In this state, the receptors recruit co-repressors that maintain a transcriptionally repressive state. Once bound to RA, RARs undergo a conformational change, releasing bound co-repressors and recruiting co-activators. Molecular coordination between RA-bound RARs, co-activators, and transcription initiation complex allows for active transcription of target genes.

A cell's chromatin state and histone modifications determine the ability of RARa to bind at RAREs and initiate transcription. This leads to differential expression of RA target genes depending on cell type-specific chromatin state.[32,33] Binding of RXRs to active enhancers is influenced by H3K27ac status,[33] as well as CTC.[34] Further, studies have noted coordinated action between lysine demethylase KDM5B, which demethylates H3K4, and RARs required for transcription.[35] Additionally, RA-induced transcription of target genes has also been shown to depend on DNA methylation status.[36] Tightly controlled expression of RARs by cell type, tissue type, brain region, and temporal progression can further tune genomic responses to RA. Therefore, the complex interplay between epigenetic modifications, expression of RARs, and presence of RA lead to differences in transcriptional output, leading to its pleiotropic roles during CNS development, as discussed further below.

Retinoic Acid in CNS development

Systemic Acquisition of RA during development

During development, the embryo obtains RA exclusively from maternal dietary VA since it cannot be synthesized *de novo*. VA is metabolized maternally to retinol bound to RBP4 or retinyl esters that are packaged into lipoproteins, which cross the maternal-fetal interface initially through the yolk sac, and later, through the placenta. Since maternal RBP4 cannot cross this barrier, free retinol is released at the interface and binds embryonic RBP4, which is first synthesized by the yolk sac until the fetal liver develops and is subsequently delivered to embryonic tissues via circulation.[37,38] As fetal development progresses, the placenta maintains controlled receptor-mediated retinol uptake, transfer and metabolism, with the developing fetal liver progressively taking over metabolic and storage functions, and becoming the primary site of RBP4 production and retinol

storage by the second trimester,[37–40] RA levels are buffered within the developing embryo, minimizing the effects of maternal fluctuations.

Beyond this systemic stabilization, RA distribution and concentration are tightly and independently regulated within embryonic tissues. At an additional layer of regulatory control, embryonic cells locally metabolize retinoids to their intermediates or their active form, RA, when required. Consequently, embryonic RA and RA signaling do not occur uniformly across the embryo, with specific cells and regions acting as discrete RA-synthesizing, RA-degrading, RA-refractory and RA-responsive domains.[41]

Embryogenesis

The involvement of retinoids in embryogenesis were first noted with teratogenic effects upon excess RA presence. This has been studied in humans in the context of pregnant mothers treated with retinoids or consuming things like liver with high levels of retinoids. The role of RA during embryogenesis is well conserved across species, and this has additionally been studied in quail,[42] chick,[43] zebrafish,[44] *Xenopus laevis,*[1] and mice.[45,46] Experiments in these models have led to an improved understanding of how RA orchestrates embryogenesis.

Because RA is a widely diffusible lipophilic molecule, tight regulation of its synthesis ensures proper signaling within developing tissues. This leads to morphogenic gradients, with a high concentration of RA synthesizing enzymes (RALDH1-3) marking one boundary, and a high concentration of RA degradation enzymes (CYP26A1, CYP26B1, CYP26C1) marking the other. The cellular expression of the RARs grants further control of RA signaling, leading to transcriptional responses to RA signaling in a subset of cells. High RA signaling during

embryogenesis instructs patterning of caudal nervous system structures, most notably the hindbrain. Here, high expression of RA-induced HOX genes in a temporally and spatially regulated manner leads to proper hindbrain development.

Global Anterior-Posterior RA Gradient

A global, linear RA gradient is established along the anterior-posterior axis during early embryogenesis. The graded distribution of RA provides positional identity, and through its interactions with other pathways, RA establishes signaling boundaries that are both essential for body axis and central nervous system patterning. RA signaling begins during gastrulation, with RA synthesis occurring in the primitive streak and paraxial mesoderm that later form the trunk of the embryo.[45,46] Expression of RALDH2, the primary RA-synthesizing enzyme, by the trunk mesoderm results in high RA concentration in this region, making it the signaling 'source'.[46] RA then diffuses out from this trunk region towards the anterior (head) and posterior (tail) ends of the embryo, which act as metabolic 'sinks', due to the expression of RA degrading CYP26 enzymes (primarily CYP26A1 in the anterior head region and the tail bud and CYP26C1 to a lesser extent in the anterior head region).[41,43,47] This global anterior-posterior gradient of RA is maintained from gastrulation through early neurulation.[44,47] Functionally, this global RA gradient acts as a trunk-organizing signal that, while specifying the primary body axis, plays a central role in coordinating and specifying anterior-posterior identity of the neural plate.[18]

During body axis extension and somitogenesis, the anterior-posterior gradient of RA from the trunk to the tailbud directly opposes a posterior-anterior gradient of FGF8 and Wnt via repression of their expression.[48] Somite formation occurs at the differentiation wavefront, where these two signaling gradients intersect.[49,50] As the body axis elongates, the RA source shifts posteriorly with

the newly formed somites, thereby dynamically following axial growth.[47] While FGF8 and Wnt signaling from the posterior end maintain tailbud cells in an undifferentiated, proliferative state, RA promotes cell differentiation and somite budding, thereby coordinating segmentation with axial growth.[48,49,51] Consequently, RA not only patterns the body axis, but it also establishes a maturation gradient that progressively drives progenitor cells from proliferation to differentiation along the elongating axis.[48,51]

RA synthesized by RALDH2 in the trunk mesoderm is a paracrine source that establishes a steep posterior to anterior RA gradient (from the neck to the midbrain) in the adjacent neural tube during early neurulation that is essential for the patterning of the central nervous system.[52,53] The early neural plate that gives rise to the neural tube has a default anterior identity that favors forebrain specification. Higher RA levels in posterior regions act as an instructive posteriorizing signal, transforming neuroectodermal tissue to specify the hindbrain and spinal cord.[24,42] In the trunk region, high RA concentrations antagonize anteriorizing signals such as BMP and Wnt signaling from the head, thereby restricting anterior neural plate expansion and ensuring proper forebrain and eye development.[51,54,55] Along the anterior-posterior axis in the neural tube, the highest concentration of RA is in the cervical region, promoting spinal cord formation and elongation in coordination with axial growth.[25,53] The midbrain and forebrain remain largely devoid of RA signaling during these stages due to the localized expression of RA degrading CYP26 enzymes. This metabolic "sink" establishes the midbrain-hindbrain boundary and preserves the anterior neural identity. Furthermore, similar to its role in the elongation of the body axis, the higher concentration of RA in the posterior neural tissues drive differentiation and neurogenesis in these regions, while its exclusion from the forebrain in the anterior region, preserves the proliferative progenitor state that is required for the expansion of the forebrain vesicles.[17,53,56,57] Thus, the global

RA gradient drives anterior-to-posterior regional specification and morphogenesis of the forebrain, midbrain, hindbrain and spinal cord within the vertebrate neural tube.

RA and specification of caudal CNS structures

The graded posterior (high) to anterior (low) RA signaling drives segmentation of the hindbrain into distinct rhombomeres through the induction of Hox genes. The Hox code and expression pattern of Hox genes reflects their position within the cluster, with genes located at the 3' end of the cluster being expressed more anteriorly than genes at the 5' end.[58] Within the hindbrain, the higher posterior RA concentration specifies posterior rhombomeres (r4 - r8), which eventually give rise to the medulla oblongata, and the relatively lower anterior RA concentration specifies the anterior rhombomeres (r1 - r3), which form part of the cerebellum.[8,46,59] The specification and patterning of the identity of each rhombomere along the anterior-posterior axis are defined by the combinatorial Hox code that is regulated by the concentration and spatiotemporal dynamics of RA exposure.[8,26,60,61] Furthermore, the spatiotemporal activity of RA signaling is modulated by CYP26 enzymes, CRABPs, and RARs. As hindbrain segmentation progresses, these proteins establish shifting RA boundaries that define discrete rhombomere-specific territories through differential regulation of segmental identity genes.[26–28] These rhombomere identities provide the spatial map required for the organization of cranial nerve nuclei and determine the migration patterns of neural crest cells for the formation of craniofacial structures.

The specific identities of spinal cord segments and the motor neuron subtypes organized in columns, with each axial position reflecting their functional identity, are determined by the Hox code which is regulated by the opposing RA and FGF/Wnt gradients. High RA in the anterior region induces the expression of Hox genes (Hoxc5, Hoxc6) that specify the cervical region of the

spinal cord.[62] Subsequently, high anterior RA levels also induce the expression of Hox genes (Hoxc6, Hoxc8) specifying the brachial region, followed by the local synthesis of RA by the RALDH2-expressing forelimb motor neurons of the lateral motor column (LMC) generated in this region.[63–65] Moving posteriorly, RA concentration decreases in the next segment, the thoracic region, concomitant with CYP26B1 expression at segmental boundaries, with the low RA concentration being permissive for the specification of this region.[64–66] Instead, FGFs (FGF8, FGF4) and GDF11 (a member of the TGF-β superfamily) emanating from the notochord and presomitic mesoderm at the posterior end of the embryo, induce hox genes specifying the thoracic region (Hoxc9).[62] At the posterior-most end of the spinal cord, where FGF (particularly FGF8), Wnt and GDF11 signals originating from the tail are the highest, a second peak of elevated RA is observed, which together induce hox genes (Hoxc10) specifying the lumbar region.[62,64,65] Furthermore, this region contains the LMC, resulting in local RA synthesis by RALDH2-expressing hindlimb motor neurons occupying this region.[62–65] Within the brachial and lumbar regions, local synthesis of RA by the early-born, ventrally projecting medial LMC motor neurons drives the molecular specification and fate switch required for the generation of the later-born lateral LMC motor neurons that project dorsally.[12,63,67,68] As development progresses, additional sources of RA arise, and local RA gradients are established and maintained by the spatiotemporal regulation of components of RA signaling. These local RA signaling domains play a wide variety of roles including patterning, neuronal differentiation, neuronal migration, cell fate specification, maturation and connectivity.

Local RA Sources in Forebrain Development

Beyond the sustained global gradient of RA, additional dynamic local gradients of RA also form during embryogenesis, some of which regulate development and patterning. This is driven by RA either directly, or indirectly via the regulation of other signaling pathways or transcription factors, and is highly spatially, temporally or target driven, which potentially underlies the pleiotropic nature of RA during development. Although RA synthesized in the trunk, which establishes the global posteriorizing gradient during development, is excluded from the head region, both earlier work in animal models and emerging evidence in human prenatal tissue have identified discrete local sources of RA in the developing head and forebrain.[15] Due to the limitations of human prenatal tissue, most mechanistic studies have been conducted in animal models. Species-specific differences, the essential role of RA in developmental processes required for survival beyond the brain, and contradictory findings across studies limit the extent to which these results can be extrapolated to human development.

Early transient anterior sources

Avian and rodent models identified early transient expression of RA synthesizing proteins during late gastrulation and early somite stages, in anterior tissues that give rise to or are in close proximity to the prospective head and forebrain.[54,69] In mice, RALDH2 is transiently expressed in the anterior neural plate and, subsequently, in the forebrain neuroepithelium and optic vesicle, while RALDH3 is transiently expressed in the frontonasal surface ectoderm overlying the anterior forebrain in both mouse and chick models.[69,70] Whether RA from these early transient sources is required for or induces forebrain patterning, acts indirectly by regulating other morphogen pathways, or is instead

correlated and required solely for optic vesicle and olfactory pit morphogenesis is widely contested.[69,71]

The meninges is a primary RA source in forebrain development

As development progresses beyond neurulation and brain vesiculation begins, additional local sources of RA within and around the developing brain emerge that have been implicated in forebrain development and patterning. Among these are the meninges that are in close contact with the developing brain that form a local RA gradient by facilitating efficient local diffusion of RA by paracrine signaling.

The meninges, a three-layered protective envelope surrounding the brain and spinal cord, forms during and immediately after neural tube closure. The meninges arise from two distinct origins, with the meninges surrounding the forebrain originating from the cranial neural crest, while the meninges surrounding the midbrain, hindbrain and spinal cord originating from the mesoderm.[72–74] These distinct developmental origins may contribute to regional differences in meningeal RA production and signaling, potentially underlying the differing roles and concentrations of RA across the forebrain, midbrain, and hindbrain.

The meninges act as an active primary source of RA for the developing brain,[17,75] forming a RA gradient with the concentration being the highest at the outer layer of the developing brain. Within the cortex, this pial surface is where cortical progenitors, radial glia, endfeet attach, with RA diffusing inward to the underlying neocortex.[17] In mice, the meninges express high levels of proteins involved in RA synthesis, primarily RDH10 and RALDH2, with peak RA signaling occurring during midgestation.[17,76–78] Meningeal maturation is accompanied by the expression of

RA metabolizing proteins in a ventral-to-dorsal pattern.[76] RALDH2 expression begins in a lateral to medial pattern in the developing cortex, until it is expressed in the entire meninges, along with weaker levels of RALDH1 expression.[17,79] RALDH2 expression extends postnally in mice before declining in adulthood, with an increase in RALDH1 expression.[17,79] Transcriptomic profiling of meningeal fibroblasts revealed brain region- and layer-specific expression patterns, with regional differences in expression levels of RA signaling proteins.[76,79] The meningeal layer specific expression pattern of RA metabolizing and transport proteins, STRA6 and RDH10 required for the first step of RA synthesis in pial fibroblasts, RALDH2 required for the second step of RA synthesis in pial and arachnoid fibroblasts, and CRABP2 in arachnoid and dural fibroblasts, and *CYP1B1* expression in pial fibroblasts, indicates the possibility of an organized and coordinated RA synthesis mechanism in the meninges.[76]

Comparative studies have demonstrated that this meningeal source of RA for the forebrain is conserved in humans.[15,76] In humans, the midfetal human meninges express all three RALDH proteins, Aldh1a1, Aldh1a2, and Aldh1a3, with Aldh1a1 showing the highest expression, followed by Aldh1a2 and then Aldh1a3.[76] Additionally, the expression of RALDH2 and CRABP2 in arachnoid fibroblasts, along with the layer-specific meningeal marker expression pattern, is conserved between mice and humans.[76]

Another candidate RA source is the choroid plexus, the secretory epithelium of the lateral ventricles, the third ventricle in the diencephalon, and the fourth ventricle in the hindbrain, that produces and circulates CSF. In mice, RA synthesizing enzymes are expressed in the choroid op which includes the expression of RALDH2 in the choroid plexus of the fourth ventricle,[80,81] RDH10 expression in the choroid plexus of embryonic mice,[15,17] and RALDH3 expression in the

choroid plexus of the lateral ventricle.[82] However, the function of the choroid plexus as an RA source for the forebrain, and in human development has not been explored.

Intrinsic Sources

Although the meninges acts as a primary RA source for the developing forebrain, intrinsic RA sources within the developing forebrain at later stages also exist. The first sustained local RA source identified in the developing forebrain, is RALDH3 expressed by newly born neurons in the subventricular zone (SVZ) in the lateral ganglionic eminence (LGE), and septum in mice.[9,70,71] This locally synthesized RA is required for the GABAergic differentiation of LGE-derived neurons occupying the striatum, and interneurons that migrate to the olfactory bulb and cortex.[9–11] Furthermore, these findings were modeled *in vitro,* where the defined exposure of human embryonic stem cells to RA was sufficient to generate GABAergic interneurons.[9] However, whether this RALDH3-dependent mechanism for the generation of LGE-derived GABAergic neurons occurs in the human ventral forebrain has not been studied.

Recent work established that a low concentration gradient of RA exists in the developing cortex during midgestation.[15] The concentration of RA was measured using ELISA in mouse, macaque and human cortical tissue, and it was found that the concentration of RA was highest in the prefrontal cortex (PFC), resulting in an anterior-to-posterior concentration gradient across the cortex. The concentration of RA was highest in humans, followed by macaque and then mouse tissue, with relatively high concentrations in the PFC across all 3 species as well as a high concentration in the inferior temporal cortex (ITC) in humans. Furthermore, primate-specific ALDH1A1 expression patterns were observed, with expression in multiple cell types in primates,

which included subplate neurons, astrocytes and incoming dopaminergic axons, while in mice, it was largely confined to the dopaminergic system only.

Forebrain patterning

Ventral forebrain

RA signaling is critical for patterning of several cortical and subcortical structures. Several studies have shown the importance of RA in specifying ventral telencephalic fate. The ganglionic eminences (caudal, medial, lateral) are particularly sensitive to RA signaling, where RA synthesizing enzymes within the LGE are required for specifying GABA-ergic interneuron fate during development, as shown in mouse.[9] Further work has shown that transcriptional activation of RA target genes is required for this ventral specification, in addition to specification of motor neurons.[12]

Dorsal Forebrain

In addition to ventral forebrain structures, recent work has strongly implicated RA signaling in patterning of the dorsal forebrain as well. RALDH2 is the primary RA-synthesizing enzyme required for forebrain development.[54] More recently, several studies have implicated RA in human specific neurodevelopmental features, such as the specification of the prefrontal cortex.[14–16] The prefrontal cortex is an expanded cortical area in humans and is responsible for orchestrating higher cognitive functions such as speech. Recent work has identified an expansion of RA signaling genes in the PFC, as well as an enrichment in RA itself in this region compared to mice.[15] This suggests a requirement of RA signaling in establishing anterior forebrain identity, which gives rise to human specific brain anatomy and functions.

Dorsal Forebrain Arealization

The noted expansion of RA signaling in the PFC suggests the involvement of RA in cortical arealization and generation of particular subtypes of cortical neurons. Specifically, recent ATAC-seq data shows that RAR binding motifs are highly enriched among open chromatin in PFC areas of the human cortex, but not in the posterior V1 region.[14] Additionally, in vitro derived cortical organoids treated with RA during differentiation exhibit a gene expression profile highly similar to that of primary human PFC, but not V1.[14] This suggests that RA signaling leads to PFC-like fate of glutamatergic neurons in cortical organoids, and that RA signaling promotes arealization and development of the PFC during human neurodevelopment.

Further work has shown that mice with conditional knockout of *Meis2*, a RA-induced gene, in cortical excitatory neurons have altered arealization of the frontal cortex.[16] PFC markers *Cbln2* and *Plxnc1* were largely absent, accompanied by a medial expansion of anterolateral motor cortex and secondary motor cortex M2 (marked by *Cyp26b1* and *Etv5*) into the PFC region, and rostral expansion of the somatosensory cortex (marked by *Bhlhe22*) into the motor cortex region. Viral anterograde tracing in *Meis2* cKO mice revealed that altered molecular identity corresponded with decreased connectivity of the medial PFC to ventral thalamic nuclei, including mediodorsal and ventromedial, as well as the basolateral amygdala.

Beyond arealization, RA is also critical for cell type specification of cortical excitatory neurons. Both Ziffra 2021[14] and Yang 2025[16] report RA-dependent expression of deep layer excitatory neuron markers, including *BCL11B, FOXP1,* and *FOXP2*. Yang 2025 further showed that RA-synthesizing enzyme *ALDH1A3* expression depends on *MEIS2*, suggesting an autoregulatory loop for RA signaling in the PFC. Interestingly, glutamatergic neurons of the PFC are particularly

vulnerable to mutations in ASD risk genes, including *BCL11B, FOXP1, FOXP2* and *MEIS2,*[83–85] and genes within the RA signaling pathway have been noted as common variants in ASD[86] (see RA signaling and disease). Coexpression analysis of ASD risk genes has consistently implicated the midgestation PFC in ASD pathobiology,[83,87] and these data suggest that RA signaling and resulting cortical arealization and cell type specification could be one potential mechanism underlying this vulnerability.

Effect of RA on neurogenesis and synaptic plasticity

Neurogenesis

A major role of RA during development is promoting differentiation.[88] RA promotes neurogenesis and differentiation in several central nervous system structures, including the striatum,[9,12,89] dentate gyrus and hippocampus,[90–92] spinal cord,[48,93] and the cortex.[17,78] RA signaling mediated by both RARs and PPARs has been shown to promote differentiation of neuronal progenitors during development in mice, where RARs induce early stages of differentiation from stem cells into neural progenitors, and PPARs induce later stages of differentiation from neural progenitors to mature neurons. In mice, the primary source of RA is the meninges.[17] When the meninges is disrupted during development, aberrant cortical neurogenesis is observed, with atypical distribution of neurons and improper differentiation of cortical progenitors.[17] This phenotype is rescued by dietary intake of vitamin A by pregnant mice, demonstrating the necessity of RA production by the meninges during cortical development. Ablating *Raldh2* expression in mice just before meninges development leads to altered cortical migration, lamination, and morphology of deep layer excitatory neurons.[78] Additionally, mutant mice lacking mesenchymal and epithelial cells required for RA production show abnormal forebrain development, including enlarged

ventricles and disrupted cortical lamination.[94] Importantly, these phenotypes resemble those seen in schizophrenia, which will be discussed in detail below.

The orphan nuclear receptor CoupTFI has been shown to mediate the effects of RA signaling during cortical development.[95] CoupTFI is known to interact with RARs, and overexpression of CoupTFI in mutant mice lacking proper meningeal development rescues cortical progenitor cell behavior and differentiation.[95] RA also promotes differentiation of inhibitory neurons in the ganglionic eminences of the ventral forebrain, where RA signaling is required for development of GABAergic cells.[9,12]

Synaptic plasticity

In addition to its canonical role as a transcriptional regulator, several recent studies have identified a translation-dependent role of RA at the synapse important for synaptic plasticity. Treatment of adult human cortical slices with atRA led to an increase of spontaneous excitatory postsynaptic current (sEPSC) amplitude driven by structural and functional remodeling of dendritic spines.[96] These effects were seen in superficial layer 2/3 pyramidal neurons. Further experiments in mice confirmed this is dependent on synaptopodin, an actin-remodeling protein associated with a subset of dendritic spines and important for orchestrating long term spine dynamics through RhoA signaling.[97,98]

RA,[99–101] as well as synaptopodin,[102,103] have been further implicated in both Hebbian and homeostatic synaptic plasticity. This has been shown to occur through RA-induced local AMPA receptor synthesis and trafficking using rat hippocampal slices and is mediated by mTORC1-signaling.[100,101] Importantly, the RARa receptor has been implicated in this pathway. Vitamin A

deficiency leads to loss of synaptic plasticity in mice[104] and treatment with pharmacological, genetic, or nutritional RA intervention showed an increase in synaptic plasticity and improvement in memory.[105] These effects on synaptic plasticity in the hippocampus are mediated by RARb and RXRg through genetic knockouts in mice,[105] highlighting the pleiotropic roles of RA signaling in the brain. This modulation of homeostatic synaptic plasticity in hippocampal CA1 and cortical neurons could be the mechanism by which therapeutic RA treatment benefits Alzheimer's patients, as discussed further below. The effects of RA signaling on synaptic plasticity are reviewed further in.[106]

Implications in neurodevelopmental conditions

Origins of RA teratogenicity

The connection between retinoic acid signaling and neurodevelopmental disorders was first noted in 1995,[107] with initial links between oral administration retinoids and adverse reactions originating even earlier in the 1980's.[108,109] Screening of pregnant women identified a higher incidence of neural crest malformations upon birth in women with elevated dietary and supplemental intake of vitamin A (more than 15,000 IU per day), when compared to pregnant women with low vitamin A consumption (less than 5,000 IU per day). The teratogenic effects of vitamin A are most likely to be observed in pregnant women who consume high levels of vitamin A within the first seven weeks of gestation. Further studies in mice verified that both high (100 mg/kg) and moderate (50 mg/kg and 25 mg/kg), as well as low (10 mg/kg) doses of vitamin A led to craniofacial, cardiac, and thymic malformations, as well as neural crest dysfunction.[110]

Classic teratogenic and genetic studies in vertebrates established RA as a canonical posteriorizing morphogen during early embryonic development and central nervous system patterning. Maternal exposure to excessive VA or RA leads to the loss of anterior neural structures, such as the forebrain, midbrain, eyes, arising from the inappropriate posteriorization of anterior neural tissue, where elevated RA levels shift the positional identity boundaries toward more posterior fates.[24,56,111,112] On the other hand, VA deficiency or loss-of-function mutations in crucial components of the RA signaling pathway results in the failure to properly develop posterior structures such as the hindbrain, spinal cord, and trunk, resulting in expanded disorganized anterior structures.[42,47,56,113] These phenotypes are strongly correlated to the RA concentration and developmental timing, indicating a tightly controlled spatial and temporal role of RA signaling.

Prenatal vitamin A levels have been implicated in neuropsychiatric disorders, with a particular vulnerability during the second trimester of pregnancy. A study utilizing early postnatal mice, a developmental period similar to second trimester development in humans, showed that RA administration led to abnormal development of the hippocampus and anterior cingulate cortex.[114]

Autism spectrum disorders

Analysis of postmortem brains from autistic individuals revealed reduced expression of RA signaling gene *RORA*, specifically in the prefrontal cortex and cerebellum.[115] Polymorphisms in *RORA* are positively correlated with ASD risk.[116,117] *RORB*, another RA signaling gene, is also associated with autistic features, in addition to epilepsy, developmental delay, and behavioral abnormalities.[118,119] *RORB* is one of the genes deleted in 9q21.13 microdeletion syndrome, which is comorbid with autism and other neurodevelopmental phenotypes.[118] Polymorphisms in the *RORB* gene have also been identified in ASD probands,[120] and whole exome sequencing identified

rare de novo mutations in *RORB* in autistic individuals,[84,85] further implicating this gene and RA signaling in ASD progression. Additionally, maternal vitamin A deficiency is associated with autism-like phenotypes in rats.[121]

In addition to genes directly involved in the RA signaling pathway (including RORA, and RORB), genes shown to be regulated by RA signaling in the developing brain have also been characterized as ASD risk genes. These downstream RA target genes include: FOXP1, BCL11B, and MEIS2. Further, ASD risk genes converge in glutamatergic excitatory neurons in the anterior-most prefrontal cortex of humans.[83] The prefrontal cortex is expanded in humans compared to mice, and genes expressed in this anatomical region have been shown to be particularly vulnerable to mutations that give rise to ASD.[87] As described above, RA signaling is also expanded and enriched in the PFC of humans compared to more posterior regions, including the somatosensory and motor cortices. The convergence of both enriched RA signaling and higher incidence of ASD mutations in humans suggests a possible link between RA signaling and genetic pathways often perturbed in ASD.

Because variants in ASD risk genes within the RA signaling pathway are typically classified as common, and variants downstream of RA signaling are frequently rare and high confidence variants, this provides a basis for an interesting and important connection between common and rare ASD variants, with RA signaling acting as a nexus. It will be interesting to determine if there is phenotypic convergence with disruption of either RA signaling or downstream RA target genes to better understand the link between common and rare variants associated with RA signaling.

Further, studies have demonstrated a link between nutrient deficiency including vitamin A and ASD,[122,123] where autistic children had significantly decreased serum retinol levels compared to neurotypical individuals.[124,125] In fact, almost 80% of autistic children surveyed in Liu 2016[124] had vitamin A deficiency. Further, there is a noted decreased expression of the RARs (RARa, RARb, RARg) in autistic children, and vitamin A supplementation not only increased expression of these receptors, but also improved in autistic traits.[125] Together, the effects of retinoids have a clear link to ASD postnatally, and the emerging genetic interactions with ASD risk genes during development suggest these traits manifest early.

Schizophrenia

A possible link between schizophrenia and retinoids has long been appreciated. In the 1990's, studies began to elucidate a connection between phenotypes seen in both vitamin A deficiency and schizophrenia, including enlarged ventricles, microcephaly, and craniofacial abnormalities.[40,126,127] A direct link between prenatal vitamin A levels and schizophrenia diagnosis was first noted in 2012, which studied a birth cohort from the Prenatal Determinants of Schizophrenia study.[128] Levels of serum vitamin A in the second trimester were correlated with the prevalence of schizophrenia in offspring members of this study, and the authors found a threefold increased risk of schizophrenia and other schizophrenia spectrum disorders.[129] Disrupted metabolite levels are also seen in people diagnosed with schizophrenia, where at-RA and serum retinol were lower than in undiagnosed individuals.[130]

Eventually, genetic connections between RA signaling and schizophrenia became clear, with identification of risk loci that contain genes within the RA signaling and biosynthetic pathway. Notably, microarray data identified lowered levels of *ALDH1A1*, an enzyme that synthesizes at-

RA from retinaldehyde, in cases of schizophrenia. More recent transcriptomic analysis from a large cohort of over 500 schizophrenic individuals[131] revealed both *RAI1* and *RERE* as high confidence genes dysregulated in schizophrenia compared to neurotypical individuals,[132] and additional work demonstrated a specific upregulation of *RAI1* in the dorsolateral prefrontal cortex of individuals with schizophrenia.[133] GWAS studies then further implicated RA signaling genes within high confidence risk loci in addition to *RERE* and *RAI1*, including *ZNF563, CYP26B1,* and *CNOT1*[134,135] (reviewed in Reay and Cairns 2020[136]). Copy number variants of *RXRA* are also associated with schizophrenia.[137] Additionally, binding of RAR:RXR heterodimers within functional RARE motifs within the promoter of dopamine receptor gene *DRD2* is required for its expression.[40] Dopaminergic signaling dysfunction is commonly seen in cases of schizophrenia[138], and the dopamine signaling pathway is altered with impaired RA signaling.[139,140] This highlights how altered expression of RA metabolic genes, RA signaling genes, and RA target genes may all be implicated in schizophrenia.

In addition to prenatal vitamin A levels, vitamin D deficiency has also been associated with schizophrenia.[141] There is a noted interaction between RARs and vitamin D receptors (VDRs), and this heterodimerization influences gene expression in response to vitamin D levels during development.[142] Therefore, decreased levels of vitamin D during development may also contribute to schizophrenia in a RAR-dependent manner, where vitamin A and vitamin D deficiency phenotypes converge as schizophrenia phenotypes.

Because vitamin A levels can be easily modulated by dietary intake or supplementation, the RA signaling pathway is an attractive candidate for therapeutic intervention to mitigate schizophrenia symptoms. This idea has long been proposed.[127,143] Two clinical trials tested the efficacy of oral

administration of bexarotene, a pan-RXR agonist, in addition to ongoing antipsychotic treatment for schizophrenia.[144,145] In both studies, individuals showed significant improvement in the Positive and Negative Symptom Scale, among other metrics. This demonstrates the utility of targeting the RA signaling pathway to treat psychiatric disorders such as schizophrenia. The role of retinoids in schizophrenia etiology is reviewed in LaMantia 1999[94] and possible treatment is reviewed in Reay 2020.[136]

Major depressive disorder

A link between retinoids and depressive symptoms has been noted since 1983.[146] The synthetic retinoid isotretinoin (Accutane) has been marketed since this time for treatment resistant acne, and is now classified as one of the top 10 drugs associated with depressive symptoms and suicidal ideation in the Food and Drug Administration's database.[147] Extensive reports have been published linking clinical symptoms related to depression and retinoid treatment in humans and is reviewed in Hu *et al.* 2020[148] and Bremner and McCaffery 2008.[149] An additional association between therapeutic retinoids and changes in hypothalamus, hippocampus,[150] and orbitofrontal cortex have been noted in humans as well (reviewed in Bennett 2011[151] and Hu *et al.* 2020[148]). Altered neuron and glia density, as well as overall structure volume, is observed in these regions. A causal relationship was more recently established between administered 13-cis-RA[152] or all-trans-RA[153,154] and depressive behavior in mice or rats. This work highlighted the disruption of the hypothalamus-pituitary-adrenal (HPA) axis upon chronic retinoid exposure. A study of 12 diagnosed compared to 12 undiagnosed individuals identified elevated RARA levels in hypothalamic tissues, and further in vitro experiments demonstrated that RA signaling through RARA acts upstream of corticotropin releasing hormone gene expression.[155] This work established a potential RA-CRH pathway by which depressive symptoms might originate.

Further, recent work leveraging a human major depressive disorder (MDD) cohort of 109 individuals identified elevated retinol serum and at-RA synthesis levels in MDD individuals compared to neurotypical counterparts.[156] Interestingly, when accounting for sex, the difference in serum retinol was only significant when comparing diagnosed versus undiagnosed males, but not females. In contrast, synthesis of at-RA was significantly higher in diagnosed females compared to undiagnosed females, but no difference was detected for males. This suggests sex differences in response to RA, which could underlie sex biased penetrance of neurological conditions such as MDD (female-biased), as well as autism and schizophrenia (male-biased).

Genome wide association studies (GWAS) of family cohorts of MDD have revealed that this condition is about 30% heritable.[157,158] The significant GWAS loci can then be functionally interrogated by massively parallel reporter assays (MPRAs), as was done in Mulvey and Dougherty 2021.[159] This study tested over 1000 SNPs from 39 GWAS loci determined to carry significant genetic liability for MDD or other psychiatric conditions, identifying 75 functional SNPs. Interestingly, these functional SNPs shared significant enrichment of RA receptor-related motifs compared to non-functional SNPs, suggesting RA involvement in MDD liability, and many overlapped RXRA binding sites identified by ChIP-seq. This study further implicated serotonergic neurons as sensitive to RA-dependent transcriptional responses, conferring MDD risk through SNPs downstream of retinoid signaling. Further, treatment of at-RA in neuroblastoma N2a cells revealed dramatic changes in the chromatin landscape, and uncovered SNP effects previously masked without treatment. This suggests that a combination of genetic risk and environmental factors such as retinoid exposure lead to MDD liability.

A complex pathway implicated in several neuropsychiatric conditions

The last four decades of research have taught us that the effects of RA in development and related conditions are remarkably complex. One question central to this work is how one signaling molecule can have such pleiotropic, widespread effects on development, and therefore lead to a myriad of neurological conditions. However, as many conditions discussed above have complex genetic contributions, implicating the RA signaling pathway presents a reasonable path forward to understanding the etiology of these conditions.

It's possible that this can be at least partially explained by: RA interaction with other morphogens during development, including FGF, SHH, and BMPs, as discussed above; RA receptor interplay with other related steroid hormone receptors such as ERs, TR, serotonin, and dopamine; and the spatiotemporal origins of RA dysfunction. It has previously been proposed that optimal RA signaling operates within a normal distribution of active RA during development, where disorders arise when the levels of RA fall to either extreme of this distribution. Prior work has hypothesized that copy number variation, or potentially other forms of genetic variation, narrow the optimal window of signaling and create a heightened sensitivity to perturbations in RA signaling. This could explain why the CNV 22q11.2 represents the strongest known genetic risk for schizophrenia, where several genes in the RA metabolic pathway are downregulated in this condition.[160] Further work highlights the shared and distinct genetic components, development timing, and brain region implicated in various disorders could shed light on convergence towards unifying molecular pathways, including RA.

Discussion

It is clear that RA signaling is critical for embryonic development across cell types, brain regions, and tissues. The pleiotropic roles of RA during development gives rise to the myriad of consequential developmental conditions when perturbed. While much progress has been made to understand the molecular biology of this pathway in a tissue specific manner, how this single pathway contributes to neurodevelopmental conditions with a range of presentations and age of onset is still unclear. There is some overlap of genes, cell types, and circuits involved in these conditions. However, it is likely that additional layers of specificity exist that contribute to condition-specific etiology. These may include cell type of origin of *de novo* mutations during development, developmental timing of mutation incidence, co-expression of other pathway-specific genes (including other transcription factors, co-repressors, or co-activators) that interact with RA signaling, or tissue- or cell type-specific chromatin architecture that impacts RA-induced transcription. Differences in baseline gene expression of RA targets, and additional impact of environmental factors or small effect noncoding variation on these genes, may also lead to variation in vulnerability and complex presentation of RA perturbation.

The work highlighted here underscores the importance of investigating the effects of RA signaling in a cell type, brain region, and developmental timing specific way to identify the contributions of this diverse signaling molecule to the condition of interest. Many high confidence neurodevelopmental condition genes are also RA target genes, so modulating RA through vitamin A in a cell type and tissue specific manner could represent an efficient therapeutic intervention that is more effective than broad, systemic exposure as has been done before. Overall, the involvement of RA in development and a variety of conditions is clear, and future work that both narrows the

scope to the cell types and developmental timing of interest, as well as broadening the scope to identify biological convergence across conditions, will help take critical steps forward as we work to understand the full developmental implications of RA.

References

1. Sive, H. L., Draper, B. W., Harland, R. M. & Weintraub, H. Identification of a retinoic acid-sensitive period during primary axis formation in Xenopus laevis. *Genes Dev.* **4**, 932–942 (1990).
2. Vermot, J. & Pourquié, O. Retinoic acid coordinates somitogenesis and left–right patterning in vertebrate embryos. *Nature* **435**, 215–220 (2005).
3. Moreno, T. A. & Kintner, C. Regulation of Segmental Patterning by Retinoic Acid Signaling during Xenopus Somitogenesis. *Dev. Cell* **6**, 205–218 (2004).
4. Mendelsohn, C. *et al.* Function of the retinoic acid receptors (RARs) during development: (II) Multiple abnormalities at various stages of organogenesis in RAR double mutants. *Development* **120**, 2749–2771 (1994).
5. Dollé, P., Ruberte, E., Leroy, P., Morriss-Kay, G. & Chambon, P. Retinoic acid receptors and cellular retinoid binding proteins: I. A systematic study of their differential pattern of transcription during mouse organogenesis. *Development* **110**, 1133–1151 (1990).
6. Stratford, T., Horton, C. & Maden, M. Retinoic acid is required for the initiation of outgrowth in the chick limb bud. *Curr. Biol.* **6**, 1124–1133 (1996).
7. Yashiro, K. *et al.* Regulation of Retinoic Acid Distribution Is Required for Proximodistal Patterning and Outgrowth of the Developing Mouse Limb. *Dev. Cell* **6**, 411–422 (2004).
8. Dupé, V. & Lumsden, A. Hindbrain patterning involves graded responses to retinoic acid signalling. *Development* **128**, 2199–2208 (2001).
9. Chatzi, C., Brade, T. & Duester, G. Retinoic Acid Functions as a Key GABAergic Differentiation Signal in the Basal Ganglia. *PLoS Biol.* **9**, e1000609 (2011).

10. Rataj-Baniowska, M. *et al.* Retinoic Acid Receptor β Controls Development of Striatonigral Projection Neurons through FGF-Dependent and Meis1-Dependent Mechanisms. *J. Neurosci.* **35**, 14467–14475 (2015).

11. Liao, W.-L. *et al.* Modular patterning of structure and function of the striatum by retinoid receptor signaling. *Proc. Natl. Acad. Sci.* **105**, 6765–6770 (2008).

12. Novitch, B. G., Wichterle, H., Jessell, T. M. & Sockanathan, S. A Requirement for Retinoic Acid-Mediated Transcriptional Activation in Ventral Neural Patterning and Motor Neuron Specification. *Neuron* **40**, 81–95 (2003).

13. Wilson, L., Gale, E., Chambers, D. & Maden, M. Retinoic acid and the control of dorsoventral patterning in the avian spinal cord. *Dev. Biol.* **269**, 433–446 (2004).

14. Ziffra, R. S. *et al.* Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).

15. Shibata, M. *et al.* Regulation of prefrontal patterning and connectivity by retinoic acid. *Nature* **598**, 483–488 (2021).

16. Yang, L. *et al.* A Retinoic Acid Autoregulatory Loop Governing Prefrontal-Motor Arealization.

17. Siegenthaler, J. A. *et al.* Retinoic Acid from the Meninges Regulates Cortical Neuron Generation. *Cell* **139**, 597–609 (2009).

18. Duester, G. Retinoic Acid Synthesis and Signaling during Early Organogenesis. *Cell* **134**, 921–931 (2008).

19. Cunningham, T. J. & Duester, G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat. Rev. Mol. Cell Biol.* **16**, 110–123 (2015).

20. De Angelis, M. T., Parrotta, E. I., Santamaria, G. & Cuda, G. Short-term retinoic acid treatment sustains pluripotency and suppresses differentiation of human induced pluripotent stem cells. *Cell Death Dis.* **9**, 6 (2018).

21. Tan, B.-T. *et al.* Retinoic acid induced the differentiation of neural stem cells from embryonic spinal cord into functional neurons in vitro.

22. Lazzeri, G. *et al.* Retinoic Acid Promotes Neuronal Differentiation While Increasing Proteins and Organelles Related to Autophagy. *Int. J. Mol. Sci.* **26**, 1691 (2025).

23. Janesick, A., Wu, S. C. & Blumberg, B. Retinoic acid signaling and neuronal differentiation. *Cell. Mol. Life Sci.* **72**, 1559–1576 (2015).

24. Durston, A. J. *et al.* Retinoic acid causes an anteroposterior transformation in the developing central nervous system. *Nature* **340**, 140–144 (1989).

25. Horton, C. & Maden, M. Endogenous distribution of retinoids during normal development and teratogenesis in the mouse embryo. *Dev. Dyn.* **202**, 312–323 (1995).

26. Sirbu, I. O., Gresh, L., Barra, J. & Duester, G. Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression. *Development* **132**, 2611–2622 (2005).

27. Hernandez, R. E., Putzke, A. P., Myers, J. P., Margaretha, L. & Moens, C. B. Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development* **134**, 177–187 (2007).

28. Sosnik, J. *et al.* Noise modulation in retinoic acid signaling sharpens segmental boundaries of gene expression in the embryonic zebrafish hindbrain. *eLife* **5**, e14034 (2016).

29. Balmer, J. E. & Blomhoff, R. Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**, 1773–1808 (2002).

30. Balmer, J. E. & Blomhoff, R. A robust characterization of retinoic acid response elements based on a comparison of sites in three species. *J. Steroid Biochem. Mol. Biol.* **96**, 347–354 (2005).

31. Chen, J.-Y. *et al.* Two distinct actions of retinoid-receptor ligands. *Nature* **382**, 819–822 (1996).

32. Kashyap, V. & Gudas, L. J. Epigenetic Regulatory Mechanisms Distinguish Retinoic Acid-mediated Transcriptional Responses in Stem Cells and Fibroblasts. *J. Biol. Chem.* **285**, 14534–14548 (2010).

33. Hamed, M. *et al.* Insights into interplay between rexinoid signaling and myogenic regulatory factor-associated chromatin state in myogenic differentiation. *Nucleic Acids Res.* **45**, 11236–11248 (2017).

34. Clément, J.-T. & Li, Q. Cistromic insight into the association of retinoid X receptor with MyoD and CTCF in proliferating myoblasts. *Sci. Rep.* **15**, 38919 (2025).

35. Zhang, Y., Liang, J. & Li, Q. Coordinated Regulation of Retinoic Acid Signaling Pathway by KDM5B and Polycomb Repressive Complex 2. *J. Cell. Biochem.* **115**, 1528–1538 (2014).

36. Angrisano, T. *et al.* Chromatin and DNA methylation dynamics during retinoic acid-induced RET gene transcriptional activation in neuroblastoma cells. *Nucleic Acids Res.* **39**, 1993–2006 (2011).

37. Quadro, L., Hamberger, L., Colantuoni, V., Gottesman, M. E. & Blaner, W. S. Understanding the physiological role of retinol-binding protein in vitamin A metabolism using transgenic and knockout mouse models. *Mol. Aspects Med.* **24**, 421–430 (2003).

38. Quadro, L. *et al.* Pathways of Vitamin A Delivery to the Embryo: Insights from a New Tunable Model of Embryonic Vitamin A Deficiency. *Endocrinology* **146**, 4479–4490 (2005).

39. Wassef, L. & Quadro, L. Uptake of Dietary Retinoids at the Maternal-Fetal Barrier. *J. Biol. Chem.* **286**, 32198–32207 (2011).

40. Goodman, A. B. Retinoid Dysregulation May Result in Abnormal Expression of Glutamic Acid Decarboxylase in Schizophrenia. *Arch. Gen. Psychiatry* **53**, 653 (1996).

41. Dubey, A., Rose, R. E., Jones, D. R. & Saint-Jeannet, J. Generating retinoic acid gradients by local degradation during craniofacial development: One cell's cue is another cell's poison. *genesis* **56**, e23091 (2018).

42. Maden, M., Gale, E., Kostetskii, I. & Zile, M. Vitamin A-deficient quail embryos have half a hindbrain and other neural defects. *Curr. Biol.* **6**, 417–426 (1996).

43. Swindell, E. C. *et al.* Complementary Domains of Retinoic Acid Production and Degradation in the Early Chick Embryo. *Dev. Biol.* **216**, 282–296 (1999).

44. Shimozono, S., Iimura, T., Kitaguchi, T., Higashijima, S. & Miyawaki, A. Visualization of an endogenous retinoic acid gradient across embryonic development. *Nature* **496**, 363–366 (2013).

45. Rossant, J., Zirngibl, R., Cado, D. & Shago, M. Expression of a retinoic acid response element-hsplacZ transgene defines speci!ic domains of transcriptional . activity during mouse embryogenesls. 13.

46. Niederreither, K., McCaffery, P., Dräger, U. C., Chambon, P. & Dollé, P. Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. *Mech. Dev.* **62**, 67–78 (1997).

47. Rhinn, M. & Dolle, P. Retinoic acid signalling during development. *Development* **139**, 843–858 (2012).

48. Del Corral, R. D. *et al.* Opposing FGF and Retinoid Pathways Control Ventral Neural Pattern, Neuronal Differentiation, and Segmentation during Body Axis Extension. *Neuron* **40**, 65–79 (2003).

49. Cunningham, T. J. *et al.* Retinoic Acid Activity in Undifferentiated Neural Progenitors Is Sufficient to Fulfill Its Role in Restricting Fgf8 Expression for Somitogenesis. *PLOS ONE* **10**, e0137894 (2015).

50. Aulehla, A. & Pourquie, O. Signaling Gradients during Paraxial Mesoderm Development. *Cold Spring Harb. Perspect. Biol.* **2**, a000869–a000869 (2010).

51. Martínez-Morales, P. L. *et al.* FGF and retinoic acid activity gradients control the timing of neural crest cell emigration in the trunk. *J. Cell Biol.* **194**, 489–503 (2011).

52. Mic, F. A., Haselbeck, R. J., Cuenca, A. E. & Duester, G. Novel retinoic acid generating activities in the neural tube and heart identified by conditional rescue of *Raldh2* null mutant mice. *Development* **129**, 2271–2282 (2002).

53. Molotkov, A., Molotkova, N. & Duester, G. Retinoic acid generated by *Raldh2* in mesoderm is required for mouse dorsal endodermal pancreas development. *Dev. Dyn.* **232**, 950–957 (2005).

54. Halilagic, A. *et al.* Retinoids control anterior and dorsal properties in the developing forebrain. *Dev. Biol.* **303**, 362–375 (2007).

55. Hamazaki, N. *et al.* Retinoic acid induces human gastruloids with posterior embryo-like structures. *Nat. Cell Biol.* **26**, 1790–1803 (2024).

56. Blumberg, B. *et al.* An essential role for retinoid signaling in anteroposterior neural patterning. *Development* **124**, 373–379 (1997).

57. Papalopulu, N. & Kintner, C. A posteriorising factor, retinoic acid, reveals that anteroposterior patterning controls the timing of neuronal differentiation in *Xenopus* neuroectoderm. *Development* **122**, 3409–3418 (1996).

58. Dasen, J. S. & Jessell, T. M. Chapter Six Hox Networks and the Origins of Motor Neuron Diversity. in *Current Topics in Developmental Biology* vol. 88 169–200 (Elsevier, 2009).

59. Linville, A., Gumusaneli, E., Chandraratna, R. A. S. & Schilling, T. F. Independent roles for retinoic acid in segmentation and neuronal differentiation in the zebrafish hindbrain. *Dev. Biol.* **270**, 186–199 (2004).

60. Maves, L. & Kimmel, C. B. Dynamic and sequential patterning of the zebrafish posterior hindbrain by retinoic acid. *Dev. Biol.* **285**, 593–605 (2005).

61. Simeone, A. *et al.* Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. **346**, (1990).

62. Liu, J.-P., Laufer, E. & Jessell, T. M. Assigning the Positional Identity of Spinal Motor Neurons: Rostrocaudal Patterning of Hox-c Expression by FGFs, Gdf11, and Retinoids. *Neuron* **32**, 997–1012 (2001).

63. Sockanathan, S. & Jessell, T. M. Motor Neuron–Derived Retinoid Signaling Specifies the Subtype Identity of Spinal Motor Neurons. *Cell* **94**, 503–514 (1998).

64. Ulven, S. M. *et al.* Identification of Endogenous Retinoids, Enzymes, Binding Proteins, and Receptors during Early Postimplantation Development in Mouse: Important Role of Retinal Dehydrogenase Type 2 in Synthesis of All-trans-Retinoic Acid. *Dev. Biol.* **220**, 379–391 (2000).

65. McCaffery, P. & Dräger, U. C. High levels of a retinoic acid-generating dehydrogenase in the meso-telencephalic dopamine system. *Proc. Natl. Acad. Sci.* **91**, 7772–7776 (1994).

66. Abu-Abed, S. *et al.* The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev.* **15**, 226–240 (2001).

67. Sockanathan, S., Perlmann, T. & Jessell, T. M. Retinoid Receptor Signaling in Postmitotic Motor Neurons Regulates Rostrocaudal Positional Identity and Axonal Projection Pattern. *Neuron* **40**, 97–111 (2003).

68. Jessell, T. M. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat. Rev. Genet.* **1**, 20–29 (2000).

69. Schneider, R. A., Hu, D., Rubenstein, J. L. R., Maden, M. & Helms, J. A. Local retinoid signaling coordinates forebrain and facial morphogenesis by maintaining FGF8 and SHH. *Development* **128**, 2755–2767 (2001).

70. Li, H. *et al.* A retinoic acid synthesizing enzyme in ventral retina and telencephalon of the embryonic mouse. *Mech. Dev.* **95**, 283–289 (2000).

71. Molotkova, N., Molotkov, A. & Duester, G. Role of retinoic acid during forebrain development begins late when Raldh3 generates retinoic acid in the ventral subventricular zone. *Dev. Biol.* **303**, 601–610 (2007).

72. Jiang, X., Iseki, S., Maxson, R. E., Sucov, H. M. & Morriss-Kay, G. M. Tissue Origins and Interactions in the Mammalian Skull Vault. *Dev. Biol.* **241**, 106–116 (2002).

73. Couly, G. F., Coltey, P. M. & Le Douarin, N. M. The developmental fate of the cephalic mesoderm in quail-chick chimeras. *Development* **114**, 1–15 (1992).

74. O'Rahilly, R. & Müller, F. The meninges in human development. *J Neuropathol Exp Neurol* **45**, 588–608 (1986).

75. Zhang, J., Smith, D., Yamamoto, M., Ma, L. & McCaffery, P. The Meninges Is a Source of Retinoic Acid for the Late-Developing Hindbrain. *J. Neurosci.* **23**, 7610–7620 (2003).

76. DeSisto, J. *et al.* Single-Cell Transcriptomic Analyses of the Developing Meninges Reveal Meningeal Fibroblast Diversity and Function. *Dev. Cell* **54**, 43-59.e4 (2020).

77. Romand, R., Kondo, T., Cammas, L., Hashino, E. & Dollé, P. Dynamic expression of the retinoic acid-synthesizing enzyme retinol dehydrogenase 10 (rdh10) in the developing mouse brain and sensory organs. *J. Comp. Neurol.* **508**, 879–892 (2008).

78. Haushalter, C., Asselin, L., Fraulob, V., Dollé, P. & Rhinn, M. Retinoic acid controls early neurogenesis in the developing mouse cerebral cortex. *Dev. Biol.* **430**, 129–141 (2017).

79. Morrison, V. E., Smith, V. N. & Huang, J. K. Retinoic Acid Is Required for Oligodendrocyte Precursor Cell Production and Differentiation in the Postnatal Mouse Corpus Callosum. *eneuro* **7**, ENEURO.0270-19.2019 (2020).

80. Yamamoto, M., McCaffery, P. & Dräger, U. C. Influence of the choroid plexus on cerebellar development: analysis of retinoic acid synthesis. *Dev. Brain Res.* **93**, 182–190 (1996).

81. Yamamoto, M., Dräger, U. C., Ong, D. E. & McCaffery, P. Retinoid-binding proteins in the cerebellum and choroid plexus and their relationship to regionalized retinoic acid synthesis and degradation. *Eur. J. Biochem.* **257**, 344–350 (1998).

82. Haskell, G. T. & LaMantia, A.-S. Retinoic Acid Signaling Identifies a Distinct Precursor Population in the Developing and Adult Forebrain. *J. Neurosci.* **25**, 7636–7647 (2005).

83. Willsey, A. J. *et al.* Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism. *Cell* **155**, 997–1007 (2013).

84. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e23 (2020).

85. 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).

86. Kumar, S. *et al.* Impaired neurodevelopmental pathways in autism spectrum disorder: a review of signaling mechanisms and crosstalk. *J. Neurodev. Disord.* **11**, 10 (2019).

87. Willsey, H. R., Willsey, A. J., Wang, B. & State, M. W. Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. *Nat. Rev. Neurosci.* **23**, 323–341 (2022).

88. Jones-Villeneuve, E. M., McBurney, M. W., Rogers, K. A. & Kalnins, V. I. Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J. Cell Biol.* **94**, 253–262 (1982).

89. Devanna, P., Middelbeek, J. & Vernes, S. C. FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. *Front. Cell. Neurosci.* **8**, (2014).

90. Jacobs, S. *et al.* Retinoic acid is required early during adult neurogenesis in the dentate gyrus. *Proc. Natl. Acad. Sci.* **103**, 3902–3907 (2006).

91. Mishra, S., Kelly, K. K., Rumian, N. L. & Siegenthaler, J. A. Retinoic Acid Is Required for Neural Stem and Progenitor Cell Proliferation in the Adult Hippocampus. *Stem Cell Rep.* **10**, 1705–1720 (2018).

92. Yu, S., Levi, L., Siegel, R. & Noy, N. Retinoic Acid Induces Neurogenesis by Activating Both Retinoic Acid Receptors (RARs) and Peroxisome Proliferator-activated Receptor β/δ (PPARβ/δ). *J. Biol. Chem.* **287**, 42195–42205 (2012).

93. Diez del Corral, R. & Morales, A. Retinoic Acid Signaling during Early Spinal Cord Development. *J. Dev. Biol.* **2**, 174–197 (2014).

94. LaMantia, A.-S. Forebrain Induction, Retinoic Acid, and Vulnerability to Schizophrenia: Insights from Molecular and Genetic Analysis in Developing Mice. *BIOL PSYCHIATRY* **46**, 19–30 (1999).

95. Harrison-Uy, S. J., Siegenthaler, J. A., Faedo, A., Rubenstein, J. L. R. & Pleasure, S. J. CoupTFI Interacts with Retinoic Acid Signaling during Cortical Development. *PLoS ONE* **8**, e58219 (2013).

96. Lenz, M. *et al.* All-trans retinoic acid induces synaptic plasticity in human cortical neurons. *eLife* **10**, e63026 (2021).

97. Asanuma, K. *et al.* Synaptopodin orchestrates actin organization and cell motility via regulation of RhoA signalling. *Nat. Cell Biol.* **8**, 485–491 (2006).

98. Yap, K. *et al.* The actin-modulating protein synaptopodin mediates long-term survival of dendritic spines. *eLife* **9**, e62944 (2020).

99. Arendt, K. L. *et al.* Calcineurin mediates homeostatic synaptic plasticity by regulating retinoic acid synthesis. *Proc. Natl. Acad. Sci.* **112**, E5744–E5752 (2015).

100. Aoto, J., Nam, C. I., Poon, M. M., Ting, P. & Chen, L. Synaptic Signaling by All-Trans Retinoic Acid in Homeostatic Synaptic Plasticity. *Neuron* **60**, 308–320 (2008).

101. Hsu, Y.-T., Li, J., Wu, D., Südhof, T. C. & Chen, L. Synaptic retinoic acid receptor signaling mediates mTOR-dependent metaplasticity that controls hippocampal learning. *Proc. Natl. Acad. Sci.* **116**, 7113–7122 (2019).

102. Vlachos, A. *et al.* Synaptopodin Regulates Plasticity of Dendritic Spines in Hippocampal Neurons. *J. Neurosci.* **29**, 1017–1033 (2009).

103. Vlachos, A. *et al.* Synaptopodin regulates denervation-induced homeostatic synaptic plasticity. *Proc. Natl. Acad. Sci.* **110**, 8242–8247 (2013).

104. Misner, D. L. *et al.* Vitamin A deprivation results in reversible loss of hippocampal long-term synaptic plasticity. *Proc. Natl. Acad. Sci.* **98**, 11714–11719 (2001).

105. Mingaud, F. *et al.* Retinoid Hyposignaling Contributes to Aging-Related Decline in Hippocampal Function in Short-Term/Working Memory Organization and Long-Term Declarative Memory Encoding in Mice. *J. Neurosci.* **28**, 279–291 (2008).

106. Chen, L., Lau, A. G. & Sarti, F. Synaptic retinoic acid signaling and homeostatic synaptic plasticity. *Neuropharmacology* **78**, 3–12 (2014).

107. Rothman, K. J., Nguyen, U.-S. D. T. & Milunsky, A. Teratogenicity of High Vitamin A Intake. *N. Engl. J. Med.* **333**, 5 (1995).

108. Bigby, M. & Stern, R. S. Adverse reactions to isotretinoin. *J. Am. Acad. Dermatol.* **18**, 543–552 (1988).

109. Windhorst, D. B. & Nigra, T. General clinical toxicology of oral retinoids. *J. Am. Acad. Dermatol.* **6**, 675–682 (1982).

110. Mulder, G. B. *et al.* Effects of excess vitamin A on development of cranial neural crest-derived structures: A neonatal and embryologic study. *Teratology* **62**, 214–226 (2000).

111. Kessel, M. & Gruss, P. Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell* **67**, 89–104 (1991).

112. I Altaba, A. R. & Jessell, T. M. Retinoic acid modifies the pattern of cell differentiation in the central nervous system of neurula stage *Xenopus* embryos. *Development* **112**, 945–958 (1991).

113. Niederreither, K., Vermot, J., Schuhbaur, B., Chambon, P. & Dollé, P. Retinoic acid synthesis and hindbrain patterning in the mouse embryo. *Development* **127**, 75–85 (2000).

114. Luo, T., Wagner, E., Crandall, J. E. & Dräger, U. C. A retinoic-acid critical period in the early postnatal mouse brain. *Biol. Psychiatry* **56**, 971–980 (2004).

115. Nguyen, A., Rauch, T. A., Pfeifer, G. P. & Hu, V. W. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, *RORA* , whose protein product is reduced in autistic brain. *FASEB J.* **24**, 3036–3051 (2010).

116. Sayad, A., Noroozi, R., Omrani, M. D., Taheri, M. & Ghafouri-Fard, S. Retinoic acid-related orphan receptor alpha (RORA) variants are associated with autism spectrum disorder. *Metab. Brain Dis.* **32**, 1595–1601 (2017).

117. The DDD Study *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).

118. Boudry-Labis, E. *et al.* A novel microdeletion syndrome at 9q21.13 characterised by mental retardation, speech delay, epilepsy and characteristic facial features. *Eur. J. Med. Genet.* **56**, 163–170 (2013).

119. Rudolf, G. *et al.* Loss of function of the retinoid-related nuclear receptor (RORB) gene and epilepsy. *Eur. J. Hum. Genet.* **24**, 1761–1770 (2016).

120. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

121. Lai, X. *et al.* Vitamin A Deficiency Induces Autistic-Like Behaviors in Rats by Regulating the RARβ-CD38-Oxytocin Axis in the Hypothalamus. *Mol. Nutr. Food Res.* **62**, 1700754 (2018).

122. Bandini, L. G. *et al.* Food Selectivity in Children with Autism Spectrum Disorders and Typically Developing Children. *J. Pediatr.* **157**, 259–264 (2010).

123. Hyman, S. L. *et al.* Nutrient Intake From Food in Children With Autism. *Pediatrics* **130**, S145–S153 (2012).

124. Liu, X. *et al.* Correlation between Nutrition and Symptoms: Nutritional Survey of Children with Autism Spectrum Disorder in Chongqing, China. *Nutrients* **8**, 294 (2016).

125. Guo, M. *et al.* Vitamin A improves the symptoms of autism spectrum disorders and decreases 5-hydroxytryptamine (5-HT): A pilot study. *Brain Res. Bull.* **137**, 35–40 (2018).

126. Goodman, A. B. Chromosomal locations and modes of action of genes of the retinoid (vitamin A) system support their involvement in the etiology of schizophrenia. *Am. J. Med. Genet.* **60**, 335–348 (1995).

127. Goodman, A. B. Three independent lines of evidence suggest retinoids as causal to schizophrenia. *Proc. Natl. Acad. Sci.* **95**, 7240–7244 (1998).

128. Susser, E. S., Schaefer, C. A., Brown, A. S., Begg, M. D. & Wyatt, R. J. The Design of the Prenatal Determinants of Schizophrenia Study. *Schizophr. Bull.* **26**, 257–273 (2000).

129. Bao, Y. *et al.* Low maternal retinol as a risk factor for schizophrenia in adult offspring. *Schizophr. Res.* **137**, 159–165 (2012).

130. Regen, F. *et al.* Clozapine modulates retinoid homeostasis in human brain and normalizes serum retinoic acid deficit in patients with schizophrenia. *Mol. Psychiatry* **26**, 5417–5428 (2021).

131. Goodman, A. B. Microarray results suggest altered transport and lowered synthesis of retinoic acid in schizophrenia. *Mol. Psychiatry* **10**, 620–621 (2005).

132. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).

133. Haybaeck, J. *et al.* Increased expression of retinoic acid-induced gene 1 in the dorsolateral prefrontal cortex in schizophrenia, bipolar disorder, and major depression. *Neuropsychiatr. Dis. Treat.* 279 (2015) doi:10.2147/NDT.S72536.

134. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

135. GERAD1 Consortium *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).

136. Reay, W. R. & Cairns, M. J. The role of the retinoids in schizophrenia: genomic and clinical perspectives. *Mol. Psychiatry* **25**, 706–718 (2020).

137. Lee, C.-H., Liu, C.-M., Wen, C.-C., Chang, S.-M. & Hwu, H.-G. Genetic copy number variants in sib pairs both affected with schizophrenia. *J. Biomed. Sci.* **17**, 2 (2010).

138. Willner, P. The dopamine hypothesis of schizophrenia: current status, future prospects. *Int Clin Psychopharmacol* **12**, 297–308 (1997).

139. Wolf, G. Vitamin A Functions in the Regulation of the Dopaminergic System in the Brain and Pituitary Gland. *Nutr. Rev.* **56**, 354–355 (2009).

140. Kręz˙el, W. *et al.* Impaired Locomotion and Dopamine Signaling in Retinoid Receptor Mutant Mice. *Science* **279**, 863–867 (1998).

141. McGrath, J. J. *et al.* Neonatal Vitamin D Status and Risk of Schizophrenia. *ARCH GEN PSYCHIATRY* **67**, (2010).

142. Schräder, M., Bendik, I., Becker-André, M. & Carlberg, C. Interaction between retinoic acid and vitamin D signaling pathways. *J. Biol. Chem.* **268**, 17830–17836 (1993).

143. Lerner, V., McCaffery, P. J. A. & Ritsner, M. S. Targeting Retinoid Receptors to Treat Schizophrenia: Rationale and Progress to Date. *CNS Drugs* **30**, 269–280 (2016).

144. Lerner, V. *et al.* Bexarotene as Add-On to Antipsychotic Treatment in Schizophrenia Patients: A Pilot Open-Label Trial. **31**, 9 (2008).

145. Lerner, V. *et al.* The Retinoid X Receptor Agonist Bexarotene Relieves Positive Symptoms of Schizophrenia: A 6-Week, Randomized, Double-Blind, Placebo-Controlled Multicenter Trial. *J. Clin. Psychiatry* **74**, 1224–1232 (2013).

146. Hazen, P. G., Carney, J. F., Walker, A. E. & Stewart, J. J. Depression--a side effect of 13-cis-retinoic acid therapy. *J Am Acad Dermatol* **9**, 278–279 (1983).

147. Barak, Y. *et al.* Affective psychosis following Accutane (isotretinoin) treatment. *Int Clin Psychopharmacol* **20**, 39–40 (2005).

148. Hu, P., Van Dam, A.-M., Wang, Y., Lucassen, P. J. & Zhou, J.-N. Retinoic acid and depressive disorders: Evidence and possible neurobiological mechanisms. *Neurosci. Biobehav. Rev.* **112**, 376–391 (2020).

149. Bremner, J. D. & McCaffery, P. The neurobiology of retinoic acid in affective disorders. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **32**, 315–331 (2008).

150. McCaffery, P., Zhang, J. & Crandall, J. E. Retinoic acid signaling and function in the adult hippocampus. *J. Neurobiol.* **66**, 780–791 (2006).

151. Bennett, M. R. The prefrontal–limbic network in depression: A core pathology of synapse regression. *Prog. Neurobiol.* **93**, 457–467 (2011).

152. O'Reilly, K. C., Shumake, J., Gonzalez-Lima, F., Lane, M. A. & Bailey, S. J. Chronic Administration of 13-Cis-Retinoic Acid Increases Depression-Related Behavior in Mice. *Neuropsychopharmacology* **31**, 1919–1927 (2006).

153. Hu, P. *et al.* All-trans retinoic acid-induced hypothalamus–pituitary–adrenal hyperactivity involves glucocorticoid receptor dysregulation. *Transl. Psychiatry* **3**, e336–e336 (2013).

154. Cai, L., Li, R. & Zhou, J.-N. Chronic all-trans retinoic acid administration induces CRF over-expression accompanied by AVP up-regulation and multiple CRF-controlling receptors disturbance in the hypothalamus of rats. *Brain Res.* **1601**, 1–7 (2015).

155. Chen, X.-N. *et al.* The Involvement of Retinoic Acid Receptor-α in Corticotropin-Releasing Hormone Gene Expression and Affective Disorders. *Biol. Psychiatry* **66**, 832–839 (2009).

156. Otto, L. R. *et al.* Retinoid homeostasis in major depressive disorder. *Transl. Psychiatry* **13**, 67 (2023).

157. Flint, J. & Kendler, K. S. The Genetics of Major Depression. *Neuron* **81**, 484–503 (2014).

158. Corfield, E. C., Yang, Y., Martin, N. G. & Nyholt, D. R. A continuum of genetic liability for minor and major depression. *Transl. Psychiatry* **7**, e1131–e1131 (2017).

159. Mulvey, B. & Dougherty, J. D. Transcriptional-regulatory convergence across functional MDD risk variants identified by massively parallel reporter assays. *Transl. Psychiatry* **11**, 403 (2021).

160. Maynard, T. M. *et al.* 22q11 Gene dosage establishes an adaptive range for sonic hedgehog and retinoic acid signaling during early development. *Hum. Mol. Genet.* **22**, 300–312 (2013).

CHAPTER 3: AUTISM MUTATIONS REWIRE PROTEIN INTERACTION NETWORKS TO DRIVE NEURODEVELOPMENTAL PATHOLOGY

Introduction

ASD is a highly heritable neurodevelopmental syndrome characterized by marked genetic heterogeneity and inter-individual clinical variability.[1-3] Over the past 15 years, whole-exome sequencing (WES) studies have identified more than 250 large-effect risk genes based on rare, often *de novo*, protein-damaging single nucleotide variants (SNVs).[4-7] These large-effect coding mutations are strongly enriched among the most severely impaired individuals with ASD and co-occurring neurodevelopmental disorders, including intellectual disability, epilepsy, minimal or absent speech, and severe motor delay.

Despite this progress, translating high-confidence ASD risk genes (hcASD) into a deeper understanding of pathophysiology–and ultimately into effective, targeted therapies for the most severe forms of the disorder – remains a major challenge.[8] This difficulty stems from extensive pleiotropy, the dynamism of the developing human brain, and limited knowledge of the physical interactions of ASD risk proteins and how mutations alter the structure and function of these protein complexes.[9,10] Two central unresolved questions are the degree to which diverse ASD proteins converge onto shared molecular complexes and how specific disease-causing mutations alter protein structure, interactions, function and, ultimately, neurodevelopment.

One promising strategy is to examine *bona fide* ASD risk genes at the level of protein–protein interactions (PPIs), and whether disease-associated coding mutations selectively disrupt or rewire these interactions rather than simply abolishing gene function.[11] If hcASD-associated mutations converge by rewiring shared protein complexes, then PPI networks should point to targetable causal mechanisms linking genetic variation to neurodevelopmental dysfunction. Testing this hypothesis requires systematic, large-scale, disease-relevant interaction maps that resolve both

wild-type (WT) interactions and mutation-specific perturbations across a substantial fraction of the ASD proteome. However, such maps have been largely lacking.

Here, using affinity purification–mass spectrometry (AP-MS), we mapped PPIs for 100 hcASD genes and 54 patient-derived pathogenic variants.[4] We generated these interaction maps in HEK293 cells to enable scalability and cross-disease comparability, and then validated key findings across multiple experimental systems, including *Xenopus tropicalis*, human induced pluripotent stem cell (iPSC)–derived neural progenitors, and forebrain organoids. This approach revealed a dual-layered architecture of molecular convergence in ASD: stable convergence of risk proteins onto shared complexes in the WT state, and convergent patterns of mutation-induced PPI rewiring.

We demonstrate that proteins interacting with these 100 ASD risk genes form a highly interconnected network that recapitulates ASD-relevant risk gene expression patterns in the developing human brain and is enriched for additional ASD risk genes, but not schizophrenia risk genes. Functional validation of key interactions confirms that these mutations perturb early neurogenesis and alter electrophysiological properties. Together, these results indicate that genetically heterogeneous ASD risk genes and distinct mutations converge onto complexes and rewire PPIs, providing a mechanistic framework linking genetic variation to protein structure and function and ultimately, to convergent neurodevelopmental pathology.

Results

Protein-protein interaction mapping reveals ASD networks

We evaluated structural and functional relationships among identified hcASD genes[4] (**Fig. 3.1A**) through the mapping of 100 Strep-tagged hcASD proteins individually expressed in HEK293T

cells using AP-MS, generating an ASD protein-protein interaction network (ASD-PPI) (**Fig. 3.1B, Fig. 3.9, A to C**). HEK293T cells provide the tractability, proteomic depth, and established benchmarks required for high-throughput PPI mapping. This approach has previously enabled the large-scale identification of *bona fide*, stable, and often stoichiometric physical interactions present across cell types for cancer, heart disease, neurodegeneration and infectious diseases.[12-19] The resulting network connects 100 hcASD proteins with 1,074 unique high-confidence interactors via 1,881 total interactions. These include 87% previously unidentified interactions, with a median of 11 interactors per hcASD (**Fig. 3.1, C and D**). These interactions reveal strong interconnectivity among hcASD proteins and convergence onto shared interactors and complexes (see below). The ASD-PPI captures a broader interaction landscape than existing PPI databases (**Fig. 3.1E**) and overlaps substantially with previously published, smaller ASD PPI networks generated in neuronal cultures or brain tissue (**Fig. 3.9D**), thus supporting its biological relevance.

ASD-PPI interactors are expressed in the human brain and enriched for ASD genetic risk

ASD-PPI interactors recapitulate core transcriptional and genetic features of ASD risk genes. In RNA-seq data from developing[20] and adult human brain,[21] the subset of interactors excluding hcASD (interactors$^{\text{no hcASD}}$, n = 1,043) is highly expressed in brain tissue and closely tracks hcASD expression patterns across pre- and postnatal development (**Fig. 3.2, A-C, Fig. 3.10A**). This subset recapitulates core expression signatures of ASD-associated genes, including elevated expression in prenatal cortex[22] and adult cortex and cerebellum (**Fig. 3.10, B, C**) and shows intermediate evolutionary constraint greater than the HEK293 proteome but lower than hcASD (**Fig. 3.10D**), suggesting they may contribute to ASD through polygenic mechanisms. ASD-associated *de novo* damaging variants are enriched in both the full set of interactors (interactors, n = 1,074) and interactors$^{\text{no hcASD}}$, but not in the rest of the HEK293T proteome (**Fig. 3.2D**), indicating that

interactors themselves are genetically associated with ASD. This enrichment magnitude is comparable to interactors from ASD networks generated in neuronal cultures[23] and from brain tissue[24,25] (**Fig. 3.10, E to G**). ASD-PPI interactors also strongly overlap with ASD risk genes identified from recent large-scale sequencing studies,[5-7] but not with genes implicated in schizophrenia,[26] supporting a degree of specificity (**Fig. 3.2E, fig. 3.10H**). To explore the potential of ASD-PPI for novel ASD gene discovery, we modeled networks of increasing size. As more hcASD are incorporated, enrichment of ASD-associated genetic variants progressively concentrates within interactors$^{\text{no hcASD}}$ (**Fig. 3.2F**) and captures a growing number of known and newly identified risk genes (hcASD+),[6] with no sign of saturation at 100 hcASD (**Fig. 3.2G**), suggesting that expanding the network could reveal additional risk genes. Together, these orthogonal analyses demonstrate that ASD-PPI captures ASD-relevant interactions.

To pinpoint ASD-relevant cell types, we examined ASD-PPI network expression in a prenatal human brain single-cell RNA sequencing (scRNAseq) atlas spanning cortical and subcortical cell types[27] (**Fig. 3.2H**). Previous studies have implicated post-mitotic excitatory and inhibitory neurons (ENs and INs).[4-6] ASD-PPI shows strongest network expression in post-mitotic ENs, with some enrichment in INs (**Fig. 3.2I**). However, expanding ASD-PPI by incorporating connections between interactors from the PPI metadatabase STRING,[28] increases the statistical weight of the highly-connected interactors in the network. This reframing from a risk gene-centric to interactome-centric view shifted the enrichment away from mature neurons and toward neural progenitor cells (NPCs) of excitatory neurons (NPC-EN) and, to a lesser extent, of inhibitory neurons (NPC-IN) (**Fig. 3.2I**). This pattern was consistent across three independent prenatal brain atlases (**Fig. 3.10I**) and recapitulated with a smaller network generated in neurons (iPSC-derived EN PPI network (iEN-PPI)[23]; **Fig. 3.2I**). These results indicate that ASD-relevant protein

interactomes converge primarily on EN generation and early differentiation, a shift from prior reports that broadly implicated post-mitotic neurons based on enrichment of hcASD genes alone.[4,6]

ASD-PPI reveals molecular convergence

Organizing hcASD and their interactors by connectivity and similarity in Gene Ontology annotations[29] (**Fig. 3.3A**) showed functional convergence consistent with previous results.[10,22,30,31] The ASD-PPI network shows high connectivity: 31% of hcASD proteins interact with each other, while 35% of interactors bind multiple hcASD (**Fig. 3.11A**), suggesting functional overlap. Consistent with this, the network exhibits increases in edge density and average node degree, measures associated with molecular convergence, while demonstrating low modularity (**Fig. 3.3B, Fig. 3.11B**). This reflects a dense, interconnected landscape across hcASD rather than segregation into isolated modules. These values are more extreme than those observed in previously published AP-MS networks organized around unified biological themes, such as tyrosine kinases (n=90[32]) or breast cancer genes (n=39[12]), indicating stronger molecular convergence in ASD-PPI (**Fig. 3.3B, and Fig. 3.11B**). Stratifying hcASD by clinical phenotype (ASD-predominant (ASD_P) vs. broad neurodevelopmental; NDD (ASD_{NDD})[4] shows no increase in interactor overlap, suggesting no functional distinction between these hcASD subgroups (**Fig. 3.11, B, C**).

Several hcASD interact with protein complexes previously linked to ASD through both known and previously unknown interactions. These include the Sin3 complex (progenitor cell proliferation and differentiation[33,34]), the Mediator complex (transcription and neural stem cell identity[35,36]) and the PAF1 complex (progenitor proliferation and neuronal migration[37-39]), with subunits CTR9 and LEO1 near or just below the threshold for high-confidence ASD association (**Fig. 3.3, C to E, and Fig. 3.11, D and E**). Although individual components of the PAF1 complex did not meet the statistical thresholds for hcASD[4,6], their high degree of connectivity to multiple

hcASD baits suggests a collective role in ASD pathobiology, consistent with recent clinical genetic evidence supporting PAF1 complex member LEO1 as an ASD risk gene.[40] Additionally, 17 hcASD interact with the AP2-associated clathrin-mediated endocytosis complex (**Fig. 3.3F**), suggesting these genes have diverse roles beyond gene regulation.

ASD-PPI is enriched for disease-relevant direct PPIs as predicted by AlphaFold

ASD-PPI includes both direct and indirect PPIs. To identify direct binding partners and their 3D interaction interfaces, we repurposed AlphaFold (AF)[41,42] to screen our ASD-PPI, which constrains the prediction of hcASD interactions across the human proteome to 1,881 experimentally supported interactions (**Fig. 3.4A**). This included both hcASD-interactor (hcASD-int; ASD-PPI) and interactor-interactor (int-int) pairs, the latter not directly assayed by our AP-MS. To identify a robust predictor of direct interactions, we evaluated several AF-derived metrics. The mean interface predicted TM-score (mean ipTM),[41] which reflects the consistency and accuracy of predicted interaction interfaces, best distinguished hcASD-int pairs from negative controls (hcASD-random), with scores above 0.5 showing a 5- to 10-fold enrichment for likely direct contacts (**Fig. 3.4B, and Fig. 3.12, A and B**). This identified 113 high-confidence direct hcASD interactions (FDR 10-20%, 49/113 previously unrecognized) (**Fig. 3.12C**) and 466 indirect interactors connected to hcASD proteins via intermediate binding partners (int-int links, **Fig. 3.4C**). ASD-PPI showed nearly twice the rate of predicted direct interactions compared to iEN-PPI,[23] suggesting that ASD-PPI is enriched for direct PPIs.

We validated a subset of predicted interactions using the NanoLuc (N2H) split luciferase binary interaction assay,[43,44] confirming 12 of 37 tested pairs (~32%) as direct binders (**Fig. 3.12D**), in range with N2H performance on known direct interactors and higher than the 10-15% typically

observed for AP-MS data.[45-47] Several pairs negative in N2H (for example, NUP155–SMPD4, SETD5–TBL1X, GNAI–TNFAIP8, DYRK1A-KIAA0232, and DYRK1A-DCAF7) were confirmed via immunohistochemistry-based colocalization in NPCs, providing orthogonal support for their physical proximity (**Fig. 3.12E, also see Fig. 3.13E**). The previously unknown PPI and AF-predicted interaction between NUP155 and SMPD4 (**Fig. 3.4G**), a protein linked to microcephaly,[48] highlights the approach's value for uncovering key interactions in neurodevelopmental conditions.

The interaction between hcASD protein DYRK1A and DCAF7 exemplifies a high-confidence AF interface prediction (**Fig. 3.4, B, D to F**). Although established,[49] its 3D structure remains unsolved. The predicted interface matches the known DCAF7-binding region of DYRK1A (residues 80-100),[50] and shows high evolutionary conservation, consistent with selective pressure to preserve functionally important binding (**Fig. 3.4, E and F**). Deletion of this region (DYRK1A$^{\Delta 80\text{-}100}$) disrupts binding to DCAF7, but not to FAM54C, which binds the DYRK1A catalytic domain (residues 156-479)[51] (**Fig. 3.12F, table S4**). These findings demonstrate the ability of AF to predict direct PPIs as well as the specific interfaces mediating these interactions. We further confirmed that AF-predicted interfaces can model the impact of patient mutations: targeted mutagenesis of the GNAI1-RIC8A interface showed that binding affinity correlates with predicted pathogenicity (see below and **Fig. 3.16E**).

To determine whether the AF-supported direct interactome is enriched in specific cell types, we evaluated its expression across the three prenatal brain scRNAseq cell atlases[27,52,53] as described above (see **Fig. 3.2I, and Fig. 3.10I**). The 113 AF-predicted direct hcASD-int pairs showed no cell type enrichment; however, the inclusion of AF-predicted int-int connections led to an enrichment in NPCs, particularly NPC-ENs (**Fig. 3.4H, Fig. 3.12G**), consistent with our STRING-

based findings (see above; **Fig. 3.2I, and Fig. 3.10I**). This pattern was recapitulated with iEN-PPI (**Fig. 3.2I, and Fig. 3.10I**). Thus, the ASD interactome converges primarily on the generation and early differentiation of the excitatory lineage.

The DCAF7-DYRK1A-KIAA0232 complex regulates neurogenesis and differentiation

To further explore molecular convergence, we focused on seven "hub interactors" that bind eight or more hcASD proteins (**Fig. 3.5A, and Fig. 3.11A**). Only the DCAF7 interactome showed strong enrichment for additional ASD risk genes (hcASD+[6]) in available interactomes,[54] suggesting hcASD converge on DCAF7 at the protein level (**Fig. 3.5B**).

DCAF7 is an adaptor protein that interacts with hcASD like DYRK1A, regulating its nuclear translocation and protein interactions (for example, with HAP1),[55] and AUTS2 and SKI to regulate neuronal lineage specification, highlighting its role in neurodevelopment.[56] Comparing interactor overlap with DCAF7 and its binding hcASD partners revealed that hcASDs DYRK1A and previously uncharacterized KIAA0232 share substantial interactor overlap with DCAF7 (**Fig. 3.5D**), and AF predicted both to be direct DCAF7 interactors. Endogenous DYRK1A AP-MS in iPSC-derived iENs recovered DCAF7 and KIAA0232 and showed considerable overlap with HEK293-derived interactors (**Fig. 3.13A and B**). Sequential AP-MS in HEK293 further confirmed that these three proteins physically interact as a complex, whose 126 shared interactors are enriched for hcASD+ (**Fig. 3.5E, and Fig. 3.13, C and D**). All three proteins localize to mitotic spindles in HEK cells and NPCs (**Fig. 3.13E**), consistent with previous findings for DYRK1A in *Xenopus* and human NPCs.[57,58] Furthermore, ASD-PPI interactors are enriched for proteins associated with mitotic spindle organization, such as centriolar satellites and centrosomes (**Fig. 3.13F**).

We evaluated whether DCAF7, DYRK1A, and KIAA0232 share neurodevelopmental functions *in vivo*. Prior work showed that DYRK1A loss in *Xenopus* disrupts the cell cycle, increases apoptosis, and reduces telencephalon size.[58] Using CRISPR-Cas9 mutagenesis in *Xenopus*, we found that knockout of *dcaf7* or *kiaa0232* phenocopies *dyrk1a* loss, reducing telencephalon size and linking this complex to an ASD-associated phenotype (**Fig. 3.5F**). To relate this to human clinical presentations, we analyzed cortical organoid scRNAseq data from idiopathic ASD patients[59]: the DCAF7-hcASD subnetwork was enriched among differentially expressed genes (DEGs) in normocephalic, but not macrocephalic, ASD (**Fig. 3.13G**). This suggests DCAF7 interactions contribute to ASD presentations without macrocephaly, consistent with the reduced forebrain growth observed in *Xenopus*.

To investigate cellular mechanisms, we used CRISPR-mediated knockdown (KD) in human iPSC-derived NPCs. DCAF7 KD increased cell death during neuronal differentiation (**Fig. 3.13, H and I**), consistent with DCAF7's role in promoting cell viability.[60] KD of either DCAF7 or KIAA0232 reduced the proportion of Ki67-positive NPCs, indicating impaired progenitor proliferation (**Fig. 3.13, J and K**). Global proteomics revealed coordinated changes following DCAF7 or KIAA0232 KD in NPCs (**Fig. 3.14A**), enriched for hcASD+ and cell cycle, chromatin, and transcriptional regulators (**Fig. 3.14, B and C**). In contrast, DYRK1A KD modestly increased Ki67-positive NPCs, (**Fig. 3.13, J and K**) consistent with dose-dependent DYRK1A effects.[61,62] Indeed, while *DYRK1A* mRNA was markedly reduced, global proteomics detected only a marginal decrease in DYRK1A protein amounts (**Fig. 3.14D**), suggesting robust post-transcriptional buffering in NPCs. Despite this buffering, the expression of critical neurogenic transcription factors - PAX6, FOXG1, and SOX2 - that are in part regulated by DYRK1A[63-65] was downregulated upon KD of all three

complex members (**Fig. 3.5, G and H**). Collectively, these data indicate that disruption of the DCAF7-DYRK1A-KIAA0232 complex impairs neurogenesis across species.

hcASD missense mutations alter PPIs and result in convergent interaction changes

We next assessed how 54 patient-derived damaging *de novo* missense mutations predicted to be highly deleterious[4] across 30 hcASD proteins reconfigure PPIs (**Fig. 3.6A**). This mutant network (ASD_{mut}-PPI) identified 253 significantly altered interactions ($p < 0.05$), both gained/strengthened and lost/weakened (**Fig. 3.6, B to D**; Methods). Lost interactions (136; 95 unique interactors) were enriched for hcASD (**Fig. 3.15A**), supporting a primary loss-of-function mechanism.[66] Mapping these interactions in human NPCs confirmed strong overlap with our HEK293T data, with shared interactors significantly enriched for both mutation-sensitive and AF-predicted direct links (**Fig. 3.6D, and Fig. 3.15B**), supporting network robustness across cellular contexts.

Given the convergence in ASD-PPI, we examined whether missense mutations in the same or across different hcASD genes cause shared protein interaction changes. In the 13 hcASD with multiple variants, 45% of altered interactions were replicated across two or more mutants (for example, SLC6A1 and FOXP1 mutants; **Fig. 3.6D**). We additionally identified 15 interactors that showed consistent changes across two different mutant hcASD; for example, three FOXP1 mutants and one FOXP2 mutant lose interaction with FOXP4 (**Fig. 3.6D, Fig. 3.15, C and D**).

Integrating ASD_{mut}-PPI with other PPI networks[23,54,67-70] (see Methods) (**Fig. 3.15E**) revealed that differential interactions consistently converge on the functional modules identified in the WT network (**Fig. 3.3, C to F**). For example, both MKX mutants (R93G, L89F) enhance interactions with the Sin3 transcriptional repression complex, whereas both MYT1L mutants (H522Q, C504R) reduce interactions with the Mediator complex, suggesting dysregulation of gene expression

regulation. Similarly, both STXBP1 mutants (R551C, A215T) and both AP2S1 mutants (R10W, G64D) show stronger interactions with vesicle transport and clathrin complexes, respectively. Collectively, these findings indicate that patient-derived mutation-driven interaction changes converge on key biological processes governing neurodevelopment.

We next leveraged AF to prioritize direct interactions where an hcASD mutation is located at the predicted interaction interface (< 10 Å, see Methods) (**Fig. 3.7A**). This identified 34 mutations across 22 hcASD at the interface of 216 interactions. 46% of interactions weakened in the mutant state are associated with a mutation at the interface, versus only 18% of strengthened interactions, suggesting that interface mutations are primarily disruptive (**Fig. 3.16A**). For example, the PPP2R5D E198K mutation directly contacts the interaction interface (1.6 Å) leading to the loss of interaction with PPP4C (**Fig. 3.7, B and C, Fig. 3.16B**). In contrast, a mutation in GNAI1 I319T more distant to its interface with RIC8A (*71*) (6.1 Å) strengthens the PPI (**Fig. 3.7, D and E, Fig. 3.16C**). Targeted mutagenesis of this residue revealed that interaction affinity correlates with predicted pathogenicity: whereas the benign variant I319V showed no change, three damaging variants altered RIC8A binding (**Fig. 3.16E**). Finally, three FOXP1 missense variants evaluated in ASD_{mut}-PPI disrupt its interaction with FOXP4 (**Fig. 3.7F**) and are predicted to be highly pathogenic (**Fig. 3.16F**). Despite residing at different structural locations - FOXP1 R513H at the DNA interface and FOXP1 L327P at the FOXP1–FOXP4 interface (**Fig. 3.7G, and Fig. 3.16D**) - all variants consistently weaken the FOXP1-FOXP4 association (**Fig. 3.7, H to I**), leading us to prioritize FOXP1 for additional functional investigation.

Collectively, our analysis of the WT and mutant networks reveals a dual-layered molecular convergence. First, the WT ASD-PPI identifies stable 'molecular crossroads'—hub interactors like DCAF7 that link disparate hcASD into unified complexes. Second, the ASD_{mut}-PPI reveals a

'functional convergence' of patient-derived variants; disparate mutations in the same genes, for example, MKX, STXBP1, and AP2S1, or different genes, for example, FOXP1 and FOXP2, drive similar directional changes in interaction strength within their respective functional modules. This suggests that although ASD risk is genetically heterogeneous, the molecular consequences of these mutations collapse onto a limited set of vulnerable protein systems.

hcASD mutations alter cortical neuron differentiation

FOXP1 encodes a forkhead box domain-containing transcription factor expressed in early brain development.[72-75] To identify neurodevelopmental processes disrupted by ASD patient variants, we generated isogenic iPSC lines with a FOXP1^{R513H} allele, selected for its predicted high pathogenicity (**Fig. 3.8, A, B; Fig. 3.16F**). We confirmed the loss of FOXP1-FOXP4 interaction in FOXP1$^{R513H/WT}$ iPSC-derived NPCs (**Fig. 3.17A**) and assessed cell type composition in organoids using immunostaining at time points that corresponds to early (D39) and late cortical neurogenesis (D101).[76] During early neurogenesis, we found a reduction in PAX6+ progenitor cells, and an increase in the abundance of both BCL11B+ layer V neurons and TBR1+ layer VI/subplate neurons in FOXP1$^{R513H/WT}$ compared to FOXP1$^{WT/WT}$ (**Fig. 3.8B**). At day 101, we found no difference in the abundance of PAX6+ cells, but a reduction in BCL11B+ cells in FOXP1$^{R513H/WT}$ organoids (**Fig. 3.8B**). The abundance of TBR1+ cells was lower in FOXP1$^{R513H/WT}$ compared to FOXP1$^{WT/WT}$, but this difference did not reach statistical significance after correcting for multiple comparisons. Concurrently, D101 FOXP1$^{R513H/WT}$ organoids exhibited an increase in SATB2+ neurons (**Fig. 3.18B**). These results suggest premature differentiation of cortical neurons in FOXP1^{R513H} forebrain organoids, consistent with prior findings across multiple *in vivo*, *in vitro* and *in silico* models.[22,59,77,78]

To assess if the mutation alters DNA binding, we performed CUT&Tag for FOXP1 and FOXP4 (**Fig. 3.17B**).[79] In FOXP1$^{WT/WT}$ organoids, 94.7% of FOXP4 peaks overlapped with FOXP1 peaks (**Fig. 3.8C**), consistent with their function as a heterodimer. In contrast, only 67.4% of FOXP4 peaks overlapped with FOXP1 bound peaks in FOXP1$^{R513H/WT}$ organoids (**Fig. 3.8C**). Although 85% of FOXP1$^{R513H/WT}$ FOXP1 peaks overlap with those observed in FOXP1$^{WT/WT}$, only 37% of FOXP1$^{R513H/WT}$ FOXP4 peaks overlapped with those in FOXP1$^{WT/WT}$, with 4,700 FOXP4 peaks in FOXP1$^{R513H/WT}$ organoids not found in controls (**Fig. 3.8D**). These findings suggest that FOXP1 missense variants impact FOXP4 function, where the loss of FOXP1-FOXP4 physical interaction leads to gain of FOXP4 binding at new genomic sites.

Using scRNAseq, we identified differentially expressed genes between the mutant and WT organoids in excitatory neurons (ENs). We identified three EN clusters along the maturation trajectory: cortical deep layer 6 and subplate neurons (EN-3: high expression of SOX5, NR4A2, and TBR1), deep layer 5/6 neurons (EN-2: high expression of FOXP2, PBX3, and MEIS2), and deep layer 5 neurons (EN-1: high expression of FOXP1, ROBO1, and BCL11B). The greatest number of differentially expressed genes were found in the EN-3 cluster (**Fig. 3.8, E and F, and Fig. 3.17, C and D**), corresponding to the population displaying premature differentiation in FOXP1$^{R513H/WT}$ organoids (**Fig. 3.8B**). DEGs in EN-3 were enriched for FOXP1 and FOXP4 targets (**Fig. 3.17E**), and upregulated genes with reduced FOXP1 binding in CUT&Tag were enriched for hcASD+ (**Fig. 3.17F**). 43% and 32% of EN-3 DEGs showed new FOXP1 and FOXP4 binding, respectively. Furthermore, 21% exhibited a 'switch' from lost FOXP1 binding to new FOXP4 binding (**Fig. 3.17, F and G**). These findings suggested a gain-of-function or loss of specificity of FOXP4 in FOXP1$^{R513H/WT}$ organoids, potentially compensating for or compounding the loss of FOXP1 binding at these loci.

To determine if the disrupted FOXP1-FOXP4 PPI could contribute to the cortical neurogenesis phenotype (**Fig. 3.17A**), we generated double-mutant iPSC lines with loss of function mutation in FOXP4 in the FOXP1$^{R513H/WT}$ background (FOXP1$^{R513H/WT}$-FOXP4^{KO}; **Fig. 3.17H**). Forebrain organoids showed deep layer neuron proportions similar to controls (**Fig. 3.8G**). These data suggest that a gain of FOXP4 function may account for the cortical neurogenesis defects in the FOXP1^{R513H} organoids. Thus, the R513H mutation represents not merely a simple loss of FOXP1 function, but a pathogenic gain-of-function for its disrupted interactor, FOXP4. Although the exact rescue mechanism needs to be determined, these data demonstrate that FOXP4 deletion suppresses the FOXP1 mutant phenotype.

To connect the cellular defects in FOXP1$^{R513H/WT}$ to functional properties, we recorded neural activity in organoids using multielectrode arrays (MEA). FOXP1$^{R513H/WT}$ organoids showed consistently higher mean firing rates compared to the isogenic control organoids (**Fig. 3.8, H and I, and Fig. 3.17I**). Additional analyses of population-level parameters did not reveal changes in the spike time tiling coefficient (STTC), inter-spike intervals (ISI), and various burst metrics (**Fig. 3.17J and K**), suggesting alterations in baseline firing rather than network-wide properties. These results suggest that altered differentiation of cortical neurons observed in FOXP1$^{R513H/WT}$ organoids show functional consequences similar to those observed in independent studies of human stem cell derived neuronal cultures harboring mutations in hcASD genes.[80,81]

Finally, we examined if iPSCs carrying the FOXP1^{L327P} mutation (**Fig. 3.18A**), which resides at the predicted FOXP1-FOXP4 heterodimer interface and also leads to a loss of FOXP1-FOXP4 interaction, would phenocopy neurodevelopmental alterations identified in FOXP1$^{R513H/WT}$ organoids. Similar to FOXP1$^{R513H/WT}$, we found an increase in BCL11B during early neurogenesis and a decrease at the later timepoint (**Fig. 3.8J**) with a concurrent increase in SATB2+ neurons

(**Fig. 3.18B**) and a reduction of PAX6+ cells in both FOXP1$^{L327P/L327P}$ and FOXP1$^{L327P/WT}$ organoids (**Fig. 3.8J**). TBR1+ cells were likewise increased in FOXP1$^{L327P/L327P}$ early and decreased in FOXP1$^{L327P/WT}$ organoids at a later timepoint (**Fig. 3.8J**). Differentially expressed genes identified using scRNAseq in organoids during late neurogenesis showed concordant effect size and directionality between heterozygous FOXP1$^{L327P/WT}$ and FOXP1$^{R513H/WT}$ genotypes in the EN-3 cluster (**Fig. 3.8K, and fig. 3.18C**). The abundance of GABAergic neurons was not altered in organoids with either FOXP1^{L327P} or FOXP1^{R513H} mutation (**Fig. 3.18D**), suggesting that the phenotypes were limited to the excitatory lineage. Thus, mutations disrupting different structural domains of the FOXP1 protein phenocopy each other by destabilizing the FOXP1–FOXP4 heterodimer, leading to convergent phenotypes of premature progenitor exhaustion and aberrant deep-layer neurogenesis. Finally, to test if cortical neurogenesis phenotypes are detected in the case of other ASD mutations, we introduced the E198K variant of high predicted pathogenicity (**Fig. 3.15E**) into the endogenous locus of *PPP2R5D.*[82] Consistent with findings for FOXP1 mutations, during early neurogenesis PPP2R5D$^{E198K/E198K}$ organoids demonstrated a decrease in PAX6+ cells and increase in BCL11B+ cells (**Fig. 3.18, E to F**).

Discussion

Despite substantial progress in defining the genetic architecture of ASD, identifying causal molecular mechanisms and therapeutic targets has remained a central challenge.[10,83] Efforts to identify convergent biology across hcASD have largely relied on transcriptomic analyses and low-throughput functional studies, leaving the proteomic landscape comparatively underexplored. In particular, whether and how genetically heterogeneous ASD risk genes organize at the level of protein complexes—and how disease-causing mutations perturb these interactions—remains poorly defined.

Here, we integrate large-scale human genetics, AP-MS, and AlphaFold-based structural prediction to map the physical interactome of proteins implicated in ASD. Mapping PPIs for 100 hcASD genes and 54 patient-derived missense variants in HEK293T cells enabled the depth, reproducibility, and scale required for comprehensive interactome mapping and systematic mutant interrogation, consistent with prior applications of this approach across diverse disease contexts.[12-19] Multiple orthogonal lines of evidence support the disease relevance of these data, including strong overlap with neuronal interaction networks, high co-expression with hcASD genes in the developing human brain, selective enrichment for ASD—but not schizophrenia—risk genes, and functional validation moving from HEK293T cells to *Xenopus* and human brain organoids.

Our data reveal a dual-layered molecular architecture underlying ASD, where ASD risk proteins converge onto stable, shared protein complexes, and independent ASD-associated mutations induce convergent patterns of interaction rewiring. Together, these layers provide a mechanistic bridge between genetic variation and neurodevelopmental dysfunction and nominate specific protein complexes, interactions and interfaces as candidate therapeutic entry points. The ASD-PPI network highlights molecular modules not readily detectable by gene-centric genetic studies alone[4,6]; identification of the PAF1 complex as a central hub illustrates how coherent protein assemblies can harbor multiple ASD-relevant components whose individual genetic signals fall below stringent significance thresholds.

Despite extensive genetic and phenotypic heterogeneity of ASD, the substantial interactor overlap among hcASD proteins suggests that a relatively limited set of molecular pathways may underlie core ASD biology, particularly for the severe end of the ASD spectrum that is markedly enriched for large-effect coding mutations.[4,8,84,85] Although deleterious variants in the same genes have a very high degree of phenotypic variability, contributing to multiple neurodevelopmental

conditions, including ASD, epilepsy, intellectual disability, and schizophrenia, the ASD network described here exhibits distinctiveness for ASD, as evidenced by the absence of enrichment for schizophrenia risk genes. These findings raise the possibility that disease specificity may emerge not simply from gene identity but from how mutations perturb protein networks within developmentally relevant contexts.

Systematic comparison of wild-type and mutant PPIs revealed recurrent gain and loss of interactions across multiple hcASD variants, highlighting shared molecular vulnerabilities across genetically distinct mutations. Integration of AlphaFold predictions enabled identification of candidate interaction interfaces affected by disease-causing mutations. Although AlphaFold was not designed for interaction discovery, its performance in this context—particularly when leveraging consistency across predictions rather than maximum scores—outperformed conventional approaches and enriched for physiologically relevant interactions. Interactions observed across multiple cellular contexts were more likely to exhibit high AlphaFold confidence and mutation sensitivity, suggesting that broadly conserved interactions may represent core functional nodes.

Functional validation in human neural progenitor cells and forebrain organoids demonstrated that mutation-induced PPI rewiring has direct consequences on neurodevelopment. FOXP1 variants disrupted FOXP1-FOXP4 interactions, altered DNA binding, and promoted ectopic recruitment of FOXP4 to novel genomic loci. Despite high expression of FOXP1 in ganglionic eminences,[86] we did not detect similar differences in GABAergic neurogenesis, potentially due to compensation by other FOXP family members.[87] Our data suggest that FOXP4, which is highly enriched in cortical but not subcortical progenitors,[27,77] is responsible for altered glutamatergic neurogenesis. Altered cortical neurogenesis has been shown in recent reports using stem cell derived models of ASD,[59,77]

and we also detected a similar trend in organoids with a missense mutation in PPP2R5D. Together these findings underscore the importance of altered timing of cortical neurogenesis to ASD (**Fig. 3.18G**), consistent with high co-expression of hcASD risk genes in these cell types,[4,22] as well as findings from postmortem brain tissue.[88]

Collectively, this study establishes a generalizable framework for mechanistic interpretation of genomic risk through PPI mapping. By integrating wild-type and mutant interactomes with structural modeling and human organoid systems, we enable causal inference at single-mutation resolution while maintaining the systemic context of the broader molecular landscape. Identification of both lost or weakened and aberrantly gained or strengthened interactions delineates complementary therapeutic strategies, including stabilization of disrupted complexes and inhibition of pathological interactions. More broadly, this platform prioritizes druggable protein interfaces and pathways, offering a rational foundation for precision therapies aimed at restoring neurodevelopmental trajectories and circuit-level function in ASD, and establishing a generalizable blueprint for mechanistic drug discovery across neurological and psychiatric disease.

Figures

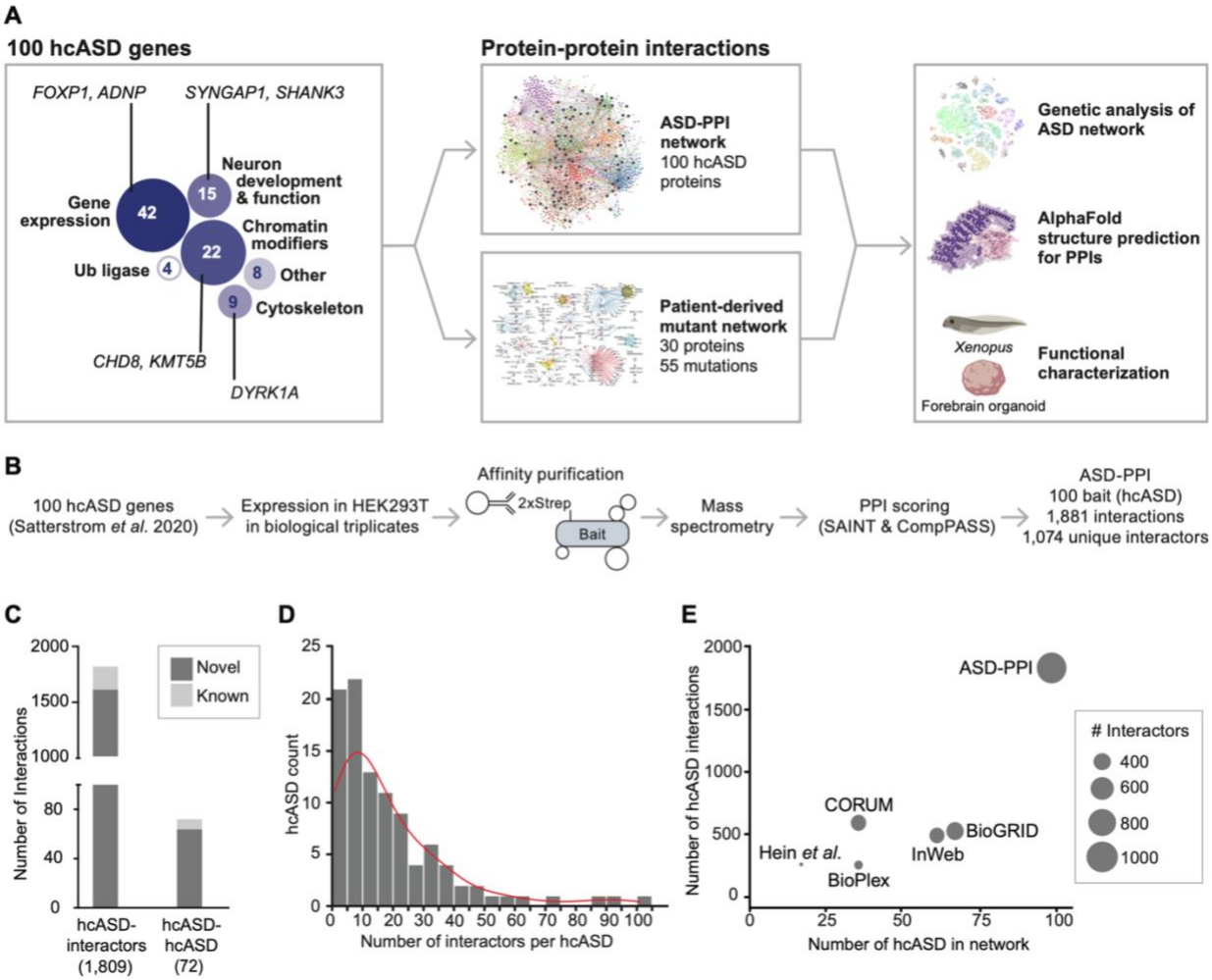

Figure 3.1 hcASD protein interaction mapping reveals novel interactors.
(A) Study overview showing prior functional annotations of 100 hcASD genes, highlighting the most significant two genes per category (FDR < 0.05, (4)), generation of wild-type (ASD-PPI) and patient-derived mutant (ASDmut-PPI) interaction networks, and downstream studies. **(B)** Workflow to generate PPI data for 100 hcASD in HEK293T cells. **(C)** The ASD-PPI network contains a total of 1,881 interactions including 1,809 interactions with non-hcASD proteins (interactors) and 72 interactions with hcASD proteins. **(D)** The distribution of the number of interactors per hcASD (median = 11). **(E)** Comparison of the number of hcASD proteins profiled and the number of hcASD-associated interactions in ASD-PPI and existing large-scale PPI datasets (63, 77, 78, 100, 101). Point size reflects the number of unique interactors for each dataset.

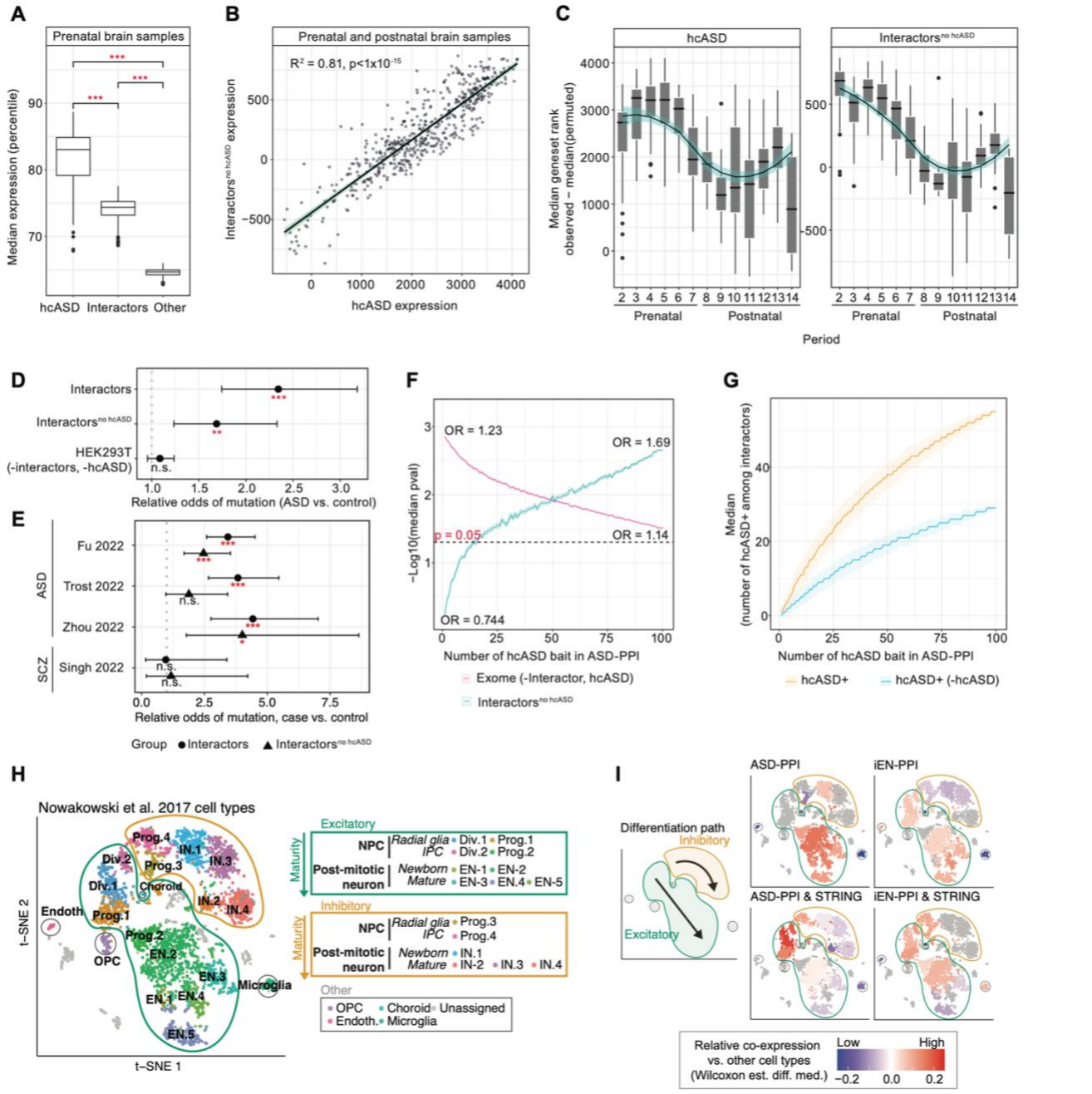

Figure 3.2 ASD-PPI interactors are expressed in brain and enriched for ASD genetic risk. **(A)** hcASD and interactors$^{no\ hcASD}$ show higher prenatal human brain expression than other HEK293T proteins (28). **(B to C)** Expression of these genes is highly correlated across 524 human brain samples (B; Pearson R2 = 0.81, p < 1 x 10-15 15) and developmental periods (C; Spearman's ρ = 16 0.946). **(D to E)** Interactors are enriched for ASD de novo damaging mutations (D) and ASD risk genes (5–7, 31) but not schizophrenia risk genes (5–7, 31) (E). **(F to G)** Increasing bait number improves ASD risk enrichment (F) and hcASD+ recovery. **(H)** t-SNE of prenatal human brain scRNAseq data (32) showing 19 cell clusters. **(I)** Relative co-expression of ASD-PPI and iEN-PPI networks across cell types. Nominally significant cell types (p < 0.05) are colored; others are gray. (Figure legend continued on next page.)

(Figure legend continued from previous page.) STRING edges shift enrichment from post-mitotic neurons toward NPC–EN. Color: network edge co-expression relative to other cell types. Abbreviations: div, dividing; EN, excitatory neuron; IN, inhibitory neuron; NPC, neuronal progenitor cell; OPC, oligodendrocyte precursor cell; prog, progenitor. Statistics: t-test (A), one-sided Fisher's exact test (D, F, G), Wilcoxon rank-sum (I). Nominally significant: $p < 0.05$; significant: Bonferroni-adjusted $p < 0.05$. Boxplots:median, IQR, ±1.5×IQR. Dot-whisker:OR ±95% CI. n.s., not significant; *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$p < 0.001$.

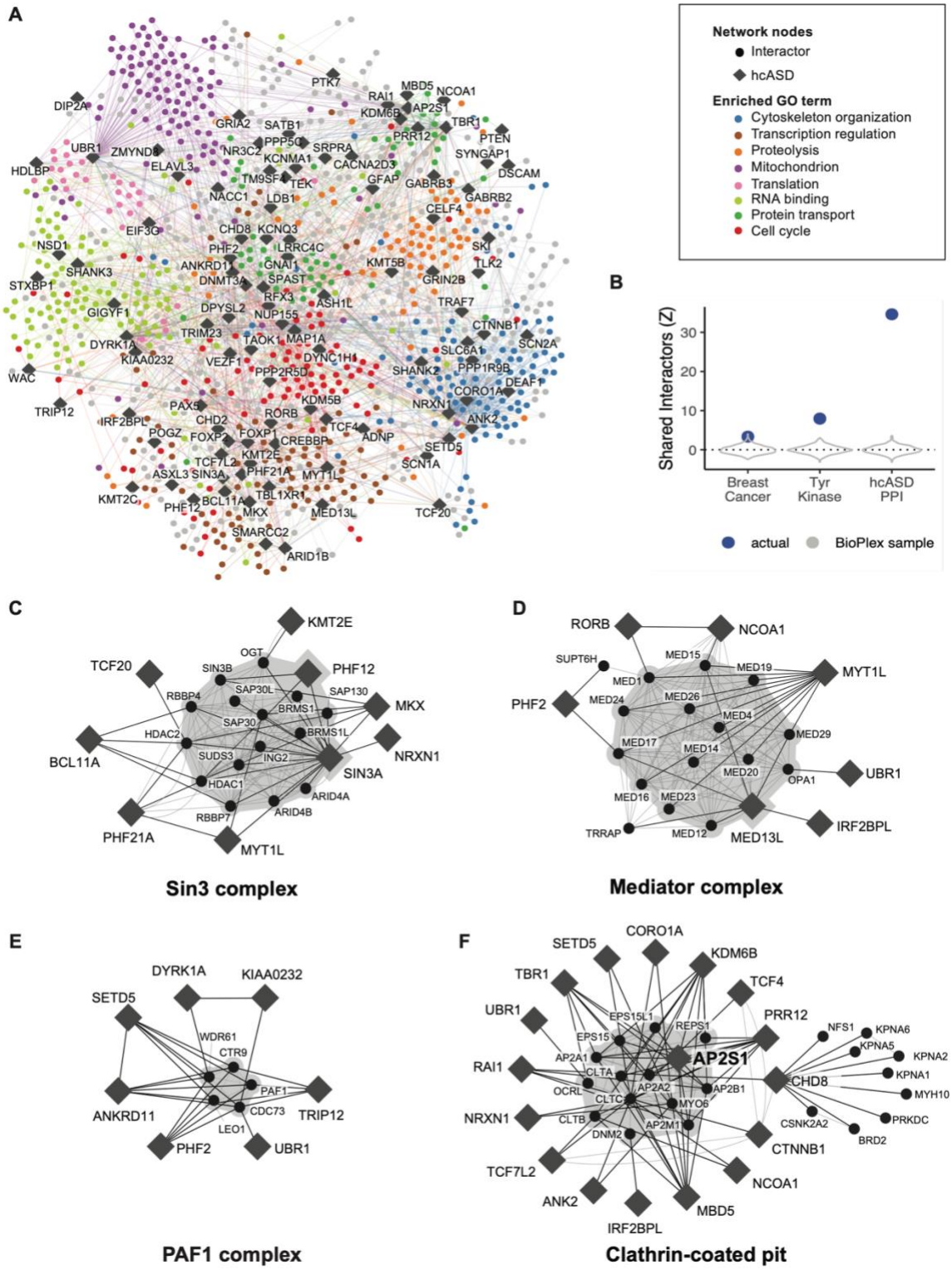

Figure 3.3 ASD-PPI demonstrates molecular convergence among hcASD.
(A) ASD-PPI network map, with hcASD and interactors positioned based on connectivity (AP-MS) and shared GO terms. Interactor color reflects (Figure legend continued on next page.)

(Figure legend continued from previous page.) representative enriched GO categories that were selected to maximize coverage while minimizing redundancy. **(B)** Interactor overlap among bait pairs in ASD-PPI, breast cancer PPI (12), and kinome PPI (42), shown as the proportion of significant pairwise overlaps (hypergeometric $p < 0.05$). Proportions were converted to Z scores separately for each network by comparison to 1,000 random degree-matched BioPlex networks (violins). ASD-PPI shows the greatest interactor overlap. **(C to F)** PPI networks showing hcASD interactions with the Sin3 (C), Mediator (D), PAF1 (E), and AP2-mediated clathrin-coated pit (F) complexes. Dark lines indicate AP-MS edges; thin gray lines indicate CORUM or STRING edges; gray shading denotes CORUM complexes.

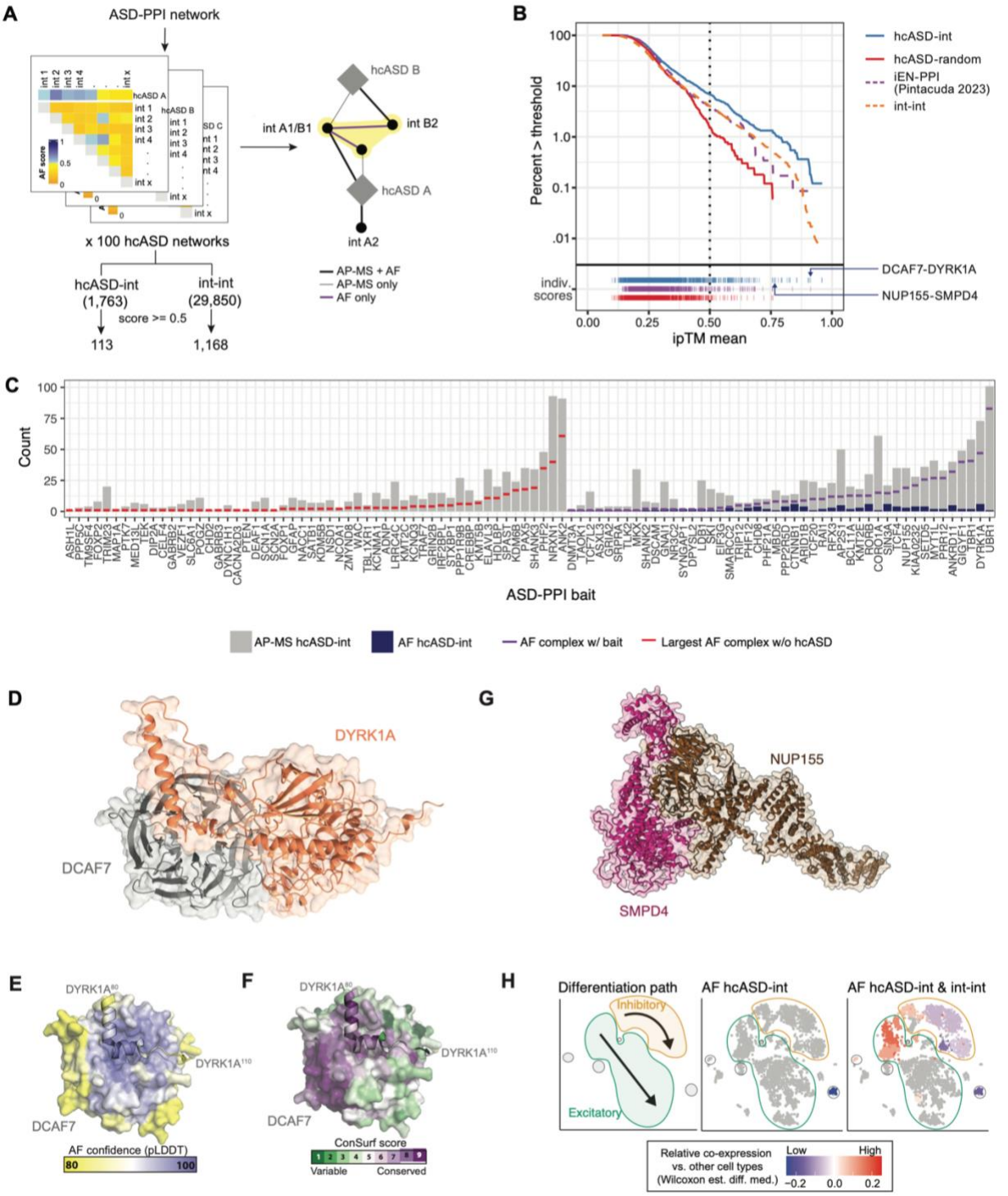

Figure 3.4 AlphaFold predicts direct PPI interaction interfaces.
(A) Workflow for AlphaFold (AF) interface predictions of hcASD-int (n=1,763) and int-int (n=29,850) pairs. Example subnetwork shows AF-predicted edges linking two hcASD via int-int interactions (purple). **(B)** Distribution of AF ipTM scores across PPI sets. hcASD-int pairs exhibit ~5-fold enrichment over random controls at ipTM >0.5 (dotted line). Higher curves indicate more predicted direct interfaces. (Figure legend continued on next page.)

(Figure legend continued from previous page.) Comparison includes iEN-PPI (Pintacuda 2023) and hcASD-random (negative control). **(C)** Interactor counts per hcASD (grey), highlighting those predicted as direct by AF (blue). Purple lines denote interactors connected to hcASDs via AF-validated direct or mediated (int-int) paths. For hcASDs lacking direct AF edges, red lines indicate the largest AF-validated interactor complex. **(D to F)** AF-predicted DYRK1A-DCAF7 structure (D) colored by model confidence (pLDDT, E) or sequence conservation (ConSurf, F), revealing a high-confidence, conserved interface. **(G)** AF-predicted NUP155-SMPD4 structure. **(H)** Relative co-expression of ASD-PPIs with or without AF int-int edges across brain cell types (Nowakowski et al. 2017, see Fig 2H). Color reflects relative network edge co-expression in each cell type compared to all others (Wilcoxon rank-sum, estimated median difference). Nominally significant cell types ($p < 0.05$) are colored; others are gray. Abbreviations: AF, AlphaFold-Multimer; int, interactor.

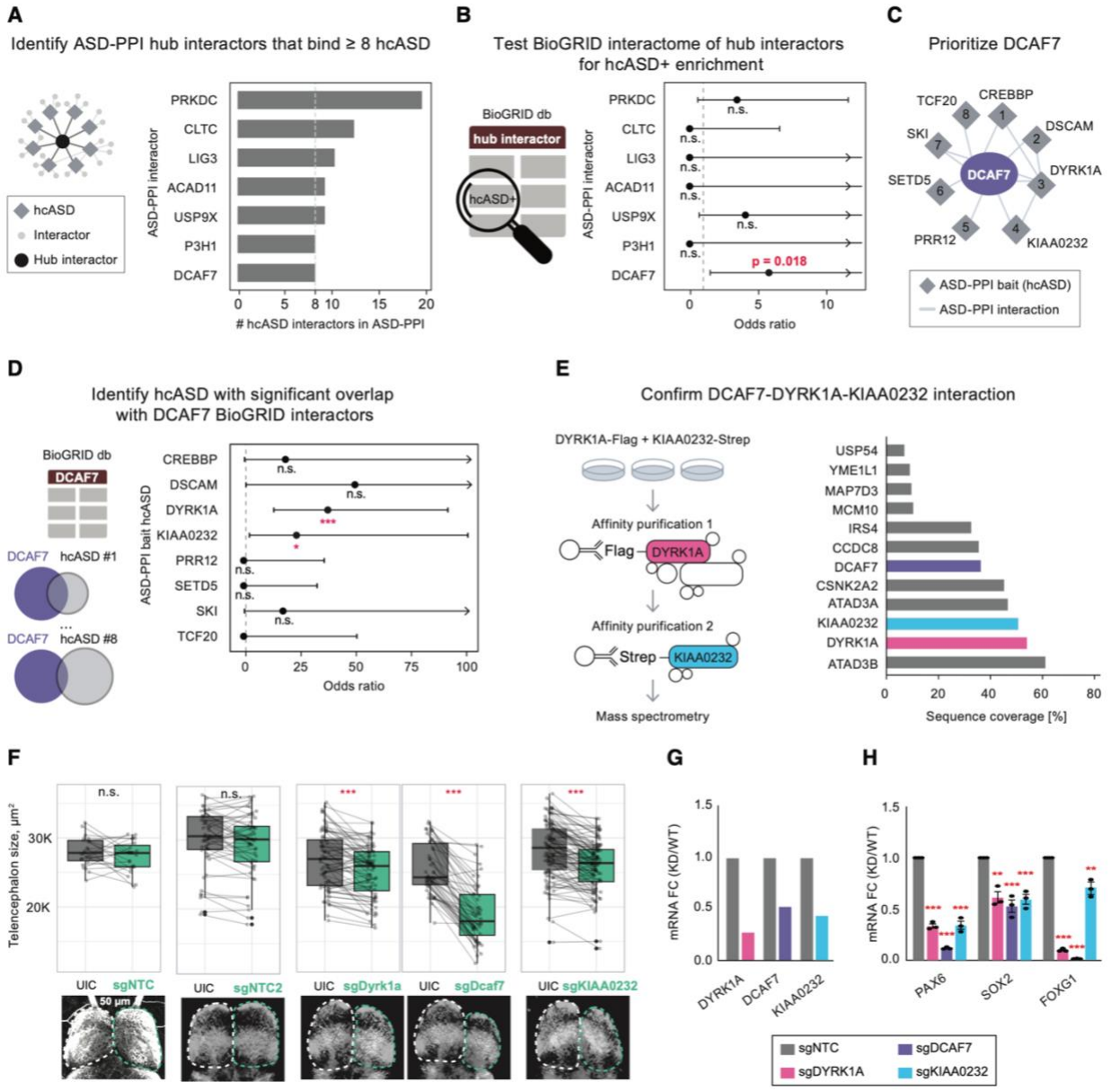

Figure 3.5 DCAF7 forms a complex with multiple hcASD.
(A) ASD-PPI "hub" interactors binding ⩾8 hcASD. **(B)** Enrichment of hcASD+ in known interactomes of hub interactors from BioGRID (63). **(C)** Hub interactor DCAF7 binds eight hcASD in ASD-PPI. **(D)** Overlap between DCAF7 BioGRID interactors and ASD-PPI interactors of hcASD bound by DCAF7. DYRK1A and KIAA0232 showed significant overlap and were prioritized. **(E)** Mass spectrometry from sequential IP of DYRK1A and KIAA0232 identifies DCAF7 as a shared interactor. **(F)** Xenopus telencephalon size following unilateral sgRNA injection targeting Dcaf7, Dyrk1a, or Kiaa0232 versus non-targeting control (NTC) shows reduced telencephalon size. Scale bar: 50 µM. **(G)** CRISPR knockdown in NPCs reduces DYRK1A (pink), DCAF7 (purple) and KIAA0232 (blue) mRNA in respective cell lines. **(H)** Quantitative RT-PCR in knockdown NPCs shows decreased PAX6, SOX2 and FOXG1 expression versus NTC. Statistics: (B and D) Fisher's exact test (one sided, greater), (F) paired t-test, (H) one-way ANOVA. P-values are unadjusted (B) (Figure legend continued on next page.)

(Figure legend continued from previous page.) or corrected using Bonferonni (D) or Dunnett (H). Dot-whisker plots (B to D): dot = OR, whiskers = 95% CI. Bar plots: bar = mean, whiskers = SEM (n=3). n.s., not significant; *0.01 < p < 0.05; **0.001 < p < 0.01; ***p < 0.001.

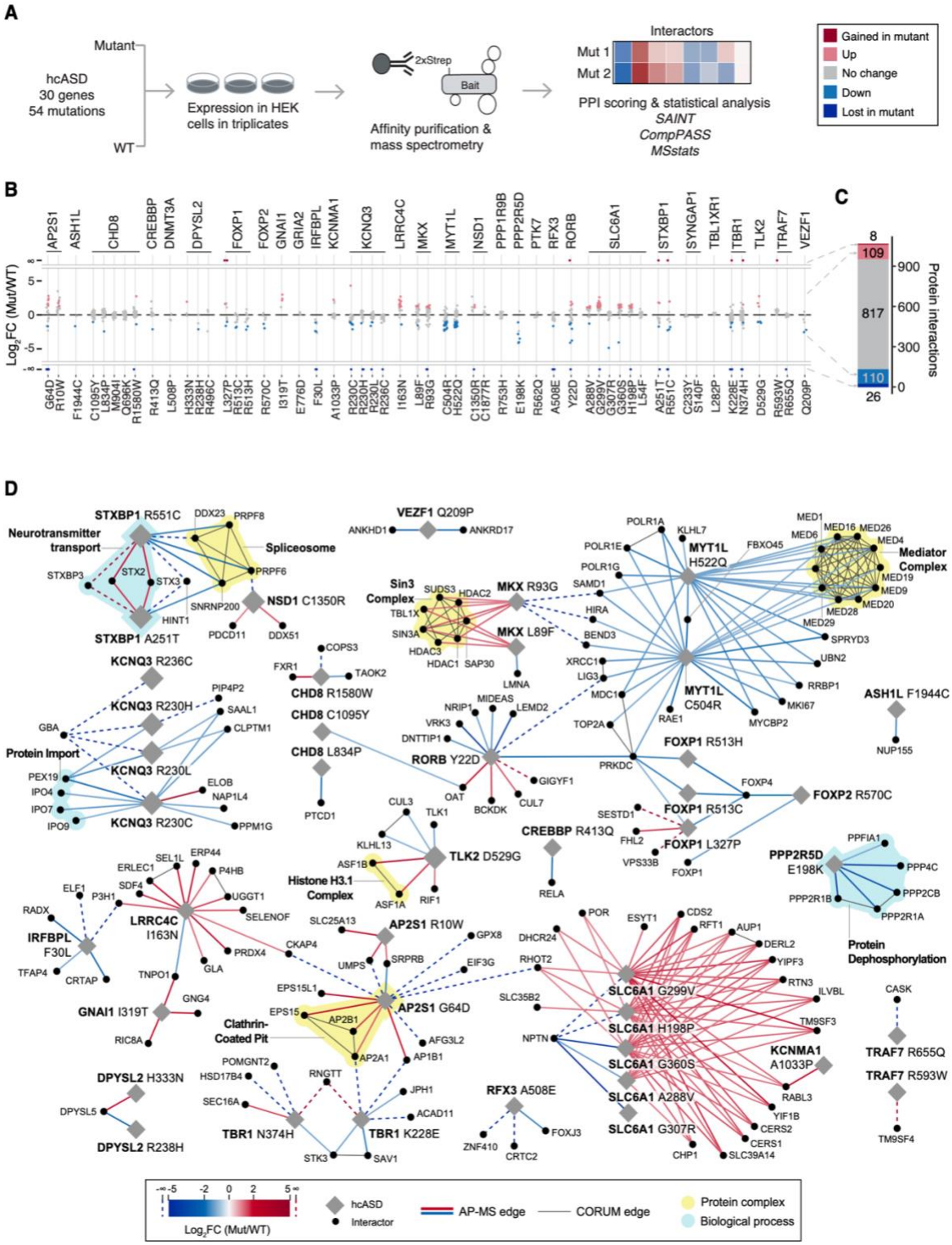

Figure 3.6 Patient-derived hcASD missense mutations alter protein interactions.
(A) Overview of the generation of the ASD patient-derived mutant network (ASDmut-PPI). **(B)** Interaction changes between mutant and wildtype (Figure legend continued on next page.)

(Figure legend continued from previous page.) hcASD, grouped by hcASD with mutations labeled on the x-axis. Significantly weakened (blue) or strengthened (red) hcASD interactions are highlighted, with mutant- or wildtype-specific interactions plotted at log2FC = ∞ and $-\infty$, respectively. **(C)** Quantification of differential protein interactions shows 117 hcASD mutant-enriched interactions (red; 109 strengthened, 8 unique to mutant) and 136 wildtype hcASD-enriched interactions (blue; 110 weakened, 26 unique to wildtype). **(D)** Network view of significant differential interactions in ASDmut-PPI. Edge color reflects interaction specificity (blue, stronger in wildtype; red, stronger in mutant), and dashed edges indicate interactions unique to mutant or wildtype. CORUM protein complexes (yellow) and Gene Ontology biological processes (blue) highlight functional modules disrupted within and across hcASD. Statistical tests (B-D): pooled t-test on intensity values with FDR estimated using Storey's q method; interactors with $|\log2FC(Mut/WT)| > 1$, FDR < 0.1, and $p < 0.05$ were considered significantly differential.

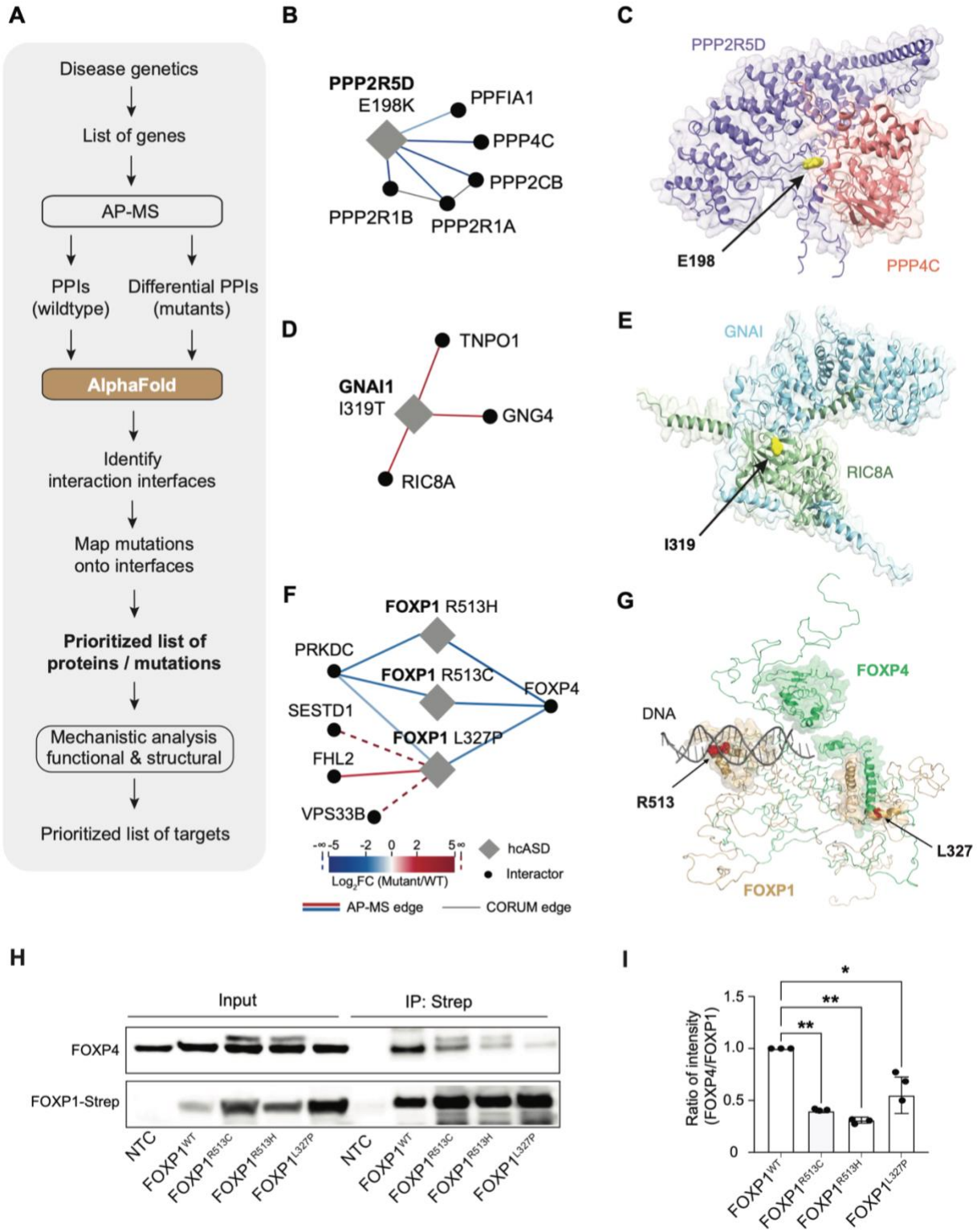

Figure 3.7 AlphaFold maps hcASD mutations to protein structures and interaction interfaces.

(A) Workflow to prioritize protein interactions affected by hcASD mutations using AP-MS and AF structure predictions. **(B, D, and F)** Differential ASDmut-PPI networks for PPP2R5D^{E198K} (B), GNAI^{I319T} (D), and FOXP1^{R513H}, FOXP1^{R513C}, (Figure legend continued on next page.)

(Figure legend continued from previous page.) and FOXP1^{L327P} (F). Edge colors and significant differential interactions are as in Fig. 6D. **(C, E, and G)** AF-predicted interactions for PPP2R5D-PPP4C with E198 at the interaction interface (C); GNAI1-RIC8A with I319 at the interaction interface (E); and model of the FOXP1 (yellow) / FOXP4 (green) complex visualizing the relative flexibility of structured DNA binding and coiled-coil domains (shown in surface) mediated by interspacing intrinsically disordered regions (IDR)(coil). Highlighted in red are FOXP1 residues R513 at the FOXP1 DNA-binding interface and L327 at the FOXP1/FOXP4 dimerization interface (G). **(H)** IP-Western blot of non-transfected control (NTC) or Strep-tagged FOXP1WT, FOXP1R513C, FOXP1R513H or FOXP1L327P in HEK293T cells shows loss of FOXP4 interaction in FOXP1 mutants. **(I)** Quantification of (H). Data shown as mean + SEM (n = 3). Statistics: one-way ANOVA with Dunnett correction; p.adj = 0.0011 (FOXP1^{R513C} 8), p.adj = 0.0018 (FOXP1^{R513H}), p.adj = 0.0141 (FOXP1^{L327P}).

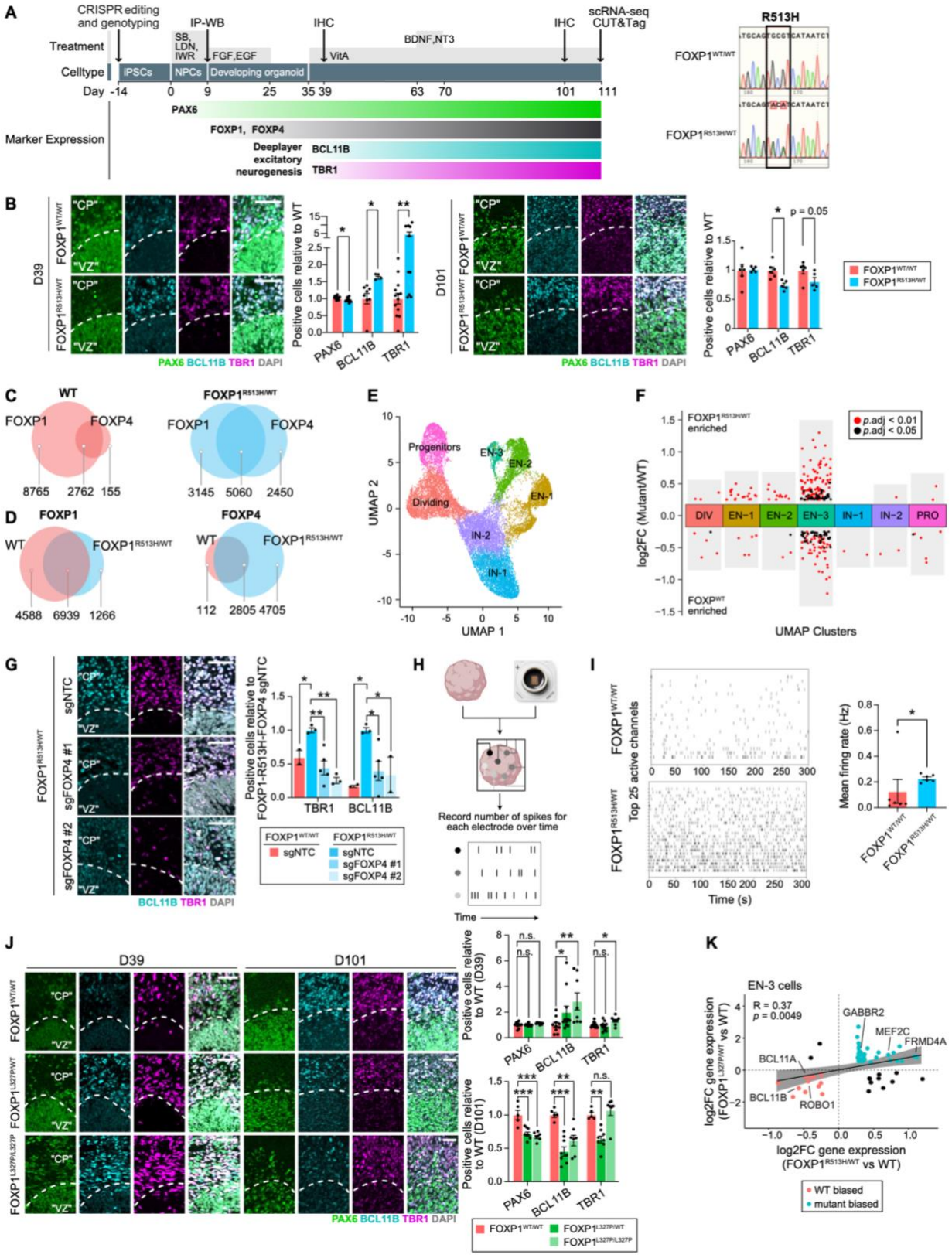

Figure 3.8 hcASD variants alters deep layer cortical neurogenesis in forebrain organoids.
(A) Isogenic iPSC generation and differentiation (Figure legend continued on next page.)

(Figure legend continued from previous page.) workflows with validation (R513H; right). **(B)** Immunohistochemistry and quantification of PAX6+, TBR1+ and BCL11B+ cells at D39 and D101 in FOXP1R513H/WT vs. WT. **(C-D)** CUT&Tag analysis showing DNA binding overlap (C) and genomic distribution (D) of FOXP1/FOXP4. (Figure legend continued on next page). (Figure legend continued from previous page). **(E-F)** scRNA-seq UMAP visualization (E) and cell-type specific DEGs (F) in D111 organoids. **(G)** Rescue of D39 differentiation phenotype in $FOXP1^{R5134H/WT}$ upon FOXP4 knockout. **(H-I)** MEA recordings (H) and representative raster plots (I) showing increased mean firing rate in $FOXP1^{R5134H/WT}$. **(J)** Immunohistochemistry and quantification of $FOXP1^{L327}$ allelic series (WT, heterozygous, homozygous) at D39 and D101. **(K)** Pearson's correlation of EN-3 DEGs between variants. Statistics: (B, G, J) Linear mixed model (D39) or linear regression (D101) followed by permutation testing and Benjamini-Hochberg (BH) correction; n (organoids)=10-13 (B, D39, D101), 2-6 (G, D39, D101) 8-11 (J, D39), 5-8 (J, D101). (F) Wilcoxon rank sum with Bonferroni correction. (I) One-sided Mann-Whitney U test, n=6 organoids. Scale bars: 50μm; CP, cortical plate; "VZ", ventricular zone-like; n.s., not significant; *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$p < 0.001$.

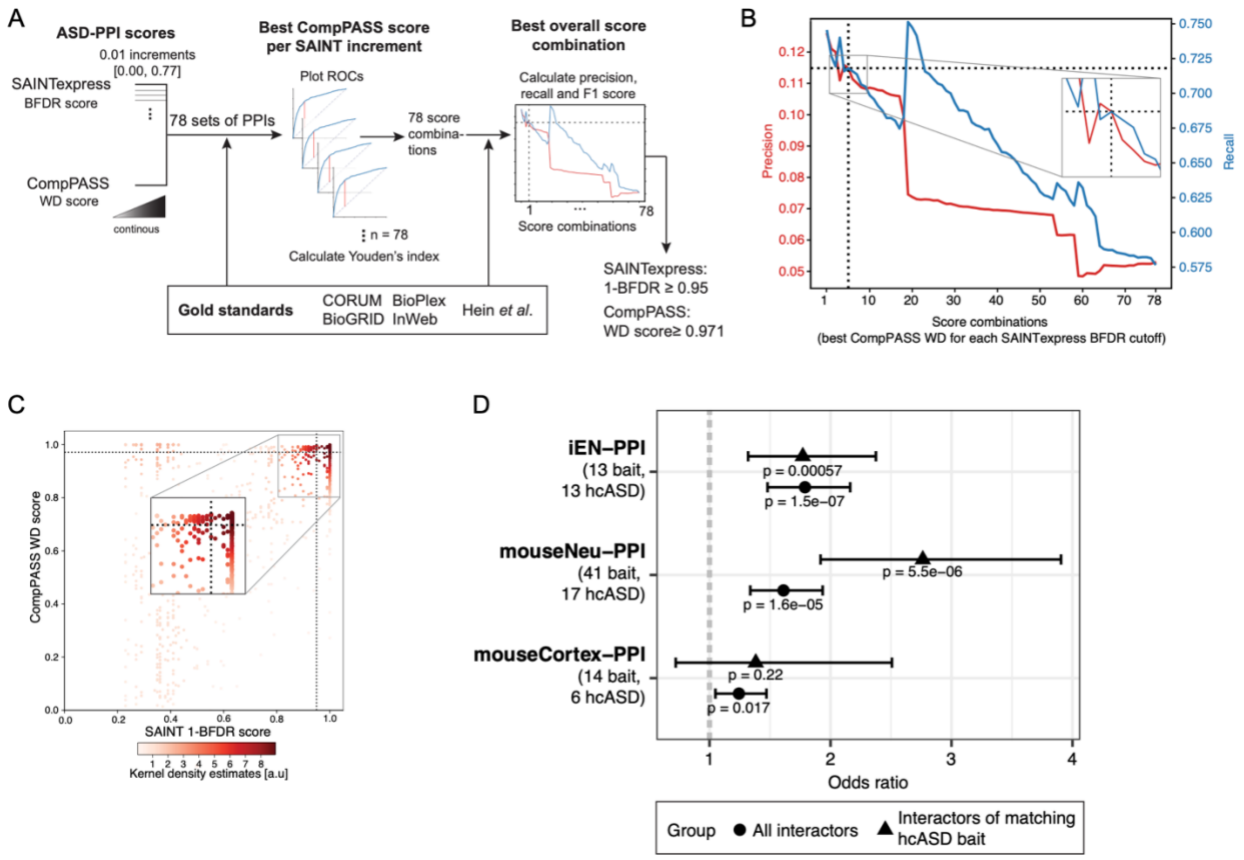

Figure 3.9 AP-MS in HEK293T cells identifies interactors that overlap with those from other PPI methods.

(A) Workflow for selecting SAINTexpress (103) and CompPASS (155) score cutoffs by maximizing sensitivity and specificity metrics for recovering known interactions from public databases (77, 78, 100, 101, 106). **(B)** Precision (red) and recall (blue) curves for known interactions across SAINTexpress and CompPASS cutoff combinations. Dotted lines indicate thresholds used for ASD-PPI. **(C)** Kernel density distribution of scores for known interactions in the unfiltered PPI dataset. Dotted lines mark the ASD-PPI cutoffs, which coincide with a high density of known interactions. **(D)** Overlap between ASD-PPI interactors and those from autism-relevant PPI datasets, including iEN-PPI (34) (13 hcASD, endogenous IP-MS in iPSC-derived excitatory neurons), mouseNeu-PPI (112) (41 ASD-associated bait/17 hcASD, proximity labeling with bait overexpressed in mouse neuron/glia co-cultures), and mouseCortex-PPI (113) (14 ASD-associated bait/6 hcASD, in vivo endogenous proximity labeling in developing mouse brain). Triangles show overlap using all interactors; circles show overlap restricted to shared hcASD bait across datasets. Fisher's exact test (one-sided, greater), unadjusted p values. Dot = OR, whiskers = 95% CI.

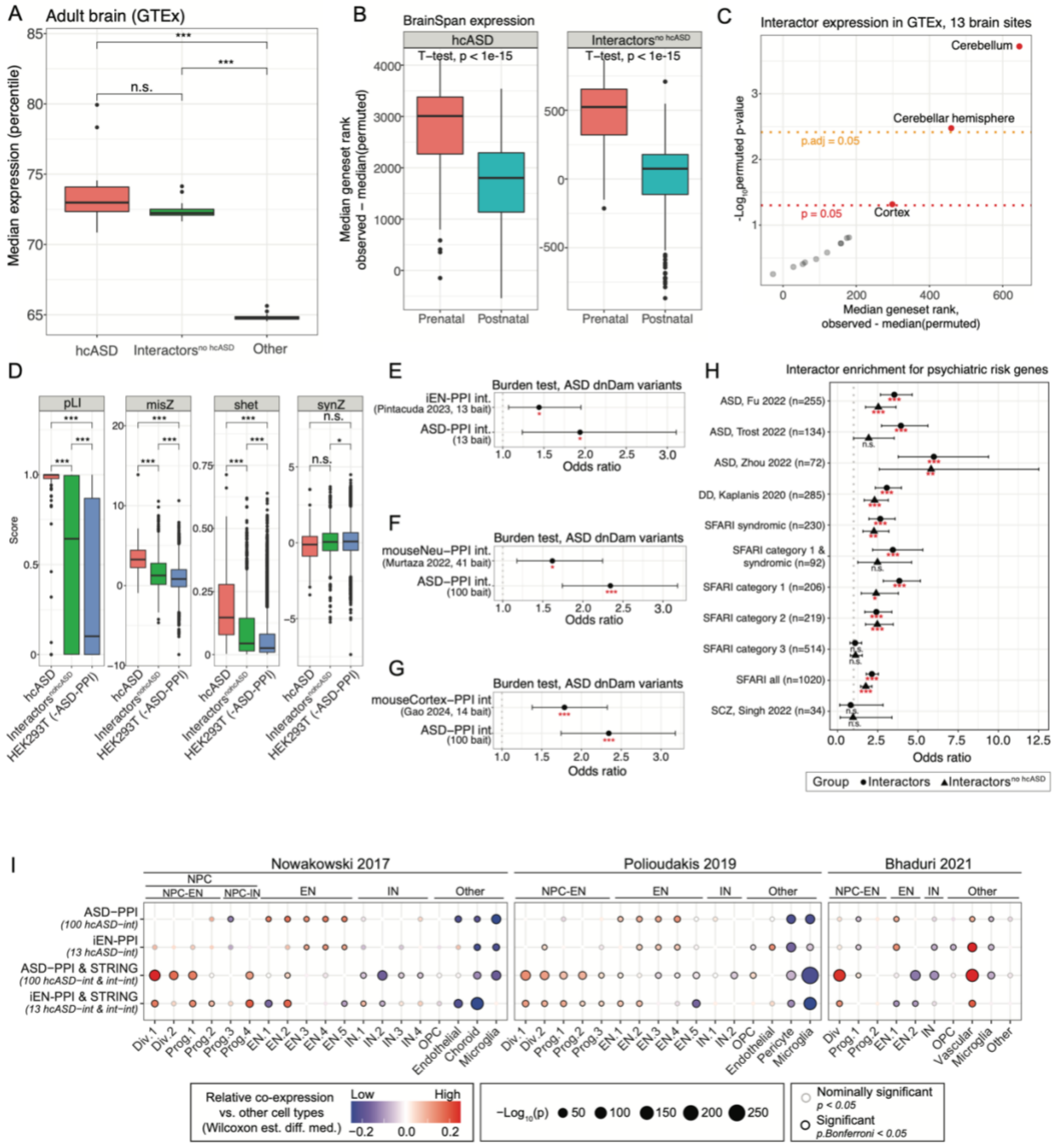

Figure 3.10 Interactors are expressed in contexts associated with ASD and enriched for ASD genetic risk.

(A) hcASD and interactorsno hcASD 2 are more highly expressed than other HEK293T proteins in adult brain tissue samples from 13 brain regions in GTEx (114) **(B)** hcASD (left) and interactors$^{no\ hcASD}$ (right) are more highly expressed in prenatal than postnatal brain samples (BrainSpan) (28). Gene set expression is measured as deviation from median rank across 100,000 permuted gene sets. **(C)** Expression of interactors$^{no\ hcASD}$ in 13 adult brain regions (114) shows significant enrichment in cerebellum and cortex compared with 100,000 permuted gene sets. Dashed lines indicate nominal (red, p = 0.05) and adjusted (orange, Bonferroni for 13 tests) significance thresholds. **(D)** hcASD and (Figure legend continued on next page.)

(Figure legend continued from previous page.) interactors$^{\text{no hcASD}}$ show significantly higher evolutionary constraint (pLI, misZ, shet) than HEK293T proteins ('HEK293T (-ASD-PPI)'). SynZ, which reflects neutral genetic variation, shows reduced or no difference. **(E to G)** Gene set burden analyses show enrichment of de novo damaging variants in ASD probands relative to unaffected siblings from the Simons Simplex Collection (4) for interactors identified in iEN-PPI (34) and interactors from the matching set of 13 hcASD from ASD-PPI (E), mouseNeu-PPI (112) (F), and mouseCortex-PPI (113) (G). **(H)** ASD-PPI interactors (circle) and interactors$^{\text{no hcASD}}$ (triangle) are enriched for ASD and developmental disorder (DD) but not schizophrenia (SCZ) risk gene sets compiled from recent sequencing studies and curated databases (see Methods). **(I)** Relative co-expression of ASD-PPI and iEN-PPI (34) across cell types from three prenatal brain single cell transcriptomic atlases, with hcASD–interactor edges defined by each PPI network and interactor–interactor edges derived from STRING. Incorporating STRING interactor–interactor edges shifts enrichment toward NPC-EN. Abbreviations: div, dividing; EN, excitatory neuron; IN, inhibitory neuron; NPC, neuronal progenitor cell; OPC, oligodendrocyte precursor cell; prog, progenitor. Statistical tests: (A to B, D, I) t-test; (C) permutation test; (E to H) Fisher's exact test, one-sided, greater; (I) Wilcoxon rank sum test. P values corrected for multiple hypothesis testing (Bonferroni). Nominally significant: raw $p < 0.05$; significant: Bonferroni-adjusted $p < 0.05$. n.s. not significant. *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$p < 0.001$.

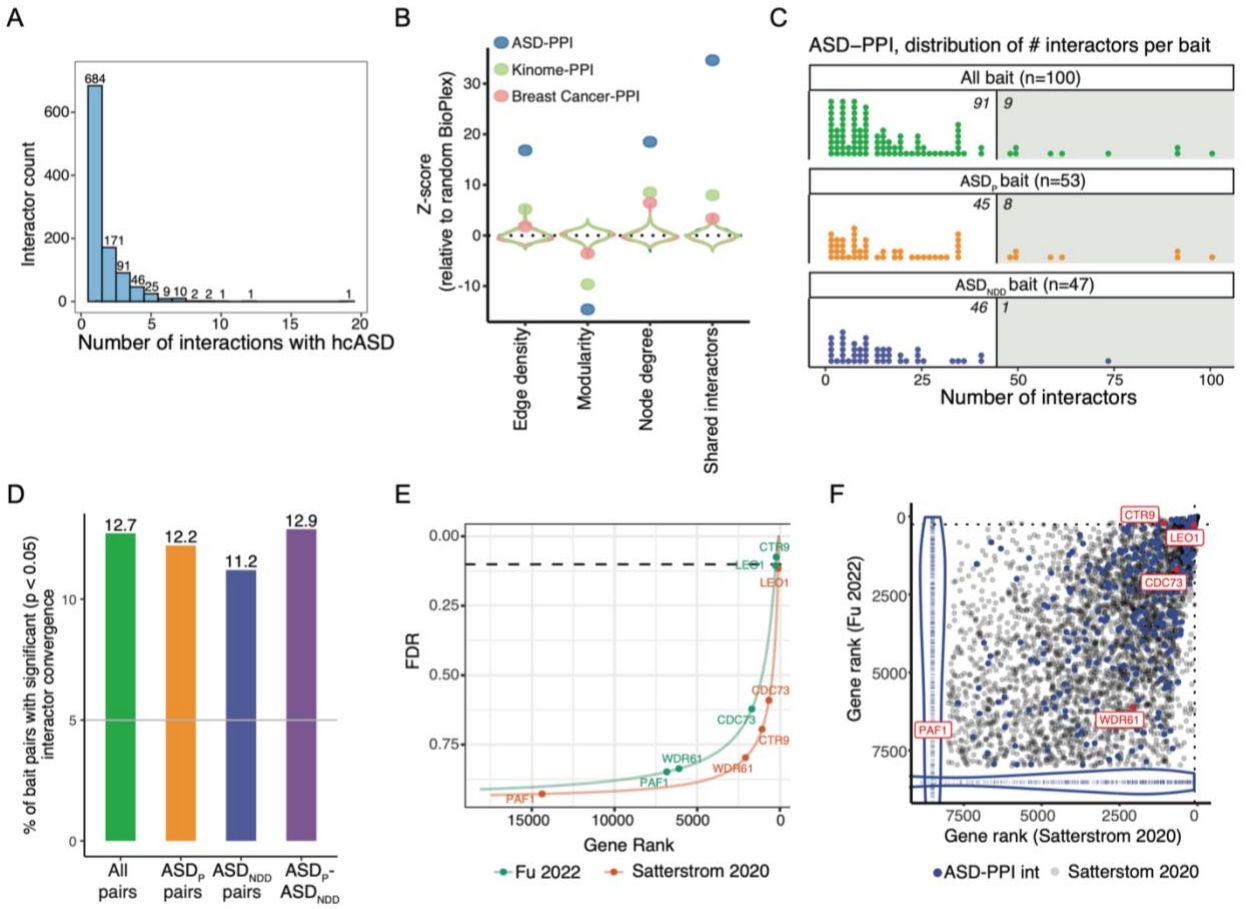

Figure 3.11 ASD-PPI captures protein complexes with multiple ASD risk genes.
(A) Distribution of hcASD interactions per ASD-PPI interactor. **(B)** Connectivity-related network measures for ASD-PPI and two previously published AP-MS networks centered on tyrosine kinase or breast cancer risk gene baits (Kaushik et al. 2020, Kim et al. 2021). Scores shown as Z-scores relative to 1,000 random degree-matched BioPlex subnetworks (violin plots). **(C)** Number of ASD-PPI interactors for all hcASD (green), hcASD that are more frequently mutated in ASD (orange, ASD-predominant; ASDP) and hcASD that are more frequently mutated in NDD (blue, ASDNDD) (4). hcASD with >45 interactors (gray) were excluded from analyses in (D). **(D)** Proportion of hcASD gene pairs with significant interactor overlap (hypergeometric test, $p < 0.05$) across indicated hcASD pair combinations. No difference was observed among categories (Kruskal–Wallis test, $p = 0.15$). **(E)** Exome-wide gene rank and FDR from Fu et al. 2022 (green) or Satterstrom et al. 2020 (orange), where lower values indicate stronger ASD association. Dashed line marks hcASD threshold (FDR = 0.1). Five PAF1 complex members are labeled; CTR9 (FDR = 0.07 (6)) and LEO1 (FDR = 0.11 (6)) fall below or near threshold in Fu et al. 2022. **(F)** Comparison of exome-wide ASD association ranks from Satterstrom et al. 2020 (x-axis) and Fu et al. 2022 (y-axis). ASD-PPI interactors (blue) are among more strongly associated genes; PAF1 complex members are labeled in red; other genes are shown in gray. Marginal violins show ASD-PPI interactor rank distributions; points show genes in top 8,000 in both studies.

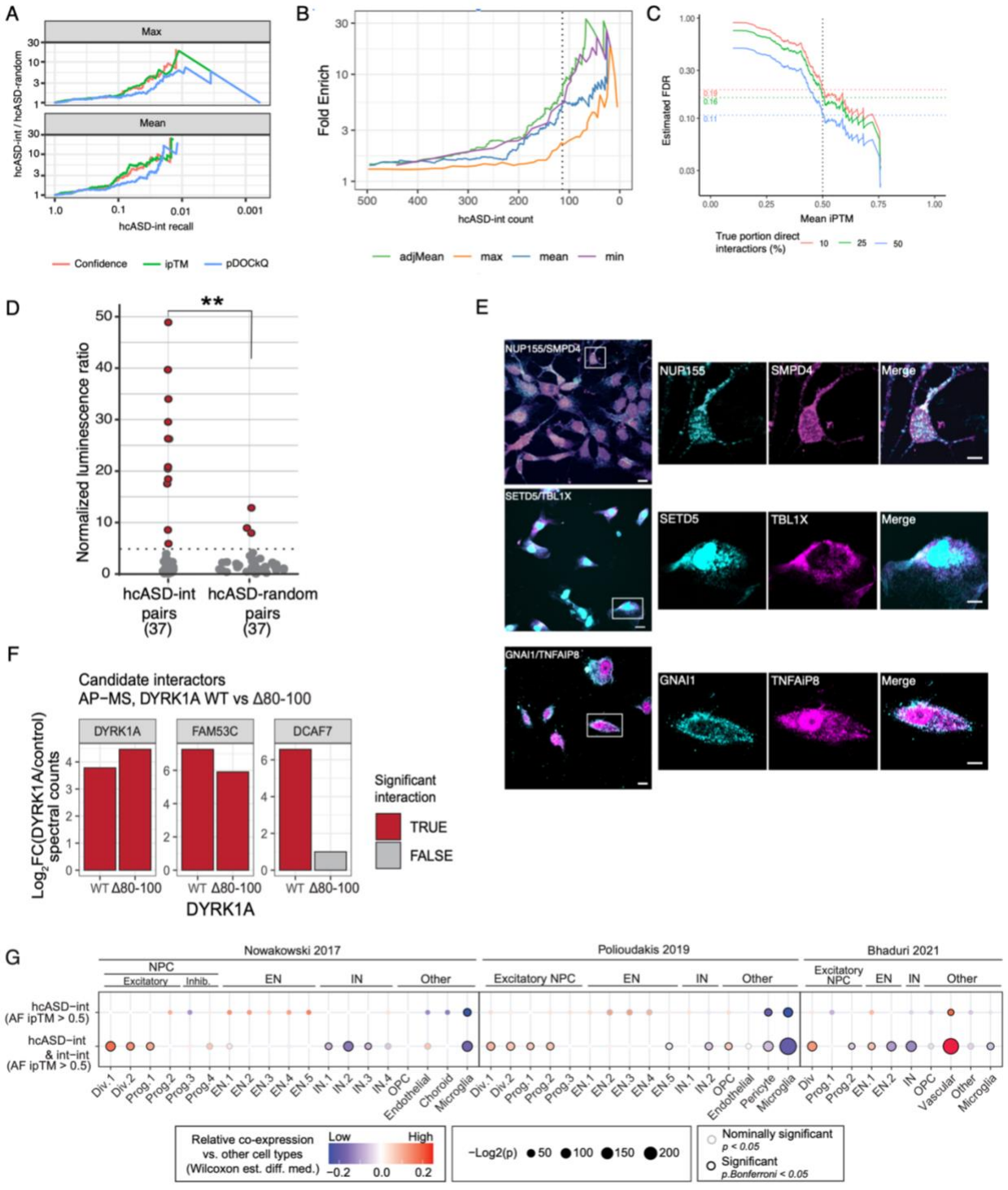

Figure 3.12 AlphaFold predicts direct protein interactions and their interfaces.
(A) Performance of AlphaFold scoring metrics (confidence, ipTM, pDockQ) in distinguishing hcASD-interactor pairs from hcASD-random pairs, shown as enrichment (y-axis) versus recall (x-axis). Scores are summarized by maximum (top) or mean (bottom) model values. Greater area under the curve reflects better performance. **(B)** Comparison of ipTM summary statistics for distinguishing hcASD-int from hcASD-random pairs, shown as enrichment of hcASD over random pairs (y-axis) versus number of hcASD pairs (Figure legend continued on next page.)

(Figure legend continued from previous page.) recalled (x-axis) across thresholds. Summary metrics include maximum, mean, minimum, and an adjusted mean (adjMean; mean minus two standard deviations across models). At stringent thresholds, where hcASD-int counts are below 100, min and adjMean perform similarly and better than mean, but they become more equivalent at hcASD-int counts greater than 100. The dotted line at count =113 marks the selected mean ipTM threshold of 0.5. **(C)** Estimated false discovery rate (FDR) of direct interactors in the hcASD-int set across mean ipTM thresholds (x-axis). FDR was estimated assuming random pairs represent non-interactors, with their pass rate approximating the false-positive rate. Estimates are shown assuming that 10%, 25%, or 50% of hcASD-int pairs represent true direct interactions. Colored horizontal lines indicate FDR values at the selected mean ipTM threshold of 0.5 (black vertical line). **(D)** Quantification of split-luciferase binary interaction assays. hcASD-int pairs were classified as direct interactions (red) if their normalized luminescence ratio (NLR) exceeded that of a fixed threshold of NLR > 5 (dotted line). **(E)** Representative confocal images showing colocalization of hcASD (cyan, NUP155, SETD5 or GNAI) and interactors (magenta, SMPD4, TBL1X and TNFAIP8), respectively. Scale bars: 20 μm (field, left), 10 μm (zoom, right). **(F)** Comparison of DYRK1A, FAM53C, and DCAF7 abundance (log2FC in spectral counts) detected by AP-MS using Strep-tagged DYRK1A or DYRK1AΔ80-100, with significant interactions (SAINT 1-BFDR $\geqslant$ 0.95) colored red. **(G)** Relative co-expression of AF-predicted direct hcASD-int (top) and a denser network including AF-predicted direct int-int (bottom) across cell types from the three prenatal brain atlases, evaluated as in fig. S2I. Incorporating interactor–interactor edges shifts enrichment toward NPC–EN populations. Wilcoxon rank sum test; nominally significant: raw $p < 0.05$; significant: Bonferroni-adjusted $p < 0.05$. Abbreviations: AF (AlphaFold-Multimer), int (interactor) div (dividing) EN (excitatory neuron), IN (inhibitory neuron), int (interactor), NPC (neuronal progenitor cell), OPC (oligodendrocyte precursor cell), prog (progenitor).

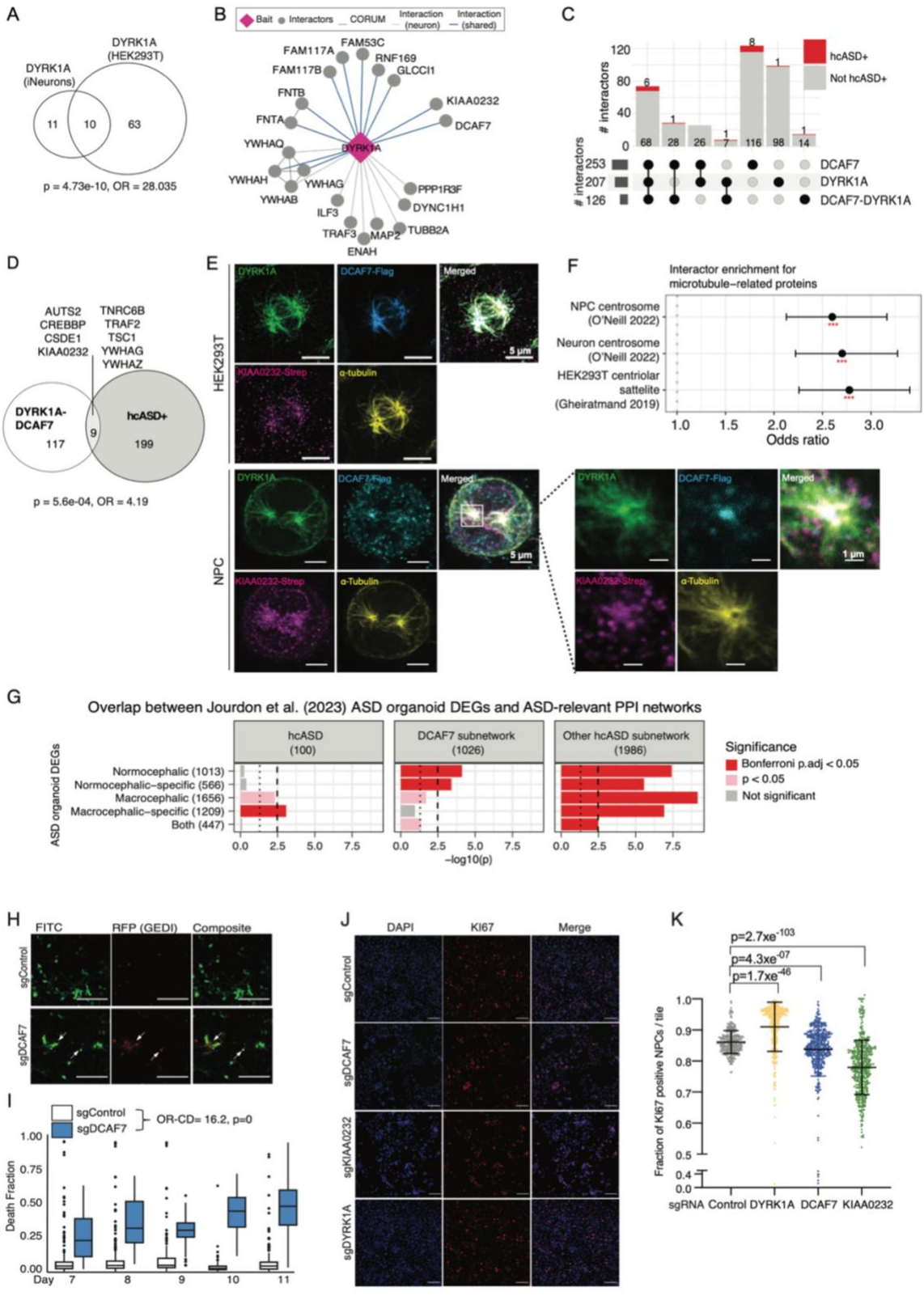

Figure 3.13 DCAF7, DYRK1A and KIAA0232 overlap functionally.
(A) Overlap of high- confidence DYRK1A interactors (Figure legend continued on next page.)

(Figure legend continued from previous page.) identified from overexpressed DYRK1A in HEK293T cells and endogenous DYRK1A in iPSC-derived excitatory neurons (iEN). Fisher's exact test (one-sided, greater). **(B)** Endogenous DYRK1A PPI network in iEN; "shared" interactions also detected from DYRK1A overexpression in HEK293T are shown in blue. **(C)** UpSet plot of interactors identified by single or sequential IP-MS for DCAF7 and DYRK1A; hcASD+ interactors are shown in red. **(D)** Enrichment of hcASD+ genes among interactors identified by sequential DCAF7–DYRK1A IP-MS (Fisher's exact test, one-sided; OR = 4.19, p = 0.000565). **(E)** Representative immunostaining showing colocalization of DYRK1A (geen), DCAF7 (blue), and KIAA0232 (magenta) with α-tubulin (yellow) at mitotic spindles in HEK293T cells and neural progenitor cells (NPCs); scale bar: 5 μm. Inset: NPC centrosome base; scale bar: 1 μm. **(F)** Enrichment of ASD-PPI interactors in three sets of proteins associated with structures important for spindle organization (centriolar satellites (130) and centrosomes (131)). Two-sided Fisher's exact tests with Bonferroni correction - HEK293T centriolar satellite: p.adj = 1 x 10-19; NPC centrosome: p.adj = 9.9 x 10-19; neuron centrosome: p.adj = 6 x 10-21. **(G)** Enrichment of hcASD and DCAF7-centered networks among differentially expressed genes (DEGs) in ASD cerebral organoids with normocephalic or macrocephalic phenotypes relative to controls (see Methods). Phenotype-specific DEGs indicate DEGs present only in the normocephalic or macrocephalic group. Network gene sets included all 100 hcASD baits ("hcASD"), DCAF7 with eight interacting hcASD baits plus their interactors ("DCAF7 subnetwork"), and the remaining baits plus their interactors ("Other hcASD subnetwork"). Dotted and dashed lines indicate nominal ($p < 0.05$) and Bonferroni-adjusted ($p.adj < 0.05$) significance (one-sided Fisher's exact test). (H) Representative images showing increased cell death following DCAF7 knockdown in iPSC-derived cortical neurons transduced with a "red genetically encoded death indicator" (RGEDI) biosensor (139) at differentiation day 13. Images show morphology (FITC, left), cell death (RFP, middle) and overlay (right). Arrows indicate dying cells. Scale bar = 100 μm. (I) NPCs with sgDYRK1A have a higher odds of cell death (OR-CD) than those with sgControl OR-CD= 16.2 (99% CI: 13.8-19.0), p=0. Plots showing the fraction of dying and dead cells over time on differentiation days 7 to 11 (n = 3 experimental replicates). The OR-CD was evaluated using a generalized linear mixed model (GLMM) (156) (see Methods). (J) Immunostaining of NPCs with control, DCAF7, KIAA0232 or DYRK1A sgRNAs. Scale bar = 100 μm. (K) Comparison of percent Ki67 positive cells in NPCs with control, DCAF7, KIAA0232 or DYRK1A sgRNAs (n=4 experimental replicates, control=332,910 cells, sgDCAF7=279,544 cells, sgDYRK1A= 345,758 cells, sgKIAA0232= 279,899 cells). P values derived from a GLMM with post-hoc Benjamini Hochberg multiple comparisons correction (144, 156, 157).

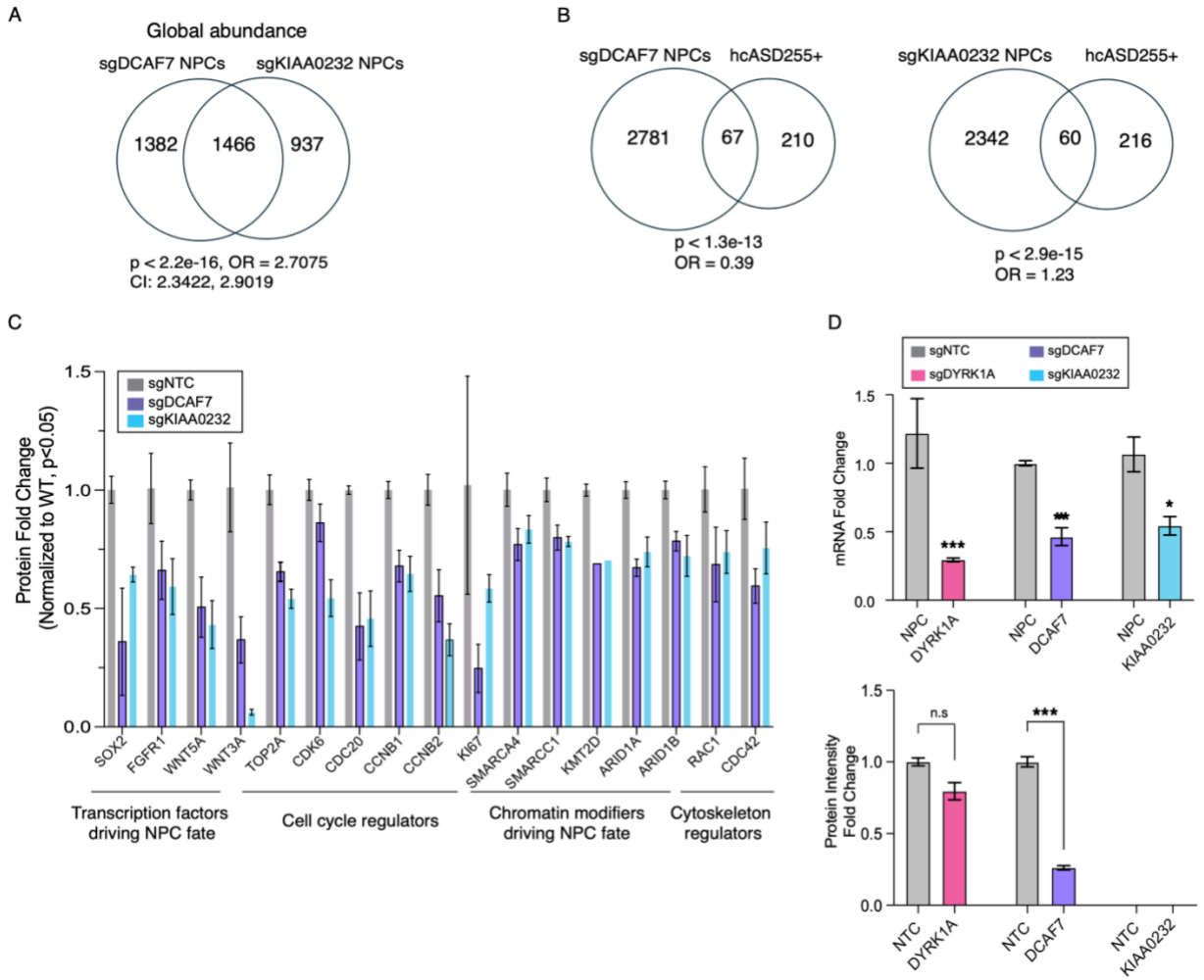

Figure 3.14 DCAF7 and KIAA0232 knockdown produce concordant proteomic and progenitor- state disruptions in NPCs.
(A) Significant overlap of differentially expressed proteins (DEPs) identified from global proteomic profiles of DCAF7 KD and KIAA0232 KD NPCs relative to wild-type controls. One-sided Fisher's enrichment test; DEPs defined by p.adj < 0.05. **(B)** DCAF7 KD and KIAA0232 KD NPC DEPs are enriched for hcASD255+. One-sided Fisher's enrichment test. **(C)** Proteins associated with maintenance of progenitor state are decreased in sgDCAF7 and sgKIAA0232. **(D)** mRNA (top) and protein (bottom) levels of DYRK1A, DCAF7 and KIAA0232 in their respective knockdown NPCs. One-way ANOVA with Dunnett's post hoc test (mRNA) or Bonferroni-corrected post hoc tests (protein), each compared to sgNTC. Significance: *p < 0.05, **p < 0.01, ***p < 0.001.

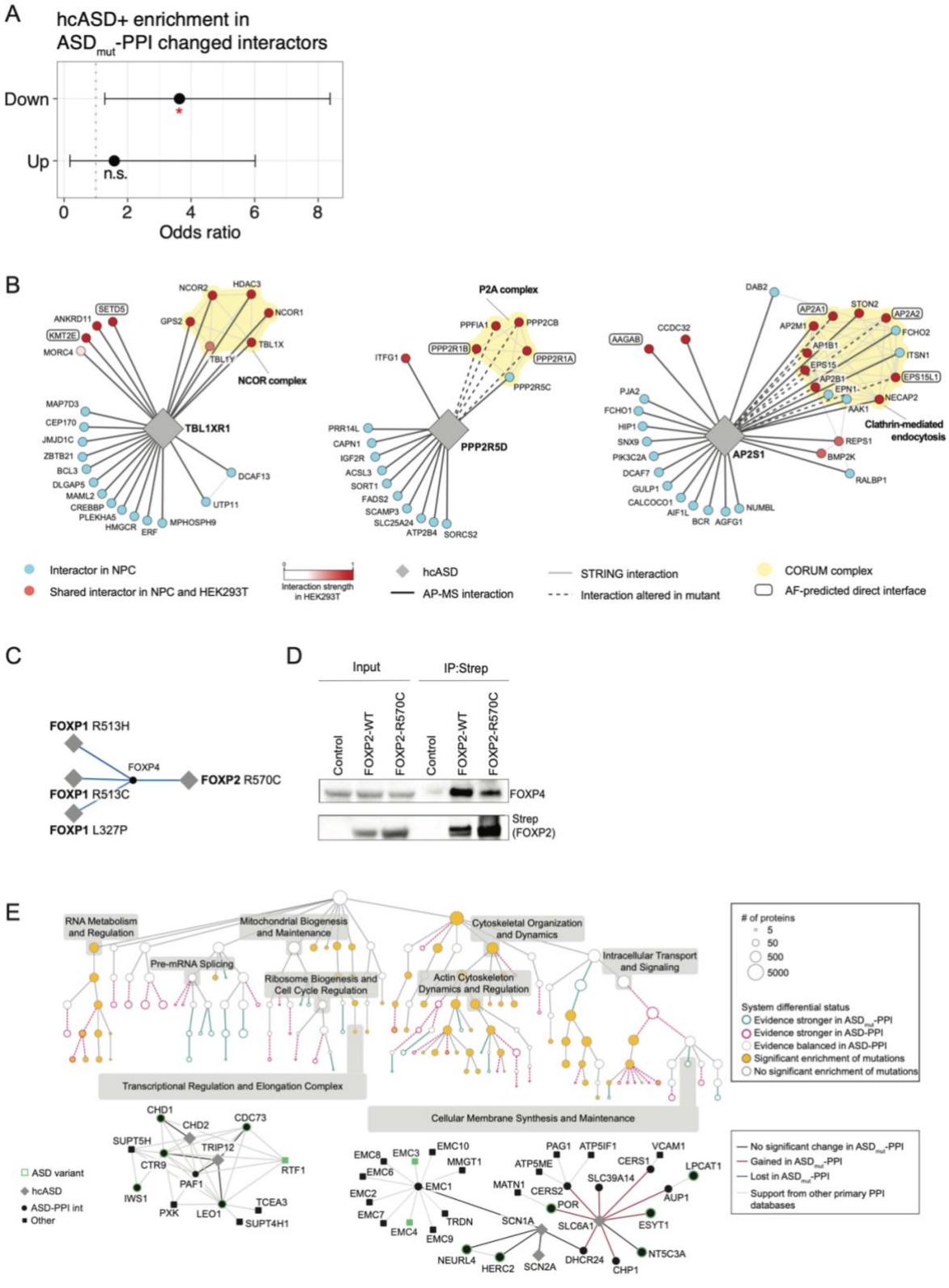

Figure 3.15 hcASD mutations result in convergent interaction changes.
(A) hcASD+ enrichment in interactors with decreased (Figure legend continued on next page.)

(Figure legend continued from previous page.) ('Down') or increased ('Up') binding to mutant versus WT hcASD in ASDmut-PPI. Two-sided Fisher's exact tests with Bonferroni correction; lost: p.adj = 0.0171; gained: p.adj = 0.738. **(B)** Neural progenitor cell (NPC) PPI networks for three hcASD (TBL1XR1, PPP2R5D and AP2S1). NPC-specific interactors are shown in blue; interactors shared with HEK293T cells are shown in red. Significant overlap between NPC and HEK293T interactors is observed for each gene (TBL1XR1: p.adj =6.756 x 10-14, PPP2R5D: p.adj =1.0302 x 10-11, AP2S1: p.adj = 1.58503 x 10-12). AF-predicted direct interactions (ipTM > 0.5) are boxed; dashed lines denote interactions altered in ASDmut-PPI. **(C)** ASDmut-PPI subnetwork showing loss of interaction (blue lines) between FOXP1 and FOXP2 patient-derived mutations and FOXP4. **(D)** IP–western blot validating reduced interaction between Strep-tagged FOXP2-R570C and FOXP4 relative to FOXP2-wt. **(E)** Multi-scale hierarchical layout of protein systems annotated by stronger support in ASDmut-PPI (green border) or ASD-PPI (pink border) and significant enrichment for ASD genetic risk (yellow fill). Circle size reflects the number of proteins in the system. Examples of systems that were not affected by ASDmut-PPI (lower left) or had differential interactions from ASDmut-PPI (lower right) are shown below. Diamonds indicate mutant hcASD, circles indicate ASDmut-PPI interactors, and squares indicate protein interactors not found in this study. Green highlights on nodes indicate genetic association with ASD (4) (e.g. observed de novo damaging variant in an ASD proband (4), see Methods). Edge colors denote interaction source and specificity.

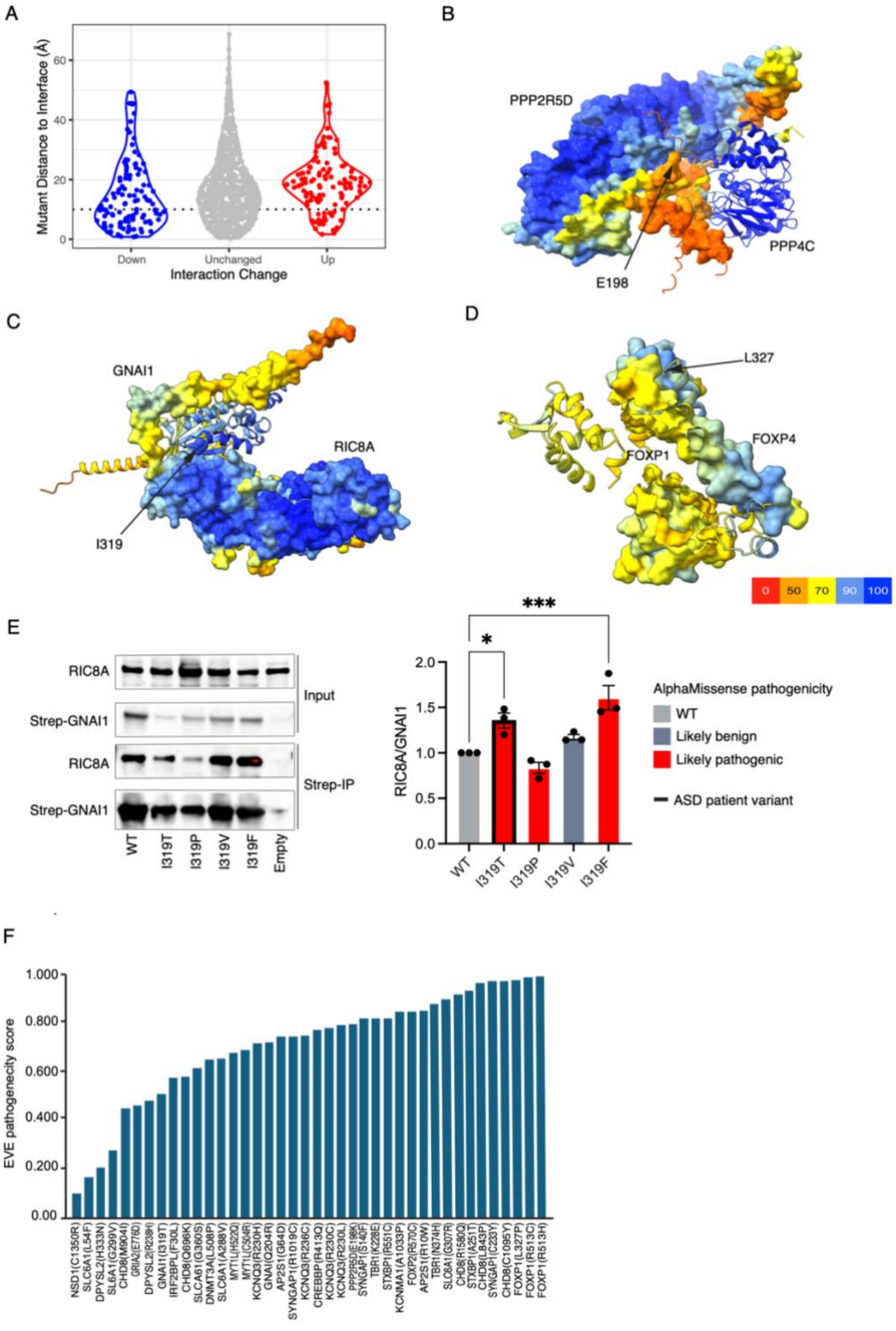

Figure 3.16 hcASD mutations near AF-predicted interfaces are associated with interaction loss.

(A) Distribution of mutant hcASD-int pairs from ASDmut-PPI based on distance of the mutated residue to the interaction interface (y axis). (Figure legend continued on next page.)

(Figure legend continued from previous page.) Interactions are classified as weakened ("down"), unchanged, or strengthened ("up") in presence of a patient-derived hcASD mutation. **(B to D)** Per-residue AlphaFold confidence scores (pLDDT) across predicted interfaces for PPP2R5D-PPP4C (B), GNAI1-RIC8A (C), and FOXP1-FOXP4 (D), where higher scores reflect higher confidence. **(E)** Western blot and quantification of GNAI-RIC8A interaction for AlphaMissense pathogenic (red) or benign (gray) mutations at residue I319. **(F)** EVE scores (150) for hcASD missense mutations in ASDmut-PPI; higher scores reflect higher predicted pathogenicity. Statistical tests: (E) One way ANOVA, $F(4, 10) = 19.57$, $p < 0.0001$, Dunnett's post hoc test for multiple comparisons; WT vs I319T p.adj = 0.0484; WT vs I319F, p.adj = 0.0001. Unpaired t-test for WT vs I319P, p = 12 0.0276.

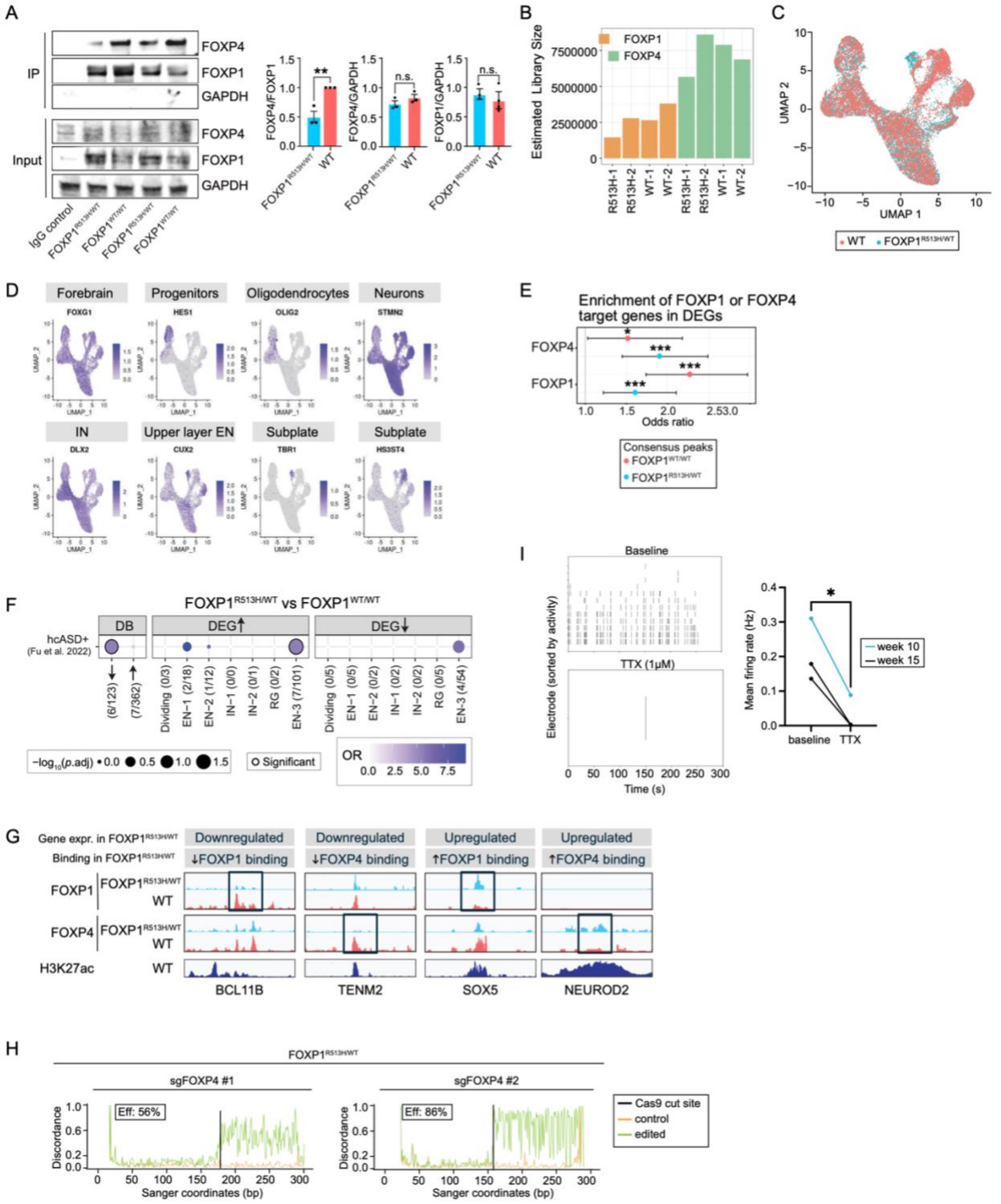

Figure 3.17 FOXP4 is necessary for altered excitatory neurogenesis phenotypes in FOXP1 R513H.

(A) Immunoprecipitation (IP)–Western blot of endogenous FOXP1 showing loss of FOXP1–FOXP4 interaction in FOXP1$^{R513H/WT}$ relative to FOXP1$^{WT/WT}$, and quantification (unpaired t-test, p = 0.0065, t = 5.205, n.s., not significant; *0.01 < p < 0.05; **0.001 < p < 0.01; ***p < 0.001). **(B)** Estimated CUT&Tag library sizes for (Figure legend continued on next page.)

(Figure legend continued from previous page.) FOXP1$^{R513H/WT}$ and ^{FOXP1WT} organoids at day 111, shown for FOXP1 binding (orange) and FOXP4 binding (green). **(C)** scRNA-seq UMAP for FOXP1R513H/WT and FOXP1WT/WT grouped by genotype. **(D)** Cell type specific markers used to annotate scRNAseq clusters. **(E)** Left, enrichment of FOXP1 and FOXP4 target genes among those differentially expressed between FOXP1$^{WT/WT}$ and FOXP1$^{R513H/WT}$ organoids. Right, enrichment of genes showing nominally significant differential FOXP1 or FOXP4 binding among EN-3 differentially expressed genes in FOXP1$^{R513H/WT}$ vs FOXP1$^{WT/WT}$. Two-sided Fisher's exact test, *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$. **(F)** Enrichment of hcASD+ (6) among FOXP1 differentially bound genes (DB, left panel) and cell-type specific differentially expressed genes (DEGs, middle panel: upregulated in FOXP1$^{R513H/WT}$, right panel: downregulated in FOXP1$^{R513H/WT}$). Two-sided Fisher's enrichment tests, p values adjusted for multiple hypothesis testing (Bonferroni, 16 tests); decreased FOXP1 binding: p.adj = 0.0038; EN-3 upregulated transcripts: p.adj = 0.012. **(G)** FOXP1 and FOXP4 binding patterns in FOXP1^{WT} and FOXP1$^{R513H/WT}$ forebrain organoid, illustrating representative examples of loss or gain of FOXP1 or FOXP4 binding associated with gene expression changes. Two genes downregulated in FOXP1$^{R513H/WT}$ (BCL11B and TENM2), and two genes upregulated in FOXP1$^{R513H/WT}$ (SOX5, NEUROD2) highlight four observed patterns of altered binding in FOXP1$^{R513H/WT}$: loss of FOXP1 binding (BCL11B); loss of FOXP4 binding (TENM2); gain of FOXP1 binding (SOX5); and gain of FOXP4 binding (NEUROD2). **(H)** Inference of CRISPR edits (ICE) analysis confirming CRISPR-mediated FOXP4 knockout (55% and 85% knockout efficiency) using two sgRNAs in the FOXP1R513H/WT background. Relative sequence discordance versus empty vector control is shown in green, with the Cas9 cut site indicated in black. (I) MEA validation comparing baseline activity to activity after a 5min treatment with 1μM TTX. Raster plots for the active electrodes within 300 seconds of the recording shown on the left, quantification with one-sided Mann Whitney U test on the right. n = 3 organoids.

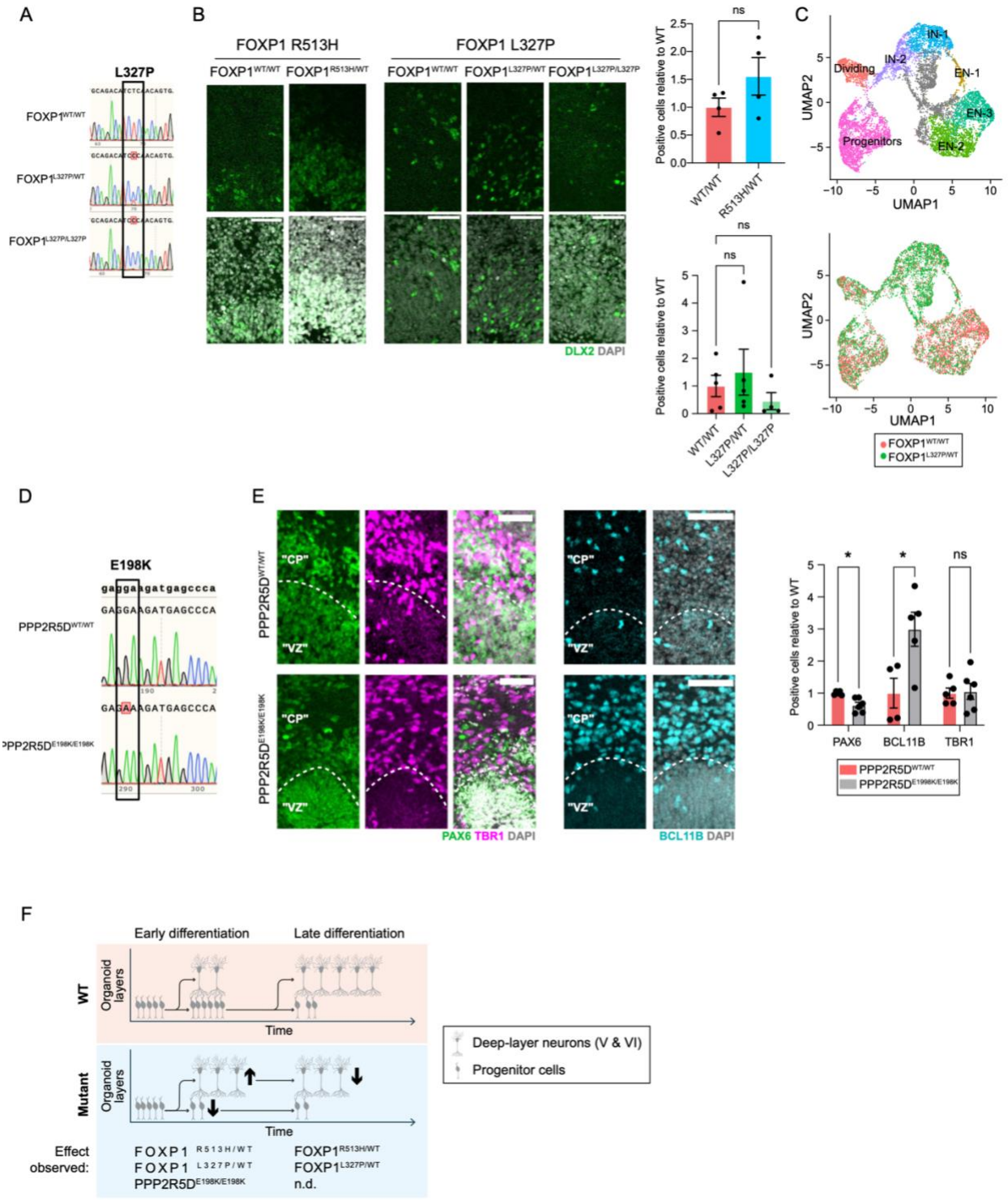

Figure 3.18 hcASD mutations alter cortical neuron differentiation.
(A) Sanger sequencing of isogenic FOXP1$^{WT/WT}$, FOXP1$^{L327P/WT}$, and FOXP1$^{L327P/L327P}$ iPSCs (WT L327: CTC, L327P: CCC). **(B)** Immunohistochemistry of inhibitory neuron progenitor marker DLX2 in day 39 organoids from independent FOXP1 R513H and L327P isogenic series, each with its one WT control. R513H series: FOXP1$^{WT/WT}$ (n=4 organoids) and FOXP1$^{R513H/WT}$ (n=4); L327P series: FOXP1$^{WT/WT}$ (n=5), (Figure legend continued on next page.)

(Figure legend continued from previous page.) FOXP1$^{L327P/WT}$ (n=5), and FOXP1$^{L327P/L327P}$ (n=4). Statistics: Linear regression followed by permutation testing and Benjamini-Hochberg FDR correction for both; n.s., not significant. **(C)** UMAP of scRNA-seq data from the FOXP1 L327P experiment (FOXP1$^{WT/WT}$ and FOXP1$^{L327P/WT}$ cells). Cell types are colored by cell type annotations transferred from the FOXP1 R513H scRNA-seq dataset (top, see Methods) or genotype (bottom, FOXP1$^{WT/WT}$ or FOXP1$^{L327P/WT}$). **(D)** Sanger sequencing of isogenic PPP2R5D iPSCs (WT E198: GAA, E198K: AAA). **(E)** Immunohistochemistry (IHC) of day 39 forebrain organoids (PPP2R5D WT or PPP2R5D$^{E198K/E198K}$ showing layer VI/subplate marker (TBR1), layer V marker (BCL11B) and progenitor marker (PAX6). Scale bar 50μm, and quantification, with linear regression with permutation testing and Benjamini Hochberg FDR correction. * $p < 0.05$, ns = not significant. **(F)** Schematic representation of the proposed mechanism by which FOXP1 and PPP2R5D variants alters deep-layer neuronal differentiation.

Materials and Methods

Cloning hcASD genes

We selected 102 hcASD genes with FDR < 0.1 reported in Satterstrom *et al.* 2020 (*4*). The coding sequence of each hcASD risk gene was cloned into a pcDNA4 plasmid with either N- or C-terminal 2xStrep tags. Where possible, the terminus position of the tags was selected to minimize interference with protein function based on the prior functional studies. The isoform selected for each hcASD risk gene was the one with highest expression in the human prenatal forebrain[20] that also preserved exon(s) harboring ASD patient-derived damaging *de novo* mutations.[4] All constructs were sequence validated. We successfully cloned 100 of the 102 hcASD genes (missing *HECTD4* and *CACNA1E*).

Cell culture

Human HEK293T cells (ATCC, CRL-3216) were cultured in Dulbecco's modified Eagle's medium (DMEM; Fisher Scientific #10566024) supplemented with 10% fetal bovine serum (FBS; Gibco, Life Technologies #10438026) and 1% penicillin–streptomycin (Corning). All cells were maintained in a humidified incubator at 37 °C with 5% CO_2. Human iPSCs (WTC11 (Conklin Lab), 13234 (Conklin Lab) were cultured on Matrigel (Corning #354230) coated plates in StemFlex Medium (Gibco #A3349401) with 100μg/ml Primocin (Invitrogen #ANTPM1). Cells were passaged 1:10 every 3 days using ReLeSR (StemCell Technologies #05873). Cells were maintained in a humidified incubator at 37°C with 5% CO_2.

Transfection

Each transfection (100 hcASD, one GFP control and one empty vector control) was carried out in a 15cm dish with 10 million HEK293T cells (70-80% confluency), with three biological replicates

per hcASD. Transfections were split in 15 batches, with three biological replicates of GFP and empty vector controls included in each batch. For each transfection, 15µg of Strep-tagged plasmids was combined with PolyJet Transfection Reagent (Fisher Scientific #504788) at a 1:3 µg:µl ratio of plasmid:transfection reagent, incubated at room temperature for 10 mins, and added dropwise to HEK293T cells. About 48h post transfection, cells were resuspended at room temperature using 10 ml Dulbecco's phosphate-buffered saline without calcium and magnesium (DPBS, Fisher Scientific #14190250) supplemented with 10 mM EDTA, followed by centrifugation at 200g, 4°C for 5 min. Cell pellets were frozen on dry ice and stored at −80°C.

Affinity purification

The cell pellets were thawed on ice and then lysed with 1 ml ice-cold lysis buffer (IP buffer (50mM Tris-HCl, pH 7.4, 150mM NaCl, 1 mM EDTA, 0.5% Nonidet P40 substitute (NP40; Fluka Analytical), cOmplete mini EDTA-free protease and PhosSTOP phosphatase inhibitor cocktails (Fisher Scientific #NC0962311, Sigma Aldrich #4906845001)). Samples were then flash-frozen on dry ice for about 10 min and partially thawed at 37°C for 30-45s before incubation for 30 min at 4°C on a tube rotator. Lysates were centrifuged at 13,000g, 4°C for 15 min to clarify lysate and pellet debris. A 50µl lysate was reserved at this point for future experiments such as Western blot. The remaining lysates either underwent automated affinity purification on the KingFisher Flex Purification System (Thermo Scientific), which was first equilibrated to 4°C in the cold room or were processed for immunoprecipitation for validation. First, MagStrep 'type3' beads (30µl; IBA Lifesciences #NC0776437) were equilibrated twice with 1ml wash buffer (IP buffer supplemented with 0.05% NP40) and incubated with 0.95ml lysate for 2 h. Beads were washed 3 times with 1 ml wash buffer and once with 1 ml IP buffer, then resuspended in 50µl denaturation–reduction buffer (2M urea, 50mM Tris-HCl pH 8.0, 1mM DTT) and 50µl 1× buffer BXT (IBA Lifesciences

#2-1042-025) and dispensed into a single 96-well KingFisher microtitre plate. Purified proteins were eluted at room temperature for 30 min with constant shaking at 1,100 rpm on a ThermoMixer C incubator.

On-bead digestion

Bead-bound proteins were denatured and reduced at 37°C for 30 min, brought back to room temperature, alkylated in the dark with 3mM iodoacetamide for 45 min and quenched with 3mM DTT for 10 min. To offset evaporation, 15µl 50 mM Tris-HCl, pH 8.0 were added and proteins were digested with 1.5µl trypsin (0.5 µg/µl; Promega #V5113) with constant shaking at 1,100 rpm and incubation at 37°C on a ThermoMixer C incubator for 4 h, and again for 1–2 h with 0.5µl additional trypsin. Resulting peptides were combined with 50µl 50mM Tris-HCl, pH 8.0 to rinse beads and then, acidified with trifluoroacetic acid (0.5% final, pH < 2.0). Desalting was conducted in a BioPureSPE Mini 96-Well Plate (20mg PROTO 300 C18; The Nest Group) according to standard protocols.

Mass spectrometry

To prepare for mass spectrometry, samples were resuspended in 4% formic acid, 2% acetonitrile solution, and separated by a reversed-phase gradient over a Nanoflow C18 column (Dr Maisch). Each sample was directly injected via an Easy-nLC 1200 (Thermo Fisher Scientific) into a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific) and analyzed with a 75 min acquisition, with all MS1 and MS2 spectra collected in the orbitrap; data were acquired using the Thermo software Xcalibur (4.2.47) and Tune (2.11 QF1 Build 3006). For all acquisitions, QCloud was used to control instrument longitudinal performance. All proteomic data were searched against the human proteome (UP000005640_9606, downloaded in October 2021) using the default settings for MaxQuant software (version 1.6.12.0)[89] with match-between-runs (MBR) feature

turned on. Briefly, the MBR algorithm annotates unidentified peaks by assessing and comparing the retention times of the identified peaks in an MS1 spectrum. Detected peptides and proteins were filtered to 1% false-discovery rate in MaxQuant, and identified proteins were then subjected to protein–protein interaction scoring with both SAINTexpress (v.3.6.3)[90] and CompPASS (version 0.0.0.9000).[91]

AP-MS data processing and identification of high-confidence interactors

In the following descriptions, we use the term "bait" to describe each tagged protein of interest, according to AP-MS conventions. A critical step in our AP-MS study is the probabilistic scoring of all quantified proteins in the dataset to identify high-confidence interaction proteins (HCIP) of the ASD genes. To do this, the scoring outputs from existing computational tools such as CompPASS (Comparative Proteomic Analysis Software Suite)[91] and SAINTexpress[90] were systematically combined. These two scoring algorithms are widely used in the proteomics community to score the quality of PPIs as distinct from background. We chose to leverage the complementary strengths of both tools to balance sensitivity and specificity. SAINTexpress uses a Bayesian framework to estimate the probability that a bait–interactor interaction is genuine by comparing baits individually against a defined control treatment, and thus is very sensitive to any increased pull-down due to overexpression of the bait. Meanwhile CompPASS assesses interactions across all baits to emphasize specificity and reproducibility, downweighting promiscuous binders. Integrating both approaches–by applying SAINTexpress and ComPASS cutoffs and considering only the overlap of proteins that passed both as true interactors–allowed us to generate a highly stringent, high-confidence PPI network.

Input files - bait, prey and interaction files - for SAINTexpress were made using artmsEvidenceToSaintExpress function in an open-sourced R package, artMS.[92] The interaction

input file for SAINTexpress was reformatted to be used as a CompPASS input file. As CompPASS requires all baits to have the same number of preys, we defined spectral counts for the union of all identified proteins across all AP-MS experiments for each bait; if the bait-prey interaction was not detected, its spectral count was assigned to be zero. This bait-prey spectral count matrix was used to search for carryover effect, which was defined as proteins with a continuous decrease in spectral counts in replicates where such proteins were used as baits in the previous sample injections. The resulting bait-prey spectral count data table after carryover removal was reformatted and used as the input file for SAINTexpress and CompPASS.

To determine the scoring cutoffs that capture the highest number of true interactions, a gold standard set of protein-protein interactions was manually defined. Gold standard interactions were extracted from publicly available large-scale PPI databases, Hein *et al.,*[68] InWeb,[93] CORUM (Core CORUM Complexes, CORUM 3.0),[94] BioPlex,[67] and BioGRID,[54] with some additional filtering steps for BioGRID and InWeb. As BioGRID contains interactions from various sources, including both experimental and predicted, the dataset was filtered to only interactions that were attained using another experimental method in addition to the AP-MS method. Similarly, InWeb databases were filtered for interactions that have a high-confidence score of > 0.95 and are identified through an experimental method.

SAINTexpress (version 3.6.3) was run batch-wise, each batch with its own empty-vector and GFP replicates as controls. The outputs were then concatenated to a single SAINTexpress output. CompPASS was run using the R package, cRomppass (https://github.com/dnusinow/cRomppass). The scoring outputs from SAINTexpress and CompPASS were merged for each interaction. The SAINTexpress Bayesian False Discovery Rate (BFDR) and CompPASS WD score (rank_WD, ranked from 0 to 1 across the dataset) were used as prediction confidence scores. Using the ranked

CompPASS WD score (rank_WD) as a predictive value and interactions found in the previously described manually curated gold standard PPI database as true positives, the optimal rank_WD score cutoff for each SAINTexpress BFDR increment was determined by calculating Youden's index. The best performing composite score of SAINTexpress and CompPASS (out of 78 composite scores) was determined by calculating precision, recall and F1 scores.

Selection and cloning of hcASD missense variants

To study protein interaction changes using AP-MS, we focused on *de novo* likely damaging missense mutations (MPC≥ 2) in hcASD genes identified from ASD probands in Satterstrom *et al*. 2020. This resulted in 87 *de novo* missense mutations in 43 hcASD genes. These missense mutations were introduced into the WT version of the pcDNA4 construct using Q5 site-directed mutagenesis. Protein interaction data was generated using the AP-MS pipeline described above. After checking for expression of the variants in HEK293T cells and further quality control, 33 variants across 13 hcASD were removed.

AP-MS data processing PPI scoring for hcASD missense variants

Following the identification and quantification of proteins using MaxQuant, high confidence interacting proteins (hcIP) were identified by running both SAINTexpress and CompPASS after carryover effects were removed. ASD-PPI refers to the AP-MS dataset generated from 100 hcASD genes. ASD_{mut}-PPI refers to AP-MS dataset generated from 54 *de novo* missense mutations in 30 hcASD genes, along with repeated AP-MS of their respective WT constructs as controls for differential interaction analysis, and empty-vector and GFP controls. SAINTexpress was run batch-wise, using empty vector and GFP constructs as controls. To have better specificity in identifying hcIPs using CompPASS, a larger input dataset was created by combining the ASD-PPI CompPASS input with ASD_{mut}-PPI input. CompPASS was then run using the CompPASS input

for ASD-PPI WT network as the background stats table. hcIPs were identified using the scoring cutoffs previously optimized from the WT ASD-PPI network: SAINTexpress score (1-BFDR) ≥ 0.95 and CompPASS score (rank_WD) ≥ 0.971. We retained only hcIPs from the mutant AP-MS experiments (including from their WT quantitative controls) to build the ASD_{mut}-PPI interactome. As a first step for quantifying differential interactions, we normalized interactor peptide intensity values to baits levels to account for variation in bait expression. This bait normalization was completed within each AP-MS run using a Tukey Median Polish normalization method on all bait peptides to calculate and remove the bait-variation across all peptides in the run. Following normalization, we quantified the changes observed in interactors between WT and mutant baits, using the R package MSstats.[95] Using the default parameters, except to disable further normalization, we used the functions *dataProcess* and *groupComparison* on the normalized intensity values to run differential analysis by t-tests per bait-prey between WT and mutant groups. *P* values were pooled across all t-tests (only 1,019 of 1,070 comparisons had sufficient observation counts for t-tests) and adjusted for multiple testing using FDR estimation by Storey's q method[96,97] (R package qvalue). We defined significant differential interactors by requiring they pass all of FDR < 0.1, $p < 0.05$ and absolute $\log_2 FC \geq 1$ (Fig. 6, A to D).

HEK293T proteome

We defined the HEK293T proteome to be the n = 11,133 proteins with detected protein expression in a published global quantitative mass spectrometry dataset.[98] We defined the protein expression level to be the average intensity based absolute quantification (iBAQ) scores for the two HEK293T experimental replicates.[98] For subsequent analyses, we defined HEK293T proteome genes to be the union of proteins with non-zero iBAQ scores and ASD-PPI interactors.

Comparison of ASD-PPI and Pintacuda et al. 2023 iEN-PPI network

Pintacuda *et al.* 2023 reported an "iEN-PPI" network for 13 hcASD genes in using IP-MS of endogenous proteins in human excitatory neurons derived from iPS cells.[23] Interactors from iEN-PPI were extracted from Supplemental Table S3 ('Interaction_Annotations' sheet).

We used Fisher's enrichment tests (one-sided, greater) to assess whether there was greater than expected overlap in the combined interactors for all n = 13 hcASD index proteins in Pintacuda 2023 and 1) the subset of ASD-PPI interactors from the matched set of n = 13 hcASD baits; and 2) all ASD-PPI interactors from n = 100 hcASD baits. We defined the universe of genes to be the set of n = 2,552 genes that are in the HEK293T proteome and in the set of detected proteins in iEN-PPI[23] (Table S3 in Pintacuda 2023, union of the interactor and non-interactor proteins in the combined network). The 2x2 contingency table was defined as: X, the number of ASD-PPI interactors that are iEN-PPI combined network interactors; Y; the number of ASD-PPI interactors that are not iEN-PPI combined network interactors; Z, the number of iEN-PPI interactors that are not ASD-PPI interactors; W, the number of genes that are not interactors in either ASD-PPI or iEN-PPI combined network.

We additionally assessed whether damaging *de novo* variants in iEN-PPI interactors or ASD-PPI subset interactors are associated with ASD. We focused on *de novo* damaging variants identified from the Simon's Simplex Collection[4]). We defined "damaging" variants to be variants that resulted in a damaging missense mutation (Polyphen Mis3 (damaging); or MPC Mis B (MPC≥2)) or PTV (frameshift, stop gained, or canonical splice site disruption) (ssc_enrichment/scripts/01_format_satterstrom2020_SSC_variants.R). We defined 2 genesets of interest:

1. iEN-PPI interactors: interactors from combined network in Pintacuda *et al.*2023 (n = 979 genes).

2. ASD-PPI subnetwork interactors -hcASD (interactors$^{\text{no hcASD}}$): ASD-PPI subset interactors, excluding 13 hcASD (n = 222 genes)

For each Fisher's enrichment test, the universe of genes was defined to be the geneset members, and the 2x2 contingency table was: X, the number of ASD probands with a damaging variant in at least one gene in the geneset; Y; the number of control siblings with a damaging variant in at least one gene in the geneset; Z, the number of ASD probands with no damaging variants in any gene in the geneset; W, the number of control siblings with no damaging variants in any gene in the geneset. We calculated the enrichment odds ratio with Fisher's exact test (one sided, alternative = greater), and adjusted *p* values for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p**number genesets assessed).

We assessed whether there is significant difference in ORs for the Pintacuda 2023 complete network and the ASD-PPI_13 subnetwork (baits + interactors) using the Breslow Day test (R DescTools::BreslowDayTest). We found no significant difference (p=0.085). (**Fig. 3.9D, S2E**).

Comparison of ASD-PPI and Murtaza et al. 2022 mouseNeu-PPI networks

Murtaza *et al.* 2022[24] reported a "mouseNeu-PPI" network generated by overexpressing baits and identifying interactors via *in vitro* proximity labeling (BioID2). mouseNeu-PPI consisted of 41 ASD genes in mouse primary neurons co cultured with glia (17 baits overlapping with ASD-PPI baits). ASD genes were selected to have non-nuclear cellular localization.

mouseNeu-PPI data was obtained from Murtaza *et al.* 2022 Table S1. We trimmed Table S1 to the 41 ASD genes, converted "Prey" interactors from mouse to human ontology (GRCm39 to

GRCh38.p13). We removed interactors that did not map to human genes and removed self-interactions, resulting in n = 807 unique interactors (original 1107). We also defined a set of "Mouse-PPI subset" interactors by trimming to interactors of the n = 17 hcASD bait.

We defined the mouse brain proteome set of background genes using data from Murtaza *et al.* 2022 Table S5. We started with the list of mouse brain proteome genes in sheet "Mouse Brain Protein List", converted the genes from mouse to human ontology (GRCm39 to GRCh38.p13), and removed genes that did not map to human genes. The final list consists of 8,678 genes (originally 11,992).

We conducted a Fisher's exact test (one sided, greater) to evaluate the overlap between ASD-PPI (100 baits) and mouseNeu-PPI (41 baits). We defined the universe of genes to be the intersection of the HEK293T proteome and mouse brain proteome (n = 7,049 genes). The 2x2 contingency table was defined by ASD-PPI interactor (yes/no) and mouseNeu-PPI interactor (yes/no). We conducted a secondary analysis in which we evaluated the overlap between interactors of only the matching n = 17 hcASD bait in mouseNeu-PPI and ASD-PPI. Statistical significance was calculated using Fisher's exact tests (one-sided, greater) and *p* values were adjusted for 2 tests (Bonferroni).

We evaluated whether mouseNeu-PPI interactors were enriched for de novo damaging variants from ASD probands. We conducted Fisher's enrichment tests as described in "Comparison of ASD-PPI and Pintacuda *et al.* 2023 iEN-PPI network" using all interactors from Murtaza mouse-PPI that map to human genes (792 interactors from 41 baits) (**Fig. 3.9D, S3.10F**).

Comparison of ASD-PPI and Gao et al. 2024 mouseCortex-PPI networks

Gao *et al.* 2024[25] report an ASD-relevant mouseCortex-PPI network generated via proximity-based endogenous proteomics in the developing mouse brain with 14 ASD genes (6 of which are hcASD).

We defined the Gao 2024 mouse brain proteome set of background genes using data from Supplementary Data S1, sheets "Batch_1_Normalized" through "Batch_9_Normalized"). For each batch, we trimmed to proteins identified by at least 1 peptide in LC-MS/MS and extracted gene symbols (column "Description"). We then mapped gene symbols from mouse to human ontology (R packages org.Hs.eg.db, org.Mm.eg.db, and Orthology.eg.db) and removed those that did not map to human genes. The final list consists of 4,727 genes (originally 5,040).

mouseCortex-PPI interactors were obtained from Table S2 (sheet "HiUGE-iBioID"). We mapped gene symbols from mouse to human ontology, removed those that did not map to human genes and self interactions. This resulted in n = 1,205 unique interactors (original 1,252). We also defined a subset of hcASD interactors by trimming to the interactors of the n = 6 hcASD bait (Ctnnb1, Lrrc4c, Scn2a, Shank2, Shank3, Syngap1).

We conducted a Fisher's exact test (one sided, greater) to evaluate the overlap between interactors from mouseCortex-PPI (14 baits) and from ASD-PPI (100 baits). We defined the universe of genes to be the intersection between the Gao 2024 mouse brain proteome and the HEK293T proteome (union of Bekker Jensen *et al.* 2017 and ASD-PPI interactors). The 2x2 contingency table was defined by ASD-PPI interactor (yes/no) and mouseCortex-PPI interactor (yes/no). We conducted a secondary analysis in which we evaluated the overlap between interactors of only the matching

n = 6 hcASD bait. Statistical significance was calculated using Fisher's exact tests (one-sided, greater) and *p* values were adjusted for 2 tests (Bonferroni).

We determined whether the mouse cortex-PPI interactors (n = 1,205 interactors from 14 baits) were enriched for damaging *de novo* mutations from ASD probands. Fisher's enrichment tests were conducted as described in "Comparison of ASD-PPI and Pintacuda *et al.* 2023 iEN-PPI network" (**Fig. 3.9D, Fig. 3.10G**).

We performed Breslow-Day tests comparing differences in OR for more ASD-associated genetic risk between external PPI dataset and ASD-PPI. This difference was not statistically significant for any of the datasets; Pintacuda 2023 iEN-PPI: $p = 0.90$; mouseNeu-PPI: $p = 0.15$; mouseCortex-PPI: $p = 0.24$.

Interactor expression in adult brain tissue (GTEx)

The Genotype-Tissue Expression (GTEx) project includes RNA-sequencing data from 54 non-diseased tissue sites across nearly 1,000 adult donors.[99] We downloaded GTEx v8 data (dbGaP Accession phs000424.v8.p2) from the GTEx website (https://www.gtexportal.org/home/datasets), which contains median gene-level TPM for samples from 54 tissue types, including 13 brain tissue sites (gtex/scripts/01_downloadGtexData.R).

We assessed whether ASD-PPI genes are more highly expressed in GTEx brain tissues compared to other HEK293T expressed proteins. We defined HEK293T expressed proteins as the union of proteins detected by global quantitative mass spectrometry[98] and ASD-PPI interactors. We defined genesets of interest to be baits, interactors (excluding hcASD102), and HEK293T expressed proteins that are not baits or interactors. For each geneset, we found the median geneset expression percentile within individual brain tissue samples and performed T-tests to assess whether these

were significantly different across genesets. *P* values were adjusted for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* x 3 tests; **Fig. 3.10A**; gtex/scripts/02_GTEx_asdPPI100_analysis.R).

We evaluated whether interactors are more highly expressed in GTEx brain samples compared to permuted genesets from the HEK293T proteome. We created 100,000 permuted genesets from a universe of n = 10,984 genes that are in the HEK293T proteome, measured in GTEx RNA-sequencing data, and not ASD-PPI baits. The probability of gene selection was weighted by HEK293T iBAQ scores[98] and permuted genesets were required to have a median HEK293T iBAQ rank within 1 quantile of that of ASD-PPI interactors. We calculated the median geneset expression rank for interactors (observed) and for the 100,000 permuted genesets (null distribution) in GTEx brain tissue samples, where higher rank reflects higher relative expression within a sample. We then calculated the permuted significance of the observed ASD-interactor GTEx expression rank within each tissue type. *P* values were adjusted for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* x 13 brain regions). We found that 3 brain tissues - cerebellum (*p*.adj = 0.0023), cerebellar hemisphere (*p*.adj = 0.0043), and cortex (*p*.adj = 0.626, *p* = 0.048) - had at least nominally significantly higher median genset rank for interactor genes compared to permuted genesets (**Fig. 3.10C**; gtex/scripts/02_GTEx_asdPPI100_analysis.R) .

hcASD and interactor expression in human brain tissue across development (BrainSpan RNAseq data)

The BrainSpan developmental RNAseq dataset ('bsRNAseq') profiled human brain tissue RNA expression across the full course of human brain development, from early prenatal stages through late adulthood.[20] BsRNAseq data (n = 524 samples) was downloaded from brainspan.org. The original 52,376 genes were trimmed to a final set of 18,552 protein-coding genes associated with

an HGNC symbol; if a HGNC symbol was associated with multiple Ensembl Gene IDs, only one was kept. Expression values were reported in reads per kilobase million (RPKM). (bsRNAseq/scripts/ 01_download_format_bsRNAseq.R)

Expression of ASD-PPI genes and other HEK293T proteins in prenatal brain tissue: We assessed whether ASD-PPI genes are expressed at relatively higher levels in BrainSpan prenatal samples compared to 293T background genes. We evaluated 3 genesets:

- hcASD: ASD-PPI bait (n = 78 evaluated in BrainSpan).

- Interactors - hcASD (interactors$^{\text{no hcASD}}$): ASD-PPI interactors, excluding hcASD (n = 1,030 evaluated in BrainSpan).

- Other: Genes in the HEK293T proteome[98] that are not in the ASD-PPI network (n = 9,862 evaluated in BrainSpan).

We restricted our analysis to the n = 237 prenatal brain samples as the prenatal period has been previously implicated in ASD.[22,100] For each geneset, we calculate the median geneset expression percentile within individual brain tissue samples. We performed T-tests to assess whether the median genset expression percentiles are significantly different between different genesets, adjusting *p* values for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * 3 comparisons; **Fig. 3.2A**; bsRNAseq/scripts/02_asdPPI_bsRNAseq_prenatal_expression.R).

Calculating the relative expression of ASD-PPI genes compared to permuted genesets: We created 100,000 permuted genesets from HEK293T proteome genes, matching the number of interactor genes. To generate the permuted genesets, we selected genes from a set of n = 10,902 genes that are 1) in the HEK293T proteome[98] (*98*) (described above), 2) measured in BrainSpan RNAseq,

and 3) not an ASD-PPI bait. The probability of gene selection was weighted by HEK293T protein expression level (median iBAQ rank), and only genesets with a median iBAQ rank within 1 quantile of the median iBAQ rank of ASD-PPI interactors were retained. We calculated the median geneset rank ('medRank') within each bsRNAseq brain tissue sample for the interactors and each of the 100,000 permuted genesets (higher medRank reflects higher geneset expression within a sample). For each brain tissue sample, we calculated the 'normalized medRank', defined as the observed medRank – median (100,000 permuted medRanks). The normalized medRank reflects the sample-level expression enrichment of interactors, with positive values reflecting higher than expected expression. As a comparator, we also calculated the sample-level expression enrichment of hcASD102 genes[4] compared to permuted genesets. (03_asdPPI_bsRNAseq_permutations.R)

Comparing expression levels of ASD-PPI baits and interactors in brain samples across development: To assess whether the relative expression of interactors$^{\text{no hcASD}}$ mirror that of hcASD102 genes across individual brain samples, we calculated the Spearman and Pearson correlations of the normalized medRanks for interactors$^{\text{no hcASD}}$ versus hcASD102 genes across n = 524 brain samples (Spearman rho = 0.884, Pearson R = 0.897. To compare relative expression in prenatal versus postnatal brain, samples were grouped as prenatal (n = 237) or postnatal (n = 287), and the normalized medRanks of interactors$^{\text{no hcASD}}$ or hcASD102 were compared using a two-sided t-test, both showing significantly higher expression in prenatal samples (p < 2.2e-16 for both). To evaluate expression correlation between interactors$^{\text{no hcASD}}$ and hcASD102 across development, bsRNAseq brain samples were grouped by developmental period (as defined in Kang *et al.* 2011[20]), with periods 1–7 representing prenatal stages and 8–15 spanning late infancy to adulthood, resulting in 13 groups (periods 2–14). For each period, the median (normalized medRank) was calculated for each geneset across samples, and correlations across periods were

computed (Spearman ρ = 0.946, Pearson R^2 = 0.81), indicating high concordance between interactor and hcASD102 expression across developmental periods (**Fig. 3.2, B and C, Fig. 3.10D**; 04_make_bsRNAseq_asdPPI_hcASD102_permutation_plots.R).

Selection coefficients for ASD-PPI genes

We evaluated whether ASD-PPI genes are more evolutionary constrained than expected by assessing several metrics, including the pLI (probability of being intolerant of a single loss-of-function variant), misZ (missense Z score, measures gene intolerance to missense variation), synZ (synonymous Z score, measures gene intolerance to synonymous variation, used as negative control), and s_het (selective effect for heterozygous PTVs).[101,102]

We obtained pLI, misZ, and synZ scores from the Genome Aggegation Database (ExAC dataset, https://storage.googleapis.com/gcp-public-data--gnomad/legacy/exac_browser/forweb_cleaned_exac_r03_march16_z_data_pLI_CNV-final.txt.gz) and s_het scores from Cassa *et al*. 2017 Supplementary Table 1.[102] We assessed whether the pLI, misZ, s_het, and synZ scores of ASD-PPI bait genes, interactor genes, and 293T proteome genes[98] were significantly different from ASD-PPI baits, interactors (excluding hcASD102), and other HEK293T proteome genes using t-tests, adjusting *p* values for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * number of t-tests) (**Fig. 3.10D**; 02_eval_asdPPI_geneConstraintMetrics.R).

Interactor enrichment for ASD genetic risk

If ASD-PPI has successfully identified ASD network genes that are ASD-relevant, interactor genes should be enriched for ASD genetic risk. We used *de novo* genetic variants previously identified from simplex family studies as a measure of ASD genetic risk[4] and conducted enrichment tests

(Fisher's exact) to assess whether interactors are enriched for ASD genetic risk compared to other genes in the HEK293T proteome or other exome genes.

We focused on *de novo* damaging variants identified from the Simon's Simplex Collection (Satterstrom *et al.* 2020 Supplementary Table 1). We defined "damaging" variants to be variants that resulted in a damaging missense mutation (Polyphen Mis3 (damaging) or MPC MisB (MPC≥2)) or PTV (frameshift, stop gained, or canonical splice site disruption; ssc_enrichment/scripts/01_format_satterstrom2020_SSC_variants.R.

We determined the association between having a damaging variant in different genesets of interest with ASD status. We defined the following genesets:

- Interactors: ASD-PPI interactors (n = 1,074)
- Interactors (-hcASD102): Interactors, excluding hcASD102 (n = 1,043)
- HEK293T (-Interactors, -hcASD102): HEK293T proteome genes, excluding interactors and hcASD102 (n = 10,045)
- Exome: all autosomal genes measured in Satterstrom *et al.* 2020 (n = 17,332)
- Exome (-hcASD102): autosomal genes measured in Satterstrom *et al.* 2020, excluding hcASD102 (n = 17,230)
- Exome (-Interactors, -hcASD102): autosomal genes measured in Satterstrom *et al*. 2020, excluding hcASD102 and interactors (n = 16,234)

For each Fisher's enrichment test, the universe of genes was defined to be the geneset members, and the 2x2 contingency table was: X, the number of ASD probands with a damaging variant in at

least one gene in the geneset; Y; the number of control siblings with a damaging variant in at least one gene in the geneset; Z, the number of ASD probands with no damaging variants in any gene in the geneset; W, the number of control siblings with no damaging variants in any gene in the geneset. We calculated the enrichment odds ratio with Fisher's exact test (one sided, alternative = greater), and adjusted *p* values for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * number genesets assessed; **Fig. 3.2D**; ssc_enrichment/scripts/03_asdPPI_geneticRisk_Satterstrom2020_SSC.R).

Downsampling analyses: To determine whether increasing the number of baits used to create ASD-PPI is associated with increased ability to identify interactors associated with ASD genetic risk, we downsampled the ASD-PPI network. Specifically, we randomly selected sets of baits (ranging from n = 1-100 baits, 1000 iterations for each set size) and trimmed the ASD-PPI network to include only interactors associated with the downsampled baits. For each downsampled network, we calculated the association between having a damaging variant in different genesets of interest with ASD status as described above. We evaluated 2 genesets: 1) Interactors (-hcASD102), and 2) Exome (-Interactors, -hcASD102), which allows us to compare the amount of ASD-associated genetic risk in interactors versus the remaining genes in the human exome. For each bait set size, we calculated the median and standard deviation of ORs across the 1000 iterations, as well as the median and standard error of *p* values across the 1000 iterations (**Fig. 3.2F**; ssc_enrichment/scripts/04_asdPPI_prey_ASDrisk_downsamplingAnalysis.R).

We additionally assessed whether increasing the number of baits used to create ASD-PPI is associated with increased ability to identify interactors that are novel hcASD genes. We defined novel hcASD genes as the n = 255 genes with FDR < 0.1 (hcASD+) in the latest omnibus WES study.[6] We downsampled the ASD-PPI network as described above. For each downsampled

network, we 1) counted the number of hcASD+ genes in the interactors and 2) conducted a Fisher's exact test (one sided, greater) to assess for enrichment of hcASD+ in interactors. The counts of the two-by-two contingency tables were: X, the number of hcASD+ that are interactors; Y, the number of hcASD+ genes that are not interactors; Z, the number of interactors that are not in hcASD+; and W the number of genes that are not interactors or hcASD+. We defined the universe of possible genes to consist of those with 1) proteins that are detected in the HEK293T proteome[98] or ASD-PPI interactors); 2) measured in Satterstrom *et al.* 2020[4]; and 3) measured in Fu *et al.* 2022.[6] For each bait set size, we calculated the median and standard deviation of OR, -log10(pvalue), and number of hcASD+ genes among interactors across the 1,000 iterations. We repeated this analysis after excluding hcASD102 genes (Satterstrom *et al.* 2020[4], FDR <0.1) from the gene universe to assess the ability of ASD-PPI to identify truly novel hcASD genes (**Fig. 3.2G**; risk_gene_enrichment/scripts/02_asdPPI_downsampling_hcASD255discovery.R).

Interactor enrichment for ASD, DD, or SCZ risk genes

We evaluated whether interactors are enriched for risk genes that have been implicated in ASD, DD, or schizophrenia. We defined several sets of risk genes, including:

- ASD, Fu *et al.* 2022[6]: n = 255 autosomal genes with TADA FDR <0.1 (hcASD+)
- ASD, Trost *et al.* 2022[5]: n = 134 autosomal genes with TADA FDR<0.1
- ASD, Zhou *et al.* 2022[7]: n = 72 genes with study-wide significance (based on 5,754 constrained genes, p<8.69E-6)
- ASD, SFARI[103]: SFARI genes are a database that endeavors to include all genes associated with ASD risk, regardless of the level or nature of evidence. SFARI genes are further divided into categories, including syndromic, category 1 (high confidence), category 2

(strong candidate), and category 3 (suggestive evidence). SFARI genes were downloaded from gene.sfari.org (09/02/2021 release), and included a total of n = 1,020 genes, of which there were n = 230 syndromic genes, n = 92 syndromic and category 1 genes, n = 206 category 1 genes, n = 219 category 2 genes, and n = 514 category 3 genes.

- DD, Kaplanis 2022[104]: n = 285 genes that are significantly associated with DD after one-sided Bonferroni correction
- SCZ, Singh 2022[26]: n =3 4 genes with TADA FDR<0.1.

We defined the HEK293T expressed genes to be the union of interactors and proteins that were detected by global quantitative mass spectrometry.[98] We defined two sets of interactors: the full set of ASD-PPI interactors ('Interactors', n = 1,074) and ASD-PPI interactors excluding hcASD ('Interactors$^{\text{no hcASD}}$, n = 1,043). We conducted two sets of enrichment tests. For the first set of enrichment tests, we assessed only risk genes from genetic studies that reported the set of genes that were evaluated as possible risk genes.[5-7,26] We defined the universe of possible genes to be those that were 1) HEK293T expressed genes; 2) measured in Satterstrom *et al.* 2020,[4] and 3) measured in the genetic study that generated the risk gene set of interest[5-7,26] (**Fig. 3.2E**). For the second set of enrichment tests (**Fig. 3.10G**), we simply defined the gene universe to be HEK293T expressed genes, as the exact gene universe considered is unavailable for the SFARI risk genesets. For both sets of enrichment tests, we conducted Fisher's exact tests (one sided, greater), in which the contingency tables were set up by interactor status (yes/no) and condition geneset status (yes/no). We corrected *p* values for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * number of genesets assessed). **(Fig. 3.2E, Fig. 3.10G**;

risk_gene_enrichment/01_asdPPI_preyOnly_riskGeneEnrichment.R; data/geneAnnotation_hek293T_ASD_DD_SCZ.csv).

Interactor enrichment for ASD common variant genetic risk

We evaluated whether ASD-PPI interactors are enriched for ASD common variant risk as reported by Grove *et al*. 2019[105] by applying geneset level Multi-marker Analysis of GenoMic Annotation (MAGMA[106]) to ASD-PPI bait, interactors, and interactors$^{no\ hcASD}$. We found no significant enrichment in any geneset.

Network co-expression in prenatal brain cells

We compared relative network co-expression in cells from three prenatal brain atlases. Briefly, for each cell type within a prenatal brain atlas, we evaluated the observed coexpression of network gene pairs and normalized this value by a background coexpression value to adjust for differences in the expected coexpression of genes present in the context from which the network was derived (i.e. for the 100 hcASD network we used all gene pairs present in the proteome of HEK293T cells[98] and for the iEN network we used all gene pairs present in the iEN proteome[23]). We then compared the distribution of observed network co-expression across cell types using Wilcoxon rank sum tests.

We defined 2 sets of networks of interest:

1. Set 1: ASD-relevant PPI networks from experimental data and STRING[28]
    a. ASD-PPI (HEK) hcASD-int: 1,143 nodes and 1,879 edges.
    b. Pintacuda 2023 (iEN) hcASD-int: 1034 nodes and 1343 edges. 1343 index-interactor interactions identified for 13 hcASD index genes from iEN cells in

Pintacuda *et al.* 2023. Interactions were extracted from Supplemental Table S3 ('Interaction_Lists' sheet, selected rows where IsInteractor column is 'T' and Index protein column was not 'COMBINED').

c. ASD-PPI (HEK) hcASD-int & int-int: 1143 nodes and 10,191 edges. Union of interactions from ASD-PPI and STRINGv11.5 (experimental score >0) for genes in ASD-PPI

d. Pintacuda 2023 (iEN) hcASD-int & int-int: 1034 nodes and 11,706 edges. Union of interactions from Pintacuda *et al.* 2023 and STRINGv11.5 (experimental score >0) for genes in Pintacuda 2023 (iEN) hcASD-int.

2. Set 2: ASD-relevant PPI networks trimmed to direct interactions supported by AF

   a. hcASD-int (AF iPTM > 0.5): 291 nodes and 284 edges. ASD-PPI hcInteractions with AF iPTM>0.5

   b. hcASD-int & int-int (AF iPTM>0.5): 933 nodes and 2,350 edges. Union of and ASD-PPI hcASD-int and int-int interactions with AF iPTM>0.5.

We used data from three prenatal brain cell atlases:

1. The Nowakowski *et al*. 2017[27] scRNAseq data (n = 4,261 cells) was downloaded from https://cortex-dev.cells.ucsc.edu/. The original set of n = 56,864 genes was trimmed to a final set of 18,803 genes by keeping only protein coding genes associated with a HGNC symbol. If a HGNC symbol was associated with multiple Ensembl gene IDs, only the first was kept. Expression values were not altered from the original data download and are in

the form of TPM. We grouped cells into the 19 cell types defined in Satterstrom *et al.* 2020[4] and excluded unassigned cells from the analysis.

2. The Polioudakis *et al*. 2019[52] scRNAseq data (n = 33,976 cells) was downloaded from http://solo.bmap.ucla.edu/shiny/webapp/. Genes were trimmed to a final set of 16548 genes by keeping only protein coding genes associated with a HGNC symbol. If a HGNC symbol was associated with multiple Ensembl gene IDs, only the first was kept. Expression values were not altered from the original data download and are in the form of UMI counts. Cell types were defined in Polioudakis *et al*. 2019, and consisted of 16 cell types: vRG, ventricular radial glia; oRG, outer radial glia; PgS, cycling progenitors (S phase); PgG2M, Cycling progenitors (G2/M phase); IP, intermediate progenitor; ExN, migrating excitatory; ExM, maturing excitatory; ExM-U, maturing excitatory upper enriched; ExDp1, excitatory deep layer 1; ExDp2, excitatory deep layer 2; InMGE, interneuron MGE; InCGE, interneuron MGE; OPC, oligodendrocyte precursor; End, endothelial cell; Per, pericyte; Mic, microglia.

3. The Bhaduri *et al*. 2021[53] scRNAseq data (neocortex, GW15-25, n = 404,218 cells) was downloaded from https://cells-test.gi.ucsc.edu/?ds=dev-brain-regions+neocortex. Cell types were defined using the "ConsensusCellType-Final" column in the meta.tsv file (n = 12 cell types). We excluded 'Excitatory Neurons' from our analyses as this cluster contains few cells (n = 61) and this cluster is not included in the Bhaduri *et al*. 2021 main text Figure 2.

scRNAseq data is limited by dropouts and resultant sparse and heterogeneous gene expression across cells. Therefore, we define gene pairs to be 'co-expressed' in a cell if there is at least 1 read

of each gene detected. We define the co-expression between gene pairs as the proportion of cells in a cell type that co-express both genes (range 0-1).

We defined two sets of background genes

1. HEK: 11,185 genes in the HEK293T proteome (union of Bekker Jensen 2017[98] and ASD-PPI genes). We used this background for ASD-PPI based networks.

2. iEN: 3,274 genes in the iEN proteome as defined in Pintacuda *et al.* 2023[23] (Pintacuda *et al.* 2023 Supplemental table 3, sheet 'Interaction_Lists', 'Gene symbol' column). We used this background for Pintacuda *et al.* 2023[23] based networks.

For each cell type context, we determined the average background co-expression by calculating the connectivity between all possible gene-gene pairs. We use this background co-expression as a normalization factor to remove biases that have to do with global differences in co-expression in each cell type. Let i and j denote genes and k denote cell type context. Define r_{ijk} as the co-expression between genes i and j in cell type k, and define $avg(r_k)$ as the average co-expression over all pairs of background genes measured in cell type k. Then define a normed co-expression as:

$$c_{ijk} = \frac{r_{ijk}}{avg(r_k)avg(r_k)}$$

Within each prenatal brain atlas, for each cell type, we calculated the normed co-expression of all observed network edges. We then evaluated whether the distribution of observed network co-expression within each individual cell type was significantly different from the base distribution (concatenated observed network co-expression of all cell types) using a two-sample Wilcoxon rank sum test (R wilcox.test). We used the Wilcoxon estimator for the difference in location as a

measure of relative difference in network co-expression between in the cell type versus base distribution). Within each single cell atlas, *p* values were corrected for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * number of cell types * number of networks evaluated) (**Fig. 3.10I and Fig. 3.12H**).

For the Nowakowski 2017 dataset, which included cells from both prefrontal cortex and ventral telencephalon and therefore includes both excitatory and inhibitory progenitors and maturing neurons, we projected the Wilcoxon location difference estimator onto t-SNE maps to visualize cell type–specific network co-expression. T-SNE coordinates were downloaded from cortex-dev.cells.ucsc.edu (**Fig. 3.2, H and I and Fig. 3.4H**).

Network layout of baits and interactors

The total network showing AP-MS detected interactions between 100 hcASD baits and their connections to 1,043 prey plus connections to other baits was laid out using t-SNE (R package Rtsne) on a weighted combination of distance based on shared Gene Ontology annotations (weight = 4) and distance based on shared AP-MS connections (weight = 1). Binary distances, equivalent to 1-Jaccard similarity, were used for both distances. For GO distances, to create a limited set of 141 GO terms that maximally cover and describe the set of 1043 interactors, the GO terms were limited to those significantly (FDR < 0.05) enriched in the full set of 1043 interactors and further limited to only those terms that were the top-enriched term (by *p* value) per at least one interactor. Enrichment was computed and scored by the function enricher in the R clusterProfiler package using GO annotations from all GO in org.Hs.eg.db (R package) and limiting GO terms to those with at least 20 and at most 500 genes. Interactors were assigned to terms according to their GO annotations. Baits were assigned to these GO terms based on significant (*p* value < 0.05) enrichment in their set of interactors using the set of 1043 as background (universe), or when no

significant enrichment existed, simply the one term (of 141) with highest number of interactors per bait. For network distances, a symmetric adjacency matrix was built where each bait and interactor were connected to itself and to each protein that it connects to by a direct AP-MS interaction, and binary distances were computed on rows of the matrix. After layout by t-SNE (default settings except is_distance = TRUE and theta = 0.0), protein two dimensional coordinates were adjusted to avoid overlapping nodes using an iterative, repulsive algorithm. The final image was formatted and drawn using the R package ggplot2, mostly with functions geom_point and geom_segment. Coloring of interactors was based on a subset of the GO terms chosen to best cover the 1043 interactors in a non-redundant fashion with a small number of terms. For this term selection, an additional GO enrichment run was completed, allowing terms to have at most 2000 genes, and including additional terms that described interactors previously undescribed by the 141 terms. Clustering rows and columns of the annotation matrix, 1043 genes x 186 terms, was performed to aid in this manual selection. In the network view, where a gene is annotated to more than one of the chosen terms, it is assigned the color whose median 2D location on the plot it is closest to (**Fig. 3.3A**).

Comparing PPI network convergence

Convergence in an AP-MS network, the interaction of multiple baits with the same interactor, was directly measured by the portion of all possible paired baits that had a statistically significant ($p < 0.05$) number of overlapping (convergent) genes as measured by a hypergeometric test with a background size of 11,169 proteins (the number of proteins in the HEK293T proteome[98]). To calculate a baseline, considering a study's network size and degree distribution of the baits, we used random samples of baits from BioPlex to match each study. For each actual bait in a study, a random bait was chosen from BioPlex that had an identical number of interactors (degree), or from

among the smallest balanced window of degree around the actual degree that included at least 100 different baits. 1,000 randomly assembled networks were used for each study, and the full all-pair hypergeometric tests were done per randomly assembled network. The portions with $p < 0.05$ on the 1,000 random subnetworks of BioPlex were used to calculate a standard deviation and mean from which we scaled these portions to Z scores (**Fig. 3.3B**). Additional measures directly or indirectly related to convergence, including edge density, average node degree, and modularity were evaluated in a similar fashion and converted to Z scores. The igraph package in R was used for network calculations including the function edge_density and the combination of functions cluster_fast_greedy and modularity to measure the maximum possible modularity per network. (**Fig. 3.11B**)

Comparison of interactor overlap between bait pairs predominantly associated with ASD or ASD/NDD

Satterstrom *et al.* 2020[4] categorized hcASD into those that were more frequently mutated in ASD (ASD-predominant; ASD_P) and those that were more frequently mutated in NDD (ASD_{NDD}). If ASD_P vs. ASD_{NDD} genes have separable molecular functions, we would expect ASD_P baits to have greater interactor overlap with other ASD_P baits (and the same for ASD_{NDD} bait pairs). We categorized ASD-PPI baits as ASD_P or ASD_{NDD} as defined in Satterstrom *et al.* 2020.[4] We evaluated the interactor overlap for the following four groups of bait pairs: 1) all ASD-PPI bait-bait pairs, 2) ASD_P - ASD_P bait pairs; 3) ASD_{NDD} - ASD_{NDD} bait pairs; and 4) ASD_P - ASD_{NDD} bait pairs. We determined the significance of interactor overlap using Fisher's exact tests (one sided, greater), where the 2x2 contingency table was defined by: bait #1 interactor (yes/no) and bait #2 interactor (yes/no). The universe of genes was defined to be the HEK293T proteome. For each of the four baits groupings, we calculated the proportion of bait pairs whose interactors had

nominally significant overlap ($p < 0.05$). We additionally used a Kruskal-Wallis rank-sum test to evaluate whether the distribution in bait-bait interactor overlap *p* values is significantly different across the 4 groups of bait pairs (R kruskal.test). We conducted these analyses using a trimmed dataset that includes only baits with fewer than 45 interactors (excludes 9 baits). Analyses using the full ASD-PPI dataset showed similar findings (data not shown). (**Fig. 3.11, B and C**)

Generating a random set of ASD-PPI bait-int

We generated a set of random interactors to use as a baseline benchmark for the ASD-PPI AF[41] pairwise interaction analysis. For each of the baits in ASD-PPI, we generated a set of random interactors that matched true interactors in protein size and for which odds of selection were weighed by HEK293T protein expression levels. Specifically, we defined HEK293T protein expression level by average iBAQ expression (Bekker-Jensen *et al*. 2017[98] Supplementary table 7, average of iBAQ scores for the two HEK293T experimental replicates) and used cDNA length reported in Satterstrom *et al.* 2020 as a proxy for protein size[4] Supplementary Table 2, "Autosomal" sheet, "cDNA" column). We divided the roughly 10,000 HEK293T expressed proteins into 5 groups based on cDNA length (cDNA_quantile). For each observed ASD-PPI interactor, we selected a random interactor from HEK293T-expressed proteins in the same cDNA quantile, with probability of selection weighted by HEK293T protein expression level. Random interactors were required to be unique at a bait level but could be duplicated across different baits.

Running AlphaFold to predict direct pairwise interactions

Pairwise protein interactions and protein complexes were predicted using AF (*107*) (AlphaFold version 2.3.1) via ColabFold[108] (version 1.5.2). As input to protein structure prediction, sequences for each protein pair were extracted from uniprot (https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteom

es/Eukaryota/UP000005640/UP000005640_9606.fasta.gz). In all cases, we utilized the WT protein sequence even when examining interactions defined by mutant-specific baits. Multiple sequence alignments (MSA) were generated from ColabFold using MMseqs2[41] (Release 14), with the following databases: Colabfold Env DB (2021_08), UniRef30 (2021_03). All input complexes were run without templates or amber relaxation for 10 recycles and with at least three random seeds. For bait-prey interactions three model replicates were run for each seed, bringing the total number of replicates to at least nine. Only one model per seed was run for the int-int set, so only three models per pair were produced. Some inputted sequences failed to return structures due to memory limits (particularly for large complexes) or MSA failures. The five different sets of pairs modeled were: bait-int produced 1,654 structures of 1,879 attempted pairs, bait-random produced 1,677 for 1,862, bait-int-mutant (those pairs only observed in the mutant network) produced 126 for 148 pairs, bait-int-Pintacuda produced 1,133 for 1,303, and int-int produced 32,890 for 35,897.

From each replicate, multiple statistics were generated. AF predicted scores (pTM, ipTM, pLDDT) were extracted directly from the AF output json and pdb files. ipTM represents a predicted measure of topological accuracy in the interface.[41] The Confidence score is a measure which includes interface and overall topological accuracy using a weighted combination of pTM and ipTM, 0.8 x ipTM + 0.2 x pTM. The pDockQ score is based on parameters fit to match true DockQ values for simulated protein complexes with resolved crystal structures, and it was calculated as described previously.[42] We assume that any observed difference in the proportion of high AF quality scores in the bait-int vs. bait-random sets results from AF modeling the direct interactions in the bait-int set more successfully. We used this to find the AF score with the greatest separation in distribution between bait-int and bait-random sets across nine generated AF models per PPI pair (**Fig. 3.4B, Fig. 3.12, A to C**).

Searching Protein Data Bank (PDB) for homologous complex structures

As a proxy for possible presence of a complexed structure for any pair in the AF training set, we searched the full PDB for structures with detectably similar sequences. BLAST+ v2.15 was used to run blastp with default parameters against a custom database built based on all sequences deposited to PDB (downloaded on 11/17/2023). Possible matches were assembled in an iterative process over all sequences in any AF pair. For each protein, all individual chains that return a BLAST match were identified. For each matching chain, the other chains within each corresponding PDB entry were then queried to identify the subset that matches to a protein that forms an AF pair to the starting protein**.**

Visualizing protein structure files

ChimeraX or PyMOL were used to generate all 3D figures of protein structures.

Calculating distances from mutations to PPI interface

Inter-chain distances in an AF model were calculated using the center-center inter-atom distances between the three-dimensional coordinates of atoms in the AF structure file. Per residue, we used the minimum distance across all atoms to any atom in the opposite chain with pLDDT at least 25.0. Only WT sequences were modeled, so the distance is from any atom in the WT residue (**Fig. 3.16A**).

Mammalian cell-based NanoLuc two-hybrid (N2H) assay

For N2H, the hcASD proteins were expressed as fusions with the F1 fragment of nanoluciferase, either at their N-terminus (N1-hcASD) or at their C-terminus (hcASD-C1). Potential interactors (int) and random prey (random) were expressed in fusion with the F2 fragment fused at their N-terminus only (N2-int or N2-random). All cDNA ORFs (ASD, int, and random) were cloned in

mN2H-dedicated vectors by Gateway reaction and sequence verified. Empty N2H vectors, expressing the unfused NanoLuc fragments, were used to determine the background luminescence generated by each protein partner separately. HEK293T cells were seeded at $6x10^4$ cells per well in 96-well, flat-bottom, cell culture microplates, and cultured in 70 µl of phenol red-free DMEM supplemented with 10% fetal calf serum at 37°C and 5% CO_2. Twenty-four hours later, cells were transfected with 100 ng of each N1/C1 and N2-expressing plasmids using linear polyethylenimine. Twenty-four hours after transfection, 30 µL of 100× diluted NanoLuc substrate (Promega, N-1110) was directly added to each well. Luciferase enzymatic activity was measured using a CentroXS luminometer (Berthold; 2-second integration time). For each protein pair (ASP-int or ASD-random) a Normalized Luminescence Ratio (NLR) is calculated as follows:

$$NLR = \frac{L_{ab}}{L_{a0} + L_{b0}}$$

Where L_{ab} is the luminescence when hcASD fusion and potential interactor fusions are co-expressed, L_{a0} is the luminescence when hcASD fusion is co-expressed with its complementary unfused NanoLuc fragment, and L_{b0} is the luminescence when the potential interactor fusion is co-expressed with its complementary unfused NanoLuc fragment. hcASD-int pairs were scored (considered) positive when their NLR was greater than 5.0, a threshold chosen to separate the bulk of the distribution of hcASD-random pairs distribution from the long tail of the hcASD-int distribution. (**Fig. 3.12, D and E**).

Evaluating residue-level evolutionary conservation with ConSurf

We used the ConSurf web server (https://consurf.tau.ac.il/consurf_index.php) to estimate the evolutionary conservation of amino acid positions for DCAF7 and DYRK1A based on the

phylogenetic relations between homologous sequences. We ran ConSurf on each protein separately with default parameters (one HMMER iteration, maximum 150 homologs, E-value cutoff of 0.0001). The sequences found were clustered by their level of identity using CD-HIT and the cutoff specified to be between 35% to 95% identity) (**Fig. 3.4F**).

Effect of DYRK1A$^{\Delta 80\text{-}100}$ on DCAF7 and FAM54C interactions

We expressed Strep-tagged DYRK1A or DYRK1A$^{\Delta 80\text{-}100}$ or empty vector (control) in HEK293T cells and conducted AP-MS as discussed above (3 replicates each for control and DYRK1A; 2 replicates for DYRK1A$^{\Delta 80\text{-}100}$). We calculated SAINT scores (spectral counts). We defined significant interactions to be those with SAINT 1-BFDR ≥0.95. We evaluated the strength of binding (log_2FC(DYRK1A/control) spectral counts) of DYRK1A or DYRK1A$^{\Delta 80\text{-}100}$ to DCAF7 and FAM54C (**Fig. 3.12E**).

Prioritizing ASD-PPI interactors by number of hcASD interactions and shared interactors with hcASD

We were interested in ASD-PPI interactors that interact with large numbers of hcASD genes as they may be members of molecular complexes upon which hcASD genes functionally converge. We defined a set of n = 7 ASD-PPI interactors ("prioritized interactors") that interacted with at least 8 baits/hcASD102 genes (**Fig. 3.5A**; asdppi_prioritizedPrey/scripts/ 01_asdPPIprey_topDegreeBait.R).

We assessed whether BioGRID interactors of these prioritized interactors are enriched for hcASD genes. Interactors were defined using BioGRID human physical multi-validated interactions.[54] The file "BIOGRID-MV-Physical-4.2.191.tab2.txt" was downloaded from https://downloads.thebiogrid.org/BioGRID/Release-Archive/BIOGRID-4.2.191/.

"Official.Symbol.Interactor.A" and interactors were defined using "Official.Symbol.Interactor.B". Interactions were trimmed to human data (Organism.Interactor.A and Organism.Interactor.B = 9606). Bait and interactor symbols were updated using limma:alias2GeneSymbolTable. For each of the n = 7 prioritized interactors, we conducted one-sided Fisher's exact tests (greater) to assess whether interactors associated with individual baits are enriched for hcASD+ genes.[6] We restricted the gene universe to genes that were measured in Fu et al. 2022. The counts of the two-by-two contingency table were: X the number of interactors that are hcASD; Y, the number of interactors that are not hcASD; Z, the number of not.interactors that are hcASD; and W, the number of not.interactors that are not hcASD. We corrected *p* values for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * 7 genes). We found that DCAF7 BioGRID interactors to be nominally enriched for hcASD+ (**Fig. 3.5B**; asdppi_prioritizedPrey/scripts/02_asdPPIprey_bioGRIDinteractors_hcASD255enrichment.R).

We were interested in identifying hcASD with interactomes that overlap with that of DCAF7. Therefore, we conducted two-sided Fisher's exact tests to evaluate whether there is significant overlap between DCAF7 interactors and the interactors of each of the 8 hcASD that bound to DCAF7 in ASD-PPI. We restricted the gene universe to the HEK293T proteome. The counts in the 2x2 contingency table were: DCAF7 interactor in BioGRID (yes/no) and hcASD interactor in ASD-PPI (yes/no). We corrected *p* values for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * 7 genes) (**Fig. 3.5D**).

Endogenous AP-MS of DYRK1A in iPSC-derived excitatory neurons

Two independent sgRNAs targeting the endogenous DYRK1A locus were designed (as below) and evaluated for cutting efficiency. A ~200nt ssODN containing a 2×Strep tag served as the HDR template. Human KOLF2.1J iPSCs (Jackson Laboratory, JIPSC001000) were transfected with

CRISPR components using Lipofectamine Stem Transfection Reagent (ThermoFisher Cat # STEM00003) according to the manufacturer's protocol. Single cell clones were isolated using a Pala Cell Sorter and Single Cell Dispenser (Bio-Techne, Cat # NI006). Single-cell clones were isolated and screened by genotyping to identify lines carrying a heterozygous 2×Strep tag insertion at the DYRK1A locus.

Sequences for gene editing:

sgRNA1: AATGTAGTCACGAGCTAGCTA

ssODN:

TAGAGAAGAGTCCCCCATGACAGGAGTTTGTGTGCAACAGAGTCCTGTAGCTAGCTC

GtggagccatccacagttcgaaaaaggtggaggttctggcggtggatcaggtggaagtgcatggtctcaccctcagtttgagaaaTG

ACTACATTGAAACTTGAGTTTGTTTCTTGTGTGTTTTTATAGAAGTGGTGTTTTTT

Validated clones were transfected with a PiggyBac construct expressing doxycycline-inducible NGN2 and puromycin resistance (Addgene, Cat #172115). Following puromycin selection, NGN2-expressing iPSCs were differentiated into glutamatergic neurons as previously described.[109] Briefly, iPSCs were plated on matrigel in pre-differentiation medium consisting of Knockout DMEM/F12 (ThermoFisher, 10829-018), N2 (1X) (Thermo Fisher, 17502-048), NT3 (10ng/ml) (PeproTech, 450-03), BDNF (10ng/ml) (PeproTech, 450-02) and laminin (1µg/ml) (Thermo Fisher, 23-017-015) along with ROCK inhibitor (10 µM) and doxycycline (2µg/ml) for 72h. Cells were then dissociated with Accutase (Stem Cell Technologies, #07920) and replated on poly-D-lysine/laminin–coated surfaces in differentiation medium containing DMEM/F12 and Neurobasal-A (Gibco, #10888022), N2 (0.5X), GlutaMax (0.5X)(Gibco, #35050061), NT3 (10ng/ml), BDNF (10ng/ml), B27 (1X)(Gibco, #12587010) and laminin. Media was refreshed

weekly and neurons were harvested on day 21 post differentiation for AP-MS as described above (**Fig. 3.13B**).

Sequential AP-MS

Sequential AP-MS of Strep-DYRK1A with either Flag-DCAF7 or Flag-KIAA0232 was carried out as noted previously.[110] Briefly, a 15 cm dish with 10 million HEK293T cells (70-80% confluency, 3 replicates each) were transfected with 10ug of Strep and Flag tagged plasmids using PolyJet as described above. After 48hr of transfections, cell pellets were collected and lysed as described above. Following cell lysis, affinity purification was carried out using anti-Flag IP beads (Sigma #M8823), and proteins were eluted from the flag beads using 0.3mg/ml 3X flag peptide. The eluted proteins were then incubated with MagStrep type 3 beads as described before followed by onbead digestion and data acquisition mass spectrometry and analysis as reported before (**Fig. 3.13E, Fig. 3.13C**).

Enrichment of ASD genes among DCAF7-DYRK1A shared interactors

Co-expression of FLAG-tagged DCAF7 and Strep-tagged DYRK1A in HEK293T cells followed by sequential (double) AP-MS identified 126 shared interactors between DYRK1A and DCAF7 (see above). We evaluated whether these shared interactors are enriched for the n = 255 hcASD+[6] using a two-sided Fisher's exact test. The setup of the 2x2 contingency table was DCAF7-DYRK1A shared interactor (yes/no) and hcASD 55(yes/no). The gene universe was restricted to the HEK293T proteome (union of Bekker-Jensen 2017, ASD-PPI interactors and single/doubleIP interactors, n = 11,180 genes). (**Fig. 3.13D**, asdppi_prioritizedPrey/scripts/03_DCAF7interactors_hcASDenrichment_FET.R)

Microscopy for DYRK1A, DCAF7, KIAA0232 colocalization in HEK293T and NPCs

HEK293T cells were transfected with DYRK1A, DCAF7, KIAA032 plasmids using Lipofectamine 3000 (Invitrogen, #L3000001) and fixed after 30h with 4% paraformaldehyde (EMS, #50980487). After permeabilization with 0.1% TritonX-100, cells were stained with anti-Flag (1:1000, Cell Signaling, 14793), anti-Strep (1:1000, IBA, 2-1507), anti-α-Tubulin (1:500, Santa Cruz, sc-53029) and anti-DYRK1A (1:500, R&D systems, AF5407). Secondary fluorescence-conjugated antibodies were used at 1:1000 (Abcam ab150177, Abcam ab150159, Thermo Fisher A31553, Thermo Fisher A32732). Cells were imaged on a Zeiss 980 LSM confocal microscope with fast airyscan with a 63X oil objective and processed in ImageJ and Adobe Illustrator (**Fig. 3.13E**).

WTC11 NPCs were maintained at 37°C in complete media consisting of DMEM/F-12 (Gibco, 11320033), 1X N2-Supplement (Gibco, 17502048), 1X B-27 Supplement (Gibco, 12587010), 1X GlutaMAX Supplement (Gibco, 35050061), 1X MEM Non-Essential Amino Acids Solution (Gibco, 11140-050), 10 ng/ml Recombinant Human FGF-2 Protein (PeproTech, 100-18B) and 10 ng/ml Recombinant Human EGF Protein (R&D Systems, 236-EG), and passaged at 100,000 cells/well using Accutase (Stemcell Technologies, 07920). Cells were seeded onto an 8-well high glass bottom chamber slide (Ibidi, 80807) that was coated with matrigel matrix (Corning, 354277) diluted with DMEM/F12. Double transfection was performed using 250 ng (equimolar) of plasmids (DCAF7-3xFLAG and KIAA0232-Strep) and 1.5 uL Lipofectamine 2000 transfection Reagent (Invitrogen, 11-668-027). 46 hours after transfection, cells were fixed in 4% paraformaldehyde in PBS at room temperature (RT) for 15 min. Cells were permeabilized for 15 min in PBST (PBS and 0.2% Triton X-100) and blocked in blocking buffer (PBS, 0.2% Triton X-100 and 2% BSA) for 3 hrs at RT. Cells were incubated in primary antibodies diluted in blocking

solution overnight at 4°C, permeabilized for 15 min at RT, and incubated in secondary antibodies in blocking solution for 1h, 20mins at RT. Stained cells were then washed three times in PBST for a total of 15 mins, followed by 3 PBS washes for a total of 15 mins before being stored at 4°C in PBS for imaging. Primary antibodies: anti-Strep (IBA 2-1507, 1:1000), anti-Flag (Cell Signaling 14793, 1:1000), anti-a-Tubulin (SCBT sc-53029, 1:1000), anti-DYRK1A (R&D Systems AF5407, 1:200). Secondary antibodies: Abcam ab150177, Thermofisher scientific A42727, A32733, A48261 (all secondaries at 1:1000). Images were acquired using a Zeiss 980 LSM confocal microscope with an oil 63X objective in confocal mode. Images were processed in Fiji (**Fig. 3.13E**).

Enrichment of microtubule-related proteins among ASD-PPI interactors

We evaluated whether ASD-PPI interactors are enriched for proteins associated with structures important for spindle organization, such as centriolar satellites[111] and centrosomes[112]

We tested three genesets:

1. HEK293T centriolar satellite: centriolar satellite proteome (HEK293T) generated using in vivo proximity labelling with 22 human satellite proteins in HEK293T cells. We downloaded interactors from Gheiratmand et al. 2019 Table EV2, updated gene names using limma::alias2SymbolTable, and removed duplicate or unmapped genes.
2. NPC centrosome: centrosome-associated proteome generated by AP-MS of 10 centrosomal baits in human iPSC-derived NPCs. Interactors were downloaded from O'Neill et al. 2022 Table S7 (NSC_sig=1); gene names were updated using limma::alias2SymbolTable.

3. Neuron centrosome: centrosome-associated proteome generated by AP-MS of 10 centrosomal baits in human iPSC-derived neurons. Interactors were downloaded from O'Neill et al. 2022 Table S7 (Neuron_sig=1); gene names were updated using limma::alias2SymbolTable.

We evaluated whether ASD-PPI interactors were significantly enriched in the genesets above. For the HEK293T centriolar satellite enrichment test, we defined the universe to be the HEK293T proteome. For the NPC and neuron centrosome genesets, we defined the universe to be autosomal coding genes (NPC and neuron proteome were not reported in O'Neill 2022). We conducted Fisher's exact tests (two-sided) using a 2x2 contingency table using the variables ASD-PPI interactor (yes/no) and microtubule-associated protein (yes/no). We corrected *p* values for multiple hypothesis testing (Bonferonni, 3 tests) (**Fig. 3.13F**).

Xenopus tropicalis CRISPR and microscopy

Obtaining Xenopus embryos and tadpoles: Ovulation of adult male and female *Xenopus tropicalis* (WT superman line obtained from the National Xenopus Resource) was induced by injection of human chorionic gonadotropin (Sigma CG10-10VL) into the dorsal lymph sac according to standard procedure[113] and in accordance with approved UCSF IACUC protocol #AN209894-00A. Natural matings and *in vitro* fertilizations were performed. Embryos and tadpoles were staged according to the Normal Table of Xenopus Laevis (Daudin).[114] Clutch mates were always used as matched controls.

Xenopus microinjections: Xenopus embryonic microinjections were performed as in.[113] Microinjections were performed at the 2-cell stage using a Narishige micromanipulator, Parker Picospritzer, and Zeiss Stemi microscopes. Injection volume was calibrated with an eye-piece

micrometer. Embryos were grown between 22–25°C in 1/9 Modified Ringer's (MR) solution, which was refreshed daily. Male and female embryos were analyzed.

Xenopus CRISPR/Cas9 genome editing: High-efficiency sgRNAs were designed, synthesized, and validated.[78] For each embryo, 3 ng of purified Cas9-NLS protein (Macrolabs, UC Berkeley), 800 pg sgRNA, and a dextran dye conjugated with Alexa-555 (Invitrogen) were injected into 1 cell of a 2-cell stage embryo. The day following injection at stages 14–20, embryos were sorted left from right according to the dye. sgRNA oligo sequences to generate sgRNAs with the EnGen sgRNA synthesis kit are as follows:

scramble: 5' TTCTAATACGACTCACTATAGTGGTTTACATGTCGACTAATGTTTTAGAGCTAGA 3' ;

DCAF7: 5' TTCTAATACGACTCACTATAGAGAGGTTCCGCTTGGCGTTGGTTTTAGAGCTAGA 3';

DYRK1A: 5' TTCTAATACGACTCACTATAGCGTTTAGGTTCTGCTGACGGGTTTTAGAGCTAGA 3';

KIAA0232: 5' TTCTAATACGACTCACTATAGAGCTCTGTTGATGCAGACGGTTTTAGAGCTAGA 3'

Xenopus whole mount immunofluorescence staining: Xenopus immunostaining was carried out as previously described[58] with a primary antibody against beta-tubulin (1:100, DSHB E7) and a secondary goat anti-mouse fluorescent antibody (1:250, Life Technologies A32723).

Microscopy and Xenopus brain size measurements: Xenopus tadpole images were acquired on a Zeiss Axio Zoom.V16 microscope with apotome and a 1X objective. Images were processed using Fiji. Telencephalon size was calculated from stereoscope images of brain immunostainings using

the freehand select and measure functions in Fiji. The injected side was compared to the noninjected side (internal control). These measurements were from two-dimensional images taken from a dorsal perspective and reflect relative size differences. Differences in telencephalon size were evaluated using paired T-test, *p* values were corrected for multiple hypothesis testing (Bonferroni, 5 tests) (**Fig. 3.5F**).

DCAF7 subnetwork enrichment for DEGs from macrocephalic vs. normocephalic ASD organoids

We tested whether genes in the DCAF7 subnetwork were preferentially associated with differentially expressed genes (DEGs) from cerebral organoids generated from iPSCs from male ASD probands with either normocephalic or macrocephalic phenotypes, relative to unaffected father controls.[59] PPI networks were defined from a combination of the full ASD-PPI and an additional AP-MS experiment with DCAF7 as bait, using a cutoff of SAINT BFDR ≤ 0.05 to define interactors. The DCAF7 subnetwork consisted of DCAF7, eight hcASD that interacted with with DCAF7 in ASD-PPI (DYRK1A, KIAA0232, DSCAM, CREBBP, SKI, TCF20, SETD5, PRR12), and their interactors. The remaining 92 hcASD baits and interactors formed the "Other hcASD subnetwork". For comparison, we also evaluated the overlap between DEGs and the 100 hcASD bait genes. DEGs were obtained from Jourdon et al. (2023) by filtering Supplementary Table 5 (T3_DEG_results) for "high" confidence genes differentially expressed in ASD versus control organoids. ASD samples were classified as normocephalic or macrocephalic. To ensure sufficient power, DEGs were pooled across cell types and organoid time points, resulting in five DEG sets: normocephalic, macrocephalic, normocephalic-specific, macrocephalic-specific, and both. Similar to the original study, we distinguish between the larger normo- and macrocephalic gene sets that share 447 DEGs (44% and 27% of all DEGs, respectively) and smaller subsets that

are specific to either the normo- or macrocephalic phenotype. The universe of genes was restricted to the union of ASD-PPI and HEK293-expressed proteins. Enrichment between network gene sets and DEG sets was tested using one-sided Fisher's exact tests (greater), with Bonferroni correction for multiple comparisons **(Fig. 3.13G)**.

Differentiation of iPSCs to NPCs

Dorsal forebrain patterned NPCs were generated from iPSCs using a small molecule protocol adapted from a published method.[115] Specifically, we plated iPSCs on Matrigel coated plates in mTeSR plus medium (STEMCELL Technologies, Cat#100-0276). For the next 3 days, we fed cells with the KSR medium (15% Knockout Serum Replacement in Knockout DMEM, 1xGlutaMAX, 1xMEM-NEAA, 0.1mM BME) containing small molecules 250nM LDN193189 (Tocris, Cat No. 6053), 10uM SB431542 (Tocris, Cat#1614) and 5uM XAV939 (Tocris, Cat#3748) (LDN/SB/XAV) for dual SMAD inhibition and Wnt inhibition. On day 4 and day 5, we fed cells with in ⅔ KSR + ⅓ N2 (DMEM/F12, 1x N2, 1x B27 -Vitamin A, 1x GlutaMAX, 1x MEM-NEAA) + LDN/SB/XAV and ⅓ KSR + ⅔ N2 + LDN/SB/XAV respectively. On day 6, we passaged cells at 1:2 using EDTA and plated cells onto Matrigel coated plates. For day 6 and day 7, cells were cultured in NPC medium (DMEM/F12, 1xN2, 1xB27 -Vitamin A, 1x GlutaMAX, 1x MEM-NEAA, 10ng/ml FGF2, 10ng/ml EGF) supplemented with 5uM XAV. On day 8, we passaged the cells at 1:3 using Accutase (STEMCELL Technologies, Cat#07920) and cultured cells in NPCs medium onwards.

CRISPR/Cas9 mediated DCAF7, DYRK1A, KIAA0232 knockdowns in neuronal progenitor cells (NPCs)

We designed non-targeting control and *DCAF7*, *DYRK1A* or *KIAA0232* targeting single guide RNA (sgRNA) using an existing pipeline (https://github.com/mhorlbeck/CRISPRiaDesign).

Individual sgRNAs were cloned into lentiviral vector pMK1334 expressing BFP (Addgene Cat# 127965). Lentiviruses carrying sgRNAs were produced in Lenti-X 293T cells (Takara Bio, Cat#632180) by transfection and concentrated using Lenti-X concentrator (Takara Bio, Cat# 631231). We transduced NPCs with the concentrated lentiviruses and selected the transduced NPCs using 4ug/ml puromycin. We then cultured the NPCs in NPC medium (DMEM/F12, 1xN2, 1xB27 -Vitamin A, 1x GlutaMAX, 1x MEM-NEAA.) on tissue culture plates coated with Martrigel (Fisher Scientific, Cat# 08-774-552). The knockdown efficiency of sgRNA was measured by qPCR (**Fig. 3.5G**). sgRNA sequences were:
sgNTC:GACCCGAGACTGCTTCCCGG
sgDCAF7:GTCATCGCGTAGACTGTCCA
sgDYRK1A:TATTCGGGCGGGAAGCGAAA
sgKIAA0232: GCGGCCCCAGCGCCACTCCG

RNA isolation, cDNA synthesis and qPCR

Total RNA was isolated from 10^6 iPSC derived NPCs (control, sgDCAF7, sgDYRK1A, sgKIAA0232) using RNeasy Mini Kits (Qiagen, #74104). The isolated RNA was quantified by measuring absorbance at 260nm and the overall purity of RNA was estimated using the ratio of measurements at 260/280nm. 500ng to 1ug of RNA was converted into cDNA using SuperScript™ IV VILO™ Master Mix with ezDNAse (ThermoFisher Scientific, #11766050) (**Fig. 3.5, G to H**)

Global abundance proteomic for DCAF7 KD, DYRK1A KD, and KIAA0023 KD NPCs

NPC pellets were harvested and lysed by two freeze-thaw cycles in a 8M urea buffer (8 M urea, 150 mM NaCl, 100 mM Tris-HCl, pH 8.0). Following total protein quantification via Protein 660 assay, cysteine residues were reduced with 5 mM dithiothreitol for 30 minutes at room temperature, alkylated with 10 mM iodoacetamide for 30 minutes at room temperature in the dark,

and excess iodoacetamide was quenched with 10 mM dithiothreitol for 30 minutes at room temperature.

Lysates were processed using a single-pot solid-phase–enhanced sample preparation (SP3) workflow with MagReSyn Amine beads (Resyn Biosciences, Cat#MR-AMN002) on a KingFisher Apex (ThermoFisher Scientific) automated platform. Beads were washed twice with HPLC-grade water (Fisher Scientific, #W71) prior to use, and 20 µL bead suspension (400 ug beads, ~ 1:4 protein:beads w/w) was added per sample. Acetonitrile (Sigma, #34851) was added to sample lysate for a final concentration of 70% acetonitrile and protein binding was performed on the KingFisher Apex using four cycles of slow mixing for 1 min with 5 minute pause in between. Bead-bound proteins were washed sequentially with three washes of 95% acetonitrile and two washes of 70% ethanol (150 µL each; 3 minutes slow mixing). Beads were released and proteins were digested on-bead in 150 µL of 50 mM triethylammonium bicarbonate (TEAB)(Sigma, #140023) containing Lys-C (Promega, #V1671) and trypsin at a 1:100 (enzyme:protein, w/w) ratio for each protease, with bead release performed at slow speed, followed by overnight incubation at 37°C with shaking. Digestion solution was acidified to quench proteolysis and peptides were recovered from beads by magnetic separation. Extracted peptides were filtered through a 0.45 um filter to remove any residual beads before being partially dried to reduce volume. While drying extracted peptides, beads were incubated with HPLC water at room temperature to allow for full recovery of peptides. This secondary wash was extracted from the beads and filtered before the two extracted peptide aliquots were combined, and dried completely in a SpeedVac prior to mass spectrometry analysis.

LC-MS/MS data independent acquisition (DIA) followed previously established protocols. Raw spectra were searched against an in-house generated spectral library built from the UniProt human

database with isoforms using Spectronaut. Search parameters utilized default settings with the exception of cross run normalization, which was disabled. A 1% false discovery rate (FDR) was applied throughout. The MSstats Normal output from Spectronaut was used in combination with the MSstats package for protein-level summarization and group comparisons.

The overlap between differentially expressed proteins (DEPs) between DCAF7 KD and KIAA0232 KD was analyzed using one way Fisher enrichment test with a 2x2 contingency table: DCAF7 KD (yes/no) and KIAA0232 KD (yes/no). We defined the universe as all the proteins observed in NPCs. p values were corrected for multiple tests of hypotheses using Bonferroni correction. Similar analysis was done for overlap between DEPs and hcASD+ (**Fig 3.14, A to B**).

Cortical neuron differentiation

Cortical neuronal differentiation was carried out as described before.[115] NPCs were thawed and grown to confluency in a 6-well dish. The cells then were dissociated using Accutase for 5 minutes at 37°C to single cell resolution and plated at ~10,000 cells/well in NPC media and 10µM ROCKi. Lentiviral constructs of either RGEDI[116] or pHR-hSyn:EGFP (Addgene #114215)[117] were added to the cell suspension at 5 MOI at the time of passage and removed after 48 hours. The NPCs were cultured in a 12-well dish to confluency in NPC media and then dissociated using Accutase for 5 minutes at 37°C to single cell resolution and plated on 5 µg/mL human laminin and 5µg/mL fibronectin-coated 384 well plates (Perkin Elmer Cat #6057308) at 12,000 cells/well in NPC media and 10µM ROCKi. 24 hours later, media was replaced with Neuronal Differentiation media consisting of: 1:1 DMEM/F12:Neurobasal Media (Gibco 21103049), 1x N2, 1x B27 -Vitamin A, 1x GlutaMAX, 0.5mM dibutyryl cAMP (Sigma D0627), 0.2mM ascorbic acid (Sigma A4403-100MG), 1µM PD0325901 (Selleck Chem S1036), 5µM SU5402 (Selleck Chem S7667), 10µM

DAPT (R&D v), 20ng/mL BDNF (R&D 248-BD). Media was refreshed at 75% every other day up to Day 18 of differentiation.

Immunocytochemistry

Fixed cells were permeabilized using a 0.1% Triton-X/PBS solution for 20 minutes at room temperature. Permeabilization solution was removed, and a 1M glycine solution added and incubated at room temperature for 20 minutes. A blocking solution of 0.1% Triton-X/PBS, 2% FBS, and 3% BSA was added after removal of the glycine solution and incubated at room temperature for 1.5 hours. Primary antibodies were diluted in blocking solution and incubated overnight at 4°C. Primary antibody was removed by washing cells 3 times with 0.1%Triton-X/PBS. Secondary antibodies were added to 1:1000 in blocking solution and incubated at room temperature and covered for 1.5 hours. Cells were spun down at 8000 RPM in a cold centrifuge during the washes. Cells were washed in PBS for 5 minutes, then once with PBS plus Hoechst at a dilution of 1:1000 in PBS and incubated for 10 minutes at room temperature, covered. The Hoechst was washed out with PBS and the cells covered in PBS for imaging.

Primary antibodies used were MAP2 (1:1000 Abcam chicken-anti MAP2 # ab5392) and Ki67 (1:200 Millipore mouse-anti Ki67 # mab4190) (**Fig. 3.13J**), NUP155 (1:500 ProteinTech, #66359-1-Ig), SMPD4 (1:500 Invitrogen #PA5-110401), SETD5 (1:500 Proteintech #66955-I-Ig), TNFAIP8 (1:500 Proteintech 15790-1-AP), GNAI (1:500 SantaCruz #sc13533) (**Fig. 3.12D**). Secondary antibodies used were goat anti-chicken 488 (Abcam #ab96947), goat anti-rabbit 488 (Invitrogen #A-11008) goat anti-mouse 555 (Invitrogen #A21426) (**Fig. 3.12D, Fig. 3.13, E to G**).

Imaging: Images taken from NPCs were generated using a Nikon Ti2-E microscope equipped with a Crest X-Light-V2 spinning disk confocal (Crest Optics), emission filters, 438/24 (DAPI), 511/20 (GFP), 560/25 (RFP), and 685/40 (Cy5), and Celeste Light Engine excitation lasers

405/477/546/638 nm used respectively, Piezo stage (Mad City Labs), and a Prime 95B 25 mm CMOS camera (Photometrics) using a Plan Apo VC 60x/1.4 Oil (Nikon). The data was captured with NIS-Elements software (v. 5.41.01 build 1709, Nikon) and processed in Fiji/ImageJ (**Fig. 3.12D**).

GEDI, longitudinal imaging, and odds ratio of cell death: Cells that contained the red GEDI biosensor were imaged on an ImageXpress Micro Confocal High-Content Imaging System from Molecular Devices for 7-10 days starting on differentiation day 7. Montages of 9 tiles were imaged in RFP (200 ms) and GFP (100 ms) channels at 20x magnification per well. Ki67- stained NPCs were imaged at 9 tiles per well in the RFP (100 ms), GFP (50 ms), and DAPI (15ms) channels at 20x per well. Images were processed using a custom workflow in Galaxy.[117] Cells from each line were assessed for being alive or dead by the GEDI[116] biosensor on different plates at multiple points in time. The change in odds of cell death (OR-CD) at any point in time. The OR-CD for one specific cell-line versus another one at any point in time was calculated using a Generalized Linear Mixed effects Model (GLMM)[118] using the glmer function in R, using the binomial probability distribution as the family argument to this function. The model includes the experiment id in which the cell was assayed modelled as a random effect, the time (as a continuous variable) at which the cell (live or dead) status was ascertained, and the following terms modelled as fixed effects; 1) the cell line from which the cell was derived and 2) the interaction between the cell line and time. These changes, as OR-CD, are derived from the model fits to the data. A similar GLMM was used to model changes in the log odds of death across lines that had gRNA targeting DCAF7 versus non-targeting gRNA. Time was either modelled as a categorical variable (non-linear model) based on establishing the linearity using the lower Akaike Information Criterion (AIC)[118] (**Fig. 3.13H and I**). Code is available on: https://github.com/gladstone-institutes/MRIDcode/tree/main

Cell Profiler and statistical analysis: The Ki67 proliferation assay used a modified Cell Profiler[119] example pipeline for colocalization (https://cellprofiler.org/examples). Images were pooled in CellProfiler and analyzed on a per-image basis. Background values were calculated per image then subtracted from the whole image. Following background correction, all objects in each DAPI image were identified as nuclei using minimum and maximum diameters per object and filtering out excessive intensity values using a Minimum Cross-Entropy thresholding method. This method identifies all possible nuclear objects within the appropriate size range. Intensity values were calculated per object then used to filter out non-nuclear objects or dead cells that display bright nuclei that have much higher intensities than live cells, which were relabeled as nuclei segments. Images that contained Ki67 staining in the FITC channel were run through a feature enhancement step that increases the signal-to-noise ratio. Enhanced images were run through segmentation to identify all FITC objects that fit Ki67 signal criteria in terms of size and signal intensity again using a Minimum Cross-Entropy thresholding method. These identified objects were renamed as Ki67 segments. Ki67 segments were related to the nuclei segments to determine how many nuclei were Ki67-positive. Finally, the number of Ki67-positive nuclei was divided by the number of total nuclei to calculate the percent of Ki67-positive nuclei. The segmentation overlays and math were exported from the program and then plotted in Prism10. We previously developed a statistical package called RMeDPower[120] in R, a complete package of statistical tools that allow evaluating the effect size and variance contribution of a set of variables within a dataset to a given response. Outliers were excluded using Rosner's test, and the data were log-transformed prior to statistical analysis with a generalized linear mixed-effects model (GLMM). Model selection was guided by QQ plot evaluation and a binomial distribution was applied, depending on the optimal fit, to obtain parameter estimates and corresponding p-values. (**Fig. 3.13, J to K**).

Evaluating whether ASD_{mut}-PPI changed interactors are enriched for ASD genes

We evaluated whether the interactors that demonstrate increased (n = 69) or decreased (n = 95) interaction with hcASD in ASD_{mut}-PPI are enriched for the n = 255 hcASD+.[6] We defined the gene universe to be HEK293T expressed proteins measured in Fu *et al.* 2022. The 2x2 contingency table used for the two-sided Fisher's enrichment tests were defined by: hcASD+ (yes/no), changed interactor (yes/no). *P* values were corrected for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * 2 tests) (**Fig. 3.15A**).

NPC-PPI

Overexpression of hcASD in NPCs: Strep tagged AP2S1, PPP2R5D and TBL1XR1 genes were subcloned from pcDNA3.1 into pLenti-puro vector. To produce the lentivirus, HEK293T cells were transfected with 5µg of pLVX-TetOne pLVX-TetOne-Puro lentiviral plasmids, 3.33µg Gag-Pol packaging construct, and 1.66 µg VSV-G envelope (Addgene #8454) using Lipofectamine 3000 (Invitrogen L3000015). Supernatant from the culture was precipitated with 8.5% PEG-6000 and 0.3M NACl and incubated for 4-6h at 4ºC. Virus was pelleted at 3500 rpm for 30 mins in a spin bucket rotor and suspended in 1ml of 1X PBS. NPCs were seeded on matrigel coated plates 1 day prior to transduction with 300ul of lentivirus along with 1X of TransPlus Viral Transduction enhancer (AllStemBio #VO50). NPCs were collected 72 h post transduction followed by lysis in Lysis buffer (50mM TrisHCl, 150mM NaCl, 0.5% NP40, 1mM EDTA, protease/phosphatase inhibitors) and subjected to AP-MS as described above.

Defining PPI from NPCs: As described before (Methods), high confidence interactors from AP-MS in NPCs were identified by calculating SAINTexpress scores (using spectral counts) and ComPASS. CompPASS scoring relies on the entire matrix of spectral counts for all baits and preys.

To enable a comparison of CompPASS results from HEK293T ASD-PPI with results from NPC, we used the HEK293T WT data as background for the NPC CompPASS analysis. We combined input data from HEK293T WT baits that were not in the NPC data with the NPC data. All proteins detected in either dataset were kept, filling in zeros for spectral counts for proteins that were unique to the other cell type. We defined significant interactions to be those with SAINTexpress 1-BFDR ≥0.95 and ranked ComPASS score ≥0.97. For displaying PPI that were detected but did not pass thresholds in HEK293T WT results, the SAINTexpress and CompPASS scores were combined into a single composite score for each PPI detected in HEK293T WT by taking the mean of 1-BFDR from SAINTexpress and rank_WD from compPASS. If an interaction was detected when both proteins were used as baits, we reported the highest composite score from either bait (**Fig. 3.15B**).

Overlap between HEK293T and NPC interactors: For each of the three hcASD with NPC PPI data, we evaluated the extent of overlap in interactors identified in HEK293T (from ASD-PPI) versus NPC cells. We conducted one-sided Fisher's enrichment tests (greater), where the gene universe was restricted to the n = 9,627 genes in the intersection of HEK293T-expressed proteins (defined as above) and genes with detected transcripts in Day 0 WTC11 NPCs.[121] The 2x2 contingency table was defined by: 1) HEK293T interactor or not; 2) NPC interactor or not. *P* values were adjusted for 3 tests (Bonferroni, **Fig. 3.15B**).

Protein system enrichment

De novo mutations: A comprehensive list of exome-wide *de novo* somatic mutations identified in the ASD cohort considered in this research was obtained from the large-scale whole exome sequencing study from Satterstrom *et al.* 2020.[4] We used the 5287 filtered probands from the

analyses and considered only functionally disrupting missense and protein truncating variant mutations. Specifically, we only kept mutations that met any of the following criteria: (1) VEP_functional_class_canonical_simplified equal to one of ["frameshift_variant", "stop_gained", "splice_acceptor_variant", "splice_donor_variant"] AND pLI ≥ 0.995; (2) MPC ≥ 2; or (3) Polyphen_prediction == “probably_damaging".

Integrated protein network: We compiled a diverse set of experimental and curated protein-protein interaction datasets. We processed each PPI network as follows: (1) standardized the naming of all the gene symbols by mapping to their standard symbols via HUGO Gene Nomenclature Committee (HGNC) ID; (2) removed interactions with no documented score; (3) removed self-interactions; and (4) removed duplicates. In addition to aforementioned processing, for the case of Mentha and BioGRID, we only kept interactions in humans (Taxon A/B == 96096 in Mentha & Organism Name Interactor A/B == "Homo sapiens" in BioGRID). We further filtered BioGRID to include only physical interactions.

The ASD-PPI and other PPI networks were integrated into a single interaction network. The integration was performed using BIONIC,[122] a graph convolutional neural network that uses graph attention networks. We used the following parameter settings {epochs: 1000, batch_size: 512, learning_rate: 0.00005, embedding_size: 512}, with the remaining parameters using the default values. BIONIC ran until the reconstruction loss stabilized.

We computed an interaction score between each pair of proteins by calculating the cosine similarity score between their embeddings. We then focused the network on the 102 identified high-risk ASD-genes and the interactors of ASD-PPI and ASD_{mut}-PPI by obtaining their top associated proteins. We further retained all ASD-PPI interactions.

Multi-scale map of protein systems: To identify a multi-scale map of systems of tightly connected proteins, we applied Hierarchical community Decoding Framework (HiDeF)[123] to the integrated network. HiDeF determines connected proteins across various thresholds and identifies systems from most stringent thresholds to least stringent, where the smaller more stringent identified systems were part of the larger less stringently identified systems. We used the following parameter setting {maxres: 50, tau: 0.75, chi: 5, alg: 'leiden'}.

Differential proteins systems: To identify systems that have stronger evidence in ASD-PPI, ASD_{mut}-PPI, or balanced, we modified the computed WT integrated network according to interactions found to have stronger support in either ASD-PPI or ASD_{mut}-PPI. We then generated another multi-scale map using the modified network using the same multiscale community detection algorithm described previously. We computationally assessed which systems have stronger evidence in ASD_{mut}-PPI or ASD-PPI by aligning the original and modified multi-scale maps using the align function developed by Dutkowski & Kramer *et al*.[124]

Systems enriched with mutated genes and disrupted interactions: To identify systems that have significant enrichment for functionally disrupting mutations, we performed hypergeometric tests on the number of mutated genes in each system. To compute the number of mutated genes, we identified genes with functionally disruptive *de novo* mutations as described in section "Mutation data". The significance of the number of mutated genes was computed using the hypergeom.sf function from the scipy package in Python. We used a cutoff of 0.05 to determine significance.

Protein system names: The protein systems were labeled using a GPT-4-based pipeline developed by Hu *et al.* 2023.[125] This pipeline assigns a set of genes with succinct literature-driven names that

summarizes their consensus functions as well as providing supportive analysis text. GPT-4 (gpt-4-1106-preview) was used for this analysis (**Fig. 3.15E**).

Modeling the FOXP1/FOXP4 interaction

To investigate the relative orientation of structured domains within the FOXP4-FOXP1 complex, we performed coarse-grained molecular dynamics simulations using CALVADOS3.[126] Non-bonded interactions were modeled using the Ashbaugh-Hatch potential for hydrophobic interactions and the Debye-Hückel potential for electrostatic interactions at an ionic strength of 0.15 M and pH 7.0. The initial configuration was derived from an AlphaFold multimer model of the FOXP4-FOXP1 complex centered within a cubic simulation box with a side length of 50 nm. The system comprised two protein chains: FOXP4 (677 residues) and FOXP1 (680 residues).

Intramolecular harmonic restraints (k = 2,000 kJ/mol/nm^2) were applied to residues in the coiled-coil dimerization domains and DNA-binding domains (FOXP4: residues 118-199, 299-375, 451-551; FOXP1: residues 118-199, 299-375, 450-539). Residues outside these domain definitions, corresponding to disordered linker regions, were unrestrained and free to sample conformations under the CALVADOS force field, allowing the structured domains to move relative to one another. Intermolecular harmonic restraints (k = 700 kJ/mol/nm^2) were applied between 162 CA-CA pairs across the FOXP4-FOXP1 interface, based on contacts within 0.9 nm in the AF model: 45 pairs within the coiled-coil domains (residues 118-199; 19 FOXP4 and 16 FOXP1 residues) and 117 pairs within the dimerization domains (residues 299-375; 38 FOXP4 and 39 FOXP1 residues). The DNA-binding domains (FOXP4 residues 451-551; FOXP1 residues 450-539) were left without intermolecular restraints.

Simulations were performed using Langevin dynamics at 293 K with a friction coefficient of 0.01 ps-1. Each simulation was run for 500,000 integration steps, with configurations saved every 500

steps, yielding 1,000 frames sampling only 5,000 ps following minimization, which was used for analysis. All-atom structures were reconstructed from selected coarse-grained frames using cg2all for visualization, and frame 300 (1500ps) is shown. The FOXP1 DNA binding domain was then aligned to DNA based on an AlphaFold3 model. (**Fig. 3.7E**)

Immunoprecipitation and western blot of Strep-tagged proteins

For immunoprecipitation, the lysate was incubated with either precleared MagStrep 'type3' beads (30 µl; IBA Lifesciences) for 4h at 4ºC, followed by washing with 1ml wash buffer (IP buffer supplemented with 0.05% NP40) three times. The resulting washed beads were mixed with 6X Laemmli SDS Sample buffer (Boston Bioproducts, Cat#BP-111R), heated at 95ºC for 5 mins and loaded onto SDS-PAGE for Western blot. Antibodies used for detection of Strep-tagged proteins:Streptactin-HRP, Iba Lifesciences #2-1502-001 (**Fig. 3.7H, Fig. 3.15C**).

Immunoprecipitation of GNAI-RIC8A

GNAI1 site-specific mutagenesis: site-directed mutagenesis of human GNAI1 cDNA contained within a Strep-tagged pcDNA4 expression construct was performed to substitute isoleucine at position 319 with threonine (I319T), proline (I319P), phenylalanine (I319F), or valine (I319V). Mutagenic primers (25–35 nucleotides) were designed with the target codon centered in each sequence:

319T (F): GACACAAAGGAAACATACACC

391T (R): AGGAAACATACACCCACTTC

319V (F): AGGAAGTATACACCCACTTC

319V (R): GACACAAAGGAAGTATACACC

319P (F): AGGAACCATACACCCACTTC

319P (R): GGTGTATGGTTCCTTTGTGTC

319F (F): AGGAATTTTACACCCACTTC

319F (R): GACACAAAGGAATTTTACACC

Mutant plasmids were generated by PCR amplification using Q5 DNA polymerase. Following amplification, parental template removal and ligation were performed using the KLD enzyme mix (NEB #M0554). Reaction products were transformed into Stellar competent cells, recovered in SOC medium for 1 hour at 37°C and plated onto LB agar with Ampicillin selection. Individual colonies were screened, and plasmids carrying the desired mutations were confirmed by whole-plasmid sequencing (Plasmidsaurus).

Immunoprecipitation of GNAI: The Strep-GNAI1 WT, I319T, I319V, I319P, I319F plasmids were transfected into HEK293 cells using PolyJet as described above. Cells were harvested after 48h, lysed, and Strep-tagged proteins were immunoprecipitated and analyzed by western blotting as described above (**Fig. 3.16E**).

EVE scores

The EVE scores for the available patient derived hcASD mutants were extracted from https://evemodel.org[127] (**Fig. 3.16F**).

Isogenic hiPSC cell lines

hiPSC CRISPR/Cas9 genome editing plasmid design: FOXP1 R513H: The closest PAM sequence to FOXP1 R513 for CRISPR/Cas9 genome editing was identified, and the corresponding sgRNA sequence (GTGCGAGTAGAAAACGTTAA) was cloned into the pX459 plasmid (Addgene #62988) using BbsI Golden Gate cloning. To construct the donor plasmid, an IDT gBlock was ordered to create the R513H mutation, as well as create a silent mutation in the PAM site to prevent further editing. ~500bp homology arms to the FOXP1 cut site were generated using PCR, and the

homology arms and gBlock were inserted into the pUC18 plasmid (Addgene #50004) linearized with HindIII and BamHI. Of note, the DNA sequence variant (GrCh37 3:71021817 C>T) referred to here as FOXP1^{R513H} (a brain specific isoform) also maps to FOXP1^{R514H}, the canonical FOXP1 isoform.

FOXP1 L327P: sgRNA sequence CCTGCATTTGTACTCTACAT was cloned into pX459 as described above. A 200 nucleotide ssODN containing the L327P edit was ordered to be used for HDR and introduction of mutation.

PPP2R5D E198K: sgRNA AGGGTGGGCTCATCTTCCTC was ordered from Synthego and complexed *in vitro* with Cas9 protein. ssODN containing the E198K edit was ordered from IDT to be used for HDR.

hiPSC CRISPR/Cas9 genome editing and genotyping: For FOXP1 mutations, eWT-1323.4 hiPSCs[128] (Conklin Lab) were incubated for one hour in StemFlex media supplemented with the CEPT small molecule cocktail[129] containing: Chroman 1 (MedChem Express #HY-15392, 50nM); Emricasan (SelleckChem #S7775, 5μM); Polyamine Supplement (Sigma Aldrich #P8483, 1:1000 diluted); and trans-ISRIB (Tocris #5284, 0.7 μM). For electroporation using the Invitrogen Neon Transfection System, cells were enzymatically lifted using TrypLE (Gibco #12605010), and counted using a hemocytometer. 100,000 cells per electroporation reaction were pelleted and resuspended in 5uL Buffer R per reaction. Separately, 2μg pX459, 1μg pUC18 (for R513H) or 5uM final concentration ssODN (for L327P), and Buffer R were mixed to 7μL. 5μL cells were added to the plasmid mix, and 10μL cells and plasmid mix were electroporated using a 10μL pipet tip (Invitrogen #MPK1025) with the following settings: 1100 V, 30ms, 1 pulse. Electroporated cells were added to plates with prewarmed StemFlex media with CEPT and incubated overnight. Starting 16 hours after electroporation, cells were incubated with media supplemented with

0.5µg/ml Puromycin for 72 hours to select for cells that received the pX459 plasmid and were more likely to be genome edited.

Seven days after electroporation, individual colonies were manually passaged, with one colony per one well of a 12 well plate. After reaching confluency, cells were passaged once more, with remaining cells lysed for gDNA extraction using the NEB Monarch Genomic DNA Isolation Kit (#T3010S). Clonal gDNA was PCR amplified to contain the CRISPR/Cas9 edit site, PCR products were purified using the Macherey Nagel NucleoSpin PCR clean up kit (#740609) and submitted for Sanger sequencing. The correctly targeted clone and two untargeted clones were expanded.

For PPP2R5D^{E198K}, transfections were performed with 8 x 10^5 cells using the 4D-Nucleofector X (Lonza) and the P3 Primary Cell 4D-Nucleofector X Kit L (Lonza). Cas proteins (IDT or Synthego) were complexed with their respective sgRNA (Synthego) and single-stranded DNA oligonucleotide (IDT) serving as a repair template. Pools of cells were either stored or single-cell sorted with NamoCell Pala (Protein Simple) into a 96-well plate. DNA was extracted using Quick Extract (VWR) and PCR amplified with AmpliTaq Gold 360 Master Mix (Fisher Scientific #4398881). Successfully amplified PCR products were Sanger sequenced and analyzed with ICE (Synthego) for editing efficiency (**Fig 3.8A and Fig. 3.18, A and D**).

FOXP4 KO rescue: For *FOXP4* KO, optimized sgRNAs for CRISPR/Cas9 were chosen using CRISPick,[130] where the top two were used for experiments (sgRNA1: gagcacagtaggctgtccgt, sgRNA2: ggctgtactccggtcatcca). Empty plasmid backbone was used as a control. Guides were cloned into the pX459 plasmid as described above. 2ug of plasmid was electroporated into either FOXP1 R513H/WT or WT/WT cells. Instead of clonal expansion, bulk populations that survived puromycin selection were used for further experiments. Editing efficiency was determined by isolating genomic DNA and PCR amplifying the region surrounding the Cas9 cut site as described

above, and submitting for Sanger sequencing. The results were run through Synthego's Inference of CRISPR Edits (ICE) online tool to predict editing efficiency of sgRNA1 and sgRNA2 compared to empty vector control for each FOXP1 R513H/WT and WT/WT (**Fig. 3.17H**).

Immunoprecipitation of FOXP1 from hiPSC derived NPCs

FOXP1$^{R513H/WT}$ and control iPSCs were cultured and differentiated into NPCs using the protocol described above. For immunoprecipitation of endogenous FOXP1, NPCs (FOXP1$^{R513H/WT}$ or control) were collected in lysis buffer containing 0.5% NP40 and sonicated. After centrifugation, the lysates were incubated with anti-FOXP1 antibody (Cell Signaling #2005) or anti-rabbit IgG (Abcam, ab172730) and Protein A/G beads, washed and prepared for Western blot (FOXP1, Cell Signaling #4402, FOXP4, FOXP2 Millipore #ABE73, Millipore, #abe74, GAPDH, Proteintech, #HRP-60004) (**Fig. 3.17A**).

hiPSC-derived cortical organoid differentiation

hiPSCs were lifted using ReLeSR and resuspended in Neural Induction Media containing GMEM (Gibco #11710035), 10% Knockout Serum Replacement (Gibco #10828028), NEAA diluted 1:100 (Gibco #11140050), Sodium Pyruvate diluted 1:100 (Gibco #11360070), 5mM 2-mercaptoethanol (Sigma Aldrich #M6250), and 100µg/mL Primocin supplemented with 5µM SB431542 (Tocris #1614), 100nM LDN-193189 (Sigma Aldrich #SML0559), and 3µM IWR1-endo (Cayman Chemicals #13659). Cells in Neural induction media were moved to 6-well low attachment plates (Corning #3471), with a media change on day 3 with CEPT, and day 6 without CEPT. From day 9-25, organoids were cultured in Maintenance Media 1: 50% DMEM/F12 with Glutamax (Gibco #10565042) and 50% Neurobasal (Gibco #21103049) with B27 without vitamin A (Gibco #12587001), N2 (Gibco #17502048), NEAA diluted 1:100, Glutamax diluted 1:200 (Gibco #35050061), and 55µM 2-mercaptoethanol supplemented with 10ng/mL each FGF

(Peprotech #100-18B) and EGF (Peprotech #100-47). Media was changed every 2-3 days. From days 26-35, media was changed without FGF and EGF. From day 35 onward, organoids were cultured in Maintenance Media 2: Maintenance Media 1 supplemented with B27 with vitamin A (Gibco #17504001), instead of B27 without vitamin A. From day 63-70, organoids were cultured in Maintenance Media 2 supplemented with 10ng/mL each of BDNF (Alomone #B-250) and NT3 (Alomone #N-260).

Cortical organoid fixation, embedding, cryosectioning

Organoids were fixed in 4% PFA in PBS for 30 minutes at room temperature, then washed with PBS three times. Organoids were then dehydrated in 30% sucrose in PBS at 4ºC overnight. Cryomolds were filled with a 1:1 mixture of 30% sucrose in PBS and OCT, and organoids were embedded in the cryomolds and frozen on dry ice before being stored at -80ºC. Fixed and frozen organoids were cryosectioned to 18µm thickness and mounted on SuperFrost Plus microscope slides (Fisher Scientific #12-550-15).

Organoid immunohistochemistry

Sectioned organoids were first rehydrated in PBS for 10 minutes and then treated with boiling 10mM sodium citrate solution pH 6 for 15 minutes. Slides were washed once with PBS and blocked for one hour at room temperature in blocking buffer of PBS with 1% Normal Donkey Serum (Jackson ImmunoResearch #017-000-121), 0.1% Gelatin, and 1% Triton X-100. Primary antibodies were diluted in blocking buffer and added to slides in a hybridization chamber, where slides were incubated at 4ºC overnight. Slides were then washed three times with PBS with 1% Triton X-100 and incubated with secondary antibodies diluted 1:1000 in blocking buffer in a hybridization chamber at room temperature for 3 hours. Slides were washed three times in PBS

with 1% Triton X-100 and coverslips were mounted using ProLong Gold Antifade Mountant (Invitrogen #P36930).

The following primary antibodies and dilutions were used: PAX6 (R&D Systems #AF8150 1:250), TBR1 (Abcam #ab183032 1:300), BCL11B (Abcam #ab18465 1:500), SATB2 (Santa Cruz Biotechnology #SC-81376 1:500), DLX2 (Santa Cruz Biotechnology #SC-393879 1:50) (**Fig. 3.8, B, G, and J, and Fig. 3.18, B, D, and F**).

Imaging and quantification

Immunostaining sections were imaged using a Leica SP8 laser scanning confocal microscope using a 20x air objective. Images were processed using Fiji and quantified using CellProfiler.[119] Images were cropped to the areas surrounding FOXG1+ve rosette structures, extending out from the base 250um for standardized comparisons. All rosette quantifications for each marker were averaged per organoid. For statistical analysis, genotype effects were tested in R using lme4 and lmerTest. For experiments with multiple differentiation batches, we used a linear mixed model (~ genotype + (1 | batch)); for organoid-level datasets from a single differentiation, we used a linear model (~ genotype). Multiple hypothesis correction was performed using Benjamini-Hochberg FDR correction. To account for small sample size and potential non-normality, genotype effects were additionally evaluated by organoid-level permutation testing. Genotype labels were permuted 10,000 times across organoids, with permutations performed within the batch where applicable. For each permutation, the same model was refit and the genotype t statistic was recorded. Empirical two-sided permutation p-values were calculated as the proportion of permuted absolute t statistics greater than or equal to the observed absolute t statistic, followed by Benjamini-Hochberg FDR correction (**Fig. 3.8, B, G, and J, and Fig. 3.18, B and E**).

Cortical organoid dissociation

Organoids were dissociated using 20 units of Papain (Worthington #LK003178) with 5% Trehalose in HBSS for 30 minutes at 37ºC. DNase was added, and organoids were incubated for another 15 minutes at 37ºC. Papain was quenched using Albumin-ovomucoid inhibitor and cells were filtered through a 40µm cell strainer. Cells were spun down, resuspended, and counted.

scRNA-seq

For FOXP1 R513H, dissociated organoids as described above were used as input for Fluent Biosciences PIP-seq T2 V4.0 kit[131] with 20,000 cells per reaction as input with two technical replicates per sample. For FOXP1 L327P, dissociated organoids were used as input for Illumina Single Cell 3' RNA Prep, T2 kits (#20135689, equivalent to Fluent Bioscience PIP-seq T2 V5.0, after Fluent acquisition by Illumina) with 10,000 cells per reaction as input. Libraries were pooled to 2nM and sequenced at 750pM final concentration with 2% PhiX spike-in. Libraries were paired-end sequenced with a targeted sequencing depth of 20,000 reads per cell on either a NextSeq 2000 or NovaSeq X.

scRNA-seq analysis

Cell by gene matrices for each sample/replicate were generated using PIPseeker and were integrated into one object using Seurat. High quality cells were subsetted by removing those with less than 500 or greater than 10,000 genes; less than 1000 UMIs; and greater than 10% mitochondrial reads. Reads were normalized using SCTransform and batch correction was done with Harmony. Cells were clustered using resolution 0.2. Clusters were manually annotated using cluster markers. Differentially expressed genes between $FOXP1^{R513H/WT}$ and $FOXP1^{WT/WT}$ in each cluster were determined using FindMarkers with the default parameters (**Fig. 3.8, E and F, Fig. 3.17 C and D**).

For L327P, preprocessing and clustering was performed as described above. R513H cell type annotations were mapped onto individual cells in the L327P object using the command FindTransferAnchors, with R513H as the reference object and L327P as the query object. Anchors were then transferred to individual cells in the L327P object using TransferData. To assess correspondence between annotations, we calculated the proportion of cells in each L327P cluster with R513H cell type labels following label transfer. Each L327P cluster was subsequently renamed according to the R513H cell type representing the highest proportion of cells within that cluster. Differentially expressed genes were then determined similarly using Seurat's FindMarkers function with default settings (Wilcoxon rank sum test with Bonferroni correction). To compare transcriptional effects across genotypes, differentially expressed genes identified within excitatory neuron subtype EN-3 in both L327P and R513H datasets were compared by computing Pearson correlation coefficients between log2 fold changes obtained in each dataset (**Fig. 3.8K and Fig. 3.18C**).

CUT&Tag

CUT&Tag was performed on dissociated organoids as previously described,[79] with some modifications. Briefly, 200,000 cells per reaction were pelleted at 600g for 3 minutes at room temperature and resuspended and fixed in PBS with 0.1% formaldehyde for 2 minutes. 1.25M glycine was added to double the molar concentration of formaldehyde and stop cross linking, and cells were spun down at 1300g at 4ºC for 3 minutes. Cells were resuspended in wash buffer (20mM HEPES pH 7.5, 150mM NaCl, 0.5mM spermidine, and 1 EDTA-free complete protease inhibitor tablet). Concanavalin A-coated beads (Fisher Scientific #NC1526856) were prepared by adding 10µL beads per reaction to bead-binding buffer (20mM HEPES pH 7.9, 10mM KCl, 1mM $CaCl_2$, and 1mM $MnCl_2$). Using a magnetic rack, binding buffer was removed, and beads were

resuspended in the binding buffer once more before removing the binding buffer again and finally resuspending in enough binding buffer for 10µL per reaction. 10µL beads were then added to cells and incubated on an end over end rotator for 10 minutes at room temperature. Wash buffer was removed from cells using a magnetic rack, and cells were resuspended in enough antibody buffer (wash buffer with 2mM EDTA, 0.1% BSA, and 0.05% digitonin) for 50µL per reaction. Primary antibodies (FOXP1 Cell Signaling Technologies #2005S, FOXP4 Millipore #ABE74, H3K27ac Cell Signaling Technologies #4353, IgG Epicypher #13-0042) were added at 1:50 dilution and samples were nutated overnight at 4C.

Using a magnetic rack, primary antibody was removed, and cells were resuspended in 100µL secondary antibody (Antibodies Online #ABIN101961) diluted in wash buffer with 0.05% digitonin. Cells were then nutated for one hour at room temperature. Secondary antibody mix was removed using a magnetic rack, and cells were washed with wash buffer with 0.05% digitonin three times. pA-Tn5 preloaded with Nextera adapters (Epicypher #15-1117) was diluted 1:20 in dig-300 buffer (20mM HEPES pH7.5, 300mM NaCl, 0.5mM spermidine, 0.015% digitonin with 1 EDTA-free complete protease inhibitor tablet) and cells were resuspended in 50µL of pA-Tn5 mix. Cells were nutated for one hour at room temperature. Using a magnetic rack, pA-Tn5 mix was removed, and cells were washed three times in dig-wash buffer. Cells were resuspended in 300µL of tagmentation buffer (dig-300 buffer with 10mM $MgCl_2$) and incubated in a 37ºC water bath for one hour. Cells were released from beads with addition of 10µL 0.5M EDTA, 3µL 10% SDS, and 2.5µL 20mg/ml proteinase K. Samples were vortexed and incubated in a heat block at 55ºC for one hour.

Fragments were purified by adding 300µL phenol:chloroform:isoamyl alcohol (25:24:1 v/v) and sample to a phase lock tube (Qiagen #129046), and spun down at 16,000g for 3 minutes. 300µL

chloroform was added to each sample and spun down once more at 16,000g for 3 minutes. The aqueous layer was added to a 1.5ml lo-bind tube with 750µL 100% ethanol and mixed by pipetting. Samples were cooled on ice and spun down at 16,000g for 15 minutes at 4ºC. Supernatant was decanted, and samples were washed once more with 1ml 100% ethanol and spun down at 16,000g for 1 minute at 4ºC. Ethanol was decanted, the pellet was air dried and resuspended in 22µL water. Libraries were amplified by mixing 21µL sample, 2µL each i5 and i7 primers (Illumina #FC-131-2001), and 25µL NEBNext High Fidelity 2X Master Mix (NEB #M0541S) and using the following PCR cycle settings: 72ºC for 5 minutes, 98ºC for 30 seconds, 98ºC for 10 seconds, 61ºC for 10 seconds, repeat steps 3-4 15x, and a final 72ºC incubation for 1 minute. Libraries were purified using SPRI Select Reagent (Beckman Coulter #B23317) and eluted in 20µL water.

Libraries were pooled to 2nM and diluted to a final concentration of 750pM with 2% PhiX spike in for sequencing using the Illumina NextSeq 2000 with a targeted read depth of ~10 million reads per sample. 2 technical replicates were used per sample.

CUT&Tag analysis

Reads were trimmed using TrimGalore, aligned to reference genomes for hg38 and *E. coli* using bowtie2 with parameters "--end-to-end --very-sensitive --no-mixed --no-discordant --phred33 -I 10 -X 700." Duplicates were marked and removed with picard. Peaks were called using MACS2 using the corresponding IgG sample as a control, with q = 0.01. DESeq2 was used to determine differential peaks between $FOXP1^{R513H/WT}$ and $FOXP1^{WT/WT}$. Wald test was used to determine significance with Benjamini-Hochberg correction. Consensus peaks were determined by using bedtools intersect to find peaks consistent in both technical replicates. For peak intersections, bedtools intersect was used to identify common and unique peaks between sets of interest (**Fig. 3.8C**).

Enrichment of FOXP1 and FOXP4 targets with FOXP1 WT vs. het DEGs

To determine enrichment of FOXP1 and FOXP4 peaks in target genes in our DEG list, we first annotated FOXP1 and FOXP4 consensus peaks, as well as peaks with nominally significant differential binding, to their nearest neighbor gene using the R packages ChIPpeakAnno (annotatePeak command) and biomaRt (getBM command, mapped to Ensembl genome assembly). We then used two-sided Fisher's exact tests to determine significant enrichment of FOXP1 or FOXP4 bound genes in DEGs (**Fig. 3.17E**).

Evaluating whether FOXP1-mutant differentially bound genes or DEGs are enriched for ASD genes

We evaluated whether FOXP1-differentially bound genes or cell type-specific differentially expressed genes (DEGs) from $FOXP1^{WT/WT}$ vs. $FOXP1^{R513H/WT}$ organoids are enriched for hcASD+.[6] We defined genes that are more strongly (CUT&Tag $p < 0.05$ & $\log_2FC>0$) or weakly ($p < 0.05$ & $\log_2FC<0$) bound by FOXP1 in $FOXP1^{R513H/WT}$ organoids. For each cell type, we defined DEGs that are upregulated in $FOXP1^{R513H/WT}$ (scRNAseq *p*.adj < 0.05 and pct.1-pct.2 >0) or downregulated (*p*.adj < 0.05 and pct.1-pct.2<0) in $FOXP1^{R513H/WT}$ organoids. We defined the gene universe to be the intersection of genes expressed in each cell type and measured in Fu *et al.* 2022. We conducted two-sided Fisher's enrichment tests, defining the 2x2 contingency table by the parameters: hcASD+ (yes/no) and differentially bound/DEG (yes/no). We corrected *p* values for multiple hypothesis testing (Bonferroni correction, *p*.adj = *p* * #comparisons) (**Fig. 3.17F**).

Multi-electrode array (MEA) recordings of organoids

High density MEA chips (Maxwell Biosystems MaxOne+) were prepared according to manufacturer instructions. Briefly, chips were equilibrated with culture media for two days in the 37ºC 5% CO_2 incubator, then coated with 0.07% Poly-Ethyleneimine (PEI, Sigma #P3143) diluted

in borate buffer (20X stock, Thermo Fisher #28341) for 1 hour in the incubator. Chips were washed three times with water, and dried for one hour at room temperature. Chips were coated with 0.02mg/ml mouse Laminin (Sigma #L2020) diluted in culture media for one hour in the incubator. Laminin was aspirated when cells were ready to plate.

Whole organoids at week 15 were placed on the chip, and media was removed immediately before recording to promote organoid-electrode contact. 5 minute recordings were done for 6 organoids each for $FOXP1^{R513H/WT}$ or $FOXP1^{WT/WT}$ at D102. Spikes were sorted using kilosort2 and the mean firing rate (Hz) was calculated for the recording duration for the top 25 active electrodes, and the average of all mean firing rates for electrodes on one chip was calculated (**Fig. 3.8, H and I**). Organoid-level mean firing rates were compared using a one-sided Mann-Whitney test, based on the a priori hypothesis that FOXP1 mutations increase neuronal excitability, consistent with reported epilepsy in individuals with FOXP1 mutations. To confirm synaptic origin of the recorded activity, we treated the organoids with 1μM TTX for 5min and compared the mean firing rates before and after TTX treatment. (**Fig. 3.17I**).

Connectivity score: To analyze the electrode-to-electrode connectivity score, we used the Spike time tiling coefficient (sttc) as described.[132] First, spike trains are calculated for a pair of electrodes, for example, A and B, from the saved spike data stored in Maxwell raw recording. Then we calculate the proportion of spikes in A (PA) that fall within +- 10ms of any spikes in B. PB is calculated in a similar fashion. To account for any correlation by chance, we compute the proportion of the whole recording time, TA and TB, that falls within +- 10ms intervals of spikes in A and B respectively. Finally, to achieve a symmetric measure, sttc is calculated as:

$$STTC = 0.5\left(\frac{PA - TB}{1 - PATB} + \frac{PB - TA}{1 - PBTA}\right)$$

The denominators are chosen so that sttc values are in the range [-1, +1] and sttc is +1 for autocorrelation and -1 when PA (or PB) = 0, TB (or TA) =1. Sttc was calculated for 25 most active electrodes and represented as 2-dimensional heatmaps. Mean connectivity score was calculated from all these pairs and sttc > 0.1 was taken as a threshold for significant connections.

All analysis was performed with custom-written code in python using the following packages: NumPy, pandas, h5py, SciPy. All comparisons were done using Mann Whitney U test. (**Fig 3.17J**).

Burst and ISI analysis: For burst analysis we selected 25 most active channels (at least 4 spikes per minute). The spike trains were segmented by a maximum inter-spike interval of 100ms to detect bursts and within each segment a burst was retained if the number of spikes exceeded 3. Burst duration was defined as the time interval between the first and last spike in the burst. Burst frequency was calculated by normalizing the number of bursts by the recording duration. We also computed the proportion of all spikes that occurred in the detected bursts. We further calculated the mean inter spike interval (ISI) over the entire recording window for the top 25 electrodes for each sample. (**Fig 3.17K**).

References

1. C. Lord, T. S. Brugha, T. Charman, J. Cusack, G. Dumas, T. Frazier, E. J. H. Jones, R. M. Jones, A. Pickles, M. W. State, J. L. Taylor, J. Veenstra-VanderWeele, Autism spectrum disorder. Nat Rev Dis Primers 6, 5 (2020).
2. B. Tick, P. Bolton, F. Happé, M. Rutter, F. Rijsdijk, Heritability of autism spectrum disorders: a meta-analysis of twin studies. J. Child Psychol. Psychiatry 57, 585–595 (2016).
3. M. J. Maenner, Z. Warren, A. R. Williams, E. Amoakohene, A. V. Bakian, D. A. Bilder, M.S. Durkin, R. T. Fitzgerald, S. M. Furnier, M. M. Hughes, C. M. Ladd-Acosta, D. McArthur, E. T. Pas, A. Salinas, A. Vehorn, S. Williams, A. Esler, A. Grzybowski, J. Hall-Lande, R. H.N. Nguyen, K. Pierce, W. Zahorodny, A. Hudson, L. Hallas, K. C. Mancilla, M. Patrick, J. Shenouda, K. Sidwell, M. DiRienzo, J. Gutierrez, M. H. Spivey, M. Lopez, S. Pettygrove, Y.D. Schwenk, A. Washington, K. A. Shaw, Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. MMWR Surveill Summ 72, 1–14 (2023).
4. F. K. Satterstrom, J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J.-Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias, Autism Sequencing Consortium, iPSYCH-Broad Consortium, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, J. D. Buxbaum, Large-Scale Exome Sequencing Study Implicates

Both Developmental and Functional Changes in the Neurobiology of Autism. Cell 180, 568–584.e23 (2020).

5. B. Trost, B. Thiruvahindrapuram, A. J. S. Chan, W. Engchuan, E. J. Higginbotham, J. L. Howe, L. O. Loureiro, M. S. Reuter, D. Roshandel, J. Whitney, M. Zarrei, M. Bookman, C. Somerville, R. Shaath, M. Abdi, E. Aliyev, R. V. Patel, T. Nalpathamkalam, G. Pellecchia, O. Hamdan, G. Kaur, Z. Wang, J. R. MacDonald, J. Wei, W. W. L. Sung, S. Lamoureux, N. Hoang, T. Selvanayagam, N. Deflaux, M. Geng, S. Ghaffari, J. Bates, E. J. Young, Q. Ding, C. Shum, L. D'Abate, C. A. Bradley, A. Rutherford, V. Aguda, B. Apresto, N. Chen, S. Desai, X. Du, M. L. Y. Fong, S. Pullenayegum, K. Samler, T. Wang, K. Ho, T. Paton, S. L. Pereira, J.-A. Herbrick, R. F. Wintle, J. Fuerth, J. Noppornpitak, H. Ward, P. Magee, A. Al Baz, U. Kajendirarajah, S. Kapadia, J. Vlasblom, M. Valluri, J. Green, V. Seifer, M. Quirbach, O. Rennie, E. Kelley, N. Masjedi, C. Lord, M. J. Szego, M. H. Zawati, M. Lang, L. J. Strug, C. R. Marshall, G. Costain, K. Calli, A. Iaboni, A. Yusuf, P. Ambrozewicz, L. Gallagher, D. G. Amaral, J. Brian, M. Elsabbagh, S. Georgiades, D. S. Messinger, S. Ozonoff, J. Sebat, C. Sjaarda, I. M. Smith, P. Szatmari, L. Zwaigenbaum, A. Kushki, T. W. Frazier, J. A. S. Vorstman, K. A. Fakhro, B. A. Fernandez, M. E. S. Lewis, R. Weksberg, M. Fiume, R. K. C. Yuen, E. Anagnostou, N. Sondheimer, D. Glazer, D. M. Hartley, S. W. Scherer, Genomic architecture of autism from comprehensive whole-genome sequence annotation. Cell 185, 4409–4427.e18 (2022).

6. J. M. Fu, F. K. Satterstrom, M. Peng, H. Brand, R. L. Collins, S. Dong, B. Wamsley, L. Klei, L. Wang, S. P. Hao, C. R. Stevens, C. Cusick, M. Babadi, E. Banks, B. Collins, S. Dodge, S. B. Gabriel, L. Gauthier, S. K. Lee, L. Liang, A. Ljungdahl, B. Mahjani, L. Sloofman, A. N. Smirnov, M. Barbosa, C. Betancur, A. Brusco, B. H. Y. Chung, E. H.

Cook, M. L. Cuccaro, E. Domenici, G. B. Ferrero, J. J. Gargus, G. E. Herman, I. Hertz-Picciotto, P. Maciel, D. S. Manoach, M. R. Passos-Bueno, A. M. Persico, A. Renieri, J. S. Sutcliffe, F. Tassone, E. Trabetti, G. Campos, S. Cardaropoli, D. Carli, M. C. Y. Chan, C. Fallerini, E. Giorgio, A. C. Girardi, E. Hansen-Kiss, S. L. Lee, C. Lintas, Y. Ludena, R. Nguyen, L. Pavinato, M. Pericak-Vance, I. N. Pessah, R. J. Schmidt, M. Smith, C. I. S. Costa, S. Trajkova, J. Y. T. Wang, M. H. C. Yu, Autism Sequencing Consortium (ASC), Broad Institute Center for Common Disease Genomics (Broad-CCDG), iPSYCH-BROAD Consortium, D. J. Cutler, S. De Rubeis, J. D. Buxbaum, M. J. Daly, B. Devlin, K. Roeder, S. J. Sanders, M. E. Talkowski, Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat. Genet. 54, 1320–1331 (2022).

7. X. Zhou, P. Feliciano, C. Shu, T. Wang, I. Astrovskaya, J. B. Hall, J. U. Obiajulu, J. R. Wright, S. C. Murali, S. X. Xu, L. Brueggeman, T. R. Thomas, O. Marchenko, C. Fleisch, S. D. Barns, L. G. Snyder, B. Han, T. S. Chang, T. N. Turner, W. T. Harvey, A. Nishida, B. J. O'Roak, D. H. Geschwind, SPARK Consortium, J. J. Michaelson, N. Volfovsky, E. E. Eichler, Y. Shen, W. K. Chung, Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. Nat. Genet. 54, 1305–1319 (2022).

8. C. Lord, T. Charman, A. Havdahl, P. Carbone, E. Anagnostou, B. Boyd, T. Carr, P. J. de Vries, C. Dissanayake, G. Divan, C. M. Freitag, M. M. Gotelli, C. Kasari, M. Knapp, P. Mundy, A. Plank, L. Scahill, C. Servili, P. Shattuck, E. Simonoff, A. T. Singer, V. Slonims, P. P. Wang, M. C. Ysrraelit, R. Jellett, A. Pickles, J. Cusack, P. Howlin, P. Szatmari, A. Holbrook, C. Toolan, J. B. McCauley, The Lancet Commission on the future of care and clinical research in autism. Lancet 399, 271–334 (2022).

9. N. Sestan, M. W. State, Lost in Translation: Traversing the Complex Path from Genomics to Therapeutics in Autism Spectrum Disorder. Neuron 100, 406–423 (2018).

10. A. J. Willsey, M. T. Morris, S. Wang, H. R. Willsey, N. Sun, N. Teerikorpi, T. B. Baum, G. Cagney, K. J. Bender, T. A. Desai, D. Srivastava, G. W. Davis, J. Doudna, E. Chang, V. Sohal, D. H. Lowenstein, H. Li, D. Agard, M. J. Keiser, B. Shoichet, M. von Zastrow, L. Mucke, S. Finkbeiner, L. Gan, N. Sestan, M. E. Ward, R. Huttenhain, T. J. Nowakowski, H. J. Bellen, L. M. Frank, M. K. Khokha, R. P. Lifton, M. Kampmann, T. Ideker, M. W. State, N. J. Krogan, The Psychiatric Cell Map Initiative: A Convergent Systems Biological Approach to Illuminating Key Molecular Pathways in Neuropsychiatric Disorders. Cell 174, 505–520 (2018).

11. H. R. Willsey, A. J. Willsey, B. Wang, M. W. State, Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. Nat. Rev. Neurosci. 23, 323–341 (2022).

12. M. Kim, J. Park, M. Bouhaddou, K. Kim, A. Rojc, M. Modak, M. Soucheray, M. J. McGregor, P. O'Leary, D. Wolf, E. Stevenson, T. K. Foo, D. Mitchell, K. A. Herrington, D. P. Muñoz, B. Tutuncuoglu, K.-H. Chen, F. Zheng, J. F. Kreisberg, M. E. Diolaiti, J. D. Gordan, J.-P. Coppé, D. L. Swaney, B. Xia, L. van 't Veer, A. Ashworth, T. Ideker, N. J. Krogan, A protein interaction landscape of breast cancer. Science 374, eabf3066 (2021).

13. A. Kratz, M. Kim, M. R. Kelly, F. Zheng, C. A. Koczor, J. Li, K. Ono, Y. Qin, C. Churas, J. Chen, R. T. Pillich, J. Park, M. Modak, R. Collier, K. Licon, D. Pratt, R. W. Sobol, N. J. Krogan, T. Ideker, A multi-scale map of protein assemblies in the DNA damage response. Cell Syst 14, 447–463.e8 (2023).

14. D. L. Swaney, D. J. Ramms, Z. Wang, J. Park, Y. Goto, M. Soucheray, N. Bhola, K. Kim, F. Zheng, Y. Zeng, M. McGregor, K. A. Herrington, R. O'Keefe, N. Jin, N. K. VanLandingham, H. Foussard, J. Von Dollen, M. Bouhaddou, D. Jimenez-Morales, K. Obernier, J. F. Kreisberg, M. Kim, D. E. Johnson, N. Jura, J. R. Grandis, J. S. Gutkind, T. Ideker, N. J. Krogan, A protein network map of head and neck cancer reveals PIK3CA mutant drug sensitivity. Science 374, eabf2911 (2021).

15. M. Eckhardt, W. Zhang, A. M. Gross, J. Von Dollen, J. R. Johnson, K. E. Franks-Skiba, D. L. Swaney, T. L. Johnson, G. M. Jang, P. S. Shah, T. M. Brand, J. Archambault, J. F. Kreisberg, J. R. Grandis, T. Ideker, N. J. Krogan, Multiple Routes to Oncogenesis Are Promoted by the Human Papillomavirus-Host Protein Network. Cancer Discov. 8, 1474–1489 (2018).

16. L. M. Judge, J. A. Perez-Bermejo, A. Truong, A. J. Ribeiro, J. C. Yoo, C. L. Jensen, M. A. Mandegar, N. Huebsch, R. M. Kaake, P.-L. So, D. Srivastava, B. L. Pruitt, N. J. Krogan, B. R. Conklin, A BAG3 chaperone complex maintains cardiomyocyte function during proteotoxic stress. JCI Insight 2 (2017).

17. S. K. Hota, K. S. Rao, A. P. Blair, A. Khalilimeybodi, K. M. Hu, R. Thomas, K. So, V. Kameswaran, J. Xu, B. J. Polacco, R. V. Desai, N. Chatterjee, A. Hsu, J. M. Muncie, A. M. Blotnick, S. A. B. Winchester, L. S. Weinberger, R. Hüttenhain, I. S. Kathiriya, N. J. Krogan, J. J. Saucerman, B. G. Bruneau, Brahma safeguards canalization of cardiac mesoderm differentiation. Nature 602, 129–134 (2022).

18. B. Gonzalez-Teran, M. Pittman, F. Felix, R. Thomas, D. Richmond-Buccola, R. Hüttenhain, K. Choudhary, E. Moroni, M. W. Costa, Y. Huang, A. Padmanabhan, M. Alexanian, C. Y. Lee, B. E. J. Maven, K. Samse-Knapp, S. U. Morton, M. McGregor, C. A.

Gifford, J. G. Seidman, C. E. Seidman, B. D. Gelb, G. Colombo, B. R. Conklin, B. L. Black, B. G. Bruneau, N. J. Krogan, K. S. Pollard, D. Srivastava, Transcription factor protein interactomes reveal genetic determinants in heart disease. Cell 185, 794–814.e30 (2022).

19. L. Zhu, K. Choudhary, B. Gonzalez-Teran, Y.-S. Ang, R. Thomas, N. R. Stone, L. Liu, P. Zhou, C. Zhu, H. Ruan, Y. Huang, S. Jin, A. Pelonero, F. Koback, A. Padmanabhan, N. Sadagopan, A. Hsu, M. W. Costa, C. A. Gifford, J. G. van Bemmel, R. Hüttenhain, V. Vedantham, B. R. Conklin, B. L. Black, B. G. Bruneau, L. Steinmetz, N. J. Krogan, K. S. Pollard, D. Srivastava, Transcription Factor GATA4 Regulates Cell Type-Specific Splicing Through Direct Interaction With RNA in Human Induced Pluripotent Stem Cell-Derived Cardiac Progenitors. Circulation 146, 770–787 (2022).

20. T. E. Tracy, J. Madero-Pérez, D. L. Swaney, T. S. Chang, M. Moritz, C. Konrad, M. E. Ward, E. Stevenson, R. Hüttenhain, G. Kauwe, M. Mercedes, L. Sweetland-Martin, X. Chen, S.-A. Mok, M. Y. Wong, M. Telpoukhovskaia, S.-W. Min, C. Wang, P. D. Sohn, J. Martin, Y. Zhou, W. Luo, J. Q. Trojanowski, V. M. Y. Lee, S. Gong, G. Manfredi, G. Coppola, N. J. Krogan, D. H. Geschwind, L. Gan, Tau interactome maps synaptic and mitochondrial processes associated with neurodegeneration. Cell 185, 712–728.e14 (2022).

21. S. Jäger, P. Cimermancic, N. Gulbahce, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H. Hernandez, G. M. Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D'Orso, J. Fernandes, M. Fahey, C. Mahon, A. J. O'Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, J. A. Young, S. K. Chanda, W. I. Sundquist, T. Kortemme, R. D. Hernandez, C. S. Craik, A. Burlingame,

A. Sali, A. D. Frankel, N. J. Krogan, Global landscape of HIV-human protein complexes. Nature 481, 365–370 (2011).

22. P. S. Shah, N. Link, G. M. Jang, P. P. Sharp, T. Zhu, D. L. Swaney, J. R. Johnson, J. Von Dollen, H. R. Ramage, L. Satkamp, B. Newton, R. Hüttenhain, M. J. Petit, T. Baum, A. Everitt, O. Laufman, M. Tassetto, M. Shales, E. Stevenson, G. N. Iglesias, L. Shokat, S. Tripathi, V. Balasubramaniam, L. G. Webb, S. Aguirre, A. J. Willsey, A. Garcia-Sastre, K. S. Pollard, S. Cherry, A. V. Gamarnik, I. Marazzi, J. Taunton, A. Fernandez-Sesma, H. J. Bellen, R. Andino, N. J. Krogan, Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. Cell 175, 1931–1945.e18 (2018).

23. J. Batra, J. F. Hultquist, D. Liu, O. Shtanko, J. Von Dollen, L. Satkamp, G. M. Jang, P. Luthra, T. M. Schwarz, G. I. Small, E. Arnett, M. Anantpadma, A. Reyes, D. W. Leung, R. Kaake, P. Haas, C. B. Schmidt, L. S. Schlesinger, D. J. LaCount, R. A. Davey, G. K. Amarasinghe, C. F. Basler, N. J. Krogan, Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. Cell 175, 1917–1930.e13 (2018).

24. D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Hüttenhain, R. M. Kaake, A. L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B. J. Polacco, H. Braberg, J. M. Fabius, M. Eckhardt, M. Soucheray, M. J. Bennett, M. Cakir, M. J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. Mac Kain, L. Miorin, E. Moreno, Z. Z. C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I. T. Kirby, J. E. Melnyk, J. S. Chorba, K. Lou, S. A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C. J. P. Mathy, T. Perica, K. B. Pilla, S. J. Ganesan, D. J.

Saltzberg, R. Rakesh, X. Liu, S. B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy-Lugo, Y. Lin, X.-P. Huang, Y. Liu, S. A. Wankowicz, M. Bohn, M. Safari, F. S. Ugur, C. Koh, N. S. Savar, Q. D. Tran, D. Shengjuler, S. J. Fletcher, M. C. O'Neal, Y. Cai, J. C. J. Chang, D. J. Broadhurst, S. Klippsten, P. P. Sharp, N. A. Wenzell, D. Kuzuoglu-Ozturk, H.-Y. Wang, R. Trenker, J. M. Young, D. A. Cavero, J. Hiatt, T. L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R. M. Stroud, A. D. Frankel, O. S. Rosenberg, K. A. Verba, D. A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H. S. Malik, D. G. Fujimori, T. Ideker, C. S. Craik, S. N. Floor, J. S. Fraser, J. D. Gross, A. Sali, B. L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K. M. Shokat, B. K. Shoichet, N. J. Krogan, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583, 459–468 (2020).

25. D. E. Gordon, J. Hiatt, M. Bouhaddou, V. V. Rezelj, S. Ulferts, H. Braberg, A. S. Jureka, K. Obernier, J. Z. Guo, J. Batra, R. M. Kaake, A. R. Weckstein, T. W. Owens, M. Gupta, S. Pourmal, E. W. Titus, M. Cakir, M. Soucheray, M. McGregor, Z. Cakir, G. Jang, M. J. O'Meara, T. A. Tummino, Z. Zhang, H. Foussard, A. Rojc, Y. Zhou, D. Kuchenov, R. Hüttenhain, J. Xu, M. Eckhardt, D. L. Swaney, J. M. Fabius, M. Ummadi, B. Tutuncuoglu, U. Rathore, M. Modak, P. Haas, K. M. Haas, Z. Z. C. Naing, E. H. Pulido, Y. Shi, I. Barrio-Hernandez, D. Memon, E. Petsalaki, A. Dunham, M. C. Marrero, D. Burke, C. Koh, T. Vallet, J. A. Silvas, C. M. Azumaya, C. Billesbølle, A. F. Brilot, M. G. Campbell, A. Diallo, M. S. Dickinson, D. Diwanji, N. Herrera, N. Hoppe, H. T. Kratochvil, Y. Liu, G. E. Merz, M. Moritz, H. C. Nguyen, C. Nowotny, C. Puchades, A. N. Rizo, U. Schulze-Gahmen, A. M. Smith, M. Sun, I. D. Young, J. Zhao, D. Asarnow, J. Biel, A. Bowen, J. R. Braxton, J.

Chen, C. M. Chio, U. S. Chio, I. Deshpande, L. Doan, B. Faust, S. Flores, M. Jin, K. Kim, V. L. Lam, F. Li, J. Li, Y.-L. Li, Y. Li, X. Liu, M. Lo, K. E. Lopez, A. A. Melo, F. R. Moss 3rd, P. Nguyen, J. Paulino, K. I. Pawar, J. K. Peters, T. H. Pospiech Jr, M. Safari, S. Sangwan, K. Schaefer, P. V. Thomas, A. C. Thwin, R. Trenker, E. Tse, T. K. M. Tsui, F. Wang, N. Whitis, Z. Yu, K. Zhang, Y. Zhang, F. Zhou, D. Saltzberg, QCRG Structural Biology Consortium, A. J. Hodder, A. S. Shun-Shion, D. M. Williams, K. M. White, R. Rosales, T. Kehrer, L. Miorin, E. Moreno, A. H. Patel, S. Rihn, M. M. Khalid, A. Vallejo-Gracia, P. Fozouni, C. R. Simoneau, T. L. Roth, D. Wu, M. A. Karim, M. Ghoussaini, I. Dunham, F. Berardi, S. Weigang, M. Chazal, J. Park, J. Logue, M. McGrath, S. Weston, R. Haupt, C. J. Hastie, M. Elliott, F. Brown, K. A. Burness, E. Reid, M. Dorward, C. Johnson, S. G. Wilkinson, A. Geyer, D. M. Giesel, C. Baillie, S. Raggett, H. Leech, R. Toth, N. Goodman, K. C. Keough, A. L. Lind, Zoonomia Consortium, R. J. Klesh, K. R. Hemphill, J. Carlson-Stevermer, J. Oki, K. Holden, T. Maures, K. S. Pollard, A. Sali, D. A. Agard, Y. Cheng, J. S. Fraser, A. Frost, N. Jura, T. Kortemme, A. Manglik, D. R. Southworth, R. M. Stroud, D. R. Alessi, P. Davies, M. B. Frieman, T. Ideker, C. Abate, N. Jouvenet, G. Kochs, B. Shoichet, M. Ott, M. Palmarini, K. M. Shokat, A. García-Sastre, J. A. Rassen, R. Grosse, O. S. Rosenberg, K. A. Verba, C. F. Basler, M. Vignuzzi, A. A. Peden, P. Beltrao, N. J. Krogan, Comparative host- coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science 370 (2020).

26. J. Hiatt, J. F. Hultquist, M. J. McGregor, M. Bouhaddou, R. T. Leenay, L. M. Simons, J. M. Young, P. Haas, T. L. Roth, V. Tobin, J. A. Wojcechowskyj, J. M. Woo, U. Rathore, D. A. Cavero, E. Shifrut, T. T. Nguyen, K. M. Haas, H. S. Malik, J. A. Doudna, A. P. May, A.

Marson, N. J. Krogan, A functional map of HIV-host interactions in primary human T cells. Nat. Commun. 13, 1752 (2022).

27. M. Bouhaddou, A.-K. Reuschl, B. J. Polacco, L. G. Thorne, M. R. Ummadi, C. Ye, R. Rosales, A. Pelin, J. Batra, G. M. Jang, J. Xu, J. M. Moen, A. L. Richards, Y. Zhou, B. Harjai, E. Stevenson, A. Rojc, R. Ragazzini, M. V. X. Whelan, W. Furnon, G. De Lorenzo, V. Cowton, A. M. Syed, A. Ciling, N. Deutsch, D. Pirak, G. Dowgier, D. Mesner, J. L. Turner, B. L. McGovern, M. L. Rodriguez, R. Leiva-Rebollo, A. S. Dunham, X. Zhong, M. Eckhardt, A. Fossati, N. F. Liotta, T. Kehrer, A. Cupic, M. Rutkowska, I. Mena, S. Aslam, A. Hoffert, H. Foussard, C. O. Olwal, W. Huang, T. Zwaka, J. Pham, M. Lyons, L. Donohue, A. Griffin, R. Nugent, K. Holden, R. Deans, P. Aviles, J. A. Lopez-Martin, J. M. Jimeno, K. Obernier, J. M. Fabius, M. Soucheray, R. Hüttenhain, I. Jungreis, M. Kellis, I. Echeverria, K. Verba, P. Bonfanti, P. Beltrao, R. Sharan, J. A. Doudna, L. Martinez-Sobrido, A. H. Patel, M. Palmarini, L. Miorin, K. White, D. L. Swaney, A. Garcia-Sastre, C. Jolly, L. Zuliani-Alvarez, G. J. Towers, N. J. Krogan, SARS-CoV-2 variants evolve convergent strategies to remodel the host response. Cell 186, 4597–4614.e26 (2023).

28. H. J. Kang, Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak, T. Guennel, Y. Shin, M. B. Johnson, Z. Krsnik, S. Mayer, S. Fertuzinhos, S. Umlauf, S. N. Lisgo, A. Vortmeyer, D. R. Weinberger, S. Mane, T. M. Hyde, A. Huttner, M. Reimers, J. E. Kleinman, N. Sestan, Spatio-temporal transcriptome of the human brain. Nature 478, 483–489 (2011).

29. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA,

Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts:, Laboratory, Data Analysis &Coordinating Center (LDACC):, NIH program management:, Biospecimen collection:, Pathology:, eQTL manuscript working group:, A. Battle, C. D. Brown, B. E. Engelhardt, S. B. Montgomery, Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).

30. A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan, M. W. State, Co-expression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155, 997–1007 (2013).

31. T. Singh, T. Poterba, D. Curtis, H. Akil, M. Al Eissa, J. D. Barchas, N. Bass, T. B. Bigdeli, G. Breen, E. J. Bromet, P. F. Buckley, W. E. Bunney, J. Bybjerg-Grauholm, W. F. Byerley, S. B. Chapman, W. J. Chen, C. Churchhouse, N. Craddock, C. M. Cusick, L. DeLisi, S. Dodge, M. A. Escamilla, S. Eskelinen, A. H. Fanous, S. V. Faraone, A. Fiorentino, L. Francioli, S. B. Gabriel, D. Gage, S. A. Gagliano Taliun, A. Ganna, G. Genovese, D. C. Glahn, J. Grove, M.-H. Hall, E. Hämäläinen, H. O. Heyne, M. Holi, D. M. Hougaard, D. P. Howrigan, H. Huang, H.-G. Hwu, R. S. Kahn, H. M. Kang, K. J. Karczewski, G. Kirov, J. A. Knowles, F. S. Lee, D. S. Lehrer, F. Lescai, D. Malaspina, S. R. Marder, S. A.

McCarroll, A. M. McIntosh, H. Medeiros, L. Milani, C. P. Morley, D. W. Morris, P. B. Mortensen, R. M. Myers, M. Nordentoft, N. L. O'Brien, A. M. Olivares, D. Ongur, W. H. Ouwehand, D. S. Palmer, T. Paunio, D. Quested, M. H. Rapaport, E. Rees, B. Rollins, F. K. Satterstrom, A. Schatzberg, E. Scolnick, L. J. Scott, S. I. Sharp, P. Sklar, J. W. Smoller, J. L. Sobell, M. Solomonson, E. A. Stahl, C. R. Stevens, J. Suvisaari, G. Tiao, S. J. Watson, N. A. Watts, D. H. Blackwood, A. D. Børglum, B. M. Cohen, A. P. Corvin, T. Esko, N. B. Freimer, S. J. Glatt, C. M. Hultman, A. McQuillin, A. Palotie, C. N. Pato, M. T. Pato, A. E. Pulver, D. St Clair, M. T. Tsuang, M. P. Vawter, J. T. Walters, T. M. Werge, R. A. Ophoff, P. F. Sullivan, M. J. Owen, M. Boehnke, M. C. O'Donovan, B. M. Neale, M. J. Daly, Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature, doi: 10.1038/s41586-022-04556-w (2022).

32. T. J. Nowakowski, A. Bhaduri, A. A. Pollen, B. Alvarado, M. A. Mostajo-Radji, E. Di Lullo, M. Haeussler, C. Sandoval-Espinosa, S. J. Liu, D. Velmeshev, J. R. Ounadjela, J. Shuga, X. Wang, D. A. Lim, J. A. West, A. A. Leyrat, W. J. Kent, A. R. Kriegstein, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science 358, 1318–1323 (2017).

33. D. Szklarczyk, A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen, C. von Mering, The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 49, D605–D612 (2021).

34. G. Pintacuda, Y.-H. H. Hsu, K. Tsafou, K. W. Li, J. M. Martín, J. Riseman, J. C. Biagini, J. K. T. Ching, D. Mena, M. A. Gonzalez-Lozano, S. B. Egri, J. Jaffe, A. B. Smit, N. Fornelos, K. C. Eggan, K. Lage, Protein interaction studies in human induced neurons

indicate convergent biology underlying autism spectrum disorders. Cell Genom 3, 100250 (2023).

35. Gene Ontology Consortium, S. A. Aleksander, J. Balhoff, S. Carbon, J. M. Cherry, H. J. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N. L. Harris, D. P. Hill, R. Lee, H. Mi, S. Moxon, C. J. Mungall, A. Muruganugan, T. Mushayahama, P. W. Sternberg, P. D. Thomas, K. Van Auken, J. Ramsey, D. A. Siegele, R. L. Chisholm, P. Fey, M. C. Aspromonte, M. V. Nugnes, F. Quaglia, S. Tosatto, M. Giglio, S. Nadendla, G. Antonazzo, H. Attrill, G. Dos Santos, S. Marygold, V. Strelets, C. J. Tabone, J. Thurmond, P. Zhou, S. H. Ahmed, P. Asanitthong, D. Luna Buitrago, M. N. Erdol, M. C. Gage, M. Ali Kadhum, K. Y. C. Li, M. Long, A. Michalak, A. Pesala, A. Pritazahra, S. C. C. Saverimuttu, R. Su, K. E. Thurlow, R. C. Lovering, C. Logie, S. Oliferenko, J. Blake, K. Christie, L. Corbani, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, C. Smith, A. Cuzick, J. Seager, L. Cooper, J. Elser, P. Jaiswal, P. Gupta, P. Jaiswal, S. Naithani, M. Lera-Ramirez, K. Rutherford, V. Wood, J. L. De Pons, M. R. Dwinell, G. T. Hayman, M. L. Kaldunski, A. E. Kwitek, S. J. F. Laulederkind, M. A. Tutaj, M. Vedi, S.-J. Wang, P. D'Eustachio, L. Aimo, K. Axelsen, A. Bridge, N. Hyka- Nouspikel, A. Morgat, S. A. Aleksander, J. M. Cherry, S. R. Engel, K. Karra, S. R. Miyasato, R. S. Nash, M. S. Skrzypek, S. Weng, E. D. Wong, E. Bakker, T. Z. Berardini, L. Reiser, A. Auchincloss, K. Axelsen, G. Argoud-Puy, M.-C. Blatter, E. Boutet, L. Breuza, A. Bridge, C. Casals-Casas, E. Coudert, A. Estreicher, M. Livia Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, C. Hulo, N. Hyka-Nouspikel, F. Jungo, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, I. Pedruzzi, L. Pourcel, S. Poux, C. Rivoire, S. Sundaram, A. Bateman, E. Bowler-Barnett, H. Bye-A-Jee, P. Denny, A. Ignatchenko, R. Ishtiaq, A. Lock, Y. Lussi, M. Magrane, M. J. Martin, S. Orchard, P. Raposo, E. Speretta,

N. Tyagi, K. Warner, R. Zaru, A. D. Diehl, R. Lee, J. Chan, S. Diamantakis, D. Raciti, M. Zarowiecki, M. Fisher, C. James-Zorn, V. Ponferrada, A. Zorn, S. Ramachandran, L. Ruzicka, M. Westerfield, The Gene Ontology knowledgebase in 2023. Genetics 224 (2023).

36. D. S. Manoli, M. W. State, Autism Spectrum Disorder Genetics and the Search for Pathological Mechanisms. Am. J. Psychiatry 178, 30–38 (2021).

37. N. Sun, N. Teyssier, B. Wang, S. Drake, M. Seyler, Y. Zaltsman, A. Everitt, N. Teerikorpi, H. R. Willsey, H. Goodarzi, R. Tian, M. Kampmann, A. J. Willsey, Autism genes converge on microtubule biology and RNA-binding proteins during excitatory neurogenesis. bioRxiv, doi: 10.1101/2023.12.22.573108 (2024).

38. N. Teerikorpi, M. C. Lasser, S. Wang, E. Kostyanovskaya, E. Bader, N. Sun, J. Dea, T. J. Nowakowski, A. Jeremy Willsey, H. R. Willsey, Ciliary biology intersects autism and congenital heart disease, bioRxiv (2024)p. 2024.07.30.602578.

39. E. Kostyanovskaya, M. C. Lasser, B. Wang, J. Schmidt, E. Bader, C. Buteo, J. Arbelaez, A. R. Sindledecker, K. E. McCluskey, O. Castillo, S. Wang, J. Dea, K. A. Helde, J. Michael Graglia, E. Brimble, D. B. Kastner, A. T. Ehrlich, Matthew W. State, A. Jeremy Willsey, H. R. Willsey, Convergence of autism proteins at the cilium, bioRxiv (2024)p. 2024.12.05.626924.

40. K. E. McCluskey, K. M. Stovell, K. Law, E. Kostyanovskaya, J. Schmidt, C. R. T. Exner, J. Dea, E. Brimble, M. W. State, A. J. Willsey, H. R. Willsey, Autism gene variants disrupt enteric neuron migration and cause gastrointestinal dysmotility, bioRxiv (2024). https://doi.org/10.1101/2024.05.28.593642.

41. A. J. Willsey, T. V. Fernandez, D. Yu, R. A. King, A. Dietrich, J. Xing, S. J. Sanders, J. D. Mandell, A. Y. Huang, P. Richer, L. Smith, S. Dong, K. E. Samocha, Tourette International Collaborative Genetics (TIC Genetics), Tourette Syndrome Association International Consortium for Genetics (TSAICG), B. M. Neale, G. Coppola, C. A. Mathews, J. A. Tischfield, J. M. Scharf, M. W. State, G. A. Heiman, De Novo Coding Variants Are Strongly Associated with Tourette Disorder. Neuron 94, 486–499.e9 (2017).

42. S. Kaushik, F. Haderk, X. Zhao, H.-M. Hu, K. N. Shah, G. M. Jang, V. Olivas, S. Nanjo, J. Jascur, V. B. Masto, D. Ciznadija, I. Sloma, E. Gross, S. L. Weinrich, J. R. Johnson, T. G. Bivona, N. J. Krogan, S. Bandyopadhyay, A tyrosine kinase protein interaction map reveals targetable EGFR network oncogenesis in lung cancer, bioRxiv (2020)p. 2020.07.02.185173.

43. D. Halder, C.-H. Lee, J. Y. Hyun, G.-E. Chang, E. Cheong, I. Shin, Suppression of Sin3A activity promotes differentiation of pluripotent cells into functional neurons. Sci. Rep. 7, 44818 (2017).

44. A. Roopra, L. Sharling, I. C. Wood, T. Briggs, U. Bachfischer, A. J. Paquette, N. J. Buckley, Transcriptional repression by neuron-restrictive silencer factor is mediated via the Sin3-histone deacetylase complex. Mol. Cell. Biol. 20, 2147–2157 (2000).

45. M. Quevedo, L. Meert, M. R. Dekker, D. H. W. Dekkers, J. H. Brandsma, D. L. C. van den Berg, Z. Ozgür, W. F. J. van IJcken, J. Demmers, M. Fornerod, R. A. Poot, Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. Nat. Commun. 10, 2669 (2019).

46. N. Ding, H. Zhou, P.-O. Esteve, H. G. Chin, S. Kim, X. Xu, S. M. Joseph, M. J. Friez, C. E. Schwartz, S. Pradhan, T. G. Boyer, Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. Mol. Cell 31, 347–359 (2008).

47. M. J. Jurynec, X. Bai, B. W. Bisgrove, H. Jackson, A. Nechiporuk, R. A. S. Palu, H. A. Grunwald, Y.-C. Su, K. Hoshijima, H. J. Yost, L. I. Zon, D. J. Grunwald, The Paf1 complex and P-TEFb have reciprocal and antagonist roles in maintaining multipotent neural crest progenitors. Development 146 (2019).

48. C. Zhang, L. A. Mejia, J. Huang, P. Valnegri, E. J. Bennett, J. Anckar, A. Jahani-Asl, G. Gallardo, Y. Ikeuchi, T. Yamada, M. Rudnicki, J. W. Harper, A. Bonni, The X-linked intellectual disability protein PHF6 associates with the PAF1 complex and regulates neuronal migration in the mammalian brain. Neuron 78, 986–993 (2013).

49. E. C. Ung, N. A. Borja, LEO1 haploinsufficiency is associated with developmental delays and autism spectrum disorder. J. Hum. Genet. 71, 109–111 (2026).

50. R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, O. Ronneberger, S. Bodenstein, M. Zielinski, A. Bridgland, A. Potapenko, A. Cowie, K. Tunyasuvunakool, R. Jain, E. Clancy, P. Kohli, J. Jumper, D. Hassabis, Protein complex prediction with AlphaFold-Multimer, bioRxiv (2022)p. 2021.10.04.463034.

51. P. Bryant, G. Pozzati, A. Elofsson, Improved prediction of protein-protein interactions using AlphaFold2. Nat. Commun. 13, 1265 (2022).

52. S. G. Choi, J. Olivet, P. Cassonnet, P.-O. Vidalain, K. Luck, L. Lambourne, K. Spirohn, I. Lemmens, M. Dos Santos, C. Demeret, L. Jones, S. Rangarajan, W. Bian, E. P. Coutant, Y. L. Janin, S. van der Werf, P. Trepte, E. E. Wanker, J. De Las Rivas, J. Tavernier, J.-C.

Twizere, T. Hao, D. E. Hill, M. Vidal, M. A. Calderwood, Y. Jacob, Maximizing binary interactome mapping with a minimal number of assays. Nat Commun 10, 3907 (2019).

53. Y.-C. Chen, S. V. Rajagopala, T. Stellberger, P. Uetz, Exhaustive benchmarking of the yeast two-hybrid system. Nat Methods 7, 667–8; author reply 668 (2010).

54. V. J. Heintz, L. Wang, D. J. LaCount, NanoLuc luciferase as a quantitative yeast two-hybrid reporter. FEMS yeast research 21 (2022).

55. X. Peng, J. Wang, W. Peng, F.-X. Wu, Y. Pan, Protein-protein interactions: detection, reliability assessment and applications. Brief Bioinform 18, 798–819 (2017).

56. P. Braun, M. Tasan, M. Dreze, M. Barrios-Rodiles, I. Lemmens, H. Yu, J. M. Sahalie, R. R. Murray, L. Roncari, A.-S. de Smet, K. Venkatesan, J.-F. Rual, J. Vandenhaute, M. E. Cusick, T. Pawson, D. E. Hill, J. Tavernier, J. L. Wrana, F. P. Roth, M. Vidal, An experimentally derived confidence score for binary protein-protein interactions. Nat Methods 6, 91–97 (2009).

57. P. Magini, D. J. Smits, L. Vandervore, R. Schot, M. Columbaro, E. Kasteleijn, M. van der Ent, F. Palombo, M. H. Lequin, M. Dremmen, M. C. Y. de Wit, M. Severino, M. T. Divizia, P. Striano, N. Ordonez-Herrera, A. Alhashem, A. Al Fares, M. Al Ghamdi, A. Rolfs, P. Bauer, J. Demmers, F. W. Verheijen, M. Wilke, M. van Slegtenhorst, P. J. van der Spek, M. Seri, A. C. Jansen, R. W. Stottmann, R. B. Hufnagel, R. J. Hopkin, D. Aljeaid, W. Wiszniewski, P. Gawlinski, M. Laure-Kamionowska, F. S. Alkuraya, H. Akleh, V. Stanley, D. Musaev, J. G. Gleeson, M. S. Zaki, N. Brunetti-Pierri, G. Cappuccio, B. Davidov, L. Basel-Salmon, L. Bazak, N. R. Shahar, A. Bertoli-Avella, G. M. Mirzaa, W. B. Dobyns, T. Pippucci, M. Fornerod, G. M. S. Mancini, Loss of SMPD4 Causes a Developmental

Disorder Characterized by Microcephaly and Congenital Arthrogryposis. Am J Hum Genet 105, 689–705 (2019).

58. D. Yu, C. Cattoglio, Y. Xue, Q. Zhou, A complex between DYRK1A and DCAF7 phosphorylates the C-terminal domain of RNA polymerase II to promote myogenesis. Nucleic Acids Res. 47, 4462–4475 (2019).

59. F. Glenewinkel, M. J. Cohen, C. R. King, S. Kaspar, S. Bamberg-Lemper, J. S. Mymryk, W. Becker, The adaptor protein DCAF7 mediates the interaction of the adenovirus E1A oncoprotein with the protein kinases DYRK1A and HIPK2. Sci. Rep. 6, 28241 (2016).

60. Y. Miyata, E. Nishida, Identification of FAM53C as a cytosolic-anchoring inhibitory binding protein of the kinase DYRK1A. Life Sci Alliance 6 (2023).

61. D. Polioudakis, L. de la Torre-Ubieta, J. Langerman, A. G. Elkins, X. Shi, J. L. Stein, C. K. Vuong, S. Nichterwitz, M. Gevorgian, C. K. Opland, D. Lu, W. Connell, E. K. Ruzzo, J. K. Lowe, T. Hadzic, F. I. Hinz, S. Sabri, W. E. Lowry, M. B. Gerstein, K. Plath, D. H. Geschwind, A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron 103, 785–801.e8 (2019).

62. A. Bhaduri, C. Sandoval-Espinosa, M. Otero-Garcia, I. Oh, R. Yin, U. C. Eze, T. J. Nowakowski, A. R. Kriegstein, An atlas of cortical arealization identifies dynamic molecular signatures. Nature 598, 200–204 (2021).

63. R. Oughtred, J. Rust, C. Chang, B.-J. Breitkreutz, C. Stark, A. Willems, L. Boucher, G. Leung, N. Kolas, F. Zhang, S. Dolma, J. Coulombe-Huntington, A. Chatr-Aryamontri, K. Dolinski, M. Tyers, The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. 30, 187–200 (2021).

64. J. Xiang, S. Yang, N. Xin, M. A. Gaertig, R. H. Reeves, S. Li, X.-J. Li, DYRK1A regulates Hap1-Dcaf7/WDR68 binding with implication for delayed growth in Down syndrome. Proc. Natl. Acad. Sci. U. S. A. 114, E1224–E1233 (2017).

65. Q. Wang, Z. Geng, Y. Gong, K. Warren, H. Zheng, Y. Imamura, Z. Gao, WDR68 is essential for the transcriptional activation of the PRC1-AUTS2 complex and neuronal differentiation of mouse embryonic stem cells. Stem Cell Res. 33, 206–214 (2018).

66. M. Lasser, N. Sun, Y. Xu, S. Wang, S. Drake, K. Law, S. Gonzalez, B. Wang, V. Drury, O. Castillo, Y. Zaltsman, J. Dea, E. Bader, K. E. McCluskey, M. W. State, A. J. Willsey, H. R. Willsey, Pleiotropy of autism-associated chromatin regulators. Development 150 (2023).

67. H. R. Willsey, Y. Xu, A. Everitt, J. Dea, C. R. T. Exner, A. J. Willsey, M. W. State, R. M. Harland, The neurodevelopmental disorder risk gene DYRK1A is required for ciliogenesis and control of brain size in Xenopus embryos. Development 147, dev189290 (2020).

68. H. R. Willsey, C. R. T. Exner, Y. Xu, A. Everitt, N. Sun, B. Wang, J. Dea, G. Schmunk, Y. Zaltsman, N. Teerikorpi, A. Kim, A. S. Anderson, D. Shin, M. Seyler, T. J. Nowakowski, R. M. Harland, A. J. Willsey, M. W. State, Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. Neuron 109, 788–804.e8 (2021).

69. A. Jourdon, F. Wu, J. Mariani, D. Capauto, S. Norton, L. Tomasini, A. Amiri, M. Suvakov, J. D. Schreiner, Y. Jang, A. Panda, C. K. Nguyen, E. M. Cummings, G. Han, K. Powell, A. Szekely, J. C. McPartland, K. Pelphrey, K. Chawarska, P. Ventola, A. Abyzov, F. M. Vaccarino, Modeling idiopathic autism in forebrain organoids reveals an imbalance of excitatory cortical neuron subtypes during early neurogenesis. Nat Neurosci 26, 1505–1515 (2023).

70. S. Frendo-Cumbo, T. Li, D. A. Ammendolia, E. Coyaud, E. M. N. Laurent, Y. Liu, P. J. Bilan, G. Polevoy, B. Raught, J. A. Brill, A. Klip, J. H. Brumell, DCAF7 regulates cell proliferation through IRS1-FOXO1 signaling. iScience 25 (2022).

71. S. Najas, J. Arranz, P. A. Lochhead, A. L. Ashford, D. Oxley, J. M. Delabar, S. J. Cook, M. J. Barallobre, M. L. Arbonés, DYRK1A-mediated Cyclin D1 Degradation in Neural Stem Cells Contributes to the Neurogenic Cortical Defects in Down Syndrome. EBioMedicine 2, 120–134 (2015).

72. A. Recasens, S. J. Humphrey, M. Ellis, M. Hoque, R. H. Abbassi, B. Chen, M. Longworth, E. J. Needham, D. E. James, T. G. Johns, B. W. Day, M. Kassiou, P. Yang, L. Munoz, Global phosphoproteomics reveals DYRK1A regulates CDK1 activity in glioblastoma cells. Cell Death Discov 7, 81 (2021).

73. V. Graham, J. Khudyakov, P. Ellis, L. Pevny, SOX2 functions to maintain neural progenitor identity. Neuron 39, 749–765 (2003).

74. N. C. Hettige, H. Peng, H. Wu, X. Zhang, V. Yerko, Y. Zhang, M. Jefri, V. Soubannier, G. Maussion, S. Alsuwaidi, A. Ni, C. Rocha, J. Krishnan, V. McCarty, L. Antonyan, A. Schuppert, G. Turecki, E. A. Fon, T. M. Durcan, C. Ernst, FOXG1 dose tunes cell proliferation dynamics in human forebrain progenitor cells. Stem Cell Reports 17, 475–488 (2022).

75. S. F. Bellmaine, D. A. Ovchinnikov, D. T. Manallack, C. E. Cuddy, A. G. Elefanty, E. G. Stanley, E. J. Wolvetang, S. J. Williams, M. Pera, Inhibition of DYRK1A disrupts neural lineage specification in human pluripotent stem cells. Elife 6 (2017).

76. K. E. Samocha, J. A. Kosmicki, K. J. Karczewski, A. H. O'Donnell-Luria, E. Pierce-Hoffman, D. G. MacArthur, B. M. Neale, M. J. Daly, Regional missense constraint improves variant deleteriousness prediction, bioRxiv (2017)p. 148353.

77. E. L. Huttlin, L. Ting, R. J. Bruckner, F. Gebreab, M. P. Gygi, J. Szpyt, S. Tam, G. Zarraga, G. Colby, K. Baltier, R. Dong, V. Guarani, L. P. Vaites, A. Ordureau, R. Rad, B. K. Erickson, M. Wühr, J. Chick, B. Zhai, D. Kolippakkam, J. Mintseris, R. A. Obar, T. Harris, S. Artavanis-Tsakonas, M. E. Sowa, P. De Camilli, J. A. Paulo, J. W. Harper, S. P. Gygi, The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell 162, 425–440 (2015).

78. M. Y. Hein, N. C. Hubner, I. Poser, J. Cox, N. Nagaraj, Y. Toyoda, I. A. Gak, I. Weisswange, J. Mansfeld, F. Buchholz, A. A. Hyman, M. Mann, A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell 163, 712–723 (2015).

79. A. Calderone, L. Castagnoli, G. Cesareni, mentha: a resource for browsing integrated protein-interaction networks. Nat. Methods 10, 690–691 (2013).

80. K. Luck, D.-K. Kim, L. Lambourne, K. Spirohn, B. E. Begg, W. Bian, R. Brignall, T. Cafarelli, F. J. Campos-Laborie, B. Charloteaux, D. Choi, A. G. Coté, M. Daley, S. Deimling, A. Desbuleux, A. Dricot, M. Gebbia, M. F. Hardy, N. Kishore, J. J. Knapp, I. A. Kovács, I. Lemmens, M. W. Mee, J. C. Mellor, C. Pollis, C. Pons, A. D. Richardson, S. Schlabach, B. Teeking, A. Yadav, M. Babor, D. Balcha, O. Basha, C. Bowman-Colin, S.-F. Chin, S. G. Choi, C. Colabella, G. Coppin, C. D'Amata, D. De Ridder, S. De Rouck, M. Duran-Frigola, H. Ennajdaoui, F. Goebels, L. Goehring, A. Gopal, G. Haddad, E. Hatchi, M. Helmy, Y. Jacob, Y. Kassa, S. Landini, R. Li, N. van Lieshout, A. MacWilliams, D.

Markey, J. N. Paulson, S. Rangarajan, J. Rasla, A. Rayhan, T. Rolland, A. San-Miguel, Y. Shen, D. Sheykhkarimli, G. M. Sheynkman, E. Simonovsky, M. Taşan, A. Tejeda, V. Tropepe, J.-C. Twizere, Y. Wang, R. J. Weatheritt, J. Weile, Y. Xia, X. Yang, E. Yeger-Lotem, Q. Zhong, P. Aloy, G. D. Bader, J. De Las Rivas, S. Gaudet, T. Hao, J. Rak, J. Tavernier, D. E. Hill, M. Vidal, F. P. Roth, M. A. Calderwood, A reference map of the human binary protein interactome. Nature 580, 402–408 (2020).

81. G. G. Tall, A. M. Krumins, A. G. Gilman, Mammalian Ric-8A (synembryn) is a heterotrimeric Galpha protein guanine nucleotide exchange factor. J. Biol. Chem. 278, 8356–8362 (2003).

82. L. Braccioli, S. J. Vervoort, Y. Adolfs, C. J. Heijnen, O. Basak, R. J. Pasterkamp, C. H. Nijboer, P. J. Coffer, FOXP1 Promotes Embryonic Neural Stem Cell Differentiation by Repressing Jagged1 Expression. Stem Cell Reports 9, 1530–1545 (2017).

83. C. A. Pearson, D. M. Moore, H. O. Tucker, J. D. Dekker, H. Hu, A. Miquelajáuregui, B. G. Novitch, Foxp1 Regulates Neural Stem Cell Self-Renewal and Bias Toward Deep Layer Cortical Fates. Cell Rep. 30, 1964–1981.e3 (2020).

84. 84. C. Bacon, M. Schneider, C. Le Magueresse, H. Froehlich, C. Sticht, C. Gluch, H. Monyer, G. A. Rappold, Brain-specific Foxp1 deletion impairs neuronal development and causes autistic- like behaviour. Mol. Psychiatry 20, 632–639 (2015).

85. X. Li, J. Xiao, H. Fröhlich, X. Tu, L. Li, Y. Xu, H. Cao, J. Qu, G. A. Rappold, J.-G. Chen, Foxp1 regulates cortical radial migration and neuronal morphogenesis in developing cerebral cortex. PLoS One 10, e0127671 (2015).

86. A. Ortiz, F. Ayhan, N. Khandelwal, E. Outland, M. Jankovic, M. Harper, G. Konopka, Cell type-specific roles of FOXP1 in the excitatory neuronal lineage during early neocortical murine development, bioRxiv (2025). https://doi.org/10.1101/2024.06.08.598089.

87. T. Kadoshima, H. Sakaguchi, T. Nakano, M. Soen, S. Ando, M. Eiraku, Y. Sasai, Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. Proc Natl Acad Sci U S A 110, 20284–20289 (2013).

88. C. Li, J. S. Fleck, C. Martins-Costa, T. R. Burkard, J. Themann, M. Stuempflen, A. M. Peer, Á. Vertesy, J. B. Littleboy, C. Esk, U. Elling, G. Kasprian, N. S. Corsini, B. Treutlein, J. A. Knoblich, Single-cell brain organoid screening identifies developmental defects in autism. Nature 621, 373–380 (2023).

89. B. Paulsen, S. Velasco, A. J. Kedaigle, M. Pigoni, G. Quadrato, A. J. Deo, X. Adiconis, A. Uzquiano, R. Sartore, S. M. Yang, S. K. Simmons, P. Symvoulidis, K. Kim, K. Tsafou, A. Podury, C. Abbate, A. Tucewicz, S. N. Smith, A. Albanese, L. Barrett, N. E. Sanjana, X. Shi, K. Chung, K. Lage, E. S. Boyden, A. Regev, J. Z. Levin, P. Arlotta, Autism genes converge on asynchronous development of shared neuron classes. Nature 602, 268–273 (2022).

90. H. S. Kaya-Okur, S. J. Wu, C. A. Codomo, E. S. Pledger, T. D. Bryson, J. G. Henikoff, K. Ahmad, S. Henikoff, CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat. Commun. 10, 1930 (2019).

91. D. Trudler, S. Ghatak, M. Bula, J. Parker, M. Talantova, M. Luevanos, S. Labra, T. Grabauskas, S. M. Noveral, M. Teranaka, E. Schahrer, N. Dolatabadi, C. Bakker, K. Lopez, A. Sultan, P. Patel, A. Chan, Y. Choi, R. Kawaguchi, P. Stankiewicz, I. Garcia-Bassets, P.

Kozbial, M. G. Rosenfeld, N. Nakanishi, D. H. Geschwind, S. F. Chan, W. Lin, N. J. Schork, R. Ambasudhan, S. A. Lipton, Dysregulation of miRNA expression and excitation in MEF2C autism patient hiPSC-neurons and cerebral organoids. Mol Psychiatry 30, 1479–1496 (2025).

92. N. Perets, L. Kerem, N. Waiskopf, N. Horesh, I. Goldman, J. Avichzer, W. Tobelaim, M. Barashi, L. David, A. Tenenbaum, Patient-derived Brain Organoids Reveal Divergent Neuronal Activity Across Subpopulations of Autism Spectrum Disorder, Neuroscience (2025). https://www.biorxiv.org/content/10.1101/2025.04.27.650821v1.

93. F. P. McCready, K. S. Pradeepan, M. Khaki, W. Wei, M. Guevara-Ferrer, N. Matusiak, B. Feng, A. Piekna, J. Martinez-Trujillo, J. Ellis, Hypersynchronous iPSC-derived SHANK2 neuronal networks are rescued by mGluR5 agonism. Stem Cell Reports 20, 102718 (2025).

94. C. Loveday, K. Tatton-Brown, M. Clarke, I. Westwood, A. Renwick, E. Ramsay, A. Nemeth, J. Campbell, S. Joss, M. Gardner, A. Zachariou, A. Elliott, E. Ruark, R. van Montfort, N. Rahman, Mutations in the PP2A regulatory subunit B family genes PPP2R5B, PPP2R5C and PPP2R5D cause human overgrowth. Human molecular genetics 24 (2015).

95. I. Iossifov, B. J. O'Roak, S. J. Sanders, M. Ronemus, N. Krumm, D. Levy, H. A. Stessman, K. T. Witherspoon, L. Vives, K. E. Patterson, J. D. Smith, B. Paeper, D. A. Nickerson, J. Dea, S. Dong, L. E. Gonzalez, J. D. Mandell, S. M. Mane, M. T. Murtha, C. A. Sullivan, M. F. Walker, Z. Waqar, L. Wei, A. J. Willsey, B. Yamrom, Y.-H. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M. C. Schatz, K. Ye, W. R. McCombie, J.

Shendure, E. E. Eichler, M. W. State, M. Wigler, The contribution of de novo coding mutations to autism spectrum disorder. Nature 515, 216–221 (2014).

96. S. J. Sanders, X. He, A. J. Willsey, A. G. Ercan-Sencicek, K. E. Samocha, A. E. Cicek, M. T. Murtha, V. H. Bal, S. L. Bishop, S. Dong, A. P. Goldberg, C. Jinlu, J. F. Keaney 3rd, L. Klei, J. D. Mandell, D. Moreno-De-Luca, C. S. Poultney, E. B. Robinson, L. Smith, T. Solli-Nowlan, M. Y. Su, N. A. Teran, M. F. Walker, D. M. Werling, A. L. Beaudet, R. M. Cantor, E. Fombonne, D. H. Geschwind, D. E. Grice, C. Lord, J. K. Lowe, S. M. Mane, D. M. Martin, E. M. Morrow, M. E. Talkowski, J. S. Sutcliffe, C. A. Walsh, T. W. Yu, Autism Sequencing Consortium, D. H. Ledbetter, C. L. Martin, E. H. Cook, J. D. Buxbaum, M. J. Daly, B. Devlin, K. Roeder, M. W. State, Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 87, 1215–1233 (2015).

97. S. V. Precious, C. M. Kelly, A. E. Reddington, N. N. Vinh, R. C. Stickland, V. Pekarik, C. Scherf, R. Jeyasingham, J. Glasbey, M. Holeiter, L. Jones, M. V. Taylor, A. E. Rosser, FoxP1 marks medium spiny neurons from precursors to maturity and is required for their differentiation. Exp Neurol 282, 9–18 (2016).

98. N. I. Ahmed, N. Khandelwal, A. G. Anderson, E. Oh, R. M. Vollmer, A. Kulkarni, J. R. Gibson, G. Konopka, Compensation between FOXP transcription factors maintains proper striatal function. Cell Rep 43, 114257 (2024).

99. T. Avino, J. J. Hutsler, Supernumerary neurons within the cerebral cortical subplate in autism spectrum disorders. Brain Res. 1760, 147350 (2021).

100. M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: the comprehensive resource of mammalian protein complexes-2019. Nucleic Acids Res. 47, D559–D563 (2019).

101. T. Li, R. Wernersson, R. B. Hansen, H. Horn, J. Mercer, G. Slodkowicz, C. T. Workman, O. Rigina, K. Rapacki, H. H. Stærfeldt, S. Brunak, T. S. Jensen, K. Lage, A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods 14, 61–64 (2017).

102. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372 (2008).

103. G. Teo, G. Liu, J. Zhang, A. I. Nesvizhskii, A.-C. Gingras, H. Choi, SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. J. Proteomics 100, 37–43 (2014).

104. M. E. Sowa, E. J. Bennett, S. P. Gygi, J. W. Harper, Defining the human deubiquitinating enzyme interaction landscape. Cell 138, 389–403 (2009).

105. artMS, Bioconductor. http://bioconductor.org/packages/artMS/.

106. R. Oughtred, C. Stark, B.-J. Breitkreutz, J. Rust, L. Boucher, C. Chang, N. Kolas, L. O'Donnell, G. Leung, R. McAdam, F. Zhang, S. Dolma, A. Willems, J. Coulombe-Huntington, A. Chatr-Aryamontri, K. Dolinski, M. Tyers, The BioGRID interaction database: 2019 update. Nucleic Acids Res. 47, D529–D541 (2019).

107. M. Choi, C.-Y. Chang, T. Clough, D. Broudy, T. Killeen, B. MacLean, O. Vitek, MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics 30, 2524–2526 (2014).

108. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100 (2003).

109. M. Krzywinski, N. Altman, Comparing samples—part II. Nature Methods 11, 355–356 (2014).

110. D. B. Bekker-Jensen, C. D. Kelstrup, T. S. Batth, S. C. Larsen, C. Haldrup, J. B. Bramsen, K. D. Sørensen, S. Høyer, T. F. Ørntoft, C. L. Andersen, M. L. Nielsen, J. V. Olsen, An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. Cell Syst 4, 587–599.e4 (2017).

111. G. Tsitsiridis, R. Steinkamp, M. Giurgiu, B. Brauner, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: the comprehensive resource of mammalian protein complexes-2022. Nucleic Acids Res. 51, D539–D545 (2023).

112. N. Murtaza, A. A. Cheng, C. O. Brown, D. P. Meka, S. Hong, J. A. Uy, J. El-Hajjar, N. Pipko, B. K. Unda, B. Schwanke, S. Xing, B. Thiruvahindrapuram, W. Engchuan, B. Trost, E. Deneault, F. Calderon de Anda, B. W. Doble, J. Ellis, E. Anagnostou, G. D. Bader, S. W. Scherer, Y. Lu, K. K. Singh, Neuron-specific protein network mapping of autism risk genes identifies shared biological mechanisms and disease-relevant pathologies. Cell Rep. 41, 111678 (2022).

113. Y. Gao, D. Shonai, M. Trn, J. Zhao, E. J. Soderblom, S. A. Garcia-Moreno, C. A. Gersbach, W. C. Wetsel, G. Dawson, D. Velmeshev, Y.-H. Jiang, L. G. Sloofman, J. D. Buxbaum, S. H. Soderling, Proximity analysis of native proteomes reveals phenotypic modifiers in a mouse model of autism and related neurodevelopmental conditions. Nat Commun 15, 6801 (2024).

114. GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).

115. N. N. Parikshak, R. Luo, A. Zhang, H. Won, J. K. Lowe, V. Chandran, S. Horvath, D. H. Geschwind, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155, 1008–1021 (2013).

116. E. Ben-David, S. Shifman, Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. Mol. Psychiatry 18, 1054–1056 (2013).

117. M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur, Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016).

118. C. A. Cassa, D. Weghorn, D. J. Balick, D. M. Jordan, D. Nusinow, K. E. Samocha, A. O'Donnell-Luria, D. G. MacArthur, M. J. Daly, D. R. Beier, S. R. Sunyaev, Estimating the

selective effects of heterozygous protein-truncating variants from human exome data. Nat. Genet. 49, 806–810 (2017).

119. B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, I. Menashe, T. Wadkins, S. Banerjee-Basu, A. Packer, SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism 4, 36 (2013).

120. J. Kaplanis, K. E. Samocha, L. Wiel, Z. Zhang, K. J. Arvai, R. Y. Eberhardt, G. Gallone, S. H. Lelieveld, H. C. Martin, J. F. McRae, P. J. Short, R. I. Torene, E. de Boer, P. Danecek, E. J. Gardner, N. Huang, J. Lord, I. Martincorena, R. Pfundt, M. R. F. Reijnders, A. Yeung, H. G. Yntema, Deciphering Developmental Disorders Study, L. E. L. M. Vissers, J. Juusola, C. F. Wright, H. G. Brunner, H. V. Firth, D. R. FitzPatrick, J. C. Barrett, M. E. Hurles, C. Gilissen, K. Retterer, Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature 586, 757–762 (2020).

121. J. Grove, S. Ripke, T. D. Als, M. Mattheisen, R. K. Walters, H. Won, J. Pallesen, E. Agerbo, O. A. Andreassen, R. Anney, S. Awashti, R. Belliveau, F. Bettella, J. D. Buxbaum, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, J. H. Christensen, C. Churchhouse, K. Dellenvall, D. Demontis, S. De Rubeis, B. Devlin, S. Djurovic, A. L. Dumont, J. I. Goldstein, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, S. Hope, D. P. Howrigan, H. Huang, C. M. Hultman, L. Klei, J. Maller, J. Martin, A. R. Martin, J. L. Moran, M. Nyegaard, T. Nærland, D. S. Palmer, A. Palotie, C. B. Pedersen, M. G. Pedersen, T. dPoterba, J. B. Poulsen, B. S. Pourcain, P. Qvist, K. Rehnström, A. Reichenberg, J. Reichert, E. B. Robinson, K. Roeder, P. Roussos, E. Saemundsen, S. Sandin, F. K. Satterstrom, G. Davey Smith, H. Stefansson, S. Steinberg, C. R. Stevens, P.

F. Sullivan, P. Turley, G. B. Walters, X. Xu, Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, BUPGEN, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 23andMe Research Team, K. Stefansson, D. H. Geschwind, M. Nordentoft, D. M. Hougaard, T. Werge, O. Mors, P. B. Mortensen, B. M. Neale, M. J. Daly, A. D. Børglum, Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. 51, 431–444 (2019).

122. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219 (2015).

123. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682 (2022).

124. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

125. A. Armon, D. Graur, N. Ben-Tal, ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J. Mol. Biol. 307, 447–463 (2001).

126. H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–50 (2016).

127. A. Ben Chorin, G. Masrati, A. Kessel, A. Narunsky, J. Sprinzak, S. Lahav, H. Ashkenazy, N. Ben-Tal, ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. Protein Sci. 29, 258–267 (2020).

128. C. B. Pantazis, A. Yang, E. Lara, J. A. McDonough, C. Blauwendraat, L. Peng, H. Oguro, J. Kanaujiya, J. Zou, D. Sebesta, G. Pratt, E. Cross, J. Blockwick, P. Buxton, L. Kinner-Bibeau, C. Medura, C. Tompkins, S. Hughes, M. Santiana, F. Faghri, M. A. Nalls, D. Vitale, S. Ballard, Y. A. Qi, D. M. Ramos, K. M. Anderson, J. Stadler, P. Narayan, J. Papademetriou, L. Reilly, M. P. Nelson, S. Aggarwal, L. U. Rosen, P. Kirwan, V. Pisupati, S. L. Coon, S. W. Scholz, T. Priebe, M. Öttl, J. Dong, M. Meijer, L. J. M. Janssen, V. S. Lourenco, R. van der Kant, D. Crusius, D. Paquet, A.-C. Raulin, G. Bu, A. Held, B. J. Wainger, R. M. C. Gabriele, J. M. Casey, S. Wray, D. Abu-Bonsrah, C. L. Parish, M. S. Beccari, D. W. Cleveland, E. Li, I. V. L. Rose, M. Kampmann, C. Calatayud Aristoy, P. Verstreken, L. Heinrich, M. Y. Chen, B. Schüle, D. Dou, E. L. F. Holzbaur, M. C. Zanellati, R. Basundra, M. Deshmukh, S. Cohen, R. Khanna, M. Raman, Z. S. Nevin, M. Matia, J. Van Lent, V. Timmerman, B. R. Conklin, K. Johnson Chase, K. Zhang, S. Funes, D. A. Bosco, L. Erlebach, M. Welzer, D. Kronenberg-Versteeg, G. Lyu, E. Arenas, E. Coccia, L. Sarrafha, T. Ahfeldt, J. C. Marioni, W. C. Skarnes, M. R. Cookson, M. E. Ward, F. T. Merkle, A reference human induced pluripotent stem cell line for large-scale collaborative studies. Cell Stem Cell 29, 1685–1702.e22 (2022).

129. S. Jäger, N. Gulbahce, P. Cimermancic, J. Kane, N. He, S. Chou, I. D'Orso, J. Fernandes, G. Jang, A. D. Frankel, T. Alber, Q. Zhou, N. J. Krogan, Purification and characterization of HIV-human protein complexes. Methods 53, 13–19 (2011).

130. L. Gheiratmand, E. Coyaud, G. D. Gupta, E. M. Laurent, M. Hasegan, S. L. Prosser, J. Gonçalves, B. Raught, L. Pelletier, Spatial and proteomic profiling reveals centrosome-independent features of centriolar satellites. EMBO J. 38, e101109 (2019).

131. A. C. O'Neill, F. Uzbas, G. Antognolli, F. Merino, K. Draganova, A. Jäck, S. Zhang, G. Pedini, J. P. Schessner, K. Cramer, A. Schepers, F. Metzger, M. Esgleas, P. Smialowski, R. Guerrini, S. Falk, R. Feederle, S. Freytag, Z. Wang, M. Bahlo, R. Jungmann, C. Bagni, G. H. H. Borner, S. P. Robertson, S. M. Hauck, M. Götz, Spatial centrosome proteome of human neural cells uncovers disease-relevant heterogeneity. Science 376, eabf9088 (2022).

132. E. J. Pearl, R. M. Grainger, M. Guille, M. E. Horb, Development of Xenopus resource centers: the National Xenopus Resource and the European Xenopus Resource Center. Genesis 50, 155–163 (2012).

133. H. L. Sive, R. M. Grainger, R. M. Harland, Early Development of Xenopus Laevis: A Laboratory Manual (CSHL Press, 2000).

134. P. D. Nieuwkoop, Normal Table of Xenopus Laevis (Daudin): A Systematical and Chronological Survey of the Development from the Fertilized Egg Till the End of Metamorphosis (Garland Science, 1994).

135. H. R. Willsey, C. R. T. Exner, Y. Xu, A. Everitt, N. Sun, B. Wang, J. Dea, G. Schmunk, Y. Zaltsman, N. Teerikorpi, A. Kim, A. S. Anderson, D. Shin, M. Seyler, T. J. Nowakowski, R. M. Harland, A. J. Willsey, M. W. State, Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. Neuron 109, 1409 (2021).

136. H. R. Willsey, P. Walentek, C. R. T. Exner, Y. Xu, A. B. Lane, R. M. Harland, R. Heald, N. Santama, Katanin-like protein Katnal2 is required for ciliogenesis and brain development in Xenopus embryos. Dev. Biol. 442, 276–287 (2018).

137. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012).

138. Y. Qi, X.-J. Zhang, N. Renier, Z. Wu, T. Atkin, Z. Sun, M. Z. Ozair, J. Tchieu, B. Zimmer, F. Fattahi, Y. Ganat, R. Azevedo, N. Zeltner, A. H. Brivanlou, M. Karayiorgou, J. Gogos, M. Tomishima, M. Tessier-Lavigne, S.-H. Shi, L. Studer, Combined small-molecule inhibition accelerates the derivation of functional cortical neurons from human pluripotent stem cells. Nat. Biotechnol. 35, 154–163 (2017).

139. J. W. Linsley, K. Shah, N. Castello, M. Chan, D. Haddad, Z. Doric, S. Wang, W. Leks, J. Mancini, V. Oza, A. Javaherian, K. Nakamura, D. Kokel, S. Finkbeiner, Genetically encoded cell-death indicators (GEDI) to detect an early irreversible commitment to neurodegeneration. Nat. Commun. 12, 5284 (2021).

140. M. K. Keaveney, H.-A. Tseng, T. L. Ta, H. J. Gritton, H.-Y. Man, X. Han, A MicroRNA-based gene-targeting tool for virally labeling interneurons in the rodent cortex. Cell Rep. 24,294–303 (2018).

141. H. Zhang, N. Lu, C. Feng, S. W. Thurston, Y. Xia, L. Zhu, X. M. Tu, On fitting generalized linear mixed-effects models for binary responses using different statistical packages. Stat Med 30, 2562–2572 (2011).

142. H. Akaike, A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723 (1974).

143. D. R. Stirling, M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, A. Goodman, CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics 22, 433 (2021).

144. M.-G. Shin, J. A. Kaye, N. Amirani, S. Lam, R. Thomas, S. Finkbeiner, RMeDPower for Biology: guiding design, experimental structure and analyses of repeated measures data for biological studies, bioRxiv (2022). https://doi.org/10.1101/2022.07.18.500490.

145. D. T. Forster, S. C. Li, Y. Yashiroda, M. Yoshimura, Z. Li, L. A. V. Isuhuaylas, K. Itto-Nakama, D. Yamanaka, Y. Ohya, H. Osada, B. Wang, G. D. Bader, C. Boone, BIONIC: biological network integration using convolutions. Nat. Methods, doi: 10.1038/s41592-022-01616-x (2022).

146. F. Zheng, S. Zhang, C. Churas, D. Pratt, I. Bahar, T. Ideker, HiDeF: identifying persistent structures in multiscale 'omics data. Genome Biol. 22, 21 (2021).

147. J. Dutkowski, M. Kramer, M. A. Surma, R. Balakrishnan, J. M. Cherry, N. J. Krogan, T. Ideker, A gene ontology inferred from molecular networks. Nat. Biotechnol. 31, 38–45 (2013).

148. M. Hu, S. Alkhairy, I. Lee, R. T. Pillich, R. Bachelder, T. Ideker, D. Pratt, Evaluation of large language models for discovery of gene set function. ArXiv (2023).

149. G. Tesei, T. K. Schulze, R. Crehuet, K. Lindorff-Larsen, Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. Proc Natl Acad Sci U S A 118 (2021).

150. J. Frazer, P. Notin, M. Dias, A. Gomez, J. K. Min, K. Brock, Y. Gal, D. S. Marks, Disease variant prediction with deep generative models of evolutionary data. Nature 599, 91–95 (2021).

151. Y. Matsumoto, Y. Hayashi, C. R. Schlieve, M. Ikeya, H. Kim, T. D. Nguyen, S. Sami, S. Baba, E. Barruet, A. Nasu, I. Asaka, T. Otsuka, S. Yamanaka, B. R. Conklin, J. Toguchida, E. C. Hsiao, Induced pluripotent stem cells from patients with human fibrodysplasia ossificans progressiva show increased mineralization and cartilage formation. Orphanet J. Rare Dis. 8, 190 (2013).

152. Y. Chen, C. A. Tristan, L. Chen, V. M. Jovanovic, C. Malley, P.-H. Chu, S. Ryu, T. Deng, P. Ormanoglu, D. Tao, Y. Fang, J. Slamecka, H. Hong, C. A. LeClair, S. Michael, C. P. Austin, A. Simeonov, I. Singeç, A versatile polypharmacology platform promotes cytoprotection and viability of human pluripotent and differentiated cells. Nat. Methods 18, 528–541 (2021).

153. J. G. Doench, N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H. W. Virgin, J. Listgarten, D. E. Root, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. 34, 184–191 (2016).

154. I. C. Clark, K. M. Fontanez, R. H. Meltzer, Y. Xue, C. Hayford, A. May-Zhang, C. D'Amato, A. Osman, J. Q. Zhang, P. Hettige, J. S. A. Ishibashi, C. L. Delley, D. W. Weisgerber, J. M. Replogle, M. Jost, K. T. Phong, V. E. Kennedy, C. A. C. Peretz, E. A. Kim, S. Song, W. Karlon, J. S. Weissman, C. C. Smith, Z. J. Gartner, A. R. Abate, Microfluidics-free single-cell genomics with templated emulsification. Nat. Biotechnol. 41, 1557–1566 (2023).

155. E. L. Huttlin, R. J. Bruckner, J. Navarrete-Perea, J. R. Cannon, K. Baltier, F. Gebreab, M. P. Gygi, A. Thornock, G. Zarraga, S. Tam, J. Szpyt, B. M. Gassaway, A. Panov, H. Parzen, S. Fu, A. Golbazi, E. Maenpaa, K. Stricker, S. Guha Thakurta, T. Zhang, R. Rad, J. Pan, D. P. Nusinow, J. A. Paulo, D. K. Schweppe, L. P. Vaites, J. W. Harper, S. P. Gygi, Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell 184, 3022–3040.e28 (2021).

156. M. Aikio, H. M. Odeh, H. J. Wobst, B. L. Lee, Ú. Chan, J. C. Mauna, K. L. Mack, B. Class, T. A. Ollerhead, A. F. Ford, E. M. Barbieri, R. R. Cupo, L. E. Drake, J. L. Smalley, Y.-T. Lin, S. Lam, R. Thomas, N. Castello, A. Baral, J. N. Beyer, M. A. Najar, J. Dunlop, A. D. Gitler, A. Javaherian, J. A. Kaye, G. M. Burslem, D. G. Brown, C. J. Donnelly, S. Finkbeiner, S. J. Moss, N. J. Brandon, J. Shorter, Opposing roles of p38α-mediated phosphorylation and PRMT1-mediated arginine methylation in driving TDP-43 proteinopathy. Cell Rep 44, 115205 (2025).

157. Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, I. Golani, Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125, 279–284 (2001).

CHAPTER 4: SEX DIFFERENCES IN THE DEVELOPING HUMAN CORTEX INTERSECT WITH GENETIC RISK OF NEURODEVELOPMENTAL DISORDERS

Introduction

Development of the human cerebral cortex is highly dynamic, and control of cortical gene expression across developmental periods, cell types, and individuals requires extensive networks of transcriptional regulators, and thousands of 'regulatory' genomic loci. The prenatally developing midgestation cerebral cortex is highly enriched for co-expression of high-confidence autism risk genes identified through studies of de novo protein-damaging variants[1–7], particularly within deep cortical layer neurons.[1,5,6] While studies of gene expression and epigenomics have begun to uncover the blueprint of gene regulatory networks that give rise to relevant brain regions and key cell types,[8–12] how these processes vary between males (XY individuals) and females (XX individuals)[13–15] has not been extensively examined. This represents an important gap relevant to our understanding of the etiology of autism, which shows a significant bias towards male diagnosis.[4,16,17]

Bulk tissue measurements of gene expression in male and female brains have emphasized the contribution of chromosome X- and Y-linked genes.[18] Consistent with this observation, X-chromosome-linked de novo gene mutations are overrepresented in females, likely due to the buffering effect of two copies of the X chromosome and hypothesized lethality in males.[3] Single-cell studies have revealed autosomal differences, with most differences specific to one or few cell types in genes with expression enriched during development.[19] There remains an unmet need to uncover possible mechanisms of sex differences in autosomal gene expression in the prenatally developing cerebral cortex. In particular, studies of gene expression regulation could reveal noncoding cis-regulatory elements (CRE) harboring pathogenic variants, as well as inherited or de novo neurodevelopmental disorder (NDD)-associated variants.[20,21]

Studies of sex differences in the developing brain have focused on brain regions with enriched hormone signaling, such as the hypothalamus and amygdala.[22] Prenatal development represents an important period of testosterone production by gonadal cells of male individuals, which peaks around mid-gestation.[23–25] Both androgen (including testosterone) and estrogen signaling have been shown to regulate the development of the cerebral cortex.[26,27] Further, androgens and estrogens have been implicated in the control of autism risk gene expression[5]. Thus, hormone signaling could represent a major source of autosomal gene expression differences between males and females during prenatal development of the human cerebral cortex. However, we currently lack a systematic assessment of genomic targets of hormone receptors in the human cerebral cortex. Characterization of hormonal, as well as nonhormonal, contributions to sex differences during development represents a potential mechanism of sex-biased penetrance of autism.

Here, we address two main hypotheses that could explain the sex-biased prevalence of autism: first, that gene expression and regulation differences lead to greater vulnerability in males compared to females; and second, that sex differences might emerge as a consequence of hormone (androgen and estrogen) signaling, where differences in hormone levels during prenatal brain development may contribute to higher susceptibility of males to autism.[28–30] To this end, we applied a comprehensive genomic approach to identify molecular signatures of sex differences in the developing human cerebral cortex at midgestation, resolving the contribution of the X and Y chromosomes, hormones, or other epigenetic differences to sex-biased gene expression. By analyzing sex-specific burdens of de novo variants, we identify an overlap between sex differences and NDD genetic risk architecture to enable systematic and comprehensive studies of mechanisms that could contribute to male-biased diagnosis rates in autism.

Results

Integrated single-nucleus multimodal analysis of the midgestation brain

We applied single-nucleus Multiome (snMultiome) analysis to measure gene expression and chromatin accessibility in 21 female and 27 male midgestation human brain samples (**Fig. 4.1a, Fig. 4.5a**), employing a pooling strategy designed to minimize technical batch effects. After removing low-quality cells, doublets, and cells with ambiguous sex, we retained 38,536 cells (19,209 female and 19,317 male) across 11 major cell types identified by enriched expression of known cell-type markers, including layer-specific excitatory neurons (ENs), medial and caudal ganglionic eminence derived inhibitory neurons (IN-MGE, IN-CGE, respectively), intermediate progenitor cells (IPCs), and radial glia (RG) (**Fig. 4.1b-c, 4.5b-e**). We captured 227,118 open chromatin regions across all cell types (**Fig. 4.1d, Fig. 4.5f-g**). To optimize genomic characterization of these regions, we further profiled putative promoter and enhancer histone marks (H3K4me3 and H3K27ac, respectively), as well as the repressive chromatin mark H3K27me3 and transcription factor CTCF by CUT&Tag[31] for a subset of male (n = 6) and female (n = 5) samples included in our snMultiome analysis (**Fig. 4.5h**). Integrating these epigenetic marks with the ATAC-seq data, we identified 17,157 promoters, 26,658 intergenic, 52,081 intronic, and 131,222 peaks at undefined genomic regions. Integrating the RNA and ATAC assays allows for identification of enriched transcription factor motifs in accessible chromatin where the associated transcription factor is also expressed in the same cell, highlighting likely gene regulatory modules in a cell-type-specific manner (**Fig. 4.1e-f, Fig. 4.5i**).

To identify potential regulators of gene expression, we predicted cell-type-specific functional enhancer-gene interactions using the single-cell Enhancer to Gene pipeline (scE2G).[32] As expected, transcriptomically similar cell types showed greater similarity among CREs (**Fig. 4.1g**). About 10-

25% of predicted functional CREs are cell-type specific (**Fig. 4.5j**). As expected, these cell-type-specific CREs were predicted to regulate known cell type markers, such as VIM in radial glia (RG), TLE4 in layer VI/subplate excitatory neurons (L6/SP), and ERBB4 in MGE inhibitory neurons (IN-MGE) (**Fig. 4.1h**).

Sex differences in gene expression

To identify sex differences in gene expression between males and females, we pseudobulked log-normalized RNA counts by cell type and used a linear mixed model to identify additional sources of gene expression variation (including batch and age, Methods).[33] We identified 943 nonredundant differentially expressed genes by sex (DEGs, 730 female biased, 205 male biased, 8 with variable sex bias with cell type) in at least one cell type (FDR < 0.05, log2FC > |0.3|) (**Fig. 4.2a, Fig. 4.6a-b**). Of the 943 total DEGs, 883 are autosomal genes (685 female-biased, 191 male-biased, 7 with variable sex bias with cell type), including the autism risk genes FOXP2 (female biased, IN-MGE), and MEF2C (female biased, IPC) that have been previously reported to show female bias in the midgestation brain[19] (**Fig. 4.2b**). Out of 738 female-biased genes, 219 (29.7%) are differentially expressed across multiple cell types (**Fig. 4.6b**), of which 16 (7.3%) are encoded on the X chromosome, including the known X chromosome inactivation (XCI) escapees DDX3X and ZFX,[34] as well as genes predicted to be subject to XCI, such as PDHA1 and NAP1L3. The other 203 genes with female-biased expression in multiple cell types are autosomal, including H3F3A and MEF2C. In contrast, 25 out of 213 (11.7%) of male-biased genes are differentially expressed in more than one cell type, with 9 out of 25 (36%) of these genes encoded on the Y chromosome (including DDX3Y, KDM5D, and ZFY) (**Fig. 4.2b, Fig. 4.6b**). With the exception of one lncRNA gene, the remaining eight out of nine male-biased Y-chromosome genes have an X-chromosome paralog, while only two genes out of 32 female-biased X genes have a

chromosome Y paralog. Both chromosome Y paralogs are male-biased in expression (DDX3X-DDX3Y, ZFX-ZFY) (**Fig. 4.2b**).

Consistent with prior bulk tissue studies18, X and Y chromosome genes were highly enriched among sex-biased DEGs across most or all cell types (**Fig. 4.6b**), with significant enrichment of X chromosome linked genes among female biased DEGs (two-sided Fisher's exact test, p = 2.9e-05, odds ratio (OR) = 2.05) and Y chromosome linked genes among male biased DEGs (two-sided Fisher's exact test, p = 3.2e-10, OR = 27.26) (**Fig. 4.2c**). Among the female-biased chrX DEGs, we find a significant under-enrichment of genes predicted to be subject to X chromosome inactivation (XCI, determined by San Roman 2023,[34] two-sided Fisher's exact test, p = 1.4e-4, OR = 0.31, **Fig. 4.2d**), as expected, with neither under nor over enrichment of X-chromosome genes predicted to escape XCI (two-sided Fisher's exact test, p = 1, OR = 0.94). Taken together, the observed gene expression differences suggest both chromosome X, as well as autosomal, contributions to female-specific gene regulation, while male-biased gene expression is predominantly driven by Y chromosome genes across all cell types.

Sex differences in chromatin accessibility do not predict differential gene expression

To identify male- and female-biased differentially accessible regions (DARs), we again applied a linear mixed model to account for covariates such as age and tissue pool. We found 522 female biased regions and 577 male biased regions (FDR < 0.05, log2FC > |1.25|) across all cell types (**Fig. 4.2e, Fig. 4.6c**). Interestingly, most male biased peaks are found at distal regulatory regions, while most female-biased peaks are found within gene bodies **(Fig. 4.6d**). The most significant differences are found on the X and Y chromosomes, respectively (**Fig. 4.2f**).

To test whether differences in chromatin accessibility are a major driver of gene expression differences in the developing cortex, we leveraged scE2G on all female cells and all male cells separately to predict CRE-gene links that could be mapped onto our DARs (**Fig. 4.6e**). We did not detect significant correlation between differential accessibility and gene expression in either females (Autosome: Pearson's R = 0.28, p = 1.99e-06; X chromosome: Pearson's R = 0.38, p = 0.048) or males (Autosome: Pearson's R = 0.14, p = 0.178, **Fig. 2g, Fig. 4.6f**), though some sex-biased DEGs have corresponding DARs, such as female-biased DDX3X and male-biased NXPH1 (**Fig. 4.6g**).

We then asked whether transcription factor (TF) motif enrichment analysis might uncover a link between sex-biased gene expression and chromatin accessibility, with TFs that are differentially expressed predicted to bind to DARs. Motifs of several TFs, including NEUROD1, BHLHA15, and NEUROG2, were enriched among male- as well as female-biased DARs, suggesting common mechanisms of sex-biased epigenomic features. Among male DARs, we also identified enrichment of NF1 and TCF21 motifs, while female DARs were enriched for YY1, MEF2C/D, and TFE3 (**Fig. 4.6h**). However, when we restrict our analysis to motifs that are female-biased (enrichment in female-biased DARs > 1 with Benjamini-Hochberg (BH)[35] p < 0.05, and enrichment in male-biased DARs < 1) and male-biased (enrichment in male-biased DARs > 1 with BH corrected p < 0.05, and enrichment in male-biased DARs < 1), we find only 12 male-biased motifs, including TBR1, and EOMES, as well as 12 female-biased motifs, including X-chromosome linked TFE3, as well as basic helix-loop-helix and SOX TFs. None of the significant 12 female biased or 12 male biased motifs are differentially expressed, but extending this analysis to nominally male-biased TF motifs without BH corrected p < 0.05, we find that 5 out of 89 (5.6%) male biased TFs are differentially expressed with female bias (CUX2, FOXP1, RORA, ZBTB18, and ZFX). Further,

ZFX, as well as 3 additional male biased TF motifs (GATA1, ZNF41, and ZNF711) are on the X chromosome (4 out of 89 nominally male-biased motifs, 4.5%). Only 2 out of 76 (2.6%) nominally female-biased TF motifs are on the X chromosome and neither are sex-biased in expression, while an additional 2 TFs are female-biased in expression (MEF2C and USF2).

In silico prediction of sex-biased gene regulatory networks

Next, we compared motifs enriched within CREs of DEGs, since DEGs and genes with DARs are not well correlated. CREs of male biased DEGs were enriched for 37 TF motifs (enrichment in CREs near male-biased DEGs > 1 and BH adjusted $p < 0.05$, enrichment in CREs near female-biased DEGs < 1), including ZBTB33 (X-linked), GLI2, and E2F7. E2F7, a TF involved in DNA damage repair[36] and cell cycle progression,[37] is the only TF with male-biased motif enrichment that is also male-biased in expression. Three out of 37 (8.1%) TFs with male biased motif enrichment are on the X chromosome (ELF4, ELK1, ZBTB33), with X-linked TF ZNF41 nominally male biased (with 61 total nominally male-biased TF motifs).

CREs of female-biased DEGs were nominally enriched for 109 motifs, with only 1 TF on the X chromosome (GATA1). Restricting our analysis to TFs with significant female-biased enrichment (BH adjusted $p < 0.05$), we identify 100 TF motifs, including PAX7, JUN, and MEF2C (**Fig. 4.2h, Fig. 4.6i**). Interestingly, MEF2C is a well-known NDD risk gene[38–40] and is the only TF with female-biased motif enrichment that shows female-biased expression, suggesting a gene regulatory network driven by MEF2C that is unique to females. To identify MEF2C targets, we used Single-cell Multiomic Enhancer-based Gene Regulatory Network Inference (scMEGA).[41] MEF2C is predicted to have the most target genes that are also female-biased DEGs, some of which are further validated by the presence of a MEF2C motif at the promoter, as with ITPR1 (n = 737 total targets, n = 222 female-biased targets, n = 10 male-biased targets, **Fig. 4.2i-j**). Taken

together, these findings suggest that female biased molecular features may in part be driven by the autosomal gene MEF2C.

While no transcription factors with male-biased motif enrichment are predicted to regulate DEGs, the male-biased DEG LHX6 is predicted to have the most male-biased DEGs among its target genes (total target genes = 467, male-biased DEG targets = 27, female-biased DEG targets = 18, **Fig. 4.6j-k**), highlighting its potential role in male-specific gene regulatory programs. However, LHX transcription factor (TF) family motifs, including LHX1, LHX2, LHX3, LHX9, are more significantly enriched in CREs near female-biased DEGs compared to those near male-biased DEGs. Interestingly, the inverse result is seen in DARs, with more significant enrichment in male-biased compared to female-biased DARs. While inconclusive, these results suggest that LHX6 might be responsible for male-biased gene regulatory networks, either through expression of LHX6 itself, binding within male-biased DARs, or promoting expression of other male-biased genes.

Estrogen and androgen receptor binding explain minimal sex differences in gene expression

To probe if differences in gene expression likely arise as a consequence of sex hormone regulation of gene expression, we employed CUT&Tag for estrogen receptor alpha (ERa, ESR1) and the androgen receptor (AR) in a subset of prenatal samples (5 males and 6 females, GW20) to identify possible genomic targets of steroid hormone signaling in the developing human cerebral cortex (**Fig. 4.6l**). We identified 2,179 ESR1 consensus peaks, and 1,161 total AR peaks. To identify candidate target genes, we assigned ESR1 and AR bound peaks to genes by proximity using ChIPpeakAnno and overlapped candidate ESR1 and AR target genes with sex-biased DEGs. Surprisingly, only 6% of female-biased DEGs contained an ESR1 peak (**Fig. 4.2k-l**), and only 6%

male-biased DEGs contained ESR1 peak, and 1% contained an AR peak. There were no differences in either ESR1 or AR peaks between male and female samples (**Fig. 4.6m**).

Genes with sex-biased enrichment for de novo mutations

We sought to better understand the genetic risk architecture of autism in males and females, seeking to identify genes or pathways that might present males or females with greater vulnerability. Towards this goal we re-analyzed whole genome sequencing (WGS) and whole exome sequencing (WES) data from parent-child trios consisting of 61,743 NDD cases (30,687 autism and 31,056 DD) from the SPARK study (iWESv2) as well as five published cohorts of autism and developmental delay (DD)[2,3,42–44] (**Fig. 4.3a**, Methods). To predict genes that harbor a significant excess of de novo mutations (DNMs, including de novo likely gene-disruptive variants and de novo missense) in male probands (n=41,961) and female probands (n=19,782) separately (**Fig. 4.3a**), we applied recently described models of DNM enrichment that estimate the number of expected DNMs by incorporating locus-specific mutation rates, the gene length, and the null expectation based on chimpanzee–human coding sequence divergence (CH model),[3,45] while the denovolyzer (DR) model estimates mutation rates based on trinucleotide context and adjusts divergence based on macaque–human gene comparisons.[3,46] 35 and 32 genes show the enrichment of DNMs specific to females and males respectively, henceforth referred to as female-biased and male-biased respectively (NDD_male32 and NDD_female35, **Fig. 4.3b**). Male-biased genes were enriched for functional categories including transcription corepressor binding and transcription coregulatory binding, whereas female-biased genes were enriched for molecular functions such as histone modifying activity and N-acetyltransferase activity (**Fig. 4.7a**). Protein-protein interaction analysis identified a significant enrichment of interactions among all male-biased NDD genes, as well as all female-biased NDD genes, suggesting potential convergent molecular phenotypes with

mutations in different sex-biased genes (**Fig. 4.7b**). After direct comparison of the DNM counts between males and females, only five genes (DDX3X, HDAC8, WDR45, MECP2, and NAA10) reached significance after correction for multiple comparisons (two-sided Fisher's exact tests with BH correction) for an excess in females when compared to males, with another 20 genes reaching nominal significance (**Fig. 4.3b**, indicated with a single asterisk). By these criteria, no genes reached significance for excess of DNMs in male probands when compared to females, consistent with prior reports.[3,47] All 5 female-biased genes are encoded on X chromosome, consistent with the likely possibility that mutations in those genes are lethal in males due to the presence of a single copy of the X chromosome[3,48,49] or lack of functional compensation from the Y-linked paralog.[50]

Intersection of NDD genes and sex-biased gene expression

Next, we asked whether intersecting sex-biased molecular signatures of cortical development and NDD risk genes could illuminate the molecular mechanisms underlying sex bias of autism/NDD penetrance. To identify general patterns, we tested for enrichment of NDD genes within our entire DEG list agnostic of cell type, including several key NDD gene sets: NDD risk genes from Wang et al.[3] (Wang_NDD615), NDD_male32 genes, NDD_female35 genes, NDD genes that enriched for DNMs in both sex groups (NDDboth97), 102 autism risk genes from Satterstrom et al.[1] (Satterstrom_ASD102), NDD risk genes from Fu et al[2] (Fu_NDD664), and SFARI genes with a score of 1 or 2 (SFARI_1003). Interestingly, both male and female biased NDD genes from this study, as well as genes without sex bias, show enrichment in female-biased DEGs (**Fig. 4.3c**). We identified significant enrichment of 6 male-biased autosomal NDD risk genes with female-biased expression: H3F3A, EEF1A2, CAMTA1, FOXP2, HNRNPD, PIK3R2 (two-sided Fisher's exact test, p = 4.2e-03, OR = 3.81) (**Fig. 4.3d-e**). This suggests that higher expression in females provides

a buffer from potential damaging variants, where these variants would lead to an NDD diagnosis in males but not females.

Further, we identified significant enrichment of 6 female-biased NDD genes in female-biased DEGs: DDX3X, PDHA1, NAA10, MEF2C, ITPR1, ATP1A3 (two-sided Fisher's exact test, p = 0.02, OR = 2.77, **Fig. 4.3d-e**). Half of the candidate genes in this category are X-linked (DDX3X, PDHA1, NAA10), as expected, since two X chromosomes in females allow for up to twice the dosage of X-linked genes compared to males. For the female-biased X-linked NDD genes that are not female biased in expression, we hypothesize mosaicism in females where the wild type X chromosome is stochastically expressed due to random XCI, causing the X chromosome carrying the DNM to be incompletely penetrant. Therefore, baseline X-linked gene expression differences, as well as X chromosome mosaicism, likely contribute to sex-specific burden of de novo variation in these genes, where females can tolerate these mutations due to the increased gene dosage.

Additionally, we identified three autosomal (MEF2C, ITPR1, ATP1A3) female-biased NDD risk genes that are female-biased in expression. ITPR1 and ATP1A3 are both predicted transcriptional targets of MEF2C (**Fig. 4.2i**) that additionally have MEF2C motifs within regulatory regions (**Fig. 4.2j**). We therefore predict that higher expression of MEF2C is required for female-biased gene regulatory programs, and that this represents a point of NDD vulnerability biased towards female diagnosis. Additionally, we find NDD risk genes with no sex bias in de novo mutations that are both male and female biased in expression (**Fig. 4.3c**).

To assess any cell-type-specific vulnerability, we next intersected our DEG list by cell type with Wang_NDD615, as well as the sex-biased NDD genes in either males (32 genes) or females (35 genes) (NDD_sexbias), and the 97 NDD genes with no sex bias at the union of the Venn diagram

(NDD_both97, **Fig. 4.3f, Fig. 4.7c**). Interestingly, DEGs in layer IV excitatory neurons show the greatest enrichment of NDD genes enriched in either males or females (adjusted p = 4.4e-03, OR = 9.82, one-sided Fisher's exact test with BH correction), while there is no significant overlap with NDD genes that are enriched in both males and females (adjusted p = 0.341, OR = 2.44); out of 65 DEGs in this cell type, three are enriched in either male or female NDD probands, DDX3X (chr X), PDHA1 (chr X), and HNRNPD (chr 4) (**Fig. 4.3g**). All three of these genes are more highly expressed in female individuals, but DDX3X and PDHA1 are enriched in female probands, while HNRNPD is enriched in male probands (**Fig. 4.3b**). Previous work has demonstrated the convergence of co-expression networks related to autism within deep layer glutamatergic neurons.[51] These results further highlight that sex differences in gene expression related to NDD pathobiology are greatest in layer IV glutamatergic neurons. Overall, we find that genes with female biased expression, but not male-biased expression, are significantly enriched for both sex-biased and non-sex-biased NDD genes (**Fig. 4.3g**).

Intersection of NDD genes and sex-biased gene regulation

We next applied this framework to differential gene regulation as a point of sex-biased vulnerability to NDD. We looked for an enrichment of NDD genes from the same datasets above (Wang_NDD615, NDD_male32, NDD_female35, NDD_both97, Satterstrom_ASD102, Fu_NDD664, and SFARI_1003) in genes with DARs, where the gene might not be expressed at the developmental time assayed, but may be differentially regulated between males and females at other points during development. Interestingly, we found that genes with male-biased DARs are enriched for NDD risk genes across multiple datasets, while there is no significant enrichment of female-biased DARs for these genes (**Fig. 4.3h**, two-sided Fisher's exact test). This significant enrichment is seen for NDD_male32 in male-biased DARs (p = 0.044, OR = 4.10), with three

genes that overlap: SHANK2, TLK2, and ZBTB20 (**Fig. 4.3i-j**). While the enrichment of NDD_female35 in female-biased DARs is not significant (p = 0.06, OR = 3.59), there are three additional genes that have both a female-biased DAR and are also female-biased NDD risk genes: DDX3X, NSD2, and PPP2R1A (**Fig. 4.3i-j**). This suggests that sex-specific accessible chromatin, and therefore gene regulation, might underlie both male and female-biased vulnerability. There is no enrichment of NDD risk genes with no sex bias (NDD_both97) in genes with female-biased DARs, while genes with male-biased DARs are moderately but significantly enriched for these genes (**Fig. 4.3h, Fig. 4.7f**). Unlike our differential gene expression analysis, we find little cell-type-specific enrichment of NDD genes in genes with DARs (**Fig. 4.7g**).

To further assess whether gene regulatory differences might lead to differences in NDD vulnerability, we looked at female- and male-biased motifs (**Fig. 4.2h, Fig. 4.6h-i**) to see if any TFs in our sex-biased NDD gene lists were also predicted to differentially regulate male and female biased genes. From our motif enrichment analysis in DARs, we identified NF1-halfsite motifs enriched in both female (p = 3.45e-5, fold enrichment = 1.1) and male (p = 1.32e-36, fold enrichment = 1.3) biased DARs, where NF1 is a female-biased NDD risk gene. Additionally, YY1 and MEF2D (similar to MEF2C) are 2 of the top 10 most female-enriched motifs (**Fig. 4.6h**), where YY1 and MEF2C are female-biased NDD risk genes. MEF2A and MEF2B are also enriched in female-biased DARs, but not male, further supporting the importance of the MEF2 TF family in female-specific gene regulatory programs, and the possibility that this might represent a point of female-biased NDD vulnerability. Interestingly, YY1 is an autosomal gene with binding heavily biased towards the inactive X chromosome,[52,53] in addition to its role as a transcriptional activator[54] highlighting how autosomal genes may regulate gene expression and chromatin state on the X chromosome. Lastly, though not a sex-biased NDD risk gene, TFE3 is also an X-linked NDD risk

gene identified in our Wang_NDD615 dataset, and one of the top 10 motifs enriched in female-biased DARs.

These findings highlight how sex-biased NDDs, while they may not be differentially expressed themselves, have the capacity to regulate sex-biased genes, and therefore lead to sex-specific effects with de novo variation. Additionally, the significant enrichment of NDD risk genes in genes with female-biased differential expression, but male-biased in accessibility, suggests that coding variation underlies female-specific protection for autosomal genes, but vulnerability on the X chromosome, while noncoding variation might underlie male-biased vulnerability. Previous work suggests that the regulatory burden of noncoding variation on the X chromosome specifically underlies male-biased vulnerability to NDD penetrance, where females are protected by the second copy of the X.

The X chromosome contributes to NDD vulnerability through sex-specific mechanisms

To further interrogate the contribution of the X chromosome to sex-biased NDD vulnerability, we first established an integrative map of genes that are differentially expressed, have sex differential accessibility, and confer NDD risk via our DNM analysis (**Fig. 4.4a**). We additionally mapped our CREs identified in bulk female samples, as well as bulk male samples. Since chromatin accessibility does not correlate with differences in expression, most DEGs on the X chromosome do not have an additional DAR, with the exception of EIF1AX and DDX3X. Of the 51 chrX DEGs, 7 are NDD risk genes: CDKL5, CNKSR2, NONO, DCX, DDX3X, NAA10, and PDHA1. 5 chrX DEGs are male-biased: AC233296.1, CNKSR2, CLIC2, PLS3, and REPS2. There are 18 total unique DARs, 15 with female bias and 3 with male bias.

While about 70% CREs identified in both female and male bulk analyses do not show any differences (48,315 shared CREs, 68,066 total bulk female CREs, 68,509 total bulk male CREs), as with CREs predicted to regulate female-biased DEG NAA10, about 30% of CREs are sex-specific (19,751 female-specific, 20,194 male-specific, **Fig. 4.4b**). For example, female-biased NDD risk genes PDHA1 and MECP2 have CREs predicted to interact with the promoter in females but not males.

Together, this comprehensive genomic map inclusive of the X chromosome highlights potential sources for sex-biased NDD vulnerability, including genes with female-biased expression that allow for tolerance to coding variation, as well as CREs that may carry noncoding mutational burden biased towards males who lack mosaicism and compensation from the wild type X chromosome (**Fig. 4.4c**).

Discussion

In the current study, we identify cell-type-specific transcriptomic and epigenomic sex differences in the developing human cerebral cortex at midgestation. In addition to the genes encoded on the X and Y chromosomes, we identify dozens of autosomal genes with female-biased expression. Consistent with a prior analysis, we identify significant enrichment of high-confidence NDD risk genes among female-biased genes.[19] Autosomal gene expression differences detected in this study highlight the unique features of prenatally developing human brain in contrast to the adult brain, where transcriptomic differences seem to be largely restricted to X and Y chromosome linked genes.[55] Interestingly, the same extent of bias among autosomal signals was not found among chromatin accessibility signatures, which were more balanced between males and females.

The prenatally developing human brain is exposed to circulating hormones, including testosterone, whose levels surge during second trimester,[23–25] and which can be converted to estrogen. Hormone signaling plays critical roles in development of brain regions relevant to social interactions,[56] but recent studies have shown that these hormones can induce changes in gene expression and progenitor cell behavior in the cerebral cortex.[26,27,57] Our study did not uncover a strong correlation between ESR1 or AR binding sites and sex biased genes within the prenatally developing cerebral cortex, but binding of other sex hormone receptors including ESR2 and progesterone receptor, as well as inter-individual variability and earlier stages of development will need to be examined in the future. Analysis of developmental regulomes implicated transcription factor MEF2C as a major regulator of sex-biased expression of hundreds of autosomal genes in females. Similarly, our analysis implicates X-linked transcription factor TFE3 as a candidate driver of chromatin accessibility differences.

By intersecting the map of molecular sex differences with genetic risk architectures of male and female individuals with autism, our study presents a resource for exploring the molecular basis of sex-biased cortical development and autism diagnosis. We highlight the role of the X chromosome as a driver of molecular sex differences in the developing cortex. The expected lethality of high penetrance X-linked NDD risk genes in males results in strong bias towards the diagnosis of coding mutations in these genes in females with autism and NDDs.[58] However, emerging evidence from studies of mutations disrupting the activity of CREs of X-linked NDD risk genes offers clues into a mechanism by which X chromosome mutations could result in autism in some males because of their smaller genetic impact on gene expression.[59] This finding is supported by recent studies implicating noncoding elements in NDD risk,[21] as well as male-specific burden of inherited variation on the X chromosome driven by female protection of two copies of the X chromosome.[59]

The majority of autism and NDD studies have overwhelmingly focused on identifying genes with an excess of protein-damaging variants,[2,3] and the contribution of noncoding mutations to autism remains underexplored. Our resource defines a set of CREs of X-linked NDD risk genes that could be used to prioritize noncoding variants for functional validation. Thus, a shift from generating familial whole-exome sequencing data to whole-genome sequencing data will be critical to testing genetic burden among CREs of NDD risk genes. The larger whole-genome sequencing datasets being generated for autism families as part of SPARK60 in particular (currently 142,357 individuals from 54,558 families in SPARK iWESv3 and 12,519 individuals from 3,394 families in SPARK iWGSv1.1, https://base.sfari.org/public-resources) will allow us to test whether DNMs in CREs located on the X chromosome and other sex-biased genes contribute to the increased male-biased diagnosis of autism. The transcriptomic and epigenomic data resource presented here provides a refined annotation of genomic loci likely to be functional in the critical brain region and period of development most highly associated with autism.[7,61]

Figures

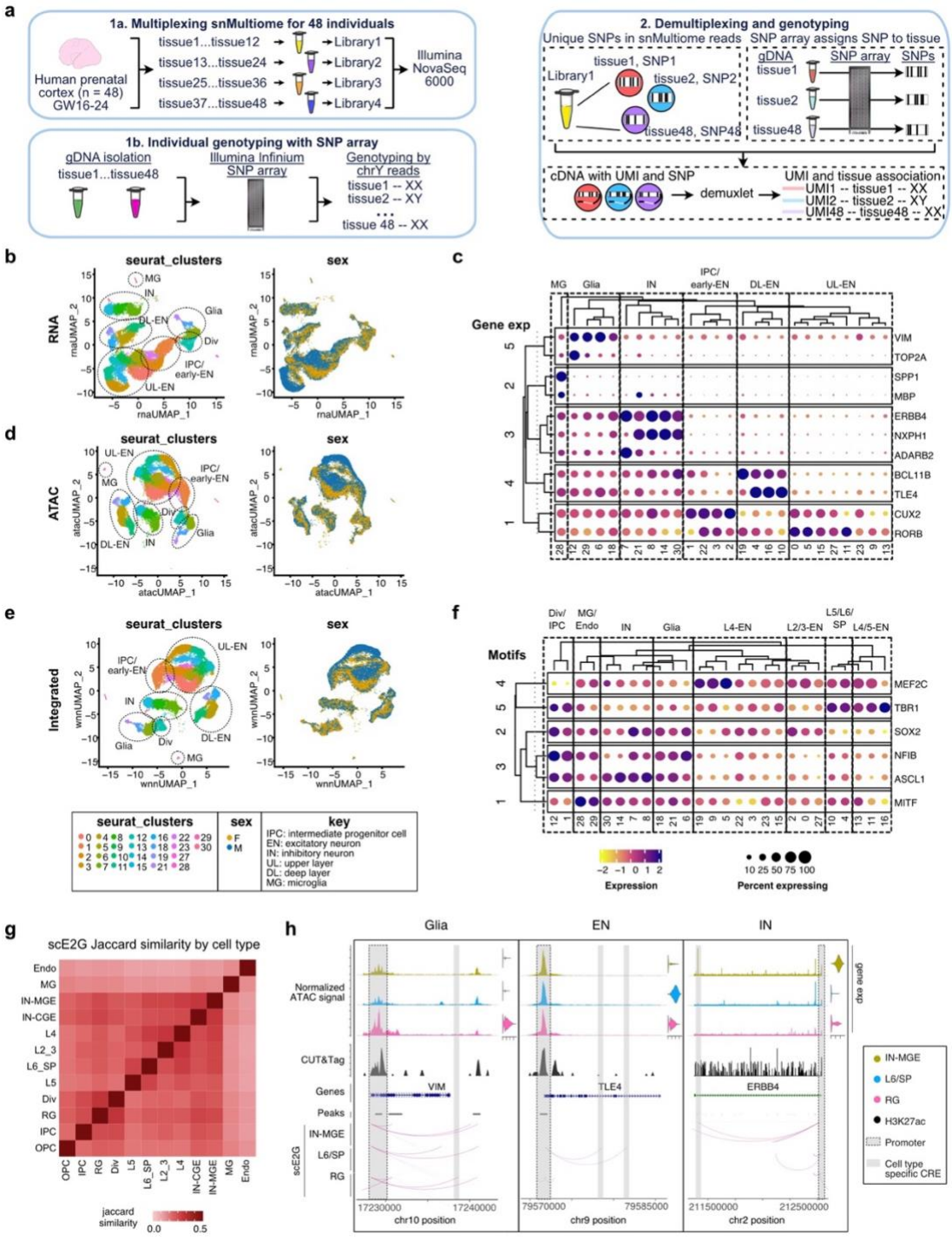

Figure 4.1 Establishing an atlas of male and female midgestation gene expression and chromatin accessibility

(a) Schematic of scMultiome (1a) and single-nucleotide (Figure legend continued on next page.)

(Figure legend continued from previous page.) polymorphism (SNP) array (1b) for 48 prenatal samples (21 female, 27 male). **(b)** Schematic of donor demuxing of scMultiome data using SNP genotype information for each individual. **(c)** UMAP of RNA assay by cell type and sex. **(d)** UMAP of ATAC assay by cell type and sex. **(e)** UMAP of integrated assay by cell type and sex. **(f)** Hierarchical dot plot of top gene expression cluster markers. **(g)** Hierarchical dot plot of top enriched motifs by cluster. **(h)** Correlation matrix of Jaccard similarity scores for CREs predicted for each cell type by the scE2G pipeline. **(i)** Coverage plots showing cell-type-specific CREs. Normalized ATAC signal, gene expression, H3K27ac CUT&Tag, and scE2G predicted CREs by cell type are shown. Curved lines represent predicted enhancer-promoter interactions. To represent the three most distinct cell types, radial glia (glia) layer 6/subplate excitatory neurons (EN), and MGE derived inhibitory neurons (IN) are shown.

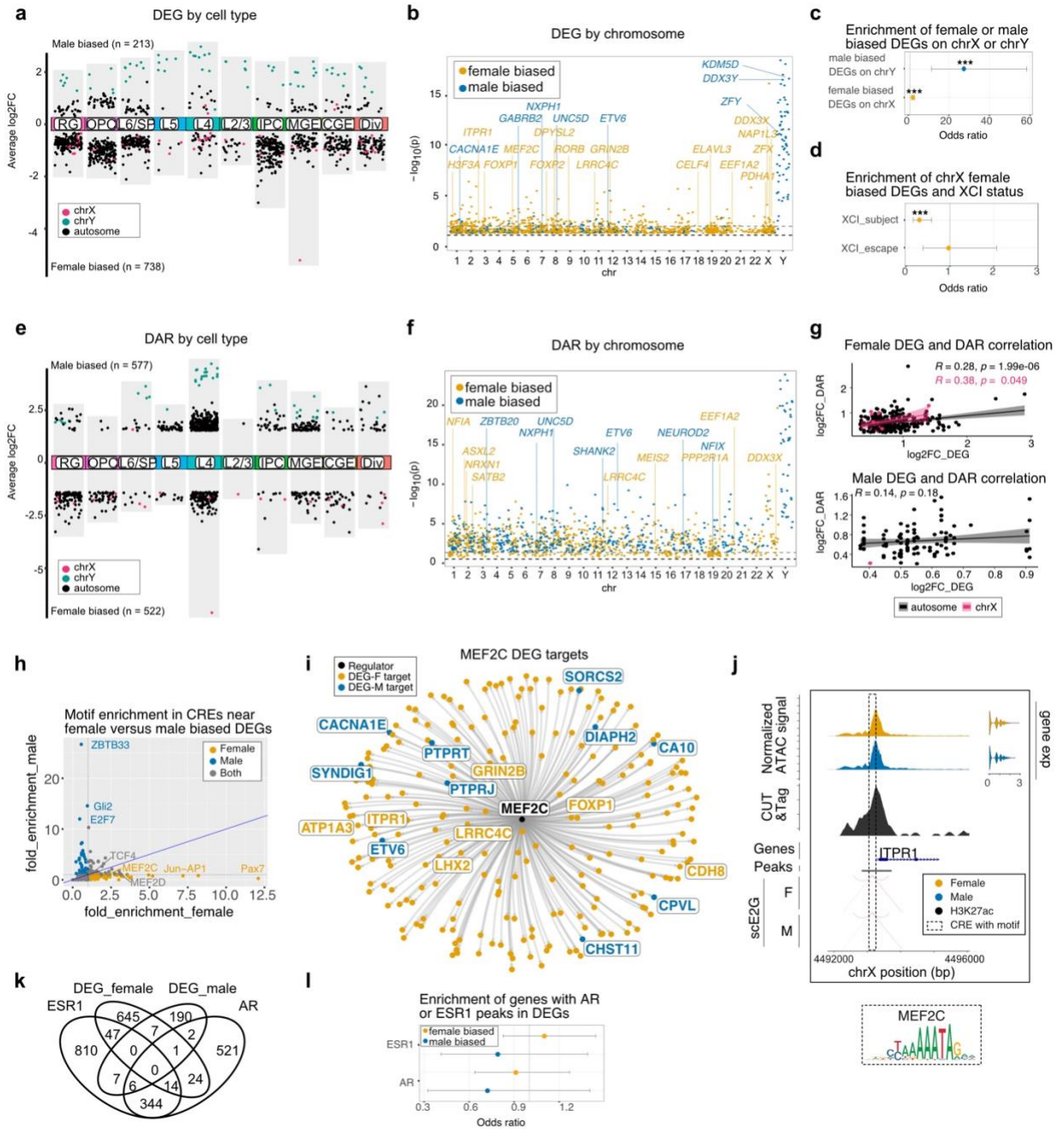

Figure 4.2 Sex-specific differences in gene expression and chromatin accessibility.
(a) Differentially expressed genes (DEGs) between male (XY, positive log2FC) and female (XX, negative log2FC) cells by cell type. FDR < 0.05. MG and Endo cell types excluded due to cell count less than 150. **(b)** Distribution of DEGs by autosome, chrX, and chrY, separated by sex-biased expression. Black dashed line indicates adjusted p = 0.05, grey dashed line indicates adjusted p = 0.01. **(c)** Enrichment of female biased genes on chrX, and male biased genes on chrY determined by two-sided Fisher's exact test. 95% confidence interval represented in grey. *** p < 0.001. **(d)** Enrichment of genes either subject to or that escape from X chromosome inactivation (XCI) among X chromosome female-biased DEGs determined by two-sided Fisher's exact test. 95% confidence interval represented in grey. (Figure legend continued on next page.)

(Figure legend continued from previous page.) **(e)** Differentially accessible regions (DARs) between male (XY, positive log2FC) and female (XX, negative log2FC) cells by cell type. Genes on autosomes are black, genes on chrX are pink, genes on chrY are green. FDR < 0.05, log2FC < |1.25|. MG and Endo cell types excluded due to cell count less than 150. **(f)** Distribution of DARs by autosome, chrX, and chrY, separated by sex-biased expression. A subset of nearest genes predicted to be regulated by DARs are labeled. Black dashed line indicates adjusted p = 0.05, grey dashed line indicates adjusted p = 0.01. **(g)** Scatter plot and Pearson's correlation for genes that are both differentially expressed (x-axis) and have a DAR (y-axis) by sex. Pearson's correlation was calculated separately for genes on autosomes (black), X chromosome (pink), and Y chromosome (green). **(h)** Scatter plot of motif enrichment for scE2G CREs for DEGs, with motif enrichment in CREs near female-biased DEGs on the x-axis, and motif enrichment in CREs near male-biased genes on the y-axis. Genes with fold enrichment > 1 in females and < 1 in males are highlighted in orange, and those with fold enrichment > 1 in males and < 1 in females are highlighted in blue. **(i)** Regulon plot for MEF2C predicted targets that are also DEGs. Targets that are female-biased DEGs are orange (n = 225), targets that are male-biased DEGs are blue (n = 7). Total targets n = 737. **(j)** Coverage plot of female biased gene *ITPR1*, with normalized ATAC tracks and gene expression by sex, as well as H3K27ac CUT&Tag coverage and scE2G CRE prediction by sex. MEF2C motif is highlighted in grey. **(k)** Venn diagram of gene overlaps between male- and female-biased DEGs, nearest neighbor genes to AR CUT&Tag peaks, and nearest neighbor genes to ESR1 CUT&Tag peaks. **(l)** Two-sided Fisher's exact test enrichment of genes with either ESR1 (top) or AR (bottom) CUT&Tag peaks in either female (orange) or male (blue) DEGs.

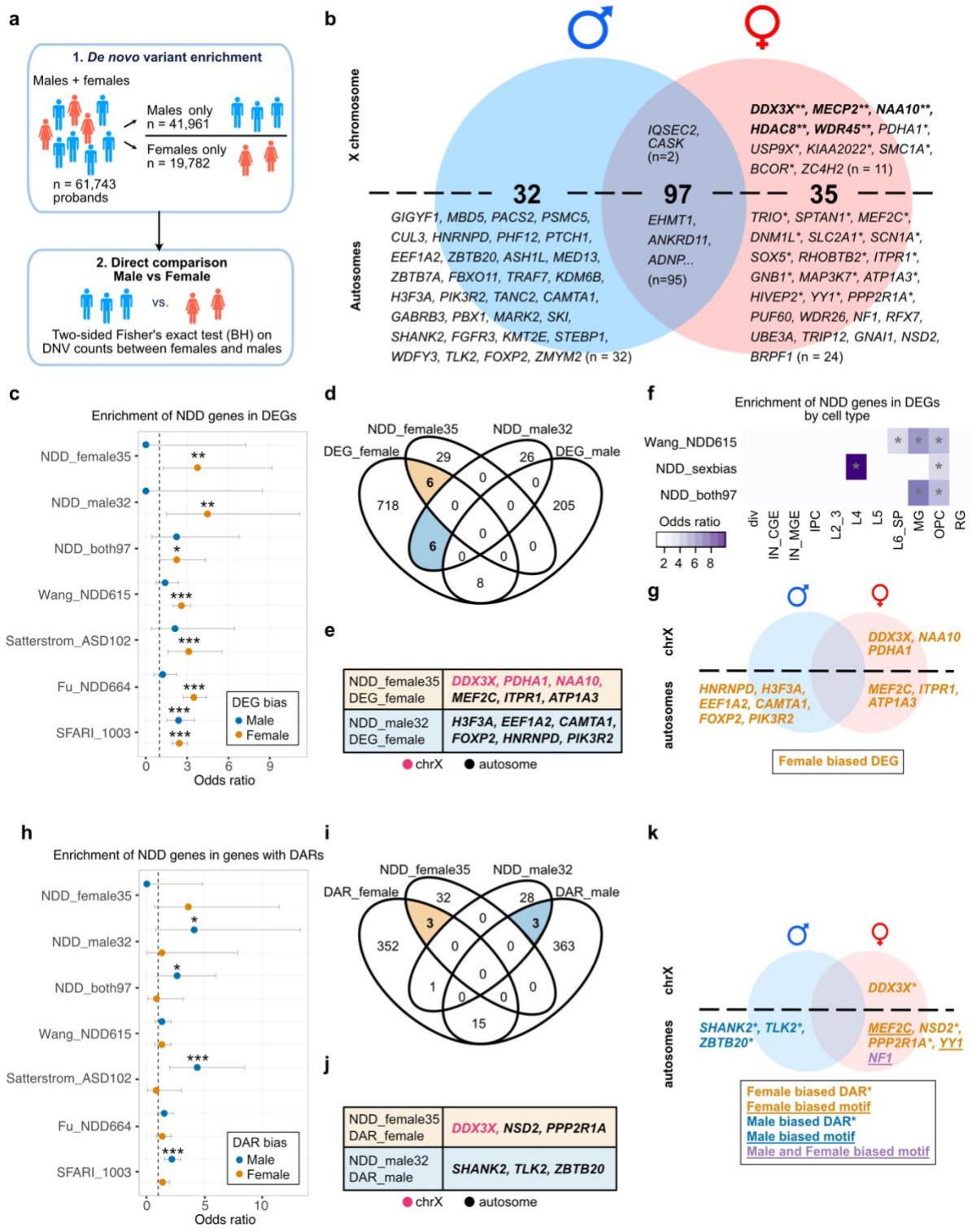

Figure 4.3 Gene expression and chromatin accessibility differences underlying NDD vulnerability

(a) Schematic of sex-biased gene discovery. *De novo* mutation (DNM) enrichment analyses (1) were performed independently in NDD males and females using CH and DR models. Sex-biased genes were further identified by two-sided Fisher's (Figure legend continued on next page.)

(Figure legend continued from previous page.) exact tests (BH correction, n=20,000 genes) of DNM counts between males and females (2). **(b)** Venn diagram of genes with potential sex bias. Genes enriched for DNMs in males (n=129) and females (n=132) are shown in blue and red circles, respectively. Genes in bold with two asterisks (**) show sex-specific significance after multiple testing corrections (a). Genes with a single asterisk (*) show nominal significance in the direct female-male comparison. **(c)** Enrichment of NDD genes from different studies and DEGs, determined by two-sided Fisher's exact test. Odds ratio shown in orange for enrichment in female biased DEGs and blue for male biased DEGs, with 95% confidence interval shown in grey. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(d)** Venn diagram of differentially expressed genes and sex-biased NDD genes. **(e)** Overlapping genes from NDD_female35 and DEG_female union (orange) and NDD_male32 and DEG_female union (blue). Overlapping genes on the X chromosome are pink, and genes on autosomes are black. **(f)** Enrichment of NDD genes and differentially expressed genes (both male and female) by cell type. Wang_NDD615 contains all NDD genes identified in Wang 2022[3]. NDD_sexbias contains 35 female biased and 32 male biased genes from (B). Asterisk indicates $p < 0.01$ (Benjamini-Hochberg correction) by two-sided Fisher's exact test. Color indicates odds ratio. **(g)** Venn diagram from (b) highlighting the sex-biased NDD genes that are also differentially expressed. **(h)** Enrichment of NDD genes from different studies and DARs, determined by two-sided Fisher's exact test. Odds ratio shown in orange for enrichment in female biased genes with DARs and blue for male biased genes with DARs, with 95% confidence interval shown in grey. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(i)** Venn diagram of nearest neighbor genes with DARs and sex-biased NDD genes. **(j)** Overlapping genes from NDD_female35 and DAR_female union (orange) and NDD_male32 and DAR_male union (blue). Overlapping genes on the X chromosome are pink, and genes on autosomes are black. **(k)** Venn diagram from (b) highlighting the sex-biased NDD genes that also either have a DAR, or have sex-biased motif enrichment.

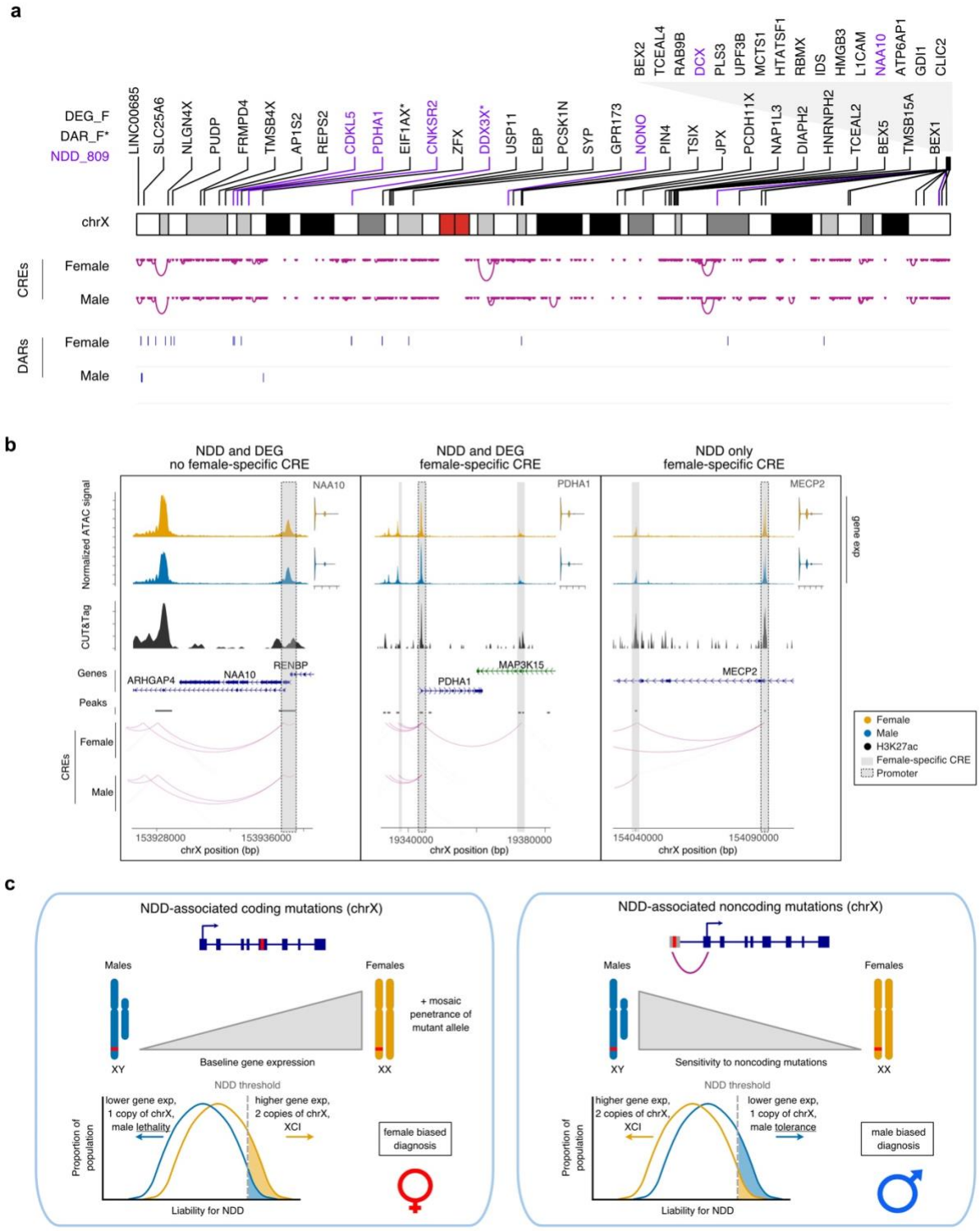

Figure 4.4 An integrated genomic map of the X chromosome
(a) Map of chromosome X including DEGs, DARs, NDD genes that are also differentially expressed, and CREs. **(b)** Three examples of chromosome X NDD genes: *NAA10*, *PDHA1*, and *MECP2*. **(c)** Proposed model of sex-specific vulnerability. X chromosome coding mutations lead to female-biased vulnerability, and noncoding mutations lead to male-biased vulnerability.

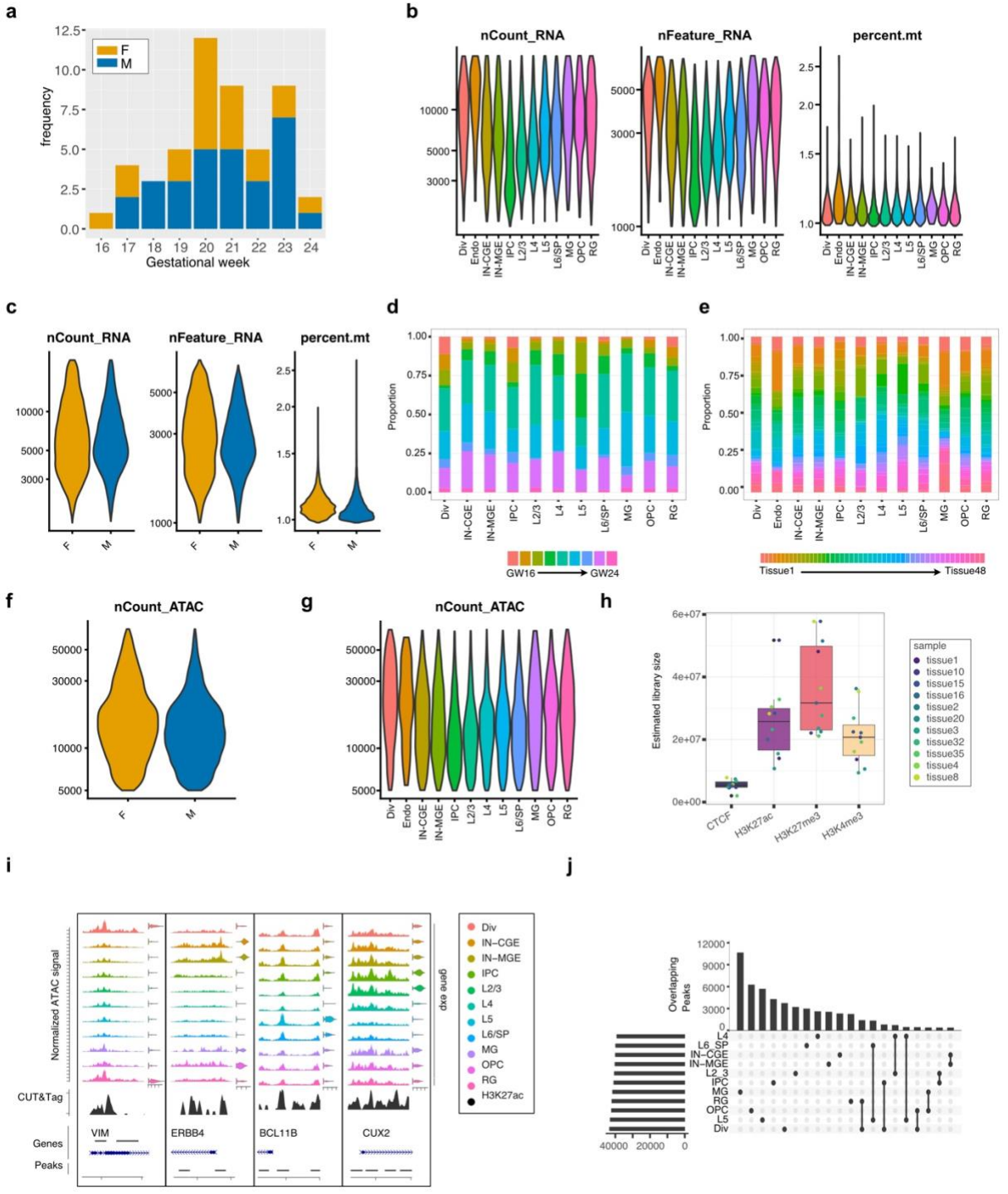

Figure 4.5 Quality control of single-nucleus Multiome and CUT&Tag samples.
(a) Histogram of samples by gestational week, by sex. **(b)** Quality control (QC) by cluster of scRNA-seq data. **(c)** QC by sex of scRNA-seq data, showing gene and UMI count, as well as mitochondrial gene percentage. **(d)** Cluster proportion graph of sample age. GW = gestational week. **(e)** Cluster proportion graph of donor. **(f)** QC by sex of scATAC-seq data. **(g)** QC by cell type of scATAC-seq data. **(h)** Estimated CUT&Tag (Figure legend continued on next page.)

(Figure legend continued from previous page.) library size for all samples (n = 6 male, n = 5 female) of H3K27ac, H3K27me3, H3K4me3, and CTCF. **(i)** Coverage plots of cluster markers VIM (radial glia progenitors), ERBB4 (inhibitory neurons), BCL11B (deep layer excitatory neurons), and CUX2 (upper layer excitatory neurons). Overlap of ATAC signal, H3K27ac signal, gene track, called ATAC peaks, and predicted peak-gene links are shown, with additional violin plots of gene expression by cell type. **(j)** Upset plot of scE2G predicted CREs by cell type.

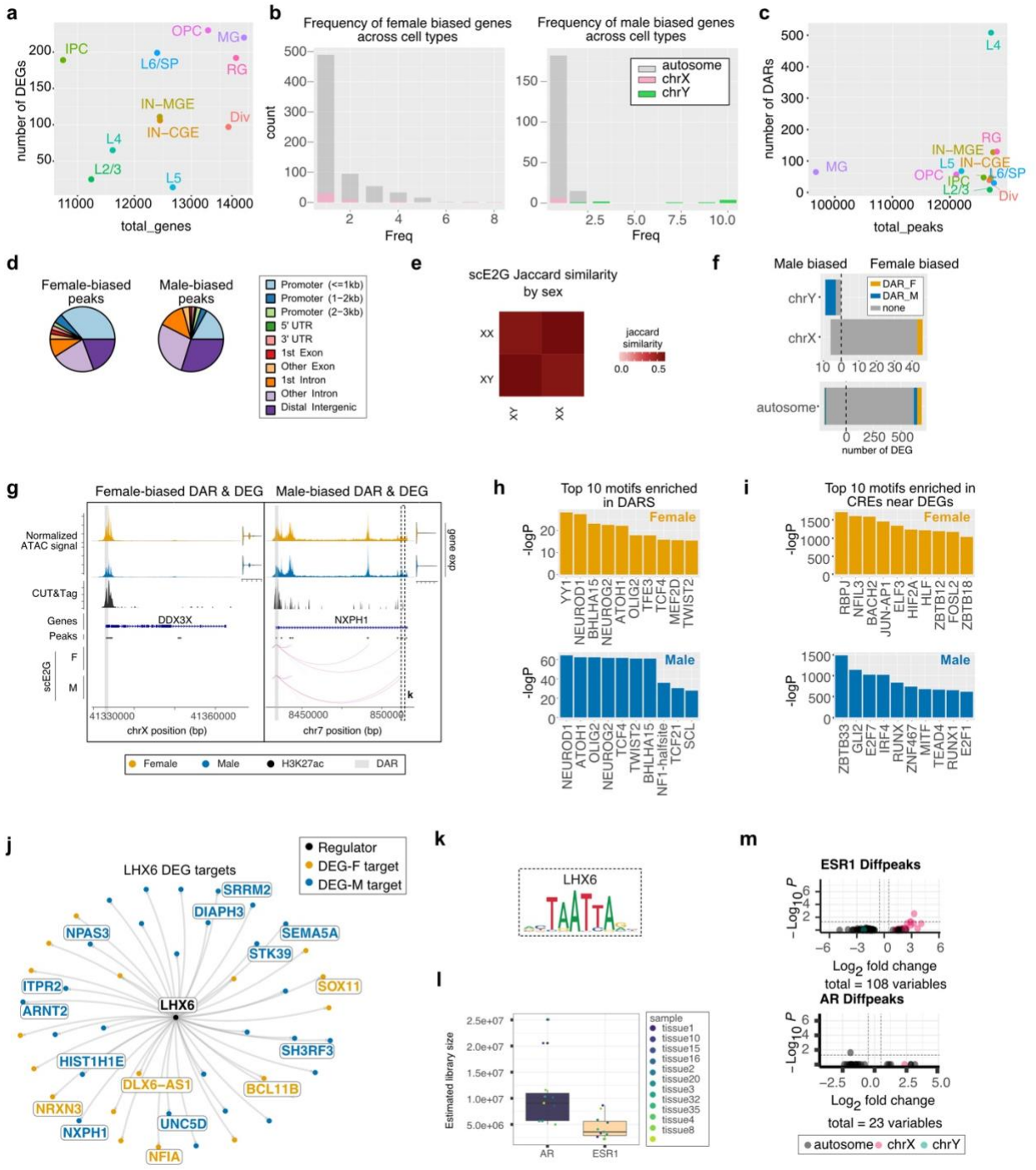

Figure 4.6 Characterizing sex biased gene expression, chromatin accessibility, motif enrichment, and gene regulatory networks.

(a) Scatter plot of total number of genes expressed (x-axis) by differentially expressed genes (y-axis) by cell type. **(b)** Histogram of DEGs by number of cell types they are differentially expressed in. Grey is autosomal, pink is chrX, and green is chrY. **(c)** Scatter plot of total number of accessible regions (x-axis) by DARs (y-axis) by cell type. **(d)** Pie chart of annotation for female-biased and male-biased DARs. (Figure legend continued on next page.)

(Figure legend continued from previous page.) **(e)** Correlation matrix of Jaccard similarity scores for CREs predicted for each sex by the scE2G pipeline. **(f)** Bar plot of female-biased (right side) and male-biased (left side) genes. Colors indicate if the DEG has a DAR. **(g)** Coverage plot of female biased gene *DDX3X* and male biased gene *NXPH1*, with normalized ATAC tracks and gene expression by sex, as well as H3K27ac CUT&Tag coverage and scE2G CRE prediction by sex. **(h)** Top 10 motifs enriched in CREs near female biased (top, orange) or male biased (bottom, blue) DEGs. **(i)** Top 10 motifs enriched in female biased (top, orange) or male biased (bottom, blue) DARs. **(j)** Regulon plot for LHX6 predicted targets that are also DEGs. Targets that are female-biased DEGs are orange (n = 18), targets that are male-biased DEGs are blue (n = 27). Total targets n = 467. **(k)** LHX6 motif, seen at *NXPH1* promoter (g). **(l)** Estimated library size for AR and ESR1 CUT&Tag across all samples. **(m)** Sex-biased ESR1 and AR peaks. Peaks shown pass nominal significance threshold, where adjusted p-value is shown.

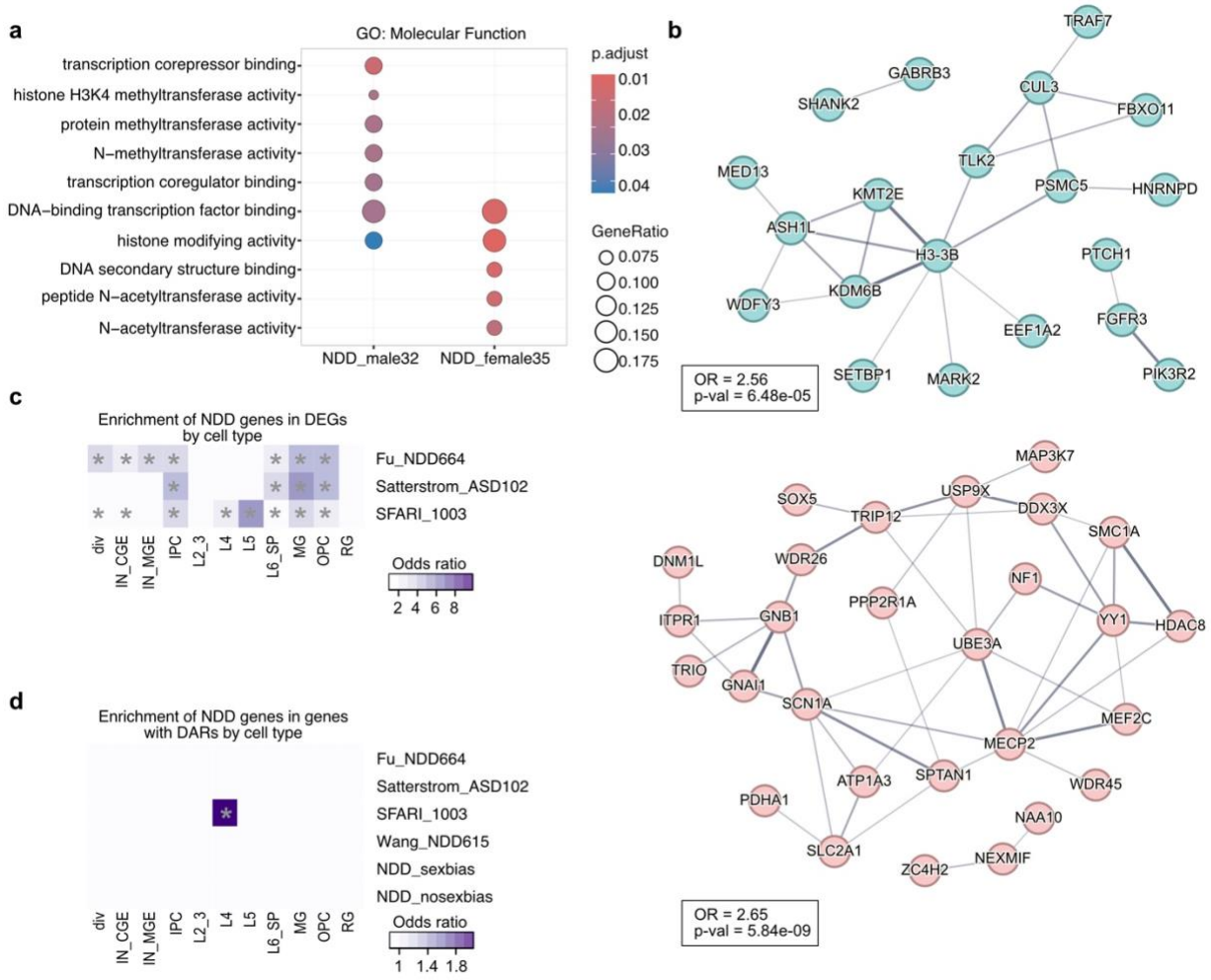

Figure 4.7 Characterizing sex-biased NDD genes.
(a) Dotplot of molecular function gene ontology for male biased NDD genes (SSD_male) and female biased NDD genes (SSD_female). Color represents BH corrected p-value. Dot size indicates gene ratio. **(b)** STRING protein interaction network of male (left) and female (right) biased NDD genes. Line thickness indicates the strength of data support from the STRING v12.0 database. **(c)** Enrichment of NDD genes from different studies in DEGs by cell type. * $p < 0.01$ (BH correction). **(d)** Enrichment of NDD genes from different studies in genes with DARs by cell type. * $p < 0.01$ (BH correction).

Materials and Methods

Tissue source

De-identified tissue samples were collected with previous patient consent in strict observance of the legal and institutional ethical regulations. Protocols were approved by the Human Gamete, Embryo, and Stem Cell Research Committee (institutional review board) at the University of California, San Francisco.

Flash freezing tissue

Isopentane was cooled with liquid nitrogen, and the tissue was submerged in the isopentane for 30 seconds. Flash frozen tissue was transferred to cryovials and stored at -80C.

Isolation of gDNA for SNP genotyping

The DNA from each sample was isolated using an adapted protocol from the Purelink Genomic DNA Kits (Fisher K182001). 15 mg of flash frozen tissue samples were shaved off from tissue blocks on dry ice using pre-chilled 60x15 mm plates and sterile blades. The shaved tissue was transported to regular ice and 210 uL of master Digestion Buffer containing 180 uL of Purelink Genomic Digestion Buffer and 30 uL of Proteinase K was added to the sample. The combined sample and master Digestion Buffer was lysed mechanically using a chilled blade and transferred to a DNA Lobind tube (Eppendorf) using a wide-bore pipette tip. The sample was incubated at 55C for 4 hours at 600 rpm with vortexing every half hour to hour. The DNA was bound to the spin column, washed, and eluted using the Purelink Genomic DNA Kit (Fisher K182001).

SNP imputation and demultiplexing

SNP QC and imputation:

All samples used for snMultiome were run on the Infinium Global Screening Array (GSAv2, Illumina #20030770). Sample sex was determined using Illumina's Genome Studio desktop application using X and Y chromosome SNPs. We started with 653,817 total SNPs, and then removed the following variants: palindromes, duplicates, relateds, out of Hardy-Weinberg Equilibrium, and monomorphs, resulting in 470,515 high quality SNPs for imputation.

We ran pre-imputation checks following https://www.chg.ox.ac.uk/~wrayner/tools/. SNPs were run through the HRC-1000G-check-bim.pl pipeline using the TOPMED reference. The output was then imputed using the TOPMED imputation server with the following parameters: R2 > 0.3, phased with Eagle v2.4, and allele frequency check on. The results from each chromosome were concatenated and then filtered for R2 > 0.95 and MAF > 0.05, resulting in 4,212,681 imputed SNPs for demultiplexing.

Demultiplexing and integration with snMultiome:

The SNPs vcf file was split into separate files containing only samples in each 10x lane using bcftools. Demuxlet (https://github.com/VincentGardeux/demuxlet) was run for each lane using the gene expression sorted bam file from cellranger-arc and the split SNP vcf file. Demuxlet outs were integrated into the snMultiome object using importDemux (dittoSeq package, https://github.com/zehualilab/DittoSeq).

Nuclei isolation

Flash frozen tissue blocks were cut in groups of twelve on dry ice using prechilled 60x15 mm plates and sterile blades. Samples were pooled, with 12 individuals per nuclei isolation. For each of the twelve tissue samples used, 5mg was obtained using this method and added to a shared tube

of frozen tissue. The shared tube was weighed after the addition of each sample to ensure equal input and individual representation.

The tube containing 60mg of pooled frozen tissue was emptied into a 2 mL glass dounce that had been pre-chilled on ice. 1 mL of homogenization buffer containing 0.2653 M Sucrose, 26.6 mM KCl, 5.31 mM MgCl2, 21.2 mM Tricine-KOH pH 7.8, Ultrapure H2O, 1 mM DTT, 0.5 mM Spermidine, 0.15 mM Spermine, 0.3% NP40, 1X Protease Inhibitor (Sigma #11873580001), and 0.6 U/uL RNAse Inhibitor was added to the dounce. The tissue was homogenized using 12 strokes from the dounce pestle A and 16 strokes from the dounce pestle B. The resulting homogenized solution was filtered through the cap of a pre-chilled sterile 5 mL Falcon round bottom tube (Corning 08-771-23) and centrifuged at 400g for 5 min at 4C in a fixed-angle centrifuge. The supernatant was removed and discarded. The remaining pellet was resuspended using 400uL of the homogenization buffer and mixed in with 1 volume of a 50% iodixanol solution, resulting in a 25% iodixanol layer. A 600uL volume of 30% iodixanol made from the 50% Iodixanol stock solution and homogenization buffer was slowly layered below the 25% layer. A 600uL volume of 40% Iodixanol solution made from the 50% Iodixanol stock solution and homogenization buffer was slowly layered below the 30% layer. The layered solution was centrifuged in a pre-chilled swinging bucket centrifuge for 20 min at 4C at 3,000g with acceleration level at 1 and deceleration level at 0. The iodixanol gradient solution post-centrifugation contained a distinct nuclei band located within the 30% iodixanol layer. The 25% layer was removed and discarded using a 200uL pipette. 200uL of the nuclei band was slowly removed and placed in a pre-chilled 1.5 mL DNA lobind tube (Eppendorf 022431021). The 200uL band collected from the gradient solution was diluted with 200uL of wash buffer containing 10mM Tris-HCL pH 7.4, 10mM NaCl, 3mM MgCl2, 1% BSA, 0.1% Tween-20, 1 mM DTT, 0.6 U/uL RNAse Inhibitor, and ultrapure H2O. The diluted

nuclei solution was counted manually using a disposable hemocytometer (Bulldog Bio DHC-N51) and Trypan Blue. The solution was then spun down using a pre-chilled fixed angle centrifuge at 500g for 5 min at 4C. The supernatant was removed and discarded. The remaining nuclei pellet was resuspended using variable volumes of Diluted Nuclei Buffer containing 1x Nuclei Buffer (10x Genomics PN-2000207), 1 mM DTT, 1 U/uL RNase Inhibitor, Nuclease-Free H2O taken from Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Kit (10x Genomics PN-1000283). The final concentration of the nuclei suspension was between 8,000 and 10,000 nuclei/uL to load into the 10x Multiome protocol.

Single-nucleus Multiome

Each nuclei extraction of pooled donors was split into four lanes of Multiome (10x Genomics Chromium Next GEM Single Cell Multiome ATAC + Gene Expression PN-1000283) with 18,000 nuclei loaded per lane, with targeted nuclei recovery of 12,000 nuclei per lane. Libraries were pooled to 4nM and sequenced on a Novaseq 6000 at a final concentration of 150pM, with targeted read depth of 20,000 reads per cell.

snMultiome QC

ATAC and RNA fastq files for each lane were processed using 10x cellranger-arc v2.1.0. Outputs for each assay were preprocessed following the Seurat Weighted Nearest Neighbor (WNN) analysis for 10x Multiome. Briefly, sc-RNA-seq data was subset to include cells with between 1000 and 25000 UMIs, between 1000 and 10000 genes, and less than 10% mitochondrial reads. Data was normalized using SCTransform. sc-ATAC-seq data was subset to include cells with between 5000 and 70000 fragments. Lanes with less than 50 cells were removed. Data was batch corrected using Harmony for the RNA and ATAC assays separately, UMAPs for each assay were generated, and scRNA- and scATAC-seq data was integrated using the command

FindMultimodalNeighbors. The WNN UMAP was generated, and clusters were identified from the SCT assay with 0.7 resolution. Demultiplexed samples with less than 50 cells were removed. Cells called as doublets by demuxlet, as well as cells with ambiguous genotype (Genome Studio genotype XY but normalized XIST gene expression > 0.2), were removed. Celltype, subclass, and class labels were transferred from Wang *et al.* 2025 using FindIntegrationAnchors on the SCT assays for both datasets, using only shared genes (cellranger reference for Wang dataset: Human GRCh38-2020-A (GENCODE v32/Ensembl98); cellranger reference for this study: Human GRCh38-2024-A (GENCODE v44/Ensembl110)).

Generation of cortex ATAC map

We supplemented the snATAC-seq data generated here with previously published snATAC-seq data to generate a map of chromatin accessibility in the fetal cortex (Ziffra 2021[11]). Using the Sinto tool (https://timoast.github.io/sinto), in conjunction with dataset-specific cell-type and donor labels, we sex & cell-type pseudobulked. For the Ziffra datasets, we utilized the following cell types: early excitatory neuron, upper-layer excitatory neuron, deep-layer excitatory neuron, CGE-derived interneuron, MGE-derived interneuron, oligodendrocyte progenitor cell, IPC, and radial glia. For the Multiome dataset from this study, we used the following cell types: layer 2/3 excitatory neuron, layer 4 excitatory neuron, layer 5 excitatory neuron, layer 6/subplate excitatory neuron, CGE-derived interneuron, MGE-derived interneuron, oligodendrocyte progenitor cell, IPC, radial glia and dividing radial glia. Peak-calling was performed on each pseudobulk using MACS2 with the following parameters: --SPMR -q 0.01 -gs --nomodel --ext-size 50 --shift 100.[63,64] Using an index script,[65] we generated a set of consensus peaks from the pseudobulked MACS2 outputs and reads from the pseudobulked datasets over the peaks were counted using the featureCounts read quantifier as part of the Subread package[66] (https://github.com/ShiLab-

Bioinformatics/subread). The consensus peaks were filtered for peaks on chromosomes 1-22, X and Y and mean FPKM above 2.

Cell type- and sex-specific peak calling and scE2G CRE predictions

Cell type- and sex-specific candidate enhancers were identified and linked to their predicted target genes within our snMultiome data using the single-cell enhancer-to-gene (scE2G)[32] method's standard multiome pipeline (https://github.com/EngreitzLab/scE2G). Pseudobulk fragment input files were generated for each analysis group using the SplitFragments function from the Signac v1.14.0 package.[67] Fragment files were sorted by coordinate using sortBed from the bedtools v2.27.1 package[68] with the '-i' flag. Samtools v1.21[69] was then used to individually zip each fragment file using the bgzip command then index each zipped file using the tabix command with the '-p bed' flag. RNA count matrix input files were generated by extracting raw counts for each analysis group from the Seurat object, ensuring that the cell IDs matched those which were present in the corresponding fragment files, then exporting each matrix to csv and compressing the resulting files using gzip. Pipeline configuration files were generated to follow the standard multiome prediction workflow using the provided 'multiome_powerlaw_v3' model. Once all files were prepared, the pipeline was run using the following command: snakemake -j1 --use-conda --configfile config/config.yaml. A threshold of 0.177 was used to generate the final CRE maps. Due to lack of power, we were not able to predict CREs on the Y chromosome.

CUT&Tag

CUT&Tag was performed on flash frozen tissue as previously described[31] with some modifications. 3mg of tissue was thawed and minced, then resuspended in PBS with 0.1% formaldehyde for 1 minute. 1.25M glycine was added to double the molar concentration of formaldehyde and stop cross linking, and cells were spun down at 1300g at 4C for 3 minutes. Tissue was resuspended in

1ml PBS and centrifuged at 2000g for 5 min at 4C. Tissue was then resuspended in wash buffer (20mM HEPES pH 7.5, 150mM NaCl, 0.5mM spermidine, and 1 EDTA-free complete protease inhibitor tablet) and homogenized using a glass dounce. Cells were transferred to a clean 1.5ml tube and centrifuged at 3000g for 3min at RT, and resuspended in 1ml wash buffer. Wash step was repeated twice more for a total of 3 washes. Concanavalin A-coated beads (Fisher Scientific #NC1526856) were prepared by adding 10μL beads per reaction to bead-binding buffer (20mM HEPES pH 7.9, 10mM KCl, 1mM CaCl2, and 1mM MnCl2). Using a magnetic rack, binding buffer was removed, and beads were resuspended in binding buffer once more before removing the binding buffer again and finally resuspending in enough binding buffer for 10μL per reaction. 10μL beads were then added to cells and incubated on an end over end rotator for 10 minutes at room temperature. Wash buffer was removed from cells using a magnetic rack, and cells were resuspended in enough antibody buffer (wash buffer with 2mM EDTA, 0.1% BSA, and 0.05% digitonin) for 50μL per reaction. Primary antibodies (H3K27me3 Cell Signaling Technologies #9733, H3K4me3 Cell Signaling Technologies #9751T, H3K27ac Cell Signaling Technologies #8173, CTCF Epicypher #13-2014, MEF2C Cell Signaling Technologies #5030T, AR Active Motif #39081, ESR1 Epicypher #13-2012, IgG Epicypher #13- 0042) were added at 1:50 dilution and samples were nutated overnight at 4C. Using a magnetic rack, primary antibody was removed, and cells were resuspended in 100μL secondary antibody (Antibodies Online #ABIN101961) diluted in wash buffer with 0.05% digitonin. Cells were then nutated for one hour at room temperature. Secondary antibody mix was removed using a magnetic rack, and cells were washed with wash buffer with 0.05% digitonin three times. pA-Tn5 preloaded with Nextera adapters (Epicypher #15-1117) was diluted 1:20 in dig-300 buffer (20mM HEPES pH7.5, 300mM NaCl, 0.5mM spermidine, 0.015% digitonin with 1 EDTA-free complete protease inhibitor tablet) and

cells were resuspended in 50µL of pA-Tn5 mix. Cells were nutated for one hour at room temperature. Using a magnetic rack, pA-Tn5 mix was removed, and cells were washed three times in dig-wash buffer. Cells were resuspended in 300µL of tagmentation buffer (dig-300 buffer with 10mM MgCl2) and incubated in a 37C water bath for one hour. Cells were released from beads with addition of 10µL 0.5M EDTA, 3µL 10% SDS, and 2.5µL 20mg/ml proteinase K. Samples were vortexed and incubated in a heat block at 55C for one hour. Fragments were purified by adding 300µL phenol:chloroform:isoamyl alcohol (25:24:1 v/v) and sample to a phase lock tube (Qiagen #129046), and spun down at 16,000g for 3 minutes at RT. 300µL chloroform was added to each sample and spun down once more at 16,000g for 3 minutes at RT. The aqueous layer was added to a 1.5ml lo-bind tube with 750µL 100% ethanol and mixed by pipetting. Samples were cooled on ice and spun down at 16,000g for 15 minutes at 4C. Supernatant was decanted, and samples were washed once more with 1ml 100% ethanol and spun down at 16,000g for 1 minute at 4C. Ethanol was decanted, the pellet was air dried, and resuspended in 22µL water. Libraries were amplified by mixing 21µL sample, 2µL each i5 and i7 primers (Illumina #FC-131-2001), and 25µL NEBNext High Fidelity 2X Master Mix (NEB #M0541S) and using the following PCR cycle settings: 72C for 5 minutes, 98C for 30 seconds, 98C for 10 seconds, 61C for 10 seconds, repeat steps 3-4 15x, and a final 72C incubation for 1 minute. Libraries were purified using SPRI Select Reagent (Beckman Coulter #B23317) and eluted in 20µL water. Libraries were pooled to 2nM and diluted to a final concentration of 750pM with 2% PhiX spike in for sequencing using the Illumina NextSeq 2000 with a targeted read depth of ~10 million reads per sample.

CUT&Tag analysis

Reads were trimmed using TrimGalore, aligned to reference genomes for hg38 and E. coli using

bowtie2 with parameters "--end-to-end --very-sensitive --no-mixed --no-discordant --phred33 -I 10 -X 700." Duplicates were marked and removed with picard. Peaks for each sample were called using MACS2 with respective IgG controls for each sample with parameters "-q 0.01 -B --SPMR --keep-dup all --nomodel." Consensus peak files were generated by concatenating bed files from all samples for each histone mark.

Differential analysis

Gene expression:

The multimodal object was randomly downsampled (scCustomize package) so male and female cells were about evenly represented among all cell types. Differentially expressed genes were identified using Dreamlet[33] (https://github.com/GabrielHoffman/dreamlet). Briefly, ribosomal and mitochondrial genes were removed. Since the same samples were run across multiple libraries, sample_library ID was used for pseudobulking by cell type with the aggregateToPseudobulk command with default parameters. Gene expression counts of the pseudobulked object were normalized with the function processAssays with default parameters, using the model: ~ genotype + scale(age) + (1 | Sample) + (1 | library) + scale(n_genes). Variance partition was run to identify variance due to each covariate. Differential analysis was run using the dreamlet function with the model: ~ genotype + scale(age) + (1|library) + scale(n_genes). Genes with FDR < 0.05 were used for downstream analysis. For DARs, thresholds of FDR < 0.05 and log2FC > |1.25| were used.

Chromatin accessibility:

Differentially accessible regions (DARs) were determined as with DEGs, using scaled log2 normalized read depth per sample per cell type instead of number of genes. Since the ATAC data is more sparse than gene expression, we normalized ATAC counts of the pseudobulked object using processAssays with minimum counts = 10 (default is 5).

DAR and DEG correlation

We leveraged our CRE-gene links identified by scE2G (methods above) for either bulk XX samples or bulk XY to link DARs to target genes. Briefly, the scE2G bedpe file with CRE-gene links was intersected with the respective DAR bed file (bulk XY scE2G with male-biased DAR, and bulk XX scE2G with female-biased DAR). We then mapped the associated gene from the scE2G file onto the DAR file to identify target genes for DARs. Then, Pearson's correlation was used to determine if there was a significant correlation between the log2FC enrichment of accessible chromatin (DARs) and gene expression (DEG) for each male and female.

Motif enrichment analysis

Bed files were used as input for Homer (https://github.com/javrodriguez/HOMER), using the findMotifsGenome command with default parameters. The knownResults predictions were used, and fold enrichment for each motif was calculated by dividing the target enrichment by the background enrichment. Sex-specific motifs were determined by calling motifs with enrichment in one sex > 1 and enrichment in the other sex < 1.

scMEGA gene regulatory network prediction

Gene regulatory network analysis was done using Single-cell Multiomic Enhancer-based Gene regulAtory network inference (scMEGA, https://github.com/CostaLab/scMEGA).[41] Using our preprocessed snMultiome object, we first determined pseudotime trajectory with ArchR[70] for computing TF binding activity with TF expression along this trajectory. Next, we created the motif slot of our ATAC assay in our snMultiome object by running getMatrixSet (using JASPAR2020 database), createMotifMatrix with hg38, and createMotifObject to be integrated into the metadata in the ATAC assay, where we subsequently used RunChromVar with BSgenome.Hsapiens.UCSC.hg38. To associate TF binding motifs with target genes, we linked

peaks to genes with SelectGenes and correlated TFs with target genes by running GetTFGeneCorrelation using the chromVar TF assay, RNA assay, and inferred pseudotime trajectory. We lastly associated TFs with genes by running motif.matching, and finally inferred gene regulatory networks with GetGRN using our matched motifs, TF-gene correlations, and peak-gene links. The output is a dataframe with TF regulator, gene target, and correlation score, where we used TF-gene predictions with a score of 0.4 or greater.

Genetic analysis of NDD cohorts and sex-biased gene discovery

Variants from the SPARK cohort were called by GATK[71] (v3.7-0-gcfedb6) and FreeBayes[72] (v1.1.0-3-g961e5f3) based on GRCh38. DNMs were discovered using developed scripts as described in Wang et al. 2022. Briefly, Candidate DNVs were defined as sites where sequencing depth greater than 9 in all trio members, and both parents were "0/0" with no alternate reads in parents (AD=",0"), while the child was "0/1" or "1/1", and child allelic balance >0.25, child genotype quality >20 (GATK) or sum of quality of the alternate observations >20 (FreeBayes), and supported by both GATK and FreeBayes call sets. DNVs in low complexity and recent repeat [73] and in outliers (samples with 8 DNMs, Poisson distribution) were filtered. The coordinates of DNMs were then converted to GRCh37 using UCSC LiftOver Tool.[74] We annotated *de novo* likely gene-disruptive (dnLGD, including frameshift, stop-gain, splice-donor and splice-acceptor) and *de novo* missense (dnMIS) variants by using Ensembl Variant Effect Predictor (VEP, v110.1)[75] and CADD score (v1.3).[76] DNMs with allele frequency > 0.1% in gnomAD exomes v2.1.1[77] were filtered. Finally, we integrated DNMs from known females and males including samples from SPARKiWESv2,[3] SPARK pilot,[42] MSSNG and AGRE cohorts,[43] ASC and SSC cohorts,[2] and DD samples.[44] DNMs that were observed in both affected and unaffected (n=9,754) were removed.

CH[45] and DR[46] models were used to identify genes with a significant excess of DNMs in 41,961 NDD males and 19,782 NDD females separately.[3] Genes were considered significantly enriched if they harbored ≥2 DNMs and passed the 5% family-wise error rate threshold ($p < 5.1E-07$, five tests) in both models. Total DNV counts for each enriched gene were compared between males and females using two-sided Fisher's exact tests, with p values adjusted for multiple testing using the Benjamini–Hochberg method[35] (n=20,000 genes). DNM counts in X-chromosomal genes were normalized by the number of X chromosomes represented in the analyzed subset of females (n = 35,076) and males (n = 31,675). This analysis was also performed for DD (17,421 males and 13,635 females) and ASD groups (24,540 ASD males and 6,147 ASD females, and 14,254 and 7806 copies of X chromosome for males and females, respectively).

GO and PPI analysis of sex-biased NDD genes

Gene ontology analysis was done using the clusterProfiler package with command compareCluster using NDD_male32 genes and NDD_female35 genes as input, with molecular function ontology assessed. STRING v12.0 database[78] was used to perform PPI network analysis. We used the multiple protein input option with default settings including "confidence for meaning of network edges", "textmining", "experiments", "databases", "co-expression", "neighborhood", "Gene Fusion", and "co-occurence" for active interaction sources", and "minimum required interaction score" of 0.400. Disconnected nodes were hidden from the network output but not from the enrichment analyses.

DEG/DAR and NDD enrichment

We defined the following variables: q = number of genes that overlap, m = number of NDD genes, n = number of total genes expressed (15,000), k = number of DEG or genes with DAR. We used a

2x2 matrix defined as (n-k, m-q, k-q, q) to be used for two-sided Fisher exact test, unless otherwise stated. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Data availability

Raw data is available at dbGaP. Processed data is available on the UCSC cell browser at https://cells.ucsc.edu/?ds=ssd-prenatal-cortex+rna.

Code availability

All scripts used for data analysis are available at https://github.com/NOW-Lab.

References

1. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e23 (2020).
2. 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).
3. Wang, T. *et al. Integrated Gene Analyses of* de Novo *Mutations from 46,612 Trios with Autism and Developmental Disorders*. http://biorxiv.org/lookup/doi/10.1101/2021.09.15.460398 (2021) doi:10.1101/2021.09.15.460398.
4. Werling, D. M. & Geschwind, D. H. Sex differences in autism spectrum disorders. *Curr. Opin. Neurol.* **26**, 146–153 (2013).
5. Willsey, H. R. *et al.* Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. *Neuron* **109**, 788-804.e8 (2021).
6. Parikshak, N. N. *et al.* Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. *Cell* **155**, 1008–1021 (2013).
7. Hartl, C. L. *et al.* Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. *Nat. Neurosci.* **24**, 1313–1323 (2021).
8. Braun, E. *et al.* Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, (2023).
9. Mannens, C. C. A. *et al.* Chromatin accessibility during human first-trimester neurodevelopment. *Nature* (2024) doi:10.1038/s41586-024-07234-1.

10. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

11. Ziffra, R. S. *et al.* Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).

12. Wang, L. *et al.* Molecular and cellular dynamics of the developing human neocortex. *Nature* (2025) doi:10.1038/s41586-024-08351-7.

13. Miyagi, M., Guthman, E. M. & Sun, S. D.-K. Transgender rights rely on inclusive language. *Science* **374**, 1568–1569 (2021).

14. Massa, M. G., Aghi, K. & Hill, M. J. Deconstructing sex: Strategies for undoing binary thinking in neuroendocrinology and behavior. *Horm. Behav.* **156**, 105441 (2023).

15. Garcia-Sifuentes, Y. & Maney, D. L. Reporting and misreporting of sex differences in the biological sciences. *eLife* **10**, e70817 (2021).

16. Werling, D. M. & Geschwind, D. H. Understanding sex bias in autism spectrum disorder. *Proc. Natl. Acad. Sci.* **110**, 4868–4869 (2013).

17. Fass, S. B. *et al.* Relationship between sex biases in gene expression and sex biases in autism and Alzheimer's disease. *Biol. Sex Differ.* **15**, (2024).

18. Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020).

19. Velmeshev, D. *et al.* Single-cell analysis of prenatal and postnatal human cortical development. *Science* **382**, (2023).

20. An, J.-Y. *et al.* Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* **362**, (2018).

21. Shin, T. *et al.* Rare variation in non-coding regions with evolutionary signatures contributes to autism spectrum disorder risk. *Cell Genomics* **4**, 100609 (2024).

22. Gegenhuber, B., Wu, M. V., Bronstein, R. & Tollkuhn, J. Gene regulation by gonadal hormone receptors underlies brain sex differences. *Nature* **606**, 153–159 (2022).

23. Abramovich, D. R. & Rowe, P. FOETAL PLASMA TESTOSTERONE LEVELS AT MID-PREGNANCY AND AT TERM: RELATIONSHIP TO FOETAL SEX. *J. Endocrinol.* **56**, 621–622 (1973).

24. Clements, J. A., Reyes, F. I., Winter, J. S. D. & Faiman, C. Studies on Human Sexual Development. III. Fetal Pituitary and Serum, and Amniotic Fluid Concentrations of LH, CG, and FSH. *J. Clin. Endocrinol. Metab.* **42**, 9–19 (1976).

25. Reyes, F. I., Winter, J. S. D. & Faiman, C. Studies on Human Sexual Development. I. Fetal Gonadal and Adrenal Sex Steroids[1]. *J. Clin. Endocrinol. Metab.* **37**, 74–78 (1973).

26. Martínez-Cerdeño, V., Noctor, S. C. & Kriegstein, A. R. Estradiol stimulates progenitor cell division in the ventricular and subventricular zones of the embryonic neocortex. *Eur. J. Neurosci.* **24**, 3475–3488 (2006).

27. Kelava, I., Chiaradia, I., Pellegrini, L., Kalinka, A. T. & Lancaster, M. A. Androgens increase excitatory neurogenic potential in human brain organoids. *Nature* **602**, 112–116 (2022).

28. Lombardo, M. V. *et al.* Fetal Testosterone Influences Sexually Dimorphic Gray Matter in the Human Brain. *J. Neurosci.* **32**, 674–680 (2012).

29. Auyeung, B., Lombardo, M. V. & Baron-Cohen, S. Prenatal and postnatal hormone effects on the human brain and cognition. *Pflüg. Arch. - Eur. J. Physiol.* **465**, 557–571 (2013).

30. Dooley, N. *et al.* Is there an association between prenatal testosterone and autistic traits in adolescents? *Psychoneuroendocrinology* **136**, 105623 (2022).

31. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).

32. Sheth, M. U. *et al.* Mapping enhancer-gene regulatory interactions from single-cell data. Preprint at https://doi.org/10.1101/2024.11.23.624931 (2024).

33. Hoffman, G. *et al.* Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. Preprint at https://doi.org/10.21203/rs.3.rs-2705625/v1 (2023).

34. San Roman, A. K. *et al.* The human inactive X chromosome modulates expression of the active X chromosome. *Cell Genomics* **3**, 100259 (2023).

35. Benjamini, Yoav & Hochberg, Yosef. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

36. Mitxelena, J. *et al.* An E2F7-dependent transcriptional program modulates DNA damage repair and genomic stability. *Nucleic Acids Res.* **46**, 4546–4559 (2018).

37. Di Stefano, L. E2F7, a novel E2F featuring DP-independent repression of a subset of E2F-regulated genes. *EMBO J.* **22**, 6289–6298 (2003).

38. Rocha, H., Sampaio, M., Rocha, R., Fernandes, S. & Leão, M. MEF2C haploinsufficiency syndrome: Report of a new MEF2C mutation and review. *Eur. J. Med. Genet.* **59**, 478–482 (2016).

39. Cooley Coleman, J. A. *et al.* Comprehensive investigation of the phenotype of *MEF2C* -RELATED disorders in human patients: A systematic review. *Am. J. Med. Genet. A.* **185**, 3884–3894 (2021).

40. Cosgrove, D. *et al.* Genes influenced by MEF2C contribute to neurodevelopmental disease via gene expression changes that affect multiple types of cortical excitatory neurons. *Hum. Mol. Genet.* **30**, 961–970 (2021).

41. Li, Z., Nagai, J. S., Kuppe, C., Kramann, R. & Costa, I. G. scMEGA: single-cell multi-omic enhancer-based gene regulatory network inference. *Bioinforma. Adv.* **3**, (2023).

42. Feliciano, P. *et al.* Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. *Npj Genomic Med.* **4**, (2019).

43. Trost, B. *et al.* Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell* **185**, 4409-4427.e18 (2022).

44. Kaplanis, J. *et al.* Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020).

45. O'Roak, B. J. *et al.* Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders. *Science* **338**, 1619–1622 (2012).

46. Ware, J. S., Samocha, K. E., Homsy, J. & Daly, M. J. Interpreting *de novo* Variation in Human Disease Using denovolyzeR. *Curr. Protoc. Hum. Genet.* **87**, (2015).

47. Turner, T. N. *et al.* Sex-Based Analysis of De Novo Variants in Neurodevelopmental Disorders. *Am. J. Hum. Genet.* **105**, 1274–1285 (2019).

48. Martin, H. C. *et al.* The contribution of X-linked coding variation to severe developmental disorders. *Nat. Commun.* **12**, 627 (2021).

49. Lennox, A. L. *et al.* Pathogenic DDX3X Mutations Impair RNA Metabolism and Neurogenesis during Fetal Cortical Development. *Neuron* **106**, 404-420.e8 (2020).

50. Nguyen, T. A. *et al.* A Cluster of Autism-Associated Variants on X-Linked NLGN4X Functionally Resemble NLGN4Y. *Neuron* **106**, 759-768.e7 (2020).

51. Willsey, A. J. *et al.* Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism. *Cell* **155**, 997–1007 (2013).

52. Chen, C. *et al.* YY1 binding association with sex-biased transcription revealed through X-linked transcript levels and allelic binding analyses. *Sci. Rep.* **6**, (2016).

53. Makhlouf, M. *et al.* A prominent and conserved role for YY1 in Xist transcriptional activation. *Nat. Commun.* **5**, (2014).

54. Pereira, M. F. *et al.* YY1 mutations disrupt corticogenesis through a cell type specific rewiring of cell-autonomous and non-cell-autonomous transcriptional programs. *Mol. Psychiatry* **30**, 3413–3429 (2025).

55. DeCasien, A. R., Auluck, P., Liu, S., Marenco, S. R. & Cookson, M. Sex effects on gene expression across the human cerebral cortex at single cell resolution.

56. Lombardo, M. V. *et al.* Sex-specific impact of prenatal androgens on social brain default mode subsystems. *Mol. Psychiatry* **25**, 2175–2188 (2020).

57. Matassa, G. *et al.* A molecular cell atlas of endocrine signalling in human neural organoids.

58. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

59. Shao, Y. *et al.* Identification and characterization of conserved noncoding *cis* -regulatory elements that impact *Mecp2* expression and neurological functions. *Genes Dev.* **35**, 489–494 (2021).

60. Feliciano, P. *et al.* SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron* **97**, 488–493 (2018).

61. Willsey, H. R., Willsey, A. J., Wang, B. & State, M. W. Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. *Nat. Rev. Neurosci.* **23**, 323–341 (2022).

62. Kang, H. M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).

63. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

64. Lee, D. *et al.* Human cardiac *cis* -regulatory elements, their cognate transcription factors, and regulatory DNA sequence variants. *Genome Res.* **28**, 1577–1588 (2018).

65. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (2020).

66. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

67. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

68. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

69. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

70. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

71. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at https://doi.org/10.1101/201178 (2017).

72. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at https://doi.org/10.48550/arXiv.1207.3907 (2012).

73. Perez, G. *et al.* The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res.* **53**, D1243–D1249 (2025).

74. Hinrichs, A. S. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).

75. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

76. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).

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

78. Szklarczyk, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

CHAPTER 5: CONCLUSIONS

Limitations of the current study

The work described here takes important steps forward as we try to understand how genetic variation contributes to NDDs. First, we identified convergent biological processes driven by protein-protein interactions that change with autism associated mutations, specifically highlighting the function of the FOXP1-FOXP4 heterodimer and the necessity of FOXP4 in the observed phenotypes. Then, we mapped paired gene expression and chromatin accessibility in developing human male and female brains, identifying differences in gene expression that are critical to sex-specific development and sex-biased vulnerability. Together, this work underscores the importance of both limiting the model to study the brain region and developmental time of interest, as well as expanding the model to look beyond gene expression and single genes implicated in NDDs.

However, there are important limitations of the current work to consider. First, Chapter 3 describes a PPI map of high confidence autism risk genes, but the initial experiments were carried out in HEK293 cells, which is not a biologically meaningful cell type. These proteins were also overexpressed, which may cause artifacts that confound the results. My contribution to the work included introducing the patient relevant mutations into the endogenous locus of FOXP1 and observing phenotypes in forebrain organoids, which addresses these caveats, but this should be expanded to include the other hcASD risk genes of interest.

In addition, while this work was done across multiple FOXP1 mutations, the number of clones of each genotype was limited (1-2), and only one parental cell line was used. Future work would benefit from leveraging multiple genetic backgrounds and multiple clones. Using experiments across organoid batches and FOXP1 mutations provides additional strength to the claims made here in lieu of multiple parental cell lines, but experimental follow up should include this as part

of the design. Additional integration of mouse models would also provide critical insights into developmental and circuit level questions that are challenging to address in vitro.

In Chapter 4, we generate a midgestation atlas of the male and female cortex, including paired chromatin accessibility and gene expression. While our goal was to generate robust statistical power to make comparisons between males and females at this time in this brain region, this analysis should be extended to additional developmental times, as well as postnatal and adult. There are atlases for this[1,2], and this data should be leveraged to draw conclusions about which sex differences persist, and which are specific to midgestation. The field is beginning to make these comparisons[1], but it is critical to include both gene expression as well as chromatin accessibility as we move towards understanding the gene regulatory networks involved in human cortex development and related disorders.

While we aimed to generate a robust atlas across hundreds of thousands of cells, technical hurdles while conducting the snMultiome assay led to lower cell numbers than we targeted. Rigorous statistical modeling ensures all claims are true despite low cell numbers, but additional experiments boosting cell number would assist with detection of sex differences in cell types with low representation, such as microglia. In the future, using FANS to enrich for these populations would also be worthwhile.

Lastly, this work focuses on the genomic contributions with little emphasis on the hormonal contribution to sex differences. While we report little overlap between androgen receptor or estrogen receptor genomic targets and differentially expression genes, additional work exposing prenatal tissue to these hormones and profiling gene expression differences could help bolster the

additional studies in this field. Though as it stands, we and others believe hormones contribute minimally to cortical sex differences.

Impact and future outlook

Despite these limitations, this work provides an important framework for studying the impact of NDD associated variation during cortex development. The platform for endogenous introduction of patient variants in iPSCs paired with forebrain organoid differentiation provides a powerful tool for investigating mutations in additional proteins beyond FOXP1, and this work is being carried out in the lab. Additionally, it is critical to understand the convergent and divergent phenotypes associated with all genetic variants in a single gene. The work here highlights the same developmental phenotypes across mutations in disparate protein domains of FOXP1, but experiments such as deep mutational scanning of this gene and others will allow us to better understand the highly heterogenous genetic contributions to NDDs, despite their uniform association with specific conditions such as autism.

One of the most interesting phenotypes observed is the necessity of FOXP4 in mediating the changes in glutamatergic neuron proportion in FOXP1 mutants. The extent to which FOXP4 is necessary for other phenotypes, including gene expression changes and altered excitability, should be investigated. FOXP1-FOXP4 sufficiency for rescuing these phenotypes in a FOXP1 mutant background will also be critical for understanding the contribution of PPI to neurodevelopment.

My hope is that the sex balanced developmental atlas is useful for the community going forward, as this has already led to several manuscripts within our collaboration[3–5]. This work establishes an integrated genomic map of the X chromosome, which we believe contributes to sex biased vulnerability dependent on the nature of the mutation (*de novo* or inherited). However, this map is

only the beginning of years of work to come, with the additional of DNA methylation in males and females, with and without autism, and with the inclusion of common NDD variants, where we primarily looked at the impact of rare variants. This work also highlights MEF2C as a critical regulator of downstream transcriptional targets, but it is still unclear what contributes to this female biased expression and regulation. Probing methylation differences and potential environmental influences (including hormones, as MEF2C is a reported androgen responsive gene) could provide insights to this end.

If I had the gift of time, I would spend an additional PhD or two in Tom's lab investigating the outstanding questions outlined here. This work is meaningful to the field of NDD genomics, but leaves many additional avenues to pursue, and I look forward to seeing where this work goes.

References

1. Velmeshev, D. *et al.* Single-cell analysis of prenatal and postnatal human cortical development. *Science* **382**, (2023).

2. Wang, L. *et al.* Molecular and cellular dynamics of the developing human neocortex. *Nature* https://doi.org/10.1038/s41586-024-08351-7 (2025) doi:10.1038/s41586-024-08351-7.

3. Sui, Y. *et al.* Using the linear references from the pangenome to discover missing autism variants. *Nat. Commun.* **17**, 1681 (2026).

4. A massively parallel reporter assay of MECP2 cis-regulatory elements reveals genetic candidates for male-biased autism.

5. Berk-Rauch, H. E. *et al.* Sex Steroid Hormone Signaling Tunes Metabolic and Neuronal Programs in Human Cortical Development.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Kelsey Hennick
55DE810F4FB6470...
Author Signature

08/05/2026
Date