Loss of DOT1L function disrupts neuronal transcription, animal behavior, and leads to a novel neurodevelopmental disorder

Marissa J. Maroni,[1,3,4] Melissa Barton,[2,3,4] Katherine Lynch,[3,4] Ashish R. Deshwar,[5,6,7,8] Philip Campbell,[10,11] Josephine Millard,[3,11] Rachel Lee,[12] Annastelle Cohen,[2,3,4] Alekh Paranjapye,[3,4] Víctor Faundes,[13] Gabriela M. Repetto,[14] Caoimhe McKenna,[15] Amelle L. Shillington,[16] Chanika Phornphutkul,[17] Grazia MS. Mancini,[18] Rachel Schot,[18,19] Tahsin Stefan Barakat,[18,19,20,21] Christopher M. Richmond,[22,23] Julie Lauzon,[24] Ahmed Ibrahim Elsayed Ibrahim,[25] Daniel Natera-de Benito,[26] Carlos Ortez,[26] Berta Estevez-Arias,[26,27] François Lecoquierre,[28] Kévin Cassinari,[28] Anne-Marie Guerrot,[28] Jonathan Levy,[29,30] Xenia Latypova,[29,30] Alain Verloes,[29,30] A. Micheil Innes,[31] Xiao-Ru Yang,[31,32] Siddharth Banka,[33,34] Katharina Vill,[35] Maureen Jacob,[36] Michael Kruer,[37,38] Peter Skidmore,[37,38,39] Carolina I. Galaz-Montoya,[37,38,40] Somayeh Bakhtiari,[37,38] Jessica L. Mester,[41] Michael Granato,[11] Karim-Jean Armache,[12] Gregory Costain,[6,7,8,9] Erica Korb[3,4*]

* Corresponding author, ekorb@pennmedicine.upenn.edu

[1]Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, PA [2]Cell and Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, PA [3]Department of Genetics, University of Pennsylvania, Philadelphia, PA [4]Epigenetics Institute, University of Pennsylvania, Philadelphia, PA [5]Program in Developmental and Stem Cell Biology, Sickkids Research Institute, Toronto, Canada, [6]Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada, [7]Department of Pediatrics, University of Toronto, Toronto, Canada, [8]Department of Molecular Genetics, University of Toronto, Toronto, Canada, [9]Program in Genetics and Genome Biology, SickKids Research Institute, Toronto, Ontario, Canada [10]Department of Psychiatry, [11]Department of Cell and Developmental Biology, [12]Skirball Institute of Biomolecular Medicine, [13]Laboratorio de Genética y Enfermedades Metabólicas, Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, [14]Rare Diseases Program, Center for Genetics and Genomics, Institute for Science and Innovation in Medicine, Facultad de Medicina, Clínica Alemana-Universidad del Desarrollo, [15] Northern Ireland Regional Genetics Service, Belfast, Northern Ireland, [16]Cincinnati Children's Hospital Medical Center, [17]Rhode Island Hospital, [18]Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands, [19]Discovery Unit, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands, [20]Whole Genome Sequencing Implementation and Research Task Force, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands, [21]ENCORE Expertise Center for Neurodevelopmental Disorders, Erasmus MC University Medical Center, Rotterdam, The Netherlands, [22]Royal Brisbane & Women's Hospital, Herston, Queensland, Australia, [23]School of Medicine, Griffith University, Gold Coast, Queensland, Australia, [24]Alberta Children's Hospital, Calgary AB Canada Department of Medical Genetics, Cummings School of Medicine, University of Calgary, Alberta Canada, [25] Brody school of medicine, [26]Neuromuscular Unit, Hospital Sant Joan de Deu, Barcelona, Spain, [27]Laboratory of Neurogenetics and Molecular Medicine, Institut de Recerca Sant Joan de Deu, Barcelona, Spain, [28] Univ Rouen Normandie, Inserm U1245 and CHU Rouen, Department of Genetics and reference center for developmental disorders, Rouen, France, [29]Department of Genetics, APHP-Robert Debré University Hospital, Paris, France, [30]Laboratoire de biologie médicale multisites Seqoia - FMG2025, Paris, France, [31]University of Calgary Department of Medical Genetics; Alberta Children's Hospital Research Institute, [32]Department of Medical Genetics, University of British Columbia, [33]Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, M13 9WL Manchester, UK, [34]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Health Innovation Manchester, M13 9WL Manchester, UK, [35]Department of Pediatric Neurology and Developmental Medicine and LMU Center for Children with Medical Complexity, Dr. von Hauner Children's Hospital, LMU Hospital, Ludwig-Maximilians-University, Munich, Germany, [36]Institute of Human Genetics, Klinikum rechts der Isar, Technical University of Munich, School of Medicine and Health, Munich, Germany, [37]Pediatric Movement Disorders Program, Division of Pediatric Neurology, Barrow Neurological Institute, Phoenix Children's Hospital, Phoenix, AZ 85016, USA, [38]Departments of Child Health, Neurology, Cellular and Molecular Medicine, and Program in Genetics, University of Arizona College of Medicine—Phoenix, Phoenix, AZ 85004, USA, [39]College of Health Solutions, Arizona State University, Tempe, Arizona, USA, [40]Genetics, GIDP PhD Program, Tucson, AZ, USA, [41]GeneDx, Gaithersburg, MD, 20877, USA

ABSTRACT

Individuals with monoallelic pathogenic variants in the histone lysine methyltransferase DOT1L display global developmental delay and varying congenital anomalies. However, the impact of monoallelic loss of *DOT1L* remains unclear. Here, we present a largely female cohort of 11 individuals with *DOT1L* variants with developmental delays and dysmorphic facial features. We found that *DOT1L* variants include missense variants clustered in the catalytic domain, frameshift, and stop-gain variants. We

demonstrate that specific variants cause loss of methyltransferase activity and therefore sought to define the effects of decreased DOT1L function. Using RNA-sequencing of cultured neurons and single nucleus RNA-sequencing of mouse cortical tissue, we found that partial *Dot1l* depletion causes sex-specific transcriptional responses and disrupts transcription of synaptic genes. Further, *Dot1l* loss alters neuron branching and expression of synaptic proteins. Lastly using zebrafish and mouse models, we found behavioral disruptions that include sex-specific deficits in mice. Overall, we define how DOT1L loss leads to neurological dysfunction by demonstrating that partial *Dot1l* loss impacts transcription, neuron morphology, and behavior across multiple models and systems.

INTRODUCTION

Neurodevelopmental disorders (NDDs) are a diverse group of highly prevalent (0.3 – 18.5%) (*1*) conditions that manifest during development and impact central nervous system functions (*2*, *3*). The spectrum of NDDs include intellectual disability, autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), communication disorders, specific learning disabilities, and motor disorders (*4*). The cause of NDDs is multifactorial, and includes both inherited and *de novo*, genetic variants with a notable overrepresentation of epigenetic regulators (*5*–*9*). One subset of epigenetic regulators, histone methyltransferases, are linked to numerous NDDs (*10*–*12*) and function by methylating histones to regulate transcription. Histone methyltransferases are critical for neurogenesis, neuronal migration, neuronal differentiation, synaptic plasticity and cognition (*13*) yet several disease-linked methyltransferases have not yet been studied in the context of neuronal function or animal behavior.

Prior exome sequencing studies identified variants in the histone methyltransferase *DOT1L* as a potential causative driver of NDDs (*6*, *12*). More recent work identified two variants in *DOT1L* in individuals displaying ADHD (*14*) and nine monoallelic (presumed) *de novo* variants of *DOT1L* were identified in individuals with global developmental delay (*15*). Complete loss of DOT1L in mouse models is embryonic lethal (*16*) while in *Drosophila*, loss of *grappa*, the *Drosophila DOT1L* ortholog, leads to developmental delay and lethality (*15*). However, *grappa* is highly divergent from DOT1L and thus does not provide an ideal model to study an emerging human disorder. Further, while two previously identified variants were proposed to be gain-of-function based on human cell based-assays (*15*) most identified variants have unclear functional consequences. Lastly, while most prior work used full and transmitted knockout models, variants are typically monoallelic and *de novo* making it difficult to define the effect of partial DOT1L disruption from existing data. Thus, the underlying mechanisms linking DOT1L to NDDs remain unclear.

DOT1L is the sole methyltransferase responsible for depositing mono-, di-, and trimethyl methyl marks on the histone-fold domain on residue 79 of histone H3 (H3K79me) (*17*, *18*). H3K79me is

enriched in gene bodies peaking after the transcription start site (*17*), with higher methyl states linked to greater transcriptional output (*19*). DOT1L interacts with RNA polymerase II (*20*) and TFIID (*21*) and recruits effector proteins such as Menin (*22*) to regulate transcription. DOT1L functions in numerous cellular processes, including development (*23*) such as in neural progenitor proliferation and differentiation in the cortex, cerebellum, and spinal cord (*24–27*) and in maintaining the transcriptional state in differentiating neural progenitors(*28–31*). Further work demonstrated that stress modulates DOT1L expression and H3K79me in the nucleus accumbens and that monoallelic loss of *Dot1l* in the midbrain disrupts synaptic and mitochondrial genes (*32*). Cumulatively, this suggests that DOT1L is critical in neuronal development and neuronal function. Despite these advances, the majority of DOT1L research has focused on biallelic loss of *DOT1L,* which does not reflect the monoallelic nature of individual variants or has not examined effects on development and behavior. Thus, the consequences of monoallelic disruptions of *DOT1L* remain poorly understood.

Here, we identified 11 individuals with monoallelic variants in *DOT1L* displaying a spectrum of neurodevelopmental phenotypes and dysmorphic facial features. Using structural protein modeling, biochemical studies, and patient-derived cells, we found that several variants cause loss of DOT1L methyltransferase activity. Utilizing a *dot1l* knockdown system in zebrafish, we identified disruptions in motor responses to sensory stimuli. Harnessing both primary cultured cortical neurons bulk RNA-sequencing and *in vivo* cortical neuronal single-nucleus RNA-sequencing in mice, we show that partial loss of *Dot1l* affects transcription of critical neuronal genes linked to synaptic function and causes sex-specific transcriptional responses. Further, cortical neurons display disruptions in neuronal morphology upon partial *Dot1l* loss. Finally, we identified behavioral alterations upon both ubiquitous and neuron-specific monoallelic loss of *Dot1l* in mice. Together, our work demonstrates that partial loss of *Dot1l* causes transcriptional disruptions impacting cognitive function and provides insight into the neurodevelopmental disruptions found in individuals with *DOT1L* variants.

RESULTS

Identification of individuals with a spectrum of neurodevelopmental disorders and DOT1L variants

Given the recent discovery of DOT1L's association with an emerging neurodevelopmental disorder, we searched specifically for individuals harboring variants in *DOT1L*. We collected a cohort of individuals through collaborating clinicians and GeneMatcher (*33*) with *DOT1L* variants identified through genome sequencing or exome sequencing. Criteria for inclusion consisted of individuals displaying developmental phenotypes with *DOT1L* variants not observed in multiple individuals from the general population and without additional known pathogenic variants.

Using these criteria, we compiled a cohort of 11 individuals with variants in *DOT1L*. All individuals have only a single monoallelic variant in *DOT1L*. Inheritance of these variants was *de novo* (6/11), maternal (1/11) or inconclusive due to one or both parents being unavailable to be sequenced (4/11). In the maternally inherited case, dysmorphic facial features were noted in the mother, but full phenotyping was not available. Variants include missense (9/11), frameshift (1/11) and stop-gain (1/11). Notably, all but one of the missense variants (8/9) are within the catalytic domain of DOT1L and affect amino acids that are rarely altered in humans (i.e. dn/ds score of <0.2), indicating intolerance to variation at these sites (Fig. 1A). According to gnomAD (v4.1.0) (*34*), DOT1L has a high probability of loss-of-function intolerance (pLI = 1, LOEUF = 0.32) and a high probability of deletion intolerance (pHaplo = 0.98). All variants were absent in gnomAD (v4.1.0) apart from one counted allele of p.L1067Dfs*66. The cohort displays a non-specific constellation of congenital anomalies, including craniofacial anomalies (10/11) such as midface hypoplasia (Table 1, Fig. 1B). Based on the information available at this time, there is no recognizable pattern of morphological differences that would suggest the diagnosis in the absence of molecular genetic testing. Additional individual phenotypes include intellectual disability (2/11), language delay (8/11), motor delay (7/11), and a diagnosis of ASD (2/11) (Table 1, Fig. 1B). Three additional individuals with variants in *DOT1L* were identified through the MSSNG (*35*) database with a diagnosis of ASD but are not included in the main cohort due to an inability to gather additional information (fig. S1A, S1B, table S2). Further, four individuals with variants in DOT1L also contained additional potential pathogenic variants or had a DOT1L variant found the general population and thus did not meet criteria for inclusion in the cohort. We include them here (fig. S1A, S1B, table S2) given that they shared some features with the main cohort and that we cannot rule out the possibility of incomplete penetrance of this disorder. Interestingly, while the cohort is not sufficiently powered to confidently determine sex enrichment and the prior smaller cohort (*15*) was split roughly equally by sex, 9 out of 11 the individuals in this cohort are female suggestive of possible sex bias.

We next determined the location of missense variants in the catalytic domain of DOT1L based on a published structure of DOT1L (PDBID: 6NJ9) (*36*) (Fig. 1C). Variants are spread throughout the catalytic domain, including regions in close proximity to the binding pocket and nucleosome interface likely to affect DOT1L function. Given that most of the variants lie within the catalytic domain, we assessed methyltransferase activity via endpoint histone methyltransferase assays. We selected two previously published variants (*15*) (p.R292C and p.E123K) one of which was reported to have no effect (p.R292C) and the other of which was proposed to increase activity (p.E123K). We also assessed the p.D157N variant based on the identification of variants at residue 157 in two unrelated individuals. Methyltransferase assays demonstrated that p.R292C and p.D157N reduced methyltransferase activity (Fig. 1D, fig. S1C). In contrast, p.E123K increased activity, as previously reported (*15*). Further, human

fibroblasts harboring the p.D157N had a decrease in all three H3K79me states compared to age- and sex-matched control fibroblasts further supportive of loss of catalytic activity in DOT1L (Fig. 1E, fig. S1D). Lastly, to determine the effect of p.D157N in an orthogonal system without the confound of different genetic backgrounds from primary human fibroblasts, we overexpressed wildtype *Dot1l* and variant *Dot1l* (p.D157N) in mouse Neuro-2A cells. Wildtype *Dot1l* increased H3K79me2 and, to a lesser extent H3K79me1/3 (Fig. 1F). However, variant *Dot1l* (p.D157N) had no detectable impact on H3K79me levels. Together, these data demonstrate that the p.D157N variant reduces catalytic activity.

In addition to functional testing of *DOT1L* variants, we noted that two variants cause early stop codons that will either result in nonsense-mediated decay or a severely truncated protein. Further, these variants truncate DOT1L upstream of nuclear localization sequences (Fig. 1A), likely preventing DOT1L from performing established nuclear functions if translated into protein. Together, this suggests that both gain- and loss-of-function variants are found within DOT1L. Given our findings that both a previously published variant and the D157N variant have reduced catalytic activity and the discovery of two patients with stop-gain variants, we chose to examine the effects of partial loss-of-function DOT1L to more broadly model individuals with *DOT1L* variants and to better understand the role of DOT1L in the brain.

Loss of dot1l in zebrafish leads to exaggerated motor behavior in response to sensory stimuli

Given the *de novo* nature of the majority of individual variants, we aimed to characterize early behavioral disruptions utilizing a system that allows for allele disruptions in the offspring of zebrafish. Additionally, zebrafish provide a vertebrate model with high genetic similarity to humans (*37*), including DOT1L (Catalytic domain: 85% identity, Whole gene: 49% identity, 57% similarity, 21% gaps) (*38*) (fig. S1E). Zebrafish also develop robust stereotypical motor movements in response to sensory stimuli (visual or acoustic) detectable within the first six days of development. Prior work demonstrated that these behaviors are sensitive to mutations in genes associated with NDDs (*39–41*), suggesting their relevance to NDD pathophysiology.

To assess behavioral roles for DOT1L in early development, we first specifically disrupted the zebrafish *dot1l* gene using a CRISPR-Cas9-approach that generates biallelic null alleles (*42*) in over 90% of animals. Briefly, we injected three guide RNAs that target non-overlapping sites along the *dot1l* gene into fertilized embryos together with Cas9 protein, generating *dot1l* 'crispants'. Control embryos were injected in parallel with three non-targeting gRNAs and Cas9. We first confirmed that each gRNAs targeted *dot1l* by sequencing (fig. S1F). *Dot1l* crispants, were viable to 6 days post-fertilization (dpf) and did not display obvious gross morphological defects (control injected n=164, *dot1l* crispant n=144, 4 independent experiments). Behavior of *dot1l* crispants was then assessed at 6dpf using a previously described pipeline that allows assessment of multiple sensorimotor behaviors including the visual motor

response, responsiveness to flashes of light or darkness, and the acoustic startle response (*39*) (Fig. 1G, fig. S1G). Compared to controls, *dot1l* crispants displayed exaggerated motor responses to multiple sensory inputs. Specifically, *dot1l* crispants displayed increased movement in response to changes in illumination, as illustrated by increased distance travelled in the visual motor response assay (*43*) (fig. S1H) and increased movement in response to flashes of darkness (*44*) (Fig. 1H). In addition, *dot1l* crispants are hypersensitive to acoustic stimuli, displaying startle responses (*45*) following stimuli that do not elicit similar responses in controls (Fig. 1I). Further, *dot1l* crispants also show an increase in acoustic startle prepulse inhibition (*46*) compared to controls (Fig. 1J). Together, these results demonstrate that zebrafish *dot1l* controls responses to visual stimuli and is required for establishing the acoustic startle threshold and acoustic startle sensorimotor gating in zebrafish. Further, they demonstrate that DOT1L loss affects early developmental behaviors.

DOT1L regulates glutamatergic synaptic gene expression

Based on the broad neurodevelopmental phenotypes observed in individuals with *DOT1L* variants, and the robust effects of DOT1L loss in zebrafish behavior, we next tested the role of DOT1L in mouse models based on highly conserved DOT1L (Catalytic domain: 96% identity, Whole gene: 84% identity, 88% similarity, 1% gaps) (*38*) (fig. S1E). We began by defining the regulation of *Dot1l* and its target histone modification H3K79me in developing mouse neurons. We leveraged mouse primary cultured neurons derived from E16.5 cortices to generate a pure neuronal population and found that both *Dot1l* and H3K79me increase throughout neuronal maturation (fig. S2A, S2B). To model partial loss of *Dot1l*, we infected primary cortical neurons with short hairpin RNAs (shRNA) targeting *Dot1l* or a non-targeting control (n.t.). We confirmed *Dot1l* loss and H3K79me depletion upon infection of *Dot1l* shRNAs (fig. S2C, S2D) demonstrating that H3K79me is dynamically regulated and requires continued DOT1L function for H3K79me deposition in developing neurons.

Given the association between H3K79me and active gene expression, we next sought to determine the effect of partial loss of *Dot1l* on transcription in primary neurons. Following *Dot1l* depletion, we performed RNA-sequencing and found widespread changes in gene expression with 677 genes significantly up-regulated and 1050 genes significantly down-regulated (Fig. 2A). Gene ontology (GO) analysis indicated an enrichment of genes involved in synaptic transmission (such as GO:0099177 and GO:0050804) in down-regulated differentially expressed genes (DEGs) and no significant enrichment of GO terms in up-regulated DEGs (Fig. 2B). Given the dysregulation of synaptic-related genes, we further interrogated differentially expressed genes using SynGO (*47*) which demonstrated enrichment for pre- and post-synaptic compartment proteins, and synaptic cleft proteins suggesting widespread disruption of expression of synaptic genes (Fig. 2C). We next asked whether the observed changes in genes related to synaptic transmission were global or specific to a class of

chemical synaptic transmission. To this end, we used gene set enrichment analysis to test for enrichment of genes related to glutamatergic, GABAergic, dopaminergic, and cholinergic synaptic transmission. Interestingly, glutamatergic transmission is enriched in down-regulated genes, while there is no significant enrichment of other classes of synaptic transmission (Fig. 2D, fig. S2E-G). In fact, down-regulated DEGs had significant overlap with the glutamatergic synaptic transmission gene set including genes such as *Gria2* and *Grin1*, two glutamate receptor subunits that are critical for appropriate levels of glutamatergic transmission throughout the brain (Fig. 2E-G). Together these findings demonstrate that H3K79me is dynamically regulated in neurons by DOT1L and that partial *Dot1l* loss disrupts expression of critical synaptic genes.

Dot1l loss impacts neuronal arborization and GluA2 levels

Given the disruption to expression of critical synaptic genes that we detected, we next tested the effect of partial *Dot1l* loss on neuronal morphology, synapses, and synaptic proteins. Neuronal branching and spine formation are critical components of neuronal maturation that allow for neuronal communication and downstream behaviors. To assess how *Dot1l* loss impacts neuronal architecture, we performed Sholl analysis on primary cortical neurons transfected with a *Dot1l* shRNA or control shRNA (Fig. 3A). *Dot1l* depleted neurons had a reduced number of intersections in comparison to controls indicative of reduced neuronal arborization (Fig. 3B,3C). In addition to neuronal branching, spine formation is critical for synapse development and essential for neuronal communication and memory consolidation. Interestingly, the spine density of *Dot1l* depleted neurons was increased (Fig. 3D). This could suggest aberrant spine development as seen in other developmental disorders (*48*) or be a compensatory mechanism to offset the loss of neuronal branching or decreased glutamatergic synapse function. Given the downregulation of genes involved in glutamatergic transmission, we also assessed whether the glutamatergic receptor subunit GluA2 protein is regulated by DOT1L in primary neurons. Using immunocytochemistry, we found that GluA2 is depleted upon *Dot1l* loss demonstrating that transcriptional disruptions functionally affect protein levels of critical synaptic genes (Fig. 3E, 3F). Conversely, overexpression of *Dot1l* in neurons did not result in significant changes to neuronal arborization or GluA2, suggesting gain-of-function variants may impact neurons through mechanisms that are distinct from loss-of-function variants (fig. S3A-F).

DOT1L regulates cortical gene expression in a sex-specific manner

Given the transcriptional disruptions in our *in vitro* primary cultured neurons and behavioral disruptions in our zebrafish model, we next sought to analyze the transcriptional effects of monoallelic loss of *Dot1l* in mice to model the monoallelic loss in individuals with *DOT1L* variants. We first examined *Dot1l* and H3K79me expression in mice during cortical development from E14 through postnatal day

28. We found that both *Dot1l* and H3K79me increase during this period, suggesting that DOT1L may play a role during this critical period of brain development (fig. S4A-D). Prior work thoroughly defined the effects of *Dot1l* loss on the transcriptome of stem cell populations and the effect of biallelic *Dot1l* loss on neurons early in development(*24–31*). However, to the best of our knowledge the effect of monoallelic loss has only been tested in the midbrain (*32*) with a focus on aging-related phenotypes and has not been tested beyond early development in brain regions relevant to the emerging disorder described here.

Given the notable increase and stabilization of *Dot1l* expression and H3K79me from P0-P28 and the lack of characterization of *Dot1l* after brain development, we assessed the transcriptional impact of monoallelic loss of *Dot1l* in 8-week-old cortical mouse tissue. We used a floxed *Dot1l* mouse model containing loxP sites flanking exon 2 of *Dot1l* that was crossed to a ubiquitously expressing Cre line under the human cytomegalovirus (CMV) minimal promoter that expresses during early embryogenesis (*49*) to generate *Dot1l*$^{floxed/+}$;CMV-Cre$^{+/-}$ (referred to as *Dot1l* HET) with littermate controls (*Dot1l*$^{+/+}$;CMV-Cre$^{+/-}$). Notably, parental lines that generated experimental cohorts included *Dot1l*$^{floxed/+}$ crossed to CMV-Cre$^{+/+}$. This ensures that parents of experimental mice have wildtype DOT1L expression to avoid effects of parental partial loss of *Dot1l* which may affect the health of offspring and to better mimic the affected individuals in which most variants are *de novo*.

We first confirmed partial loss of *Dot1l* in *Dot1l* HET cortical tissue as expected (fig. S4E). Given that prior work established effects of complete *Dot1l* loss on neurogenesis and cortical layer development and our data demonstrating that partial loss of *Dot1l* robustly affects gene expression within neurons, we harnessed single nucleus RNA-sequencing to capture both changes in cell type identity and changes in gene expression in cortical tissue (Fig. 4A). Using 3 male and 3 female animals for both control and *Dot1l* HET, we identified 25 clusters that include 10 excitatory neurons clusters (Slc17a7+), 7 inhibitory neuron clusters (Gad2+), 2 microglia clusters (Ctss+ and Ptprc+), an astrocyte cluster (Gja1 + and Gnb4+), and an oligodendrocyte cluster (Mog+, Enpp6+, and Opalin +) (fig. S4F). Interestingly, we did not find altered proportions of neuronal cell types and only a modest increase in microglia cell types in *Dot1l* HET mice in comparison to control suggesting that partial *Dot1l* loss is not sufficient to alter cortical neuron identity as occurs following complete *Dot1l* deletion (*24*) (Fig. 4B). However, we found widespread disruption of gene expression across most excitatory and inhibitory neuron clusters and modest changes in non-neuronal clusters (Fig. 4C and fig. S4G, S4H). We detected the greatest effects in excitatory neuron clusters (Fig. 4C) and thus examined the effect of *Dot1l* loss on excitatory clusters as a whole. We found 880 significantly down-regulated genes and 310 genes significantly up-regulated upon *Dot1l* loss *in vivo* fitting with culture data demonstrating more genes are decreased in gene expression following partial *Dot1l* loss (Fig. 4D). Gene ontology analysis indicated an enrichment of genes involved in synaptic function (GO:0099072, GO:1903421, GO:0050804) in

down-regulated DEGs and no significant enrichment of GO terms in up-regulated DEGs (Fig. 4E). Examining the excitatory cluster with the most DEGs (Ex_L2/3_1), showed similar effects with 602 significantly down-regulated genes and 221 genes significantly up-regulated (Fig. 4F). Gene ontology analysis of downregulated genes again indicated disruption of genes involved in synaptic function (GO:0050803, GO:0050804, GO:0099003, GO:0099536), a feature that was echoed in inhibitory neuron clusters as well (Fig. 4G and fig. S4I, S4J).

Given that the majority of individuals within the cohort were female, we interrogated whether there may be sex-specific transcriptional alterations upon monoallelic *Dot1l* loss. To parse sex-specific effects, we separated male and female cells and found the sexes were equally represented in each cluster (Fig. 4H). We again detected widespread gene expression changes in both male and female excitatory neuron clusters (Fig. 4I). However, we detected slightly more down-regulated genes in female neurons with 312 uniquely down-regulated in female and 222 genes uniquely down-regulated in males, with a similar effect in up-regulated genes (Fig. 4J, 4K and fig. S4K). Interestingly, we detected *Dot1l*-sensitive genes for which female neurons showed decreased expression compared to males even in control tissue suggesting baseline transcriptional differences in female neurons may contribute to different responses to *Dot1l* monoallelic loss. These findings demonstrate that there are both shared and sex-specific transcriptional programs down-regulated upon monoallelic *Dot1l* loss and that female neurons may be more sensitive to loss of *Dot1l* loss due to underlying differences in transcriptional states. Finally, we compared *in vivo* and *in vitro* RNA-sequencing gene sets. Genes unique to each system were identified as expected due to the difference in methods (whole cell analysis in the *in vitro* system versus nuclei-specific analysis in the *in vivo* system) and due to the differences in length and method of DOT1L depletion (5-day knockdown verses long-term genetic depletion). However, despite these differences, we identified 69 down-regulated genes shared between our *in vitro Dot1l* shRNA dataset and the *in vivo Dot1l* cKO dataset suggesting shared transcriptional disruptions between even highly distinct models of partial *Dot1l* loss (fig. S4L).

Monoallelic Dot1l loss alters early vocalization development and sociability

Previous studies using homozygous *Dot1l* mouse knockouts indicate that it is essential for hematopoiesis (*50*), cardio myocyte function (*51*), and neural progenitor proliferation and differentiation in the cortex, cerebellum, and spinal cord (*24–27*). However, defining the developmental and behavioral responses to monoallelic loss of *Dot1l* is critical to understand the implications of monoallelic variants in affected individuals. To the best of our knowledge such work has not been performed outside of one publication noting that heterozygous germline knockout mice were normal and fertile (*16*).

As previously reported (*16*), *Dot1l* HETs are viable and generated in approximately expected Mendelian ratios (fig. S5A, S5B). To assess impacts of monoallelic *Dot1l* loss on early development,

we tracked developmental milestones including physical landmarks, and sensorimotor development in *Dot1l* HET and controls. Male *Dot1l* HET had no differences in weight but had delayed development of the visual placing response, a measurement of sensory development (Fig. 5A, fig. S5C). Female *Dot1l* HET pups weighed more than controls but had no delayed development (Fig. 5A, fig. S5C). Given the language delay seen in 8/11 individuals with *DOT1L* variants, and previous studies demonstrating ultrasonic vocalizations (USVs) changes in various NDD mouse models (*52*), we measured USVs in pups during 5 minutes of maternal separation at P6. Male *Dot1l* HET pups had significantly higher decibel calls, and a greater percentage of chevron type calls (Fig. 5B-E). Female *Dot1l* HET mice had decreased total USV calls with no differences in call characteristics (Fig. 5B-E). Finally, we found both male and female *Dot1l* HET mice were slower to complete a negative geotaxis assay where mice are placed face down on an angled platform to assess early motor and vestibular development (Fig. 5F).

Next, we performed a battery of behavioral assays to assess motor and cognitive function in juvenile *Dot1l* HET and controls. We found no impairments in gross motor function in *Dot1l* HET mice in an open field assay (Fig. 5G, fig. S5D). Further, there was no evidence of anxiety-related behaviors measured using percent of time spent in open arms of the elevated zero maze and percent of time spent in the center of the open field assay (Fig. 5H, fig. S5E, S5F). We also detected no changes in working memory in *Dot1l* HET in comparison to controls measured using percent of spontaneous alternations completed in a Y maze (Fig. 5I, fig. S5G). To assess sociability, we performed the social choice assay where mice explore the 3-chamber arena with one chamber holding a rock, one chamber holding a mouse, and a neutral center chamber. Female *Dot1l* HET had a reduced time spent with the mouse measured using a discrimination index (time spent with mouse – time spent with rock / total interaction time) indicating sex-specific social behavior changes (Fig. 5J). Together, these data demonstrate that *Dot1l* HET mice have sex-dependent deficits in sensorimotor function, vocalization development, and sociability.

Neuronal Dot1l loss alters early vocalization development and sociability

Given the behavioral alterations seen in *Dot1l* HET, we next used a forebrain neuron specific *Dot1l* conditional knockout (cKO) mouse to assess whether behavioral alterations in the *Dot1l* HET mice can be attributed to monoallelic loss of *Dot1l* specifically in forebrain neurons. We confirmed partial loss of *Dot1l* and H3K79me in *Dot1l* cKO cortical tissue as expected (fig. S6A, S6B). We again recorded USVs in P6 pups and found both male and female *Dot1l* cKO mice had altered frequency of calls in comparison to controls and male *Dot1l* cKO mice had reduced down calls suggesting that DOT1L in forebrain neurons contributes to early vocalization deficits observed in ubiquitous *Dot1l* HET mice (Fig. 6A-D). We did not find any developmental delays or weight alterations suggesting effects observed in *Dot1l* HETs are independent of DOT1L function in neurons (fig. S6C-E). Similarly to *Dot1l* HET mice,

Dot1l cKO had no motor activity or anxiety-related impairments (Fig. 6E, fig. S6F-I). In contrast, female *Dot1l* cKO had increased spontaneous alternations with no change observed in males (Fig. 6F, fig. S6J). Notably, we again found a sex-specific sociability deficit in the 3-chamber social test in female *Dot1l* cKO similar to *Dot1l* HET (Fig. 6G), indicating that DOT1L loss in neurons contributes to this effect. We also found long-term memory deficits in male *Dot1l* cKO mice in contextual fear conditioning (Fig. 6H). While no significant change was observed in females in contextual fear conditioning, freezing rates were low in control female mice, so it is possible that we lacked the dynamic range to detect differences. These data indicate that forebrain neuron-specific monoallelic depletion of *Dot1l* recapitulated sociability deficits in female *Dot1l* HETs and caused changes in vocalization behavior suggesting that expression of *Dot1l* in neurons contributes to specific behavioral alterations.

DISCUSSION

Here, we identified 11 individuals with *DOT1L* variants and NDDs, including developmental delays, ASD, and intellectual disability. We confirmed two missense variants disrupt methyltransferase activity of DOT1L which, along with two other variants that result in early stop codons, suggest that monoallelic loss-of-function of DOT1L can lead to observed phenotypes. We found dynamic regulation of H3K79me in post-mitotic cortical neurons and widespread transcriptional disruptions upon partial *Dot1l* loss in excitatory neurons both *in vitro* and *in vivo*. Further, loss of *Dot1l* alters neuron arborization, spine density, and expression of synaptic genes. In addition, we found that *dot1l* depletion in zebrafish increases activity in response to multiple sensory inputs. Finally, we show that both ubiquitous and neuron-specific monoallelic loss of *Dot1l* cause sex-specific vocalization disruptions and sociability deficits, with additional motor development deficits observed in full-body heterozygous mice. Together, this work demonstrates that partial *Dot1l* loss can lead to an emerging neurodevelopmental disorder and disrupt transcription, neuron morphology, and behavior.

We provide the first functional testing of two loss-of-function variants that indicate monoallelic loss of *DOT1L* causes neurological dysfunction. Recent work proposed that gain-of-function variants in *DOT1L* contribute to the neurodevelopmental disorders based on modeling in flies and human HEK293T cells (*15*). However, the fly ortholog of *DOT1L*, *grappa*, is highly divergent from human DOT1L (Catalytic domain: 65% identity, Whole gene: 24% identity, 35% similarity, 28% gaps) (*38*) and expression of wildtype human DOT1L did not rescue developmental defects caused by *grappa* loss. Thus, we sought to characterize additional *DOT1L* variants and model them in homologous systems. We confirmed one of the previously described variants does indeed function as a gain-of-function (p.E123K) supporting prior findings (*15*). However, we also found that other missense variants disrupt methyltransferase activity through multiple approaches. This, combined with modeling of partial *Dot1l* loss in multiple systems suggests that loss-of-function variants have profound functional consequences

and are also likely causative in the identified neurodevelopmental disorder. Together, these findings place DOT1L in a growing group of epigenetic regulators for which either increases or decreases in function or expression can lead to neurodevelopment disorders (*53*, *54*).

These findings point toward likely molecular changes and cell types responsible for the resulting phenotypes. Given the ample work illustrating the importance of DOT1L in early corticogenesis (*24*, *27*), it is noteworthy that we did not detect major changes in cortical neuron identify following loss of just a single copy of *Dot1l*. Rather, we found robust changes in transcriptional programs, particularly in excitatory neurons, resulting in downregulation of genes related to synaptic function. In addition, we performed in-depth behavioral characterization of mice with monoallelic *Dot1l* loss in all tissue and in forebrain neurons. The behavioral overlap of altered early vocalization and sociability deficits suggests that DOT1L loss in neurons is at least partly responsible for behavioral deficits. Further, given that we detected sex-specific effects on both gene expression and behavior, our findings indicate that partial DOT1L loss has divergent effects based on sex. This is particularly intriguing given that our cohort includes more females than males, although whether this trend will be sustained as additional individuals are identified remains to be determined. Together, this work builds upon the previous modeling of *Dot1l* loss in the brain by indicating that partial loss of *Dot1l* is sufficient to cause changes in transcription, neuron maturation, and behavior and identifies transcriptional pathways and cell types that likely contribute to these deficits.

Several notable questions remain that will be critical to understanding the role of DOT1L in the brain and in neurodevelopmental disorders. Dynamic regulation of H3K79me is evident in primary cortical neuronal upon partial *Dot1l* loss fitting with prior work in the midbrain (*32*). However, it is unclear whether histone variant replacement or demethylase activity is responsible for H3K79me removal in the brain. Future work establishing where the deposition of this mark occurs in the neuronal genome and how this is affected by partial DOT1L loss will also be important for understanding its role in neuronal transcription. There is also conflicting evidence on whether methylation of H3K79 is required for DOT1L to fulfill its role in neuronal differentiation (*29–31*). Given that the majority of variants lie in the catalytic domain, our work suggests that H3K79me is important in neuron function. However, our findings also allow for a critical function for H3K79me to emerge *after* neuronal differentiation. Notably, several of the phenotypes that we detected in mouse models were evident in neuron-specific monoallelic loss of *Dot1l* suggesting this cell type is particularly sensitive to DOT1L dosage. Although we characterized the impact of partial *Dot1l* loss in the brain, whether transcriptional states of other cell types are also impacted in both mouse and human systems will be critical to understanding the role of DOT1L in contributing to developmental disorders. Lastly, predicted loss-of-function alleles for DOT1L are found within gnomAD that are not associated with notable phenotypes suggesting incomplete penetrance or the potential for attenuation of the disorder described through other unknown factors.

In summary, this work examines the impact of partial loss of *Dot1l* spanning from the transcriptional level to the behavioral level. This research provides insights into the effect of variants on DOT1L and the neuronal changes that may contribute to phenotypes observed in DOT1L loss-of-function variant individuals. Further, our findings expand on our understanding of DOT1L by demonstrating that disruption of a single copy of *Dot1l* is sufficient to disrupt neuronal function and contributes to an emerging neurodevelopmental disorder.

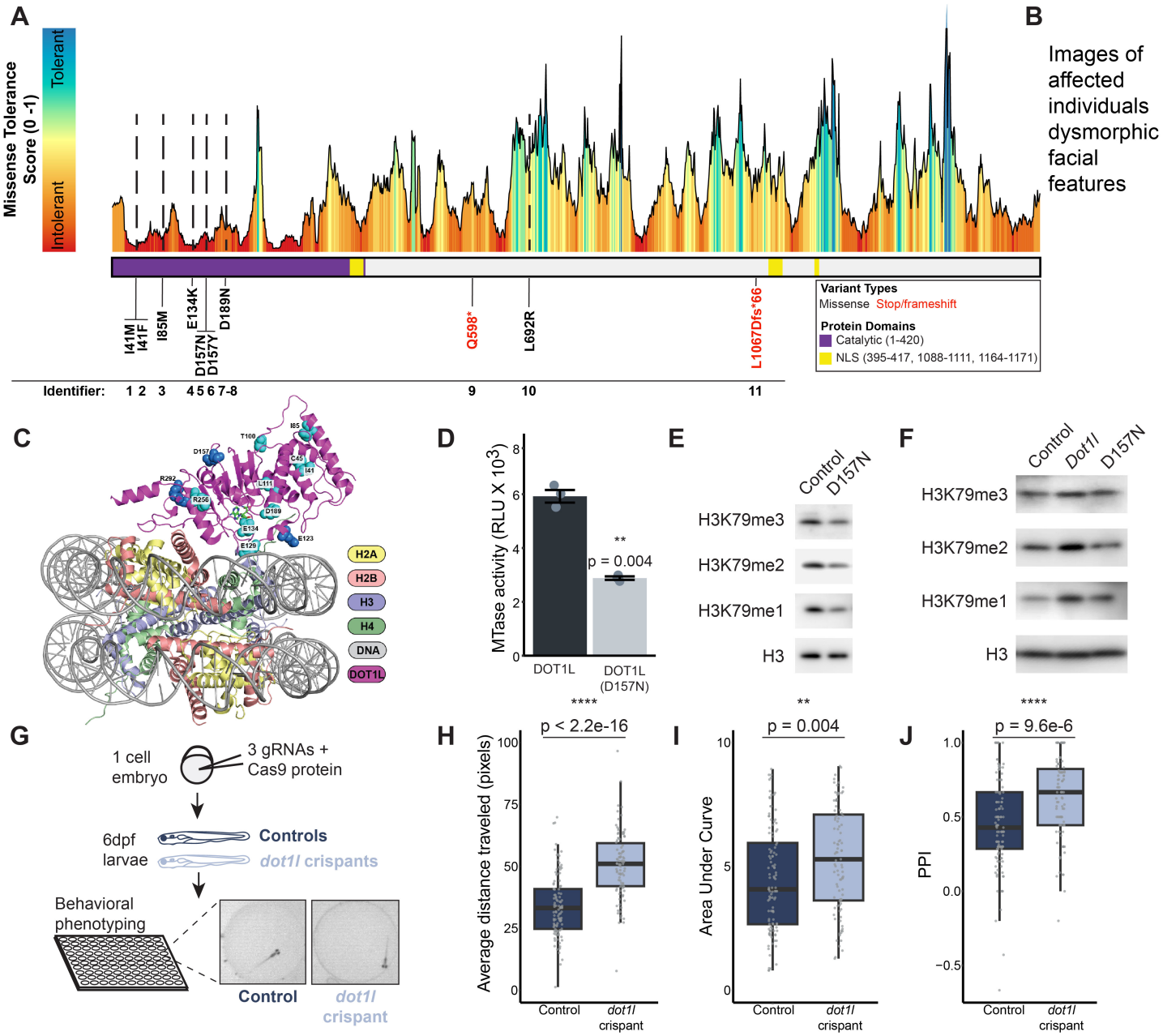


Figure 1. Variants in DOT1L are associated with a human neurodevelopmental disorder and dot1l loss in zebrafish alters motor responses to sensory stimuli. (A) The landscape of missense tolerance of DOT1L from Metadome. Schematic of DOT1L protein domains (purple = catalytic and yellow = nuclear localization signal) and locations of *DOT1L* variants (black = missense variant, red = stop/frameshift variant). (B) Images of dysmorphic facial features in individuals with *DOT1L* variants. (C) Structure of the DOT1L nucleosome complex (PDBID: 6NJ9) highlighting the position of variants (in cyan and blue spheres). Blue spheres represent variants tested by enzymatic assays in this study. DOT1L is in magenta, DNA in grey, histone H2A in pale yellow, histone H2B in red salmon, histone H3 in marine blue, and histone H4 in green. For clarity, ubiquitin is not shown. SAM is shown in green sticks. (D) Methyltransferase activity of human DOT1L or variant DOT1L (D157N) on unmodified nucleosomes. Graph shows mean ± SE (n = 3/condition, unpaired two-tailed t-test). RLU = relative light units. (E) H3K79 methyl levels in patient-derived fibroblasts from individual 5 (D157N variant) beside an age- and sex-matched control. (F) H3K79 methyl levels in Neuro-2A cells with variant *Dot1l* (D157N) overexpressed, wildtype *Dot1l*, or empty vector control. (G) Schematic of Cas9 targeting of *dot1l* and subsequent behavioral phenotyping paradigm. (H) Average distance traveled during dark flashes in *dot1l* crispant and control zebrafish (control: n = 113, *dot1l* crispant: n = 98; 3 independent experiments, unpaired two-tailed t-test). (I) Area under the curve of response to increasing stimulus intensity in *dot1l* crispant and control zebrafish (AU = arbitrary units, control: n = 113 control, *dot1l* crispant: n = 95; 3 independent experiments, Kruskal-Wallis Test). (J) Prepulse inhibition (PPI) to a medium intensity prepulse acoustic stimuli in *dot1l* crispant and control zebrafish (control: n = 113 control, *dot1l* crispant: n = 95; 3 independent experiments, Kruskal-Wallis Test). All box plot bounds indicate the 25th and 75th percentiles, the black line shows the median, and whiskers extend to the minimum and maximum value that are no further than 1.5 * interquartile range. **p<0.01, ****p<0.0001.

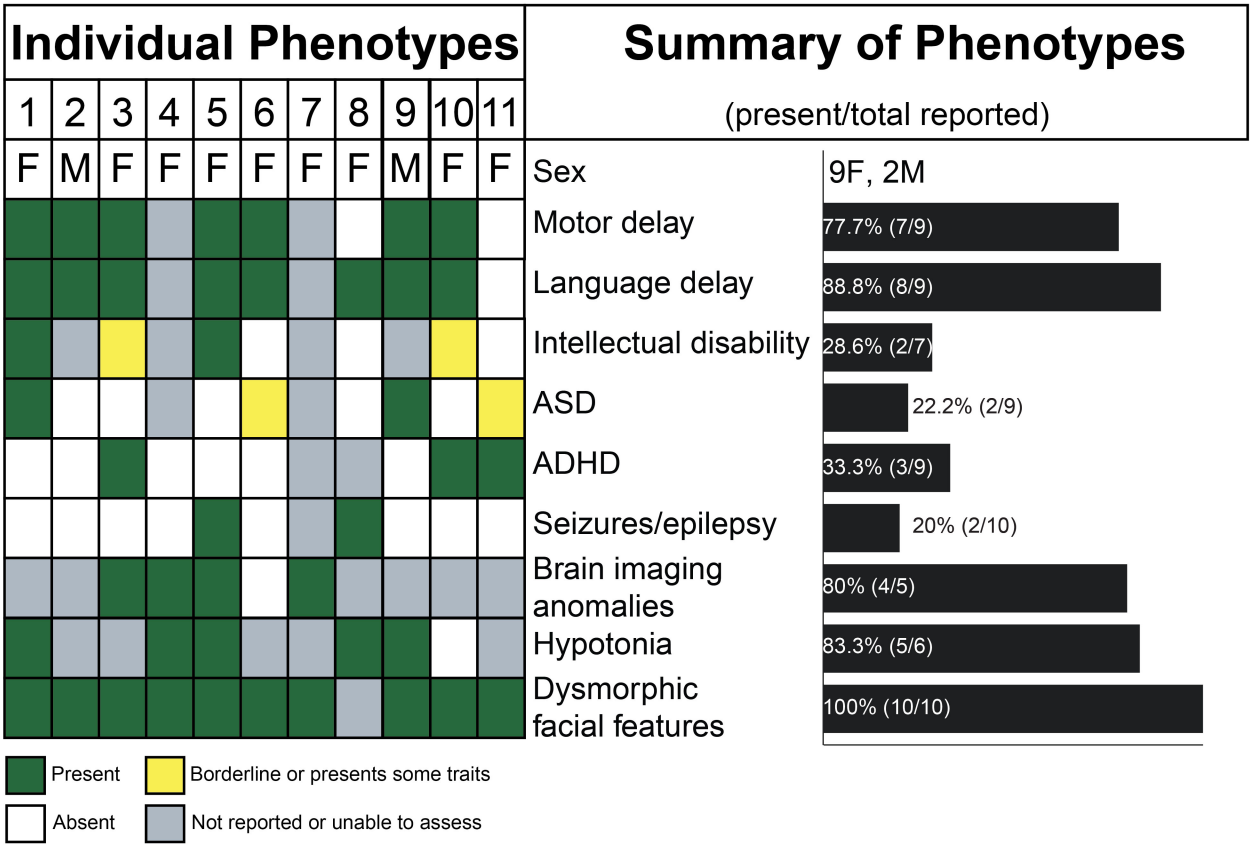


Table 1. Summary of clinical findings in affected individuals. *DOT1L* variant individuals' phenotypic spectrum and prevalences. ASD = autism spectrum disorder. ADHD = attention deficit hyperactivity disorder.

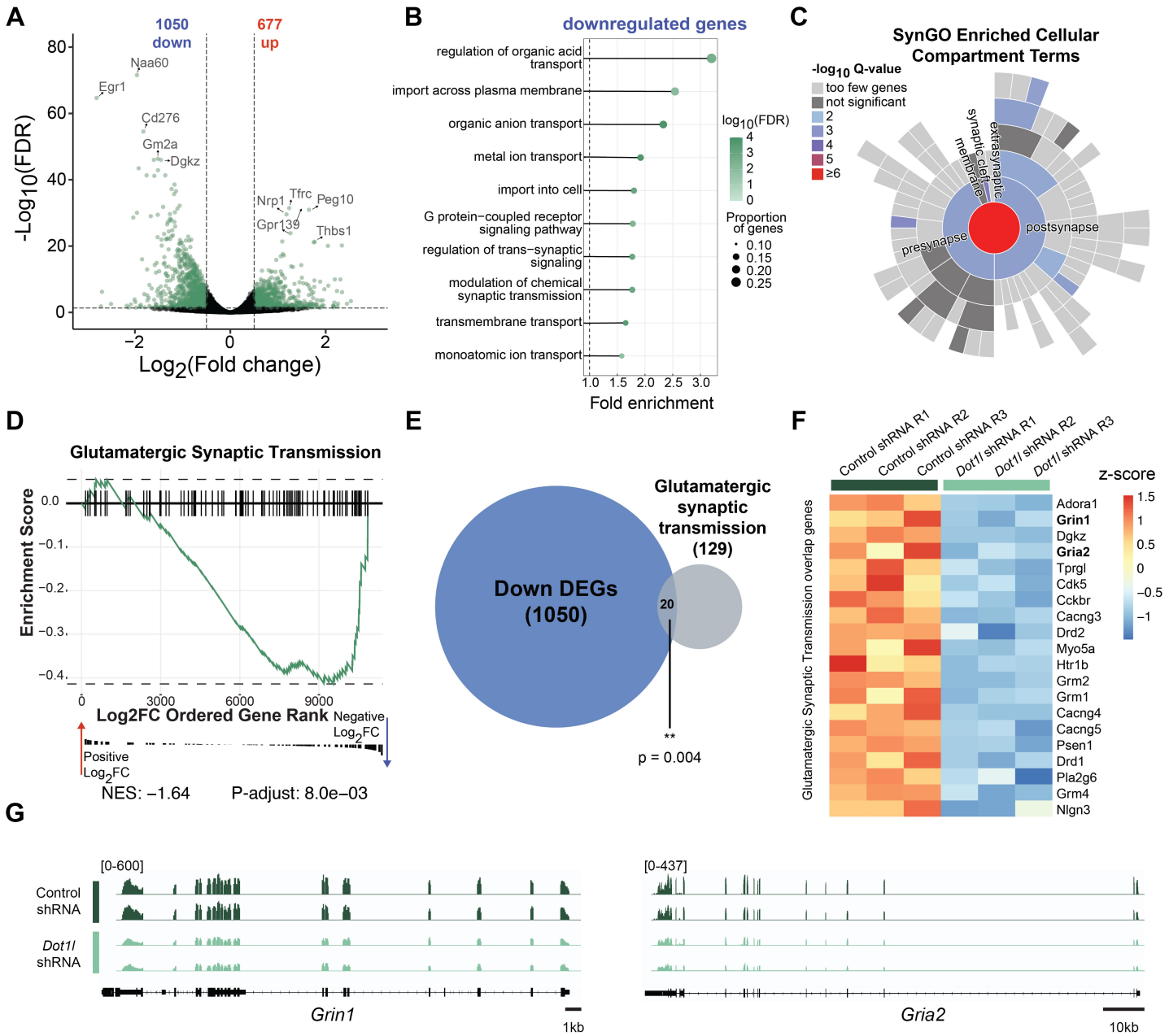


Figure 2. *Dot1l* promotes glutamatergic synaptic gene expression in neurons. (A) Volcano plot showing differentially expressed genes (DEGs) from primary cortical neurons infected with *Dot1l* shRNA or non-targeting control shRNA (n = 3 biological replicates/condtion). Green dots represent genes with log2(FoldChange) >= 0.5 and FDR <= 0.05. (B) Biological process gene ontology analysis down-regulated genes. (C) SynGO analysis of cellular compartment in down-regulated DEGs. (D) Gene set enrichment analysis of genes involved in glutamatergic synaptic transmission. NES indicates normalized enrichment score. (E) Overlap of down-regulated DEGs and glutamatergic synaptic transmission genes (hypergeometric test). (F) Heatmap of the 20 glutamatergic synaptic transmission genes that are down-regulated DEGs (R1-R3 denotes biological replicates). (G) RNA-seq genome browser tracks RNA-seq for *Grin1* and *Gria2*. **p<0.01.

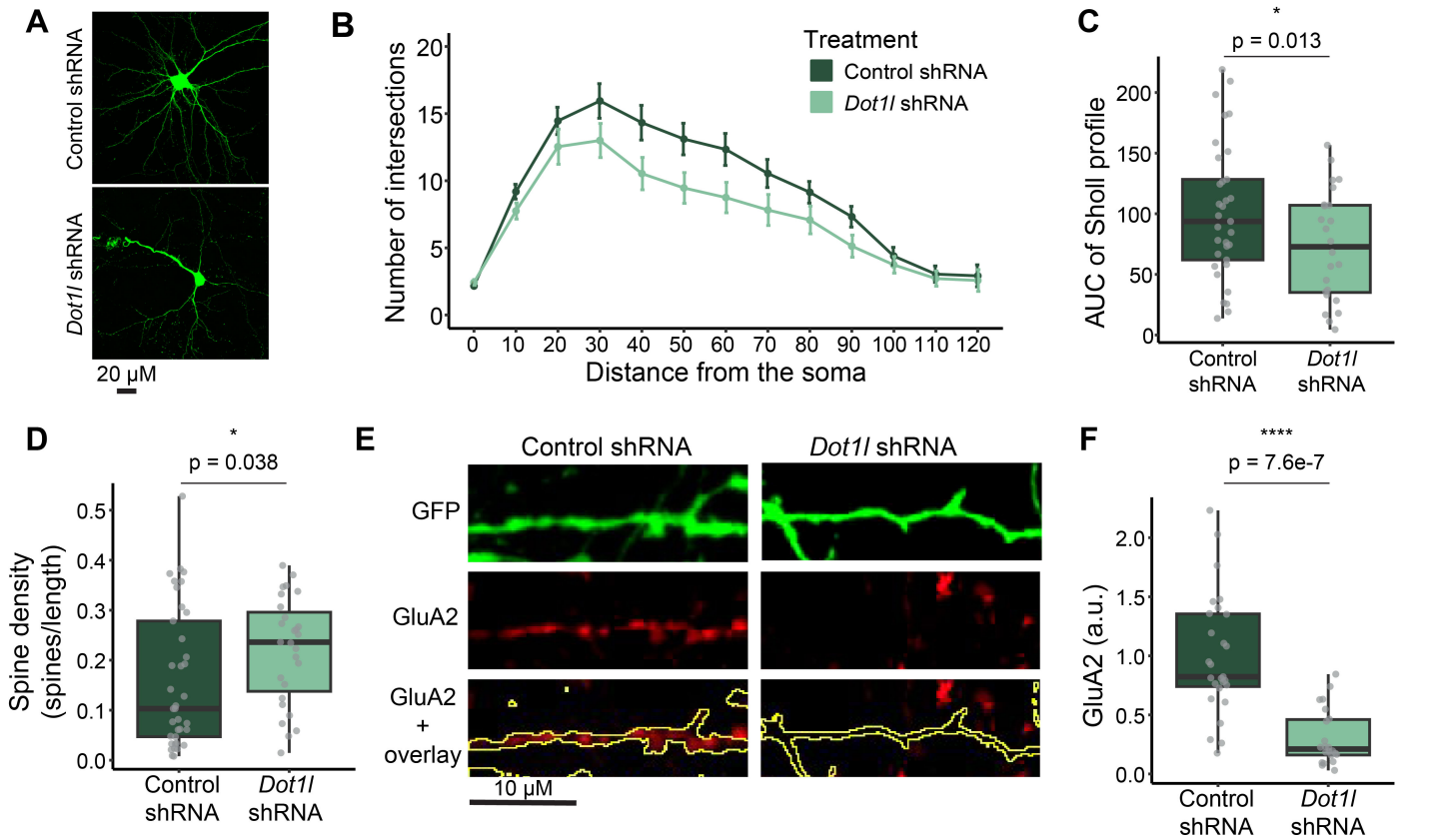


Figure 3. *Dot1l* loss impacts neuronal arborization and GluA2 levels. (A) Representative images of primary cortical neurons transfected with *Dot1l* shRNA or non-targeting control. Scale bar = 20 µm. (B) Number of branch intersections per radius of shRNA transfected primary cortical neurons (control: n = 33 neurons from 5 biological replicates, *Dot1l* shRNA: n = 24 neurons from 5 biological replicates). (C) Area under the curve (AUC) quantification of (B) (control: n = 33 neurons from 5 biological replicates, *Dot1l* shRNA: n = 24 neurons from 5 biological replicates, mixed effect model). (D) Spine density of shRNA-transfected primary cortical neurons (control: n = 33 neurons from 5 biological replicates, *Dot1l* shRNA: n = 23 neurons from 5 biological replicates, Kruskal-Wallis Test). (E) Representative images of GluA2 staining in shRNA-transfected primary cortical neurons. (F) Quantification of GluA2 staining in shRNA transfected primary cortical neurons. (control: n = 27 neurons from 3 biological replicates, *Dot1l* shRNA: n = 22 neurons from 3 biological replicates, Kruskal-Wallis Test). All box plot bounds indicate the 25th and 75th percentiles, the black line shows the median, and whiskers extend to the minimum and maximum value.*p<0.05, ****p<0.0001.

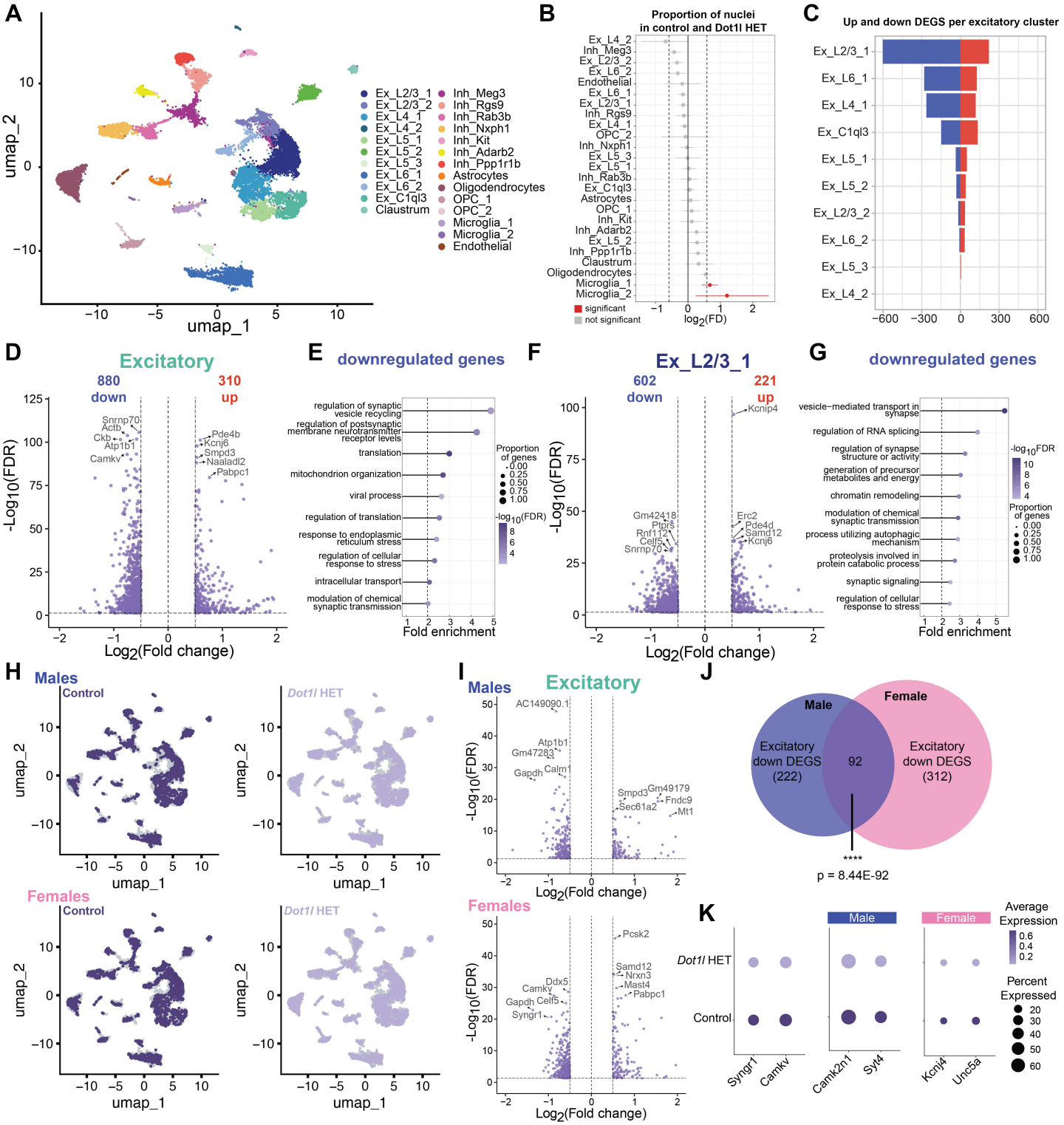


Figure 4. DOT1L regulates cortical gene expression in a sex-specific manner. (A) UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) of single nucleus RNA sequencing from 8-week-old *Dot1l* HET and control mouse cortex (n = 6 cortices per group[3 males, 3 females). (B) Proportion of nuclei in *Dot1l* HET and control in each cluster. Red indicates fold difference > 1.8 and FDR < 0.05. (C) Number of up- and down-regulated differentially expressed genes (DEGs) in excitatory clusters. (D) Volcano plot showing DEGs from *Dot1l* HET and control combined excitatory clusters. (E) Biological process gene ontology analysis down-regulated DEGs from combined excitatory clusters. (F) Volcano plot showing DEGs from *Dot1l* HET and control from Ex_L2/3_1. (G) Biological process gene ontology analysis down-regulated DEGs from Ex_L2/3_1. (H) UMAP clustering with dots representing nuclei from *Dot1l* HET or control separated by sex. (I) Volcano plots showing DEGs from male or female *Dot1l* Het and control from combined excitatory clusters. (J) Overlap of male and female down DEGs from combined excitatory clusters (hypergeometric test). (K) Dot plots showing overall down DEGs (Syngr1, Camkv) and sex-specific down DEGs (Camk2n1, Syt4, Kcnj4, Unc5a) from combined excitatory clusters. ****p < 0.00001.

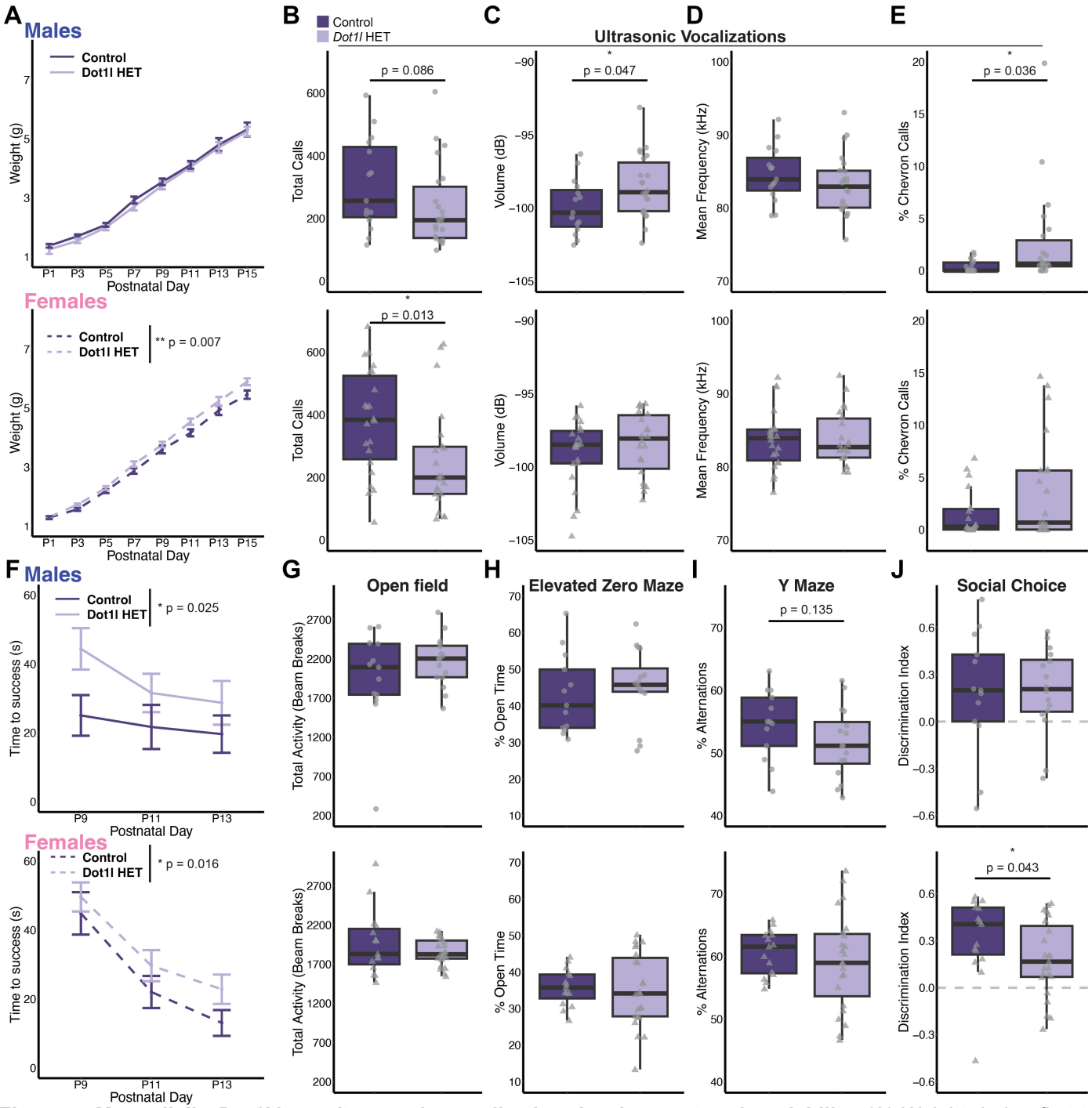


Figure 5. Monoallelic *Dot1l* loss alters early vocalization development and sociability. (A) Weight during first two weeks post-birth in male and female mice (male[control: n = 13, *Dot1l* HET: n = 16] ; female[control: n = 15, *Dot1l* HET: n = 23], repeated measures ANOVA). (B) Total number of ultrasonic vocalizations (USVs) in P6 male and female pups (male[control: n = 15, *Dot1l* HET: n = 22]; female[control: n = 22, *Dot1l* HET: n = 22], Kruskal-Wallis Test). (C) Average volume in decibels (dB) of USVs in male and female pups (unpaired two-tailed t-test). (D) Average frequency in kilohertz (kHz) of calls in male and female pups. (E) Percent chevron calls out of total calls in male and female pups (Kruskal-Wallis Test). (F) Time to right self in negative geotaxis assay during development in male and female pups (male[conrol: n = 13, *Dot1l* HET: n = 16]; female[control: n = 15, *Dot1l* HET: n = 23], repeated measures ANOVA). (G) Total activity measured from beam breaks during 10-minute open field assay in 4 week old male and female mice (male[control: n = 13, *Dot1l* HET: n = 16]; female[control: n = 15, *Dot1l* HET: n = 23], Kruskal-Wallis Test). (H) Percent of time spent in open arms of the elevated zero maze in male and female mice (unpaired two-tailed t-test). (I) Percent spontaneous alternations out of the total number of triads possible in male and female mice. (unpaired two-tailed t-test). (J) Discrimination index between interaction time with the mouse or rock cylinder during the three-chamber social choice assay in male and female mice (Kruskal-Wallis Test). All box plot bounds indicate the 25th and 75th percentiles, the black line shows the median, and whiskers extend to the minimum and maximum value that are no further than 1.5 * interquartile range. *p<0.05, **p<0.01.

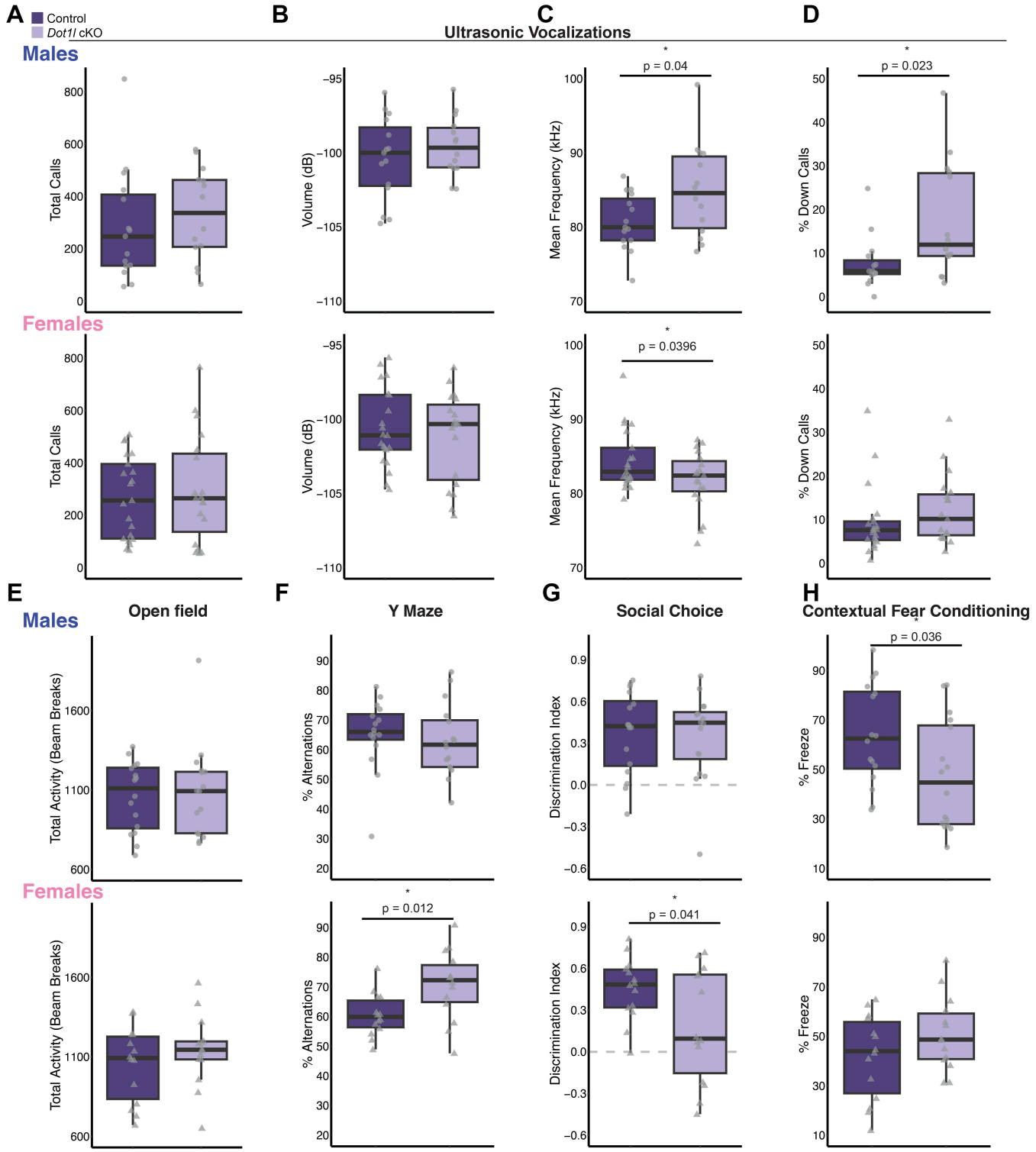


Figure 6. Monoallelic loss of *Dot1l* in forebrain neurons alters language development and sociability. (A) Total number of ultrasonic vocalizations (USVs) in P6 male and female pups (male[control: n = 15, *Dot1l* cKO: n = 14]; female[control: n = 21, *Dot1l* cKO: n = 19], Kruskal-Wallis Test). (B) Average volume in decibels (dB) of USVs in male and female pups (unpaired two-tailed t-test). (C) Average frequency in kilohertz (kHz) of calls in male and female pups (unpaired two-tailed t-test). (D) Percent of down calls out of total calls in male pups (Kruskal-Wallis Test). (E) Total activity measured from beam breaks during 10-minute open field assay in 4 week old male and female mice (male[control: n = 16, *Dot1l* cKO: n = 16]; female[control: n = 14, *Dot1l* cKO: n = 14], Kruskal-Wallis Test). (F) Percent of spontaneous alternations out of the total number of triads possible in male and female mice (unpaired two-tailed t-test). (G) Discrimination index between interaction time with the mouse or rock cylinder during the three-chamber social choice assay in male and female mice (unpaired two-tailed t-test). (H) Freezing activity after 2 weeks post-contextual fear conditioning in male and female mice (unpaired two-tailed t-test). All box plot bounds indicate the 25th and 75th percentiles, the black line shows the median, and whiskers extend to the minimum and maximum value that are no further than 1.5 * interquartile range. *p<0.05.

METHODS

Experimental design

The goal of this study was to examine the impact of monoallelic *DOT1L* variants in the brain. To accomplish this, we identified 11 individuals with variants in *DOT1L* with a spectrum of neurodevelopmental disorders. We assessed the methyltransferase activity of two previously published variants and one variant identified in our cohort to better understand how variants impact DOT1L function. In zebrafish, we defined early developmental behavioral disruptions upon *dot1l* loss. We then evaluated the impact of partial DOT1L loss on neuronal transcription and neuronal maturation in mouse primary cultured neurons. Further, we examined how monoallelic loss of *Dot1l* impacts transcription and behavior in mice using ubiquitous and forebrain specific depletion models.

Study Participants

Identification of DOT1L Variants

Variants in DOT1L were identified through connections made through collaborating clinicians, GeneMatcher **(*33*)**, Deciphering Developmental Disorders Research Study, GeneDx, and the MSSNG (www.mss.ng) database. The first individual of interest (Individual E) was identified through a prior publication **(*6*, *12*)**. The remainder of participants were identified through GeneMatcher apart from the following: individual 2, individual 6, and individual F (identified through pre-existing collaboration), individual 3, 4, 8 (GeneDx), individuals A-C (MSSNG database). The initial GeneMatcher entry was made on May 23, 2023, and all matches were considered in this study submitted until April 2024 (table S3). Variants are reported according to Human Genome Variation Society **(*55*)** nomenclature in reference to the DOT1L transcript (NM_032482.3). Allele counts were gathered from gnomAD (v4.1.0), TOPMed Bravo, and RGC Million Exome Variant Browser (table S1, table S2). Pathogenicity of missense variants was predicted using an aggregation of the following databases: Metadome **(*56*)**, Revel **(*57*)**, and AlphaMissense **(*58*)** (table S1, table S2). Variants p.I85M (SCV004169212), E134K (SCV004169195), and Gln598* (SCV003804054) are available on ClinVar.

Ethical Statement

Human subject studies were approved consistent with the principles of research ethics and the legal requirements of the lead clinician authors' jurisdiction(s) (The Hospital for Sick Children, Canada). Voluntary, informed consent was obtained from human participants, consistent with the institutional principles of research ethics and the legal requirements of each referring author's jurisdiction. Ethical approvals were obtained for participation, phenotyping, sample collection and generation/derivation of affected individual and control fibroblasts (IRB#16-013278_AM118, Children's Hospital of

Philadelphia). The authors also confirm that human research participants provided written informed consent for publication of the images in Fig. 1.

Methyltransferase Activity

Expression and Purification of DOT1L and Mutants

DOT1L and mutants were expressed and purified as previously described **(*59*)**. Briefly, the proteins were expressed in BL21 One Shot (DE3) (ThermoFisher) E. coli cells. They were grown at 37°C until reaching an OD600 equal to 0.6-0.8 and then were induced using 0.5mM IPTG for 3 hours at 37°C. The cells were harvested (Sorvall LYNX6000) and then lysed (AvestinEmulsiflexC3) (Lysis buffer: 500mM NaCl, 50mM Tris-HCl pH 8.0, 5% Glycerol, 5mM Imidazole, 2mM BME, 1x Protease Inhibitor). Lysate was incubated with Ni-NTA Beads (Qiagen). Protein was eluted (Elution buffer: 500mM NaCl, 50mM Tris-HCl pH 8.0, 5% Glycerol, 300mM Imidazole, 2mM BME) and cleaved by TEV protease overnight in dialysis (Dialysis buffer: 75mM NaCl, 20mM Tris pH 8.0, 5% Glycerol, 2mM BME). Sample was then purified over a HiTrap SP HP column (Cytvia) (Buffer A: 75mM NaCl, 25mM HEPES pH 7.5, 5% Glycerol, 2mM BME) and eluted with a linear salt gradient (75mM to 1000mM NaCl); and then further purified over HiLoad Superdex 200 16/600 size exclusion column (GE Healthcare) (150mM NaCl, 10mM HEPES 7.5, 2mM DTT). Protein was concentrated, then flash frozen in liquid nitrogen and stored in a -80°C freezer.

Purification of Widom 601 DNA

Widom 601 DNA was transformed into DH5a competent E. Coli cells (NEB) from a plasmid containing 8 copies of 147bp repeats flanked by EcoRV sites **(*60*)**. The cells were grown overnight at 37°C, then harvested and lysed. The DNA was then further purified using established protocols **(*61*)**.

Expression and Purification of Xenopus Histones

Xenopus laevis histones H2A, H2B, H3 and H4 were expressed and purified using previously published protocols **(*61*)**. Briefly, the histone constructs were cloned in a pET-3 vector and grown in pLysS (DE3) cells (NEB) to an OD600 of 0.6-0.8 and induced at 0.5mM IPTG at 37°C for 3 hours. The protein was then extracted from inclusion bodies and purified over a size exclusion column Sephacryl S200 (Cytvia) followed by an SP anion exchange column (Tosoh). Proteins were then dialyzed in 1mM BME and then lyophilized using a Vertis Sentry lyophilizer.

Reconstitution of Nucleosomes

Unmodified nucleosome was assembled as previously described**(*60*, *61*)**. First equimolar ratios of unfolded histones H2A, H2B, H3, and H4 were mixed and dialyzed in refolding buffer. The assembled

octamer was then purified over a size exclusion chromatography column Superdex 200 26/600(GE Healthcare) using refolding buffer. Nucleosomes were assembled by combining an equimolar quantity of octamer and Widom 601 DNA, followed by an overnight salt gradient dialysis using a peristaltic pump (Gilson).

Endpoint Methylation Assay

The endpoint methylation assay was performed as described **(*59*)**. Assays on DOT1L and mutants with unmodified nucleosome were done in three replicates. Briefly, in methyltransferase buffer, 250nM and 125nM of DOT1L or mutants were combined with 1µM of nucleosome. In a volume of 20ml, the reaction was incubated at 30°C for 30min. The reaction was stopped using 5ml of 0.5% TFA. SAH production was determined using a MTase-Glo methyltransferase kit (Promega). The luminescence was measured using an EnSpire 2300 Multilabel plate reader (Perkin Elmer).

Western blotting

Protein lysates or histone samples were mixed with 5X Loading Buffer (5% SDS, 0.3M Tris pH 6.8, 1.1mM Bromophenol blue, 37.5% glycerol), boiled for 10 minutes, and cooled on ice. Protein was resolved by 4%–20% Tris-glycine or 16% Tris-glycine SDS-PAGE, followed by transfer to a 0.45-µm PVDF membrane for immunoblotting. Membranes were blocked for 1 hour at RT in 5% milk in 0.1% TBST and probed with primary antibody overnight at 4C. Membranes were incubated with secondary antibody for 1 hour at RT.

Cell culture

Human Fibroblasts

Individual 5 fibroblasts and age- and sex-matched control fibroblasts were donated from collaborating clinicians. Fibroblasts were cultured in DMEM (with $4.5\,\mathrm{g\,L^{-1}}$ glucose, L-glutamine and sodium pyruvate) supplemented with 15% FBS (Sigma-Aldrich, F2442-500ML) and 1% penicillin-streptomycin (Gibco, 15140122).

Neuro-2A cells

Neuro-2A cells were obtained from the American Type Culture Collection (ATCC), cultured in DMEM (with $4.5\,\mathrm{g\,L^{-1}}$ glucose, L-glutamine and sodium pyruvate) supplemented with 10% FBS (Sigma-Aldrich, F2442-500ML) and 1% penicillin-streptomycin (Gibco, 15140122) and maintained free of mycoplasma. N2A transfections were performed in DMEM using lipofectamine 2000 (Life

Technologies, 11668027). Lipofectamine and DNA complexes were left on for overnight. Cells were harvested for analysis 2 days after transfection.

Primary neuronal culture

Cortices were dissected from E16.5 C57BL/6J embryos and cultured in neurobasal medium (Gibco 21103049) supplemented with B27 (Gibco 17504044), GlutaMAX (Gibco 35050061), penicillin-streptomycin (Gibco 15140122) in TC-treated twelve or six-well plates coated with 0.05 mg/mL Poly-D-lysine (Sigma-Aldrich A-003-E). At 3 DIV, neurons were treated with 0.5 µM AraC. Transfections were performed using lipofectamine 2000 (Life Technologies, 11668027). Neurons were put in a 1mM kynurenic acid solution during transfection to prevent excitotoxicity. Lipofectamine and DNA complexes were left on neurons for 15 minutes. Transfections were performed at 8 to 12 DIV for constructs expressing DOT1L and cells were fixed two to three days later. shRNA transfections were performed at 9 to 12 DIV and fixed three to four days later. Neuronal infections were transduced overnight with lentivirus containing the constructs described below. Virus was removed the following day, and neurons were cultured for 5-7 days.

Constructs

The GFP control plasmid was obtained from Addgene, pLenO-CMV-MCS-GFP-SV-puro (Addgene plasmid# 73582). The pET28-MHL-DOT1L (1-420) was received from the Armache lab originally purchased from Addgene (Addgene plasmid# 40736). The pDSV-DOT1l-HA-Flag-mRFP-nls and empty pDSV-mRFP-nls plasmids were received from the Vogel lab. The Sun1-GFP plasmid was a gift from Jeremy Nathans Lab, pCDNA3-CMV-Sun1-GFP-6xMyc. *Dot1l* shRNA and control Luciferase shRNAs were inserted into the pLKO.1 vector backbone (Addgene plasmid# 10878). *Dot1l* shRNA target sequences were as follows:

1. *Dot1l* shRNA 1: CCGGGTCCAGTTTGTACTGTCAATACTCGAGTATTGACAGTACAAACTGGACTTTTTG
2. Dot1l shRNA 2: CCGGCCTCGGTTTACACAGCTTCAACTCGAGTTGAAGCTGTGTAAACCGAGGTTTTTG
3. *Dot1l* shRNA 3: CCGGCGGCAGAATCGTATCCTCAAACTCGAGTTTGAGGATACGATTCTGCCGTTTTTG

DOT1L mutants were generated using sited directed mutagenesis using Pfu Turbo HotStart DNA polymerase (Agilent, 600322, for R292C and E123K) or NEB Q5 Polymerase (M0491S, for D157N), and primers were created using the DNA-based primer design feature of the online PrimerX tool or

manually creating primers using ~15 bp overlap strategy. Plasmid sequences were verified through Sanger sequencing and/or Plasmidsaurus long read sequencing.

Lentiviral production

HEK293T cells were cultured in high-glucose DMEM growth medium (with 4.5 g L^{-1} glucose, L-glutamine and sodium pyruvate), 10% FBS (Sigma-Aldrich F2442-500ML), and 1% penicillin-streptomycin (Gibco 15140122). Calcium phosphate transfection was performed with Pax2 and VSVG packaging plasmids. Viral media was removed 2 hours after transfection and collected at 48 and 72 hours later. Viral media was passed through a 0.45-µM filter and precipitated for 48 hours with PEG-it solution (40% PEG-8000 [Sigma-Aldrich P2139-1KG], 1.2 M NaCl [Fisher Chemical S271-1]). Viral particles were pelleted and resuspended in 200µL PBS.

RNA-sequencing

Library preparation & sequencing

RNA was isolated using Zymo Quick-RNA Miniprep Plus Kit (R1057). Libraries were generated using the Illumina TruSeq stranded mRNA library prep kit (Illumina 20020595). Prior to sequencing libraries were quantified by qPCR using a KAPA Library Quantification Kit (Roche 07960140001). Libraries were sequenced on an Illumina NextSeq 500/550; reads (75-bp read length, single end). Data can be accessed under the following GEO accession number: GSE279978.

Data processing and analysis

Reads were mapped to Mus musculus genome build mm10 with Star (v2.7.9a). The R packages DESeq2 (*62*) (v1.34.0) and limma (v3.50.3) via edgeR (v3.36.0) were used to perform differential gene expression analysis. We defined genes as differentially expressed where FDR < 0.05 and an absolute log2 fold change > 0.5. Volcano plots were generated using Enhanced Volcano. IGV tools (*63*) (v2.12.3) was used to generate genome browser views.

Gene ontology

PANTHER (*64*, *65*) (v18.0) was used to perform an overrepresentation test against the biological process complete ontology using default parameters. SynGO (*47*) was used for synaptic gene ontologies and overrepresentation tests of differentially expressed genes. All expressed genes (defined as any gene that did not have an NA p adjusted value in the DeSeq2 output and did not have an NA gene name) was used as a background gene list.

Revigo

Revigo (*66*) was used to remove redundant terms and gather a concise list based on a published protocol(*67*). In brief, the Panther output of the Biological Process gene ontology terms and their associated FDR-corrected p-values were input into Revigo. Revigo input parameters used were: size of resulting list – small; remove obsolete GO terms – yes; species – Mus Musculus; semantic similarity measure – Resnik. Revigo output was then filtered using the following conditions: reference genes within a gene ontology term <= 3000, dispensability < 0.2, and fold enrichment > 1. The resulting top 10 gene ontology terms based on FDR-corrected p-values were displayed.

GSEA

The R package FGSEA (*68*) was used to perform pre-ranked gene set enrichment analysis (GSEA) based on log2 fold changes obtained from DESeq2 differential expression analysis. Genes without a defined adjusted p-value and genes with a base mean < 100 were removed prior to running GSEA. The GSEA (*69*, *70*) database (https://www.gsea-msigdb.org/gsea/index.jsp) was used for synaptic transmission based gene sets.

Immunocytochemistry

GluA2 antibody (Synaptic Systems: 182103) was added to the media of live cells and incubated for 45 minutes. Cells were fixed in 4% PFA for 10 min and washed with PBS. Cells were blocked in blocking solution (PBS with 3% BSA and 2% serum) for at least 1 hour. Cell coverslips were then incubated with secondary antibody for 1 hour at room temperature. For detection of GluA2, Goat anti-Rabbit Alexa Fluor™ 647 (Thermo Fisher, A-21244; 1:500) was added to the secondary antibody solution. Nuclei were stained with DAPI (1:1,000 in PBS) for 10 min with washing in PBS. Coverslips were mounted onto microscope slides using ProLong Gold antifade reagent (Thermo Fisher).

Image acquisition

Cells were imaged on an upright Leica DM 6000, TCS SP8 laser scanning confocal microscope with 405-nm, 488-nm, 552-nm and 638-nm lasers. The microscope uses two HyD detectors and three PMT detectors. The objective used was a ×63 HC PL APO CS2 oil objective with an NA of 1.40. Type F immersion liquid (Leica) was used for oil objectives. Images were 175.91 × 171.91 μM^2, 1,024 × 1,024 pixels and 16 bits per pixel. Coverslips were imaged with a *z* stack through the neuron.

Image analysis

Sholl, Spine, and GluA2 Analysis

Images were analyzed using ImageJ (v2.14.0/1.54f) software. A singular z stack image's maximum projection of the GFP channel was generated. The image was traced in Simple Neurite Tracer (SNT) (*71*) and the Sholl analysis feature (*72*) was used to generate a data table with number of intersections per step size (radius step size = 10 μM). An R script (https://zenodo.org/records/1158612) was used generate graphs and summary statistics using a mixed effect model (*73*). Spine density was quantified from the maximum projection image. The three largest neurite branches were measured and projections from these branches were counted. Spines had to be > 0.4 μM and < 8 μM in order to be counted based off previous literature (*74*, *75*). Spine density for each branch was calculated as number of spines/branch length and this value was averaged together for each neuron imaged. To quantify GluA2 levels, an in-house macro was created. In short, this macro creates an outline of the imaged neuron from each individual stack using the GFP channel and then measures the fluorescent intensity in that stack from the far-red channel used to stain GluA2. The fluorescent intensity of each image is normalized to the average intensity of the control transfected neurons.

Single nuclei RNA-sequencing (snRNAseq)

Nuclei Isolation

For each biological replicate, one cortical hemisphere of a mouse was dissected, and flash frozen in liquid nitrogen and stored at −80 °C. The nuclei isolation procedure used was modified from (*76*, *77*). Tissue was homogenized in douncers using a loose pestle (~10-15 strokes) in 1.2 mL of homogenization buffer supplemented with 1 mM DTT, 0.15 mM spermine, 0.5 mM spermidine, RNasin ® Plus Ribonuclease Inhibitor (Promega N2611), and EDTA-free protease inhibitor (Roche). A 5% IGEPAL-630 solution was added (107ul), and the homogenate was further homogenized with the tight pestle (~10-15 strokes). The sample was then mixed with 1.3 mL of 50% iodixanol density medium (Sigma D1556) and added to a polypropylene thin wall tube (13.2 mL, Beckman and Coulter, 331372). The sample was then underlaid with a gradient of 30% and 40% iodixanol, and centrifuged at 10,000 x g for 18 minutes (no brake) in a swinging bucket centrifuge at 4 °C. Nuclei from control and *Dot1l* HET mice were individually counted (3 males, 3 females per genotype) and proportionally combined with all other biological replicates within each genotype. These samples were washed 3 times in DPBS and spun at 1000g for 5 min. Samples were resuspended in 1X Nuclei Buffer (10x Genomics PN-2000153 / 2000207) at a concentration of approximately 5,000 nuclei/ul for subsequent library preparation. All steps were performed on ice or at 4°C.

Library preparation & sequencing

For the generation of ATAC and Gene Expression libraries, the 10X Genomics Chromium Next GEM Single Cell Multiome ATAC + Gene Expression (CG000338 Rev F) protocol was followed. Briefly, 16,100 nuclei from each sample underwent a transposition reaction before being loaded on the 10X genomics Chromium controller to target 10,000 recovered nuclei per sample. The resulting barcoded transposed DNA and barcoded cDNA were then used to generate ATAC and gene expression libraries, respectively, following the manufacturer's guidelines. Quality control was performed during library preparation using an Agilent Bioanalyzer and a Thermo Fisher Qubit. Prior to sequencing libraries were quantified by qPCR using a KAPA Library Quantification Kit (Roche 07960140001). Libraries were sequenced on an Illumina NextSeq 1000, using 28 cycles for Read 1, 10 cycles for the i7 index, 10 cycles for the i5 index, and 90 cycles for Read 2. Data can be accessed under the following GEO accession number: GSE279978.

Preprocessing of snRNAseq data

Paired end sequencing reads were processed using 10X Genomics Cellranger v5.0.1. Reads were aligned to the mm10 genome optimized for single cell sequencing through a hybrid intronic read recovery approach (*78*). In short, reads with valid barcodes were trimmed by TSO sequence, and aligned using STAR v2.7.1 with MAPQ adjustment. Intronic reads were removed, and high-confidence mapped reads were filtered for multimapping and UMI correction. Empty GEMs were also removed as part of the pipeline. Initial dimensionality reduction and clustering was performed prior to processing to enable batch correction and removal of cell free mRNA using SoupX (*79*). Raw expression matrices with counted, individual nuclei UMI and genes were used for subsequent steps and filtering by QC metrics.

Clustering and merging by genotype and comparison

Raw matrices for each individual genotype were converted to Seurat objects using Seurat 5.0.1 and filtered to remove UMIs with thresholds of > 200 minimum features, < 5% mitochondrial reads, and < 5% ribosomal reads. Each genotype (control and *Dot1l* HET, each containing 6 biological replicates) were merged to generate an object for the subsequent steps. Each dataset was normalized (NormalizeData) using the default scale factor of 10000, variable selection (FindVariableFeatures) was performed using 2000 features, then scaled and centered (ScaleData) using all features without regressing any variables. Dimensionality reduction with PCA (RunPCA) used the first 30 principal components and the nearest-neighbor graph construction (FindNeighbors) used the first 10 dimensions. Clustering (FindClusters) was next performed using a resolution of 0.5 before layers corresponding to each genotype were integrated (IntegrateLayers) using CCA Integration with a k

weight of 60 and then rejoined (JoinLayers). The dataset per condition was then dimensionally reduced using the integrated CCA at with 30 dimensions (RunUMAP) and the same resolution of 0.5.

Marker gene identification

To identify marker genes for each cluster, differential expression analysis was performed using the Seurat function FindAllMarkers. Differentially expressed genes that were expressed at least in 25% cells within the cluster and with a fold change more than 0.5 (log scale) were considered marker genes. Cell identity was determined using well-established marker genes for major cortical cell types. Marker gene analysis led to the identification of 17 cortical neuron clusters (10 excitatory, 7 inhibitory), 1 subcortical neuron cluster, and 7 non-neuronal clusters. Neuronal clusters were annotated according to the cortical layer they occupy, or—if unidentifiable by cortical layer—according to the gene most differentially expressed in that cluster relative to all other excitatory or inhibitory neuronal clusters.

Differential gene expression analysis and parsing sex of nuclei

Differential gene expression analysis between control and *Dot1l* HET groups was performed using the Seurat function FindMarkers (min.pct =.001, logfc.threshold = 0.5) with a MAST test. Genes with an adjusted p-value < 0.05 and an absolute log2 fold change > 0.5 were considered differentially expressed between control and *Dot1l* HET. The sex of the nuclei was determined based on the following parameters: females were categorized as nuclei with Xist expression at or above the 70th percentile in comparison to total nuclei and with no expression of Eif2s3y or Ddx3y; males were categorized as nuclei with Xist expression below 70th percentile in comparison to total nuclei and non-zero expression of Eif2s3y or Ddx3y.

Gene ontology

PANTHER (*64*, *65*) (v19.0) was used to perform an overrepresentation test against the biological process complete ontology using default parameters. All expressed genes (defined as any gene within the current Seurat object subset with a min.pct = .001) was used as a background gene list.

Revigo

Revigo (*66*) was used to remove redundant terms and gather a concise list based on a published protocol (*67*). In brief, the Panther output of the Biological Process gene ontology terms and their associated FDR-corrected p-values were input into Revigo. Revigo input parameters used were: size of resulting list – small; remove obsolete GO terms – yes; species – Mus Musculus; semantic similarity measure – Resnik. Revigo output was then filtered using the following conditions: reference genes within a gene ontology term <= 1000, dispensability < 0.1, and fold enrichment > 2. The resulting top 10 gene ontology terms based on FDR-corrected p-values were displayed.

Mice

A floxed *Dot1l* mouse line crossed with the Sun1-sfGFP line (JAX Strain #:030952) was received from Tanja Vogel. In brief, the floxed *Dot1l* mouse line is floxed at exon 2 causing a frameshift that results in an early stop codon and nonfunctional gene product (C57BL6/J background). The *Dot1l* mouse line was originally obtained from the Knockout Mouse Project (KOMP). Heterozygous floxed *Dot1l* mice (*Dot1l*$^{floxed/+}$;Sun1-sfGFP$^{+/+}$) were crossed to the NEX-Cre line (*80*) for neuron-specific behavioral testing. *Dot1l* cKO were *Dot1l*$^{floxed/+}$;Sun1-sfGFP$^{+/-}$;NEX-Cre$^{+/-}$ and controls were *Dot1l*$^{+/+}$;Sun1-sfGFP$^{+/-}$;NEX-Cre$^{+/-}$. Heterozygous *Dot1l* mice that did not also harbor alleles from the Sun1-sfGFP line (*Dot1l*$^{floxed/+}$; Sun1-sfGFP$^{-/-}$) were crossed to the CMV-Cre line (JAX Strain #:006054) for ubiquitous monoallelic *Dot1l* loss behavioral testing. *Dot1l* HET were *Dot1l*$^{floxed/+}$; CMV-Cre$^{+/-}$ and controls were *Dot1l*$^{+/+}$;CMV-Cre$^{+/-}$. All mice were housed in a 12-hour light-dark cycle and fed a standard diet. All experiments were conducted in accordance with and approval of the IACUC at the University of Pennsylvania.

Behavioral assays

Behavioral cohorts

Male and female controls (*Dot1l*$^{+/+}$; Sun1-sfGFP$^{+/-}$;NEX-Cre$^{+/-}$ or *Dot1l*$^{+/+}$;CMV-Cre$^{+/-}$), *Dot1l* HET (*Dot1l*$^{floxed/+}$;CMV-Cre$^{+/-}$), and *Dot1l* cKO (*Dot1l*$^{floxed/+}$;Sun1-sfGFP$^{+/-}$;NEX-Cre$^{+/-}$) mice were tested in the behavioral tests described below:

For control and *Dot1l* HET mice, two cohorts were generated a month apart from each other and used for developmental milestone testing from P1 – P18. Additionally, these mice were used at 4 weeks old at the onset of behavioral testing which included: elevated zero maze, open field, Y maze, social choice, and fear conditioning, in that order. The breakdown of these cohorts was as follows: Cohort 1[control: male = 4, female = 10, *Dot1l* HET: male = 2, female = 10], Cohort 2 control: male = 9, female = 5, *Dot1l* HET: male = 14, female = 13]. A third cohort of mice were used for ultrasonic vocalizations at P6 – P7. The breakdown was as follows: *Dot1l* HET cohort[litter = 13, control: male = 15, female = 22, *Dot1l* HET: male = 22, female = 22].

For control and *Dot1l* cKO mice, a cohort was generated for developmental milestone testing from P1 – P18 [control: male = 6, female = 4, *Dot1l* cKO: male = 11, female = 9]. A separate cohort [control: male = 16, female = 14, *Dot1l* cKO: male = 16, female = 14] of mice was used at 4 weeks old at the onset of behavioral testing which included: open field, Y maze, social choice, and fear conditioning, in that order. A third cohort of mice was used at four weeks old at the onset of behavioral testing for elevated zero maze [control: male = 9, female = 14, *Dot1l* cKO: male = 10, female = 14]. A fourth cohort of mice used for ultrasonic vocalizations at P6 – P7 [litter = 10, control: male = 15, female

= 21, *Dot1l cKO* male = 14, female = 19]. For all behavioral testing, the experimenter was blinded to genotype of the mice.

Ultrasonic vocalizations

Multiple litters were used for both the *Dot1l* HET and *Dot1l* cKO cohorts. Pups at approximately P6 – P7 were individually placed into a soundproof chamber with fresh bedding. A Condenser ultrasound microphone (Avisoft-Bioacoustics CM16/CMPA, part #40011) microphone and UltraSoundGate 116H (Avisoft Bioacoustics, part # 41163, 41164) recording device was used with Avisoft-RECORDER USGH software. The recording sessions were 5 minutes in length and recorded with the following parameters: sampling rate = 375000 Hz, range = 15 – 180 kHz, and min whistle duration = 5 ms. USVs were analyzed using a MATLAB based software, VocalMat (*81*). Mice with fewer than 50 calls were excluded.

Elevated zero maze

The elevated zero apparatus consists of a circular shaped platform raised approximately 16 inches above the floor. Two opposing quadrants have raised walls (wall height = 4 inches, circle width = 2 inches) without a ceiling leaving these closed quadrants open to overhead light. The two remaining opposing quadrants were open (wall height = 0.25 inches). Mice were placed into a closed quadrant and allowed to freely explore for 5 minutes. The entire testing session was recorded, and videos were analyzed using ANY-maze software.

Open field

Mice were placed into an empty arena (15 inches x 15 inches) and allowed to freely explore for 10 minutes. Activity was measured using beam breaks recorded using Photobeam Activity System Open Field software (San Diego Instruments) and percent of center activity was quantified as number of beam breaks in the center / total beam breaks *100.

3-chamber social choice assay

The social choice test was carried out in a three-chambered apparatus, consisting of a center chamber and two outer chambers. Before the start of the test and in a counter-balanced manner, one end chamber was designated the social chamber, into which a stimulus mouse would be introduced, and the other end chamber was designed the nonsocial chamber. Two identical, clear Plexiglas cylinders with multiple holes to allow for air exchange were placed in each end chamber. In the habituation phase of the test, the experimental mouse freely explores the three chambers with empty cue cylinders in place for 10 min. Immediately following habituation, an age- and sex-matched stimulus mouse was placed in the cylinder in the social chamber while a rock was simultaneously placed into the other

cylinder in the nonsocial chamber. The experimental mouse was tracked during the 10 min habituation and 10 min social choice phases. All testing was recorded, and videos were analyzed manually.

Y maze

The Y maze test was performed on a Y shaped apparatus composed of 3 enclosed arms equidistant apart (3 in wide x 5 in wall x 15 in long). Mice were handled for 2 minutes each on 3 consecutive days immediately prior to the onset of testing. For Y maze testing mice were placed at the distal end of the arm closest to the experimenter of the Y maze apparatus. Mice were allowed to freely explore for 8 minutes. Entries into each arm were defined as all four paws of the mouse entering. A spontaneous alternation was defined as a consecutive entry into each of the 3 arms without returning to the arm that the mouse had been in immediately prior. Spontaneous alternation triads over the total number of possible triads based on the total entries were calculated as spontaneous alternations/ (total entries - 2). All testing was recorded, and videos were analyzed manually.

Contextual and cued fear conditioning

Mice were handled for 2 minutes each the day immediately prior to the onset of testing. On training day, mice were placed in individual chambers for 2 minutes followed by a loud tone lasting 30 second that co-terminated with a 2 second, 1.25-mA foot shock. One minute later mice received another tone-shock pairing and were then left undisturbed for an additional 1 minute in the chamber before being returned to their home cage. Freezing behavior, defined as no movement except for respiration, was determined before and after the tone-shock pairings and scored by MedAssociates VideoFreeze software. To test for context-dependent learning, we placed mice back into the same testing boxes 24 hours later for a total of 5 minutes without any tone or shock, and again measured the total time spent freezing. Following an additional 24 hours, we tested for cue-dependent fear memory by placing the mice into a novel chamber consisting of altered flooring, wall-panel inserts, and vanilla scent. After 2 minutes in the chamber, the cue tone was played for a total of 3 minutes, and the total time spent freezing during the presentation of this cue tone was recorded. Long-term contextual and cued fear memory were again tested with the same protocol at 14 days (contextual) or 15 days (cued) post-training.

Zebrafish experiments

Experiments were conducted on 6 dpf larval zebrafish (Danio rerio, TLF strain) raised in E3 medium at 29 °C on a 14:10 h light cycle. At this developmental stage the sex of the organism is not yet determined. Breeding adult zebrafish were maintained at 28 °C on a 14:10 h light cycle. Crispants were generated as described(*42*) by Kroll et al. Three gRNAs targeting three different regions across the dot1l locus

were designed using ChopChop v3 (https://chopchop.cbu.uib.no/). Custom Alt-R CRISPR-Cas9 crRNAs (IDT) were annealed with tracrRNA (IDT, #1072533) to form gRNAs which were subsequently complexed with Cas9 protein (IDT, #1081061) to make the final ribonucleoprotein (RNP) complex. Three non-targeting crRNAs (IDT, #1072544, 1072545, 1072546) were used to make the RNP for controls. Single-cell wildtype (TLF) zebrafish embryos were then microinjected within 15minutes of fertilization with 1nl of RNP mix containing 357pg (10.1fmol) of each gRNA and 5029pg (30.5fmol) of Cas9. Embryos displaying acute toxicity or damage from microinjection were removed from analysis. The remaining embryos were raised to 6dpf at which point they were arrayed on a 100-well plate and multiple sensorimotor behaviors including the visual motor response, responsiveness to flashes of light or darkness, and the acoustic startle response were assessed as described previously(*39*). To confirm that each of the three gRNA-Cas9 RNP complexes was able to target the predicted dot1l locus and cause mutations, genomic DNA was also extracted from dot1l crispants at 6dpf. The predicted target sites were amplified by PCR using primers that flank the region, and the PCR product was then sent for Sanger sequencing. Each of the three RNPs caused mutations at the predicted target site that were not present in control injected embryos.

Sequences:

	crRNA Sequence	GRCz11 Locus	Forward primer	Reverse primer
dot1l crRNA #1	TGAGGTGGACCACCAGTTGA	chr22:20812272	GGAAGGAACGAATAGCCAGTACAA	GCTCCACCAACACAGATCCCT
dot1l crRNA #2	TTGTGGAGCTGTCCCCTCTG	chr22:20809189	CAACCTCATTGAGAGACTGGAG	CACTGTACACCAAGAGGAACAAC
dot1l crRNA #3	GAAGCGCGGGCGACCCAAGA	chr22:20806459	TCAGCAGGCGCAGCATGATTC	GGCCGCTGAGAGGGTCTTG

Statistical analysis

All statistical analyses were performed using readily available code in R. Number of replicates and details of statistical tests are reported in figure legends. Shapiro-Wilk's method was used to test for normality of a given dataset. Detailed information on statistical tests as well as all relevant test statistics can be found in table S4.

REFERENCES

1. M. A. Megan Cleaton, A. Kirby, Why Do We Find it so Hard to Calculate the Burden of Neurodevelopmental Disorders. *J. Child. Dev. Disord.* **04** (2018).

2. L. Francés, J. Quintero, A. Fernández, A. Ruiz, J. Caules, G. Fillon, A. Hervás, C. V. Soler, Current state of knowledge on the prevalence of neurodevelopmental disorders in childhood according to the DSM-5: a systematic review in accordance with the PRISMA criteria. *Child Adolesc. Psychiatry Ment. Health* **16**, 27 (2022).

3. L. Straub, B. T. Bateman, S. Hernandez-Diaz, C. York, B. Lester, K. L. Wisner, C. J. McDougle, P. B. Pennell, K. J. Gray, Y. Zhu, E. A. Suarez, H. Mogun, K. F. Huybrechts, Neurodevelopmental Disorders Among Publicly or Privately Insured Children in the United States. *JAMA Psychiatry* **79**, 232–242 (2022).

4. Daniel Axelrad, Kristen Adams, Farah Chowdhury, Louis D'Amico, Erika Douglass, Gwendolyn Hudson, Erica Koustas, Juleen Lam, Alyson Lorenz, Gregory Miller, Kathleen Newhouse, Onyemaechi Nweke, Doreen Cantor Paster, Julie Sturza, Kari Weber, America's Children and the Environment, Third Edition.

5. H. A. F. Stessman, B. Xiong, B. P. Coe, T. Wang, K. Hoekzema, M. Fenckova, M. Kvarnung, J. Gerdts, S. Trinh, N. Cosemans, L. Vives, J. Lin, T. N. Turner, G. Santen, C. Ruivenkamp, M. Kriek, A. van Haeringen, E. Aten, K. Friend, J. Liebelt, C. Barnett, E. Haan, M. Shaw, J. Gecz, B.-M. Anderlid, A. Nordgren, A. Lindstrand, C. Schwartz, R. F. Kooy, G. Vandeweyer, C. Helsmoortel, C. Romano, A. Alberti, M. Vinci, E. Avola, S. Giusto, E. Courchesne, T. Pramparo, K. Pierce, S. Nalabolu, D. G. Amaral, I. E. Scheffer, M. B. Delatycki, P. J. Lockhart, F. Hormozdiari, B. Harich, A. Castells-Nobau, K. Xia, H. Peeters, M. Nordenskjöld, A. Schenck, R. A. Bernier, E. E. Eichler, Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49**, 515–526 (2017).

6. J. F. McRae, S. Clayton, T. W. Fitzgerald, J. Kaplanis, E. Prigmore, D. Rajan, A. Sifrim, S. Aitken, N. Akawi, M. Alvi, K. Ambridge, D. M. Barrett, T. Bayzetinova, P. Jones, W. D. Jones, D. King, N. Krishnappa, L. E. Mason, T. Singh, A. R. Tivey, M. Ahmed, U. Anjum, H. Archer, R. Armstrong, J. Awada, M. Balasubramanian, S. Banka, D. Baralle, A. Barnicoat, P. Batstone, D. Baty, C. Bennett, J. Berg, B. Bernhard, A. P. Bevan, M. Bitner-Glindzicz, E. Blair, M. Blyth, D. Bohanna, L. Bourdon, D. Bourn, L. Bradley, A. Brady, S. Brent, C. Brewer, K. Brunstrom, D. J. Bunyan, J. Burn, N. Canham, B. Castle, K. Chandler, E. Chatzimichali, D. Cilliers, A. Clarke, S. Clasper, J. Clayton-Smith, V. Clowes, A. Coates, T. Cole, I. Colgiu, A. Collins, M. N. Collinson, F. Connell, N. Cooper, H. Cox, L. Cresswell, G. Cross, Y. Crow, M. D'Alessandro, T. Dabir, R. Davidson, S. Davies, D. de Vries, J. Dean, C. Deshpande, G. Devlin, A. Dixit, A. Dobbie, A. Donaldson, D. Donnai, D. Donnelly, C. Donnelly, A. Douglas, S. Douzgou, A. Duncan, J. Eason, S. Ellard, I. Ellis, F. Elmslie, K. Evans, S. Everest, T. Fendick, R. Fisher, F. Flinter, N. Foulds, A. Fry, A. Fryer, C. Gardiner, L. Gaunt, N. Ghali, R. Gibbons, H. Gill, J. Goodship, D. Goudie, E. Gray, A. Green, P. Greene, L. Greenhalgh, S. Gribble, R. Harrison, L. Harrison, V. Harrison, R. Hawkins, L. He, S. Hellens, A. Henderson, S. Hewitt, L. Hildyard, E. Hobson, S. Holden, M.

Holder, S. Holder, G. Hollingsworth, T. Homfray, M. Humphreys, J. Hurst, B. Hutton, S. Ingram, M. Irving, L. Islam, A. Jackson, J. Jarvis, L. Jenkins, D. Johnson, E. Jones, D. Josifova, S. Joss, B. Kaemba, S. Kazembe, R. Kelsell, B. Kerr, H. Kingston, U. Kini, E. Kinning, G. Kirby, C. Kirk, E. Kivuva, A. Kraus, D. Kumar, V. K. A. Kumar, K. Lachlan, W. Lam, A. Lampe, C. Langman, M. Lees, D. Lim, C. Longman, G. Lowther, S. A. Lynch, A. Magee, E. Maher, A. Male, S. Mansour, K. Marks, K. Martin, U. Maye, E. McCann, V. McConnell, M. McEntagart, R. McGowan, K. McKay, S. McKee, D. J. McMullan, S. McNerlan, C. McWilliam, S. Mehta, K. Metcalfe, A. Middleton, Z. Miedzybrodzka, E. Miles, S. Mohammed, T. Montgomery, D. Moore, S. Morgan, J. Morton, H. Mugalaasi, V. Murday, H. Murphy, S. Naik, A. Nemeth, L. Nevitt, R. Newbury-Ecob, A. Norman, R. O'Shea, C. Ogilvie, K.-R. Ong, S.-M. Park, M. J. Parker, C. Patel, J. Paterson, S. Payne, D. Perrett, J. Phipps, D. T. Pilz, M. Pollard, C. Pottinger, J. Poulton, N. Pratt, K. Prescott, S. Price, A. Pridham, A. Procter, H. Purnell, O. Quarrell, N. Ragge, R. Rahbari, J. Randall, J. Rankin, L. Raymond, D. Rice, L. Robert, E. Roberts, J. Roberts, P. Roberts, G. Roberts, A. Ross, E. Rosser, A. Saggar, S. Samant, J. Sampson, R. Sandford, A. Sarkar, S. Schweiger, R. Scott, I. Scurr, A. Selby, A. Seller, C. Sequeira, N. Shannon, S. Sharif, C. Shaw-Smith, E. Shearing, D. Shears, E. Sheridan, I. Simonic, R. Singzon, Z. Skitt, A. Smith, K. Smith, S. Smithson, L. Sneddon, M. Splitt, M. Squires, F. Stewart, H. Stewart, V. Straub, M. Suri, V. Sutton, G. J. Swaminathan, E. Sweeney, K. Tatton-Brown, C. Taylor, R. Taylor, M. Tein, I. K. Temple, J. Thomson, M. Tischkowitz, S. Tomkins, A. Torokwa, B. Treacy, C. Turner, P. Turnpenny, C. Tysoe, A. Vandersteen, V. Varghese, P. Vasudevan, P. Vijayarangakannan, J. Vogt, E. Wakeling, S. Wallwark, J. Waters, A. Weber, D. Wellesley, M. Whiteford, S. Widaa, S. Wilcox, E. Wilkinson, D. Williams, N. Williams, L. Wilson, G. Woods, C. Wragg, M. Wright, L. Yates, M. Yau, C. Nellåker, M. Parker, H. V. Firth, C. F. Wright, Deciphering Developmental Disorders Study, Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).

7. T. Wang, K. Hoekzema, D. Vecchio, H. Wu, A. Sulovari, B. P. Coe, M. A. Gillentine, A. B. Wilfert, L. A. Perez-Jurado, M. Kvarnung, Y. Sleyp, R. K. Earl, J. A. Rosenfeld, M. R. Geisheker, L. Han, B. Du, C. Barnett, E. Thompson, M. Shaw, R. Carroll, K. Friend, R. Catford, E. E. Palmer, X. Zou, J. Ou, H. Li, H. Guo, J. Gerdts, E. Avola, G. Calabrese, M. Elia, D. Greco, A. Lindstrand, A. Nordgren, B.-M. Anderlid, G. Vandeweyer, A. Van Dijck, N. Van der Aa, B. McKenna, M. Hancarova, S. Bendova, M. Havlovicova, G. Malerba, B. D. Bernardina, P. Muglia, A. van Haeringen, M. J. V. Hoffer, B. Franke, G. Cappuccio, M. Delatycki, P. J. Lockhart, M. A. Manning, P. Liu, I. E. Scheffer, N. Brunetti-Pierri, N. Rommelse, D. G. Amaral, G. W. E. Santen, E. Trabetti, Z. Sedláček, J. J. Michaelson, K. Pierce, E. Courchesne, R. F. Kooy, M. Nordenskjöld, C. Romano, H. Peeters, R. A. Bernier, J. Gecz, K. Xia, E. E. Eichler, Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nat. Commun.* **11**, 4932 (2020).

8. J. L. Ronan, W. Wu, G. R. Crabtree, From neural development to cognition: unexpected roles for chromatin. *Nat. Rev. Genet.* **14**, 347–359 (2013).

9. B. Mossink, M. Negwer, D. Schubert, N. Nadif Kasri, The emerging role of chromatin remodelers in neurodevelopmental disorders: a developmental perspective. *Cell. Mol. Life Sci.*, doi: 10.1007/s00018-020-03714-5 (2020).

10. S. Parkel, J. P. Lopez-Atalaya, A. Barco, Histone H3 lysine methylation in cognition and intellectual disability disorders. *Learn. Mem.* **20**, 570–579 (2013).

11. J.-H. Kim, J. H. Lee, I.-S. Lee, S. B. Lee, K. S. Cho, Histone Lysine Methylation and Neurodevelopmental Disorders. *Int. J. Mol. Sci.* **18**, 1404 (2017).

12. V. Faundes, W. G. Newman, L. Bernardini, N. Canham, J. Clayton-Smith, B. Dallapiccola, S. J. Davies, M. K. Demos, A. Goldman, H. Gill, R. Horton, B. Kerr, D. Kumar, A. Lehman, S. McKee, J. Morton, M. J. Parker, J. Rankin, L. Robertson, I. K. Temple, S. Adam, C. du Souich, A. M. Elliott, A. Lehman, J. Mwenifumbo, T. N.

Nelson, C. van Karnebeek, J. M. Friedman, J. F. McRae, S. Clayton, T. W. Fitzgerald, J. Kaplanis, E. Prigmore, D. Rajan, A. Sifrim, S. Aitken, N. Akawi, M. Alvi, K. Ambridge, D. M. Barrett, T. Bayzetinova, P. Jones, W. D. Jones, D. King, N. Krishnappa, L. E. Mason, T. Singh, A. R. Tivey, M. Ahmed, U. Anjum, H. Archer, R. Armstrong, J. Awada, M. Balasubramanian, S. Banka, D. Baralle, A. Barnicoat, P. Batstone, D. Baty, C. Bennett, J. Berg, B. Bernhard, A. P. Bevan, M. Bitner-Glindzicz, E. Blair, M. Blyth, D. Bohanna, L. Bourdon, D. Bourn, L. Bradley, A. Brady, S. Brent, C. Brewer, K. Brunstrom, D. J. Bunyan, J. Burn, N. Canham, B. Castle, K. Chandler, E. Chatzimichali, D. Cilliers, A. Clarke, S. Clasper, J. Clayton-Smith, V. Clowes, A. Coates, T. Cole, I. Colgiu, A. Collins, M. N. Collinson, F. Connell, N. Cooper, H. Cox, L. Cresswell, G. Cross, Y. Crow, M. D'Alessandro, T. Dabir, R. Davidson, S. Davies, D. de Vries, J. Dean, C. Deshpande, G. Devlin, A. Dixit, A. Dobbie, A. Donaldson, D. Donnai, D. Donnelly, C. Donnelly, A. Douglas, S. Douzgou, A. Duncan, J. Eason, S. Ellard, I. Ellis, F. Elmslie, K. Evans, S. Everest, T. Fendick, R. Fisher, F. Flinter, N. Foulds, A. Fry, A. Fryer, C. Gardiner, L. Gaunt, N. Ghali, R. Gibbons, H. Gill, J. Goodship, D. Goudie, E. Gray, A. Green, P. Greene, L. Greenhalgh, S. Gribble, R. Harrison, L. Harrison, V. Harrison, R. Hawkins, L. He, S. Hellens, A. Henderson, S. Hewitt, L. Hildyard, E. Hobson, S. Holden, M. Holder, S. Holder, G. Hollingsworth, T. Homfray, M. Humphreys, J. Hurst, B. Hutton, S. Ingram, M. Irving, L. Islam, A. Jackson, J. Jarvis, L. Jenkins, D. Johnson, E. Jones, D. Josifova, S. Joss, B. Kaemba, S. Kazembe, R. Kelsell, B. Kerr, H. Kingston, U. Kini, E. Kinning, G. Kirby, C. Kirk, E. Kivuva, A. Kraus, D. Kumar, V. K. A. Kumar, K. Lachlan, W. Lam, A. Lampe, C. Langman, M. Lees, D. Lim, C. Longman, G. Lowther, S. A. Lynch, A. Magee, E. Maher, A. Male, S. Mansour, K. Marks, K. Martin, U. Maye, E. McCann, V. McConnell, M. McEntagart, R. McGowan, K. McKay, S. McKee, D. J. McMullan, S. McNerlan, C. McWilliam, S. Mehta, K. Metcalfe, A. Middleton, Z. Miedzybrodzka, E. Miles, S. Mohammed, T. Montgomery, D. Moore, S. Morgan, J. Morton, H. Mugalaasi, V. Murday, H. Murphy, S. Naik, A. Nemeth, L. Nevitt, R. Newbury-Ecob, A. Norman, R. O'Shea, C. Ogilvie, K.-R. Ong, S.-M. Park, M. J. Parker, C. Patel, J. Paterson, S. Payne, D. Perrett, J. Phipps, D. T. Pilz, M. Pollard, C. Pottinger, J. Poulton, N. Pratt, K. Prescott, S. Price, A. Pridham, A. Procter, H. Purnell, O. Quarrell, N. Ragge, R. Rahbari, J. Randall, J. Rankin, L. Raymond, D. Rice, L. Robert, E. Roberts, J. Roberts, P. Roberts, G. Roberts, A. Ross, E. Rosser, A. Saggar, S. Samant, J. Sampson, R. Sandford, A. Sarkar, S. Schweiger, R. Scott, I. Scurr, A. Selby, A. Seller, C. Sequeira, N. Shannon, S. Sharif, C. Shaw-Smith, E. Shearing, D. Shears, E. Sheridan, I. Simonic, R. Singzon, Z. Skitt, A. Smith, K. Smith, S. Smithson, L. Sneddon, M. Splitt, M. Squires, F. Stewart, H. Stewart, V. Straub, M. Suri, V. Sutton, G. J. Swaminathan, E. Sweeney, K. Tatton-Brown, C. Taylor, R. Taylor, M. Tein, I. K. Temple, J. Thomson, M. Tischkowitz, S. Tomkins, A. Torokwa, B. Treacy, C. Turner, P. Turnpenny, C. Tysoe, A. Vandersteen, V. Varghese, P. Vasudevan, P. Vijayarangakannan, J. Vogt, E. Wakeling, S. Wallwark, J. Waters, A. Weber, D. Wellesley, M. Whiteford, S. Widaa, S. Wilcox, E. Wilkinson, D. Williams, N. Williams, L. Wilson, G. Woods, C. Wragg, M. Wright, L. Yates, M. Yau, C. Nellåker, M. Parker, H. V. Firth, C. F. Wright, D. R. FitzPatrick, J. C. Barrett, M. E. . Hurles, S. Banka, Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. *Am. J. Hum. Genet.* **102**, 175–187 (2018).

13. K. D. Wilson, E. G. Porter, B. A. Garcia, Reprogramming of the epigenome in neurodevelopmental disorders. *Crit. Rev. Biochem. Mol. Biol.* **57**, 73–112 (2022).

14. A. B. Arnett, E. Harstad, M. O'Connell, K. Hayes, S. Brewster, W. Barbaresi, R. N. Doan, Rare De Novo and Inherited Genes in Familial and Nonfamilial Pediatric Attention-Deficit/Hyperactivity Disorder. *JAMA Pediatr.* **178**, 81–84 (2024).

15. Z. Nil, A. R. Deshwar, Y. Huang, S. Barish, X. Zhang, S. Choufani, P. Le Quesne Stabej, I. Hayes, P. Yap, C. Haldeman-Englert, C. Wilson, T. Prescott, K. Tveten, A. Vøllo, D. Haynes, P. G. Wheeler, J. Zon, C. Cytrynbaum, R. Jobling, M. Blyth, S. Banka, A. Afenjar, C. Mignot, F. Robin-Renaldo, B. Keren, O. Kanca, X. Mao, D. J. Wegner, K. Sisco, M. Shinawi, M. F. Wangler, R. Weksberg, S. Yamamoto, G. Costain, H. J. Bellen,

Rare *de novo* gain-of-function missense variants in *DOT1L* are associated with developmental delay and congenital anomalies. *Am. J. Hum. Genet.* **110**, 1919–1937 (2023).

16. B. Jones, H. Su, A. Bhat, H. Lei, J. Bajko, S. Hevi, G. A. Baltus, S. Kadam, H. Zhai, R. Valdez, S. Gonzalo, Y. Zhang, E. Li, T. Chen, The Histone H3K79 Methyltransferase Dot1L Is Essential for Mammalian Development and Heterochromatin Structure. *PLOS Genet.* **4**, e1000190 (2008).

17. H. Vlaming, F. van Leeuwen, The upstreams and downstreams of H3K79 methylation by DOT1L. *Chromosoma* **125**, 593–605 (2016).

18. F. Frederiks, M. Tzouros, G. Oudgenoeg, T. van Welsem, M. Fornerod, J. Krijgsveld, F. van Leeuwen, Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. *Nat. Struct. Mol. Biol.* **15**, 550–557 (2008).

19. D. J. Steger, M. I. Lefterova, L. Ying, A. J. Stonestrom, M. Schupp, D. Zhuo, A. L. Vakoc, J.-E. Kim, J. Chen, M. A. Lazar, G. A. Blobel, C. R. Vakoc, DOT1L/KMT4 Recruitment and H3K79 Methylation Are Ubiquitously Coupled with Gene Transcription in Mammalian Cells. *Mol. Cell. Biol.* **28**, 2825–2839 (2008).

20. S.-K. Kim, I. Jung, H. Lee, K. Kang, M. Kim, K. Jeong, C. S. Kwon, Y.-M. Han, Y. S. Kim, D. Kim, D. Lee, Human Histone H3K79 Methyltransferase DOT1L Methyltransferase Binds Actively Transcribing RNA Polymerase II to Regulate Gene Expression*,. *J. Biol. Chem.* **287**, 39698–39709 (2012).

21. A. Wu, J. Zhi, T. Tian, A. Cihan, M. A. Cevher, Z. Liu, Y. David, T. W. Muir, R. G. Roeder, M. Yu, DOT1L complex regulates transcriptional initiation in human erythroleukemic cells. *Proc. Natl. Acad. Sci.* **118**, e2106148118 (2021).

22. J. Lin, Y. Wu, G. Tian, D. Yu, E. Yang, W. H. Lam, Z. Liu, Y. Jing, S. Dang, X. Bao, J. W. H. Wong, Y. Zhai, X. D. Li, Menin “reads” H3K79me2 mark in a nucleosomal context. *Science* **379**, 717–723 (2023).

23. A. T. Nguyen, Y. Zhang, The diverse functions of Dot1 and H3K79 methylation. *Genes Dev.* **25**, 1345–1358 (2011).

24. H. Franz, A. Villarreal, S. Heidrich, P. Videm, F. Kilpert, I. Mestres, F. Calegari, R. Backofen, T. Manke, T. Vogel, DOT1L promotes progenitor proliferation and primes neuronal layer identity in the developing cerebral cortex. *Nucleic Acids Res.* **47**, 168–183 (2019).

25. P. P. Bovio, H. Franz, S. Heidrich, T. Rauleac, F. Kilpert, T. Manke, T. Vogel, Differential Methylation of H3K79 Reveals DOT1L Target Genes and Function in the Cerebellum In Vivo. *Mol. Neurobiol.* **56**, 4273–4287 (2019).

26. A. Gray de Cristoforis, F. Ferrari, F. Clotman, T. Vogel, Differentiation and localization of interneurons in the developing spinal cord depends on DOT1L expression. *Mol. Brain* **13**, 85 (2020).

27. A. Cheffer, M. Garcia-Miralles, E. Maier, I. Akol, H. Franz, V. S. V. Srinivasan, T. Vogel, DOT1L deletion impairs the development of cortical parvalbumin-expressing interneurons. *Cereb. Cortex* **33**, 10272–10285 (2023).

28. D. Roidl, N. Hellbach, P. P. Bovio, A. Villarreal, S. Heidrich, S. Nestel, B. A. Grüning, U. Boenisch, T. Vogel, DOT1L Activity Promotes Proliferation and Protects Cortical Neural Stem Cells from Activation of ATF4-DDIT3-Mediated ER Stress In Vitro. *STEM CELLS* **34**, 233–245 (2016).

29. F. Ferrari, L. Arrigoni, H. Franz, A. Izzo, L. Butenko, E. Trompouki, T. Vogel, T. Manke, DOT1L-mediated murine neuronal differentiation associates with H3K79me2 accumulation and preserves SOX2-enhancer accessibility. *Nat. Commun.* **11**, 5200 (2020).

30. K. Cao, M. Ugarenko, P. A. Ozark, J. Wang, S. A. Marshall, E. J. Rendleman, K. Liang, L. Wang, L. Zou, E. R. Smith, F. Yue, A. Shilatifard, DOT1L-controlled cell-fate determination and transcription elongation are independent of H3K79 methylation. *Proc. Natl. Acad. Sci.* **117**, 27365–27373 (2020).

31. B. Appiah, C. L. Fullio, C. Ossola, I. Bertani, E. Restelli, A. Cheffer, M. Polenghi, C. Haffner, M. Garcia-Miralles, P. Zeis, M. Treppner, P. Bovio, L. Schlichtholz, A. Mas-Sanchez, L. Zografidou, J. Winter, H. Binder, D. Grün, N. Kalebic, E. Taverna, T. Vogel, DOT1L activity affects neural stem cell division mode and reduces differentiation and ASNS expression. *EMBO Rep.* **24**, e56233 (2023).

32. H. J. Van Heesbeen, L. Von Oerthel, P. M. De Vries, C. M. R. J. Wagemans, M. P. Smidt, Neuronal Dot1l Activity Acts as a Mitochondrial Gene-Repressor Associated with Human Brain Aging via H3K79 Hypermethylation. *Int. J. Mol. Sci.* **24**, 1387 (2023).

33. N. Sobreira, F. Schiettecatte, D. Valle, A. Hamosh, GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Hum. Mutat.* **36**, 928–930 (2015).

34. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski, B. M. Neale, M. J. Daly, D. G. MacArthur, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

35. R. K. C Yuen, D. Merico, M. Bookman, J. L Howe, B. Thiruvahindrapuram, R. V. Patel, J. Whitney, N. Deflaux, J. Bingham, Z. Wang, G. Pellecchia, J. A. Buchanan, S. Walker, C. R. Marshall, M. Uddin, M. Zarrei, E. Deneault, L. D'Abate, A. J. S. Chan, S. Koyanagi, T. Paton, S. L. Pereira, N. Hoang, W. Engchuan, E. J. Higginbotham, K. Ho, S. Lamoureux, W. Li, J. R. MacDonald, T. Nalpathamkalam, W. W. L. Sung, F. J. Tsoi, J. Wei, L. Xu, A.-M. Tasse, E. Kirby, W. Van Etten, S. Twigger, W. Roberts, I. Drmic, S. Jilderda, B. M. Modi, B. Kellam, M. Szego, C. Cytrynbaum, R. Weksberg, L. Zwaigenbaum, M. Woodbury-Smith, J. Brian, L. Senman, A. Iaboni, K. Doyle-Thomas, A. Thompson, C. Chrysler, J. Leef, T. Savion-Lemieux, I. M. Smith, X. Liu, R. Nicolson, V. Seifer, A. Fedele, E. H. Cook, S. Dager, A. Estes, L. Gallagher, B. A. Malow, J. R. Parr, S. J. Spence, J. Vorstman, B. J. Frey, J. T. Robinson, L. J. Strug, B. A. Fernandez, M. Elsabbagh, M. T. Carter, J. Hallmayer, B. M. Knoppers, E. Anagnostou, P. Szatmari, R. H. Ring, D. Glazer, M. T. Pletcher, S. W. Scherer, Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).

36. E. J. Worden, N. A. Hoffmann, C. W. Hicks, C. Wolberger, Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* **176**, 1490-1501.e12 (2019).

37. K. Howe, M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C. Churcher, C. Scott, J. C. Barrett, R. Koch, G.-J. Rauch, S. White, W. Chow, B. Kilian, L. T. Quintais, J. A. Guerra-Assunção, Y. Zhou, Y. Gu, J. Yen, J.-H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd, E. Kenyon, S.

Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Elliott, G. Threadgold, G. Harden, D. Ware, S. Begum, B. Mortimore, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Lloyd, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. D. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C. M. Dooley, Z. Ersan-Ürün, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn, J. Konantz, M. Konantz, M. Oberländer, S. Rudolph-Geiger, M. Teucke, C. Lanz, G. Raddatz, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, S. C. Schuster, N. P. Carter, J. Harrow, Z. Ning, J. Herrero, S. M. J. Searle, A. Enright, R. Geisler, R. H. A. Plasterk, C. Lee, M. Westerfield, P. J. de Jong, L. I. Zon, J. H. Postlethwait, C. Nüsslein-Volhard, T. J. P. Hubbard, H. R. Crollius, J. Rogers, D. L. Stemple, The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).

38. Y. Hu, I. Flockhart, A. Vinayagam, C. Bergwitz, B. Berger, N. Perrimon, S. E. Mohr, An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**, 357 (2011).

39. P. D. Campbell, I. Lee, S. Thyme, M. Granato, Mitochondrial proteins encoded by the 22q11.2 neurodevelopmental locus regulate neural stem and progenitor cell proliferation. *Mol. Psychiatry* **28**, 3769–3781 (2023).

40. H. Weinschutz Mendes, U. Neelakantan, Y. Liu, S. E. Fitzpatrick, T. Chen, W. Wu, A. Pruitt, D. S. Jin, P. Jamadagni, M. Carlson, C. M. Lacadie, K. D. Enriquez, N. Li, D. Zhao, S. Ijaz, C. Sakai, C. Szi, B. Rooney, M. Ghosh, I. Nwabudike, A. Gorodezky, S. Chowdhury, M. Zaheer, S. McLaughlin, J. M. Fernandez, J. Wu, J. A. Eilbott, B. Vander Wyk, J. Rihel, X. Papademetris, Z. Wang, E. J. Hoffman, High-throughput functional analysis of autism genes in zebrafish identifies convergence in dopaminergic and neuroimmune pathways. *Cell Rep.* **42**, 112243 (2023).

41. M. E. S. Capps, A. J. Moyer, C. L. Conklin, V. Martina, E. G. Torija-Olson, M. C. Klein, W. C. Gannaway, C. C. S. Calhoun, M. D. Vivian, S. B. Thyme, Diencephalic and Neuropeptidergic Dysfunction in Zebrafish with Autism Risk Mutations. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.01.18.576309.

42. F. Kroll, G. T. Powell, M. Ghosh, G. Gestri, P. Antinucci, T. J. Hearn, H. Tunbak, S. Lim, H. W. Dennis, J. M. Fernandez, D. Whitmore, E. Dreosti, S. W. Wilson, E. J. Hoffman, J. Rihel, A simple and effective F0 knockout method for rapid screening of behaviour and other complex phenotypes. *eLife* **10**, e59683 (2021).

43. F. Emran, J. Rihel, J. E. Dowling, A Behavioral Assay to Measure Responsiveness of Zebrafish to Changes in Light Intensities. *J. Vis. Exp. JoVE*, 923 (2008).

44. H. A. Burgess, M. Granato, Modulation of locomotor activity in larval zebrafish during light adaptation. *J. Exp. Biol.* **210**, 2526–2539 (2007).

45. C. B. Kimmel, J. Patterson, R. O. Kimmel, The development and behavioral characteristics of the startle response in the zebra fish. *Dev. Psychobiol.* **7**, 47–60 (1974).

46. H. A. Burgess, M. Granato, Sensorimotor Gating in Larval Zebrafish. *J. Neurosci.* **27**, 4984–4994 (2007).

47. F. Koopmans, P. van Nierop, M. Andres-Alonso, A. Byrnes, T. Cijsouw, M. P. Coba, L. N. Cornelisse, R. J. Farrell, H. L. Goldschmidt, D. P. Howrigan, N. K. Hussain, C. Imig, A. P. H. de Jong, H. Jung, M. Kohansalnodehi, B. Kramarz, N. Lipstein, R. C. Lovering, H. MacGillavry, V. Mariano, H. Mi, M. Ninov, D. Osumi-Sutherland, R. Pielot, K.-H. Smalla, H. Tang, K. Tashman, R. F. G. Toonen, C. Verpelli, R. Reig-Viader, K. Watanabe, J. van Weering, T. Achsel, G. Ashrafi, N. Asi, T. C. Brown, P. De Camilli, M. Feuermann, R. E. Foulger, P. Gaudet, A. Joglekar, A. Kanellopoulos, R. Malenka, R. A. Nicoll, C. Pulido, J. de Juan-Sanz, M. Sheng, T. C. Südhof, H. U. Tilgner, C. Bagni, À. Bayés, T. Biederer, N. Brose, J. J. E. Chua, D. C. Dieterich, E. D. Gundelfinger, C. Hoogenraad, R. L. Huganir, R. Jahn, P. S. Kaeser, E. Kim, M. R. Kreutz, P. S. McPherson, B. M. Neale, V. O'Connor, D. Posthuma, T. A. Ryan, C. Sala, G. Feng, S. E. Hyman, P. D. Thomas, A. B. Smit, M. Verhage, SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234.e4 (2019).

48. P. Penzes, M. E. Cahill, K. A. Jones, J.-E. VanLeeuwen, K. M. Woolfrey, Dendritic spine pathology in neuropsychiatric disorders. *Nat. Neurosci.* **14**, 285–293 (2011).

49. F. Schwenk, U. Baron, K. Rajewsky, A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).

50. A. T. Nguyen, J. He, O. Taranova, Y. Zhang, Essential role of DOT1L in maintaining normal adult hematopoiesis. *Cell Res.* **21**, 1370–1373 (2011).

51. A. T. Nguyen, B. Xiao, R. L. Neppl, E. M. Kallin, J. Li, T. Chen, D.-Z. Wang, X. Xiao, Y. Zhang, DOT1L regulates dystrophin expression and is critical for cardiac function. *Genes Dev.* **25**, 263–274 (2011).

52. M. Premoli, M. Memo, S. A. Bonini, Ultrasonic vocalizations in mice: relevance for ethologic and neurodevelopmental disorders studies. *Neural Regen. Res.* **16**, 1158 (2021).

53. E. S. Na, E. D. Nelson, E. T. Kavalali, L. M. Monteggia, The Impact of MeCP2 Loss- or Gain-of-Function on Synaptic Plasticity. *Neuropsychopharmacology* **38**, 212–219 (2013).

54. C. Gracia-Diaz, Y. Zhou, Q. Yang, R. Maroofian, P. Espana-Bonilla, C.-H. Lee, S. Zhang, N. Padilla, R. Fueyo, E. A. Waxman, S. Lei, G. Otrimski, D. Li, S. E. Sheppard, P. Mark, M. H. Harr, H. Hakonarson, L. Rodan, A. Jackson, P. Vasudevan, C. Powel, S. Mohammed, S. Maddirevula, H. Alzaidan, E. A. Faqeih, S. Efthymiou, V. Turchetti, F. Rahman, S. Maqbool, V. Salpietro, S. H. Ibrahim, G. di Rosa, H. Houlden, M. N. Alharbi, N. A. Al-Sannaa, P. Bauer, G. Zifarelli, C. Estaras, A. C. E. Hurst, M. L. Thompson, A. Chassevent, C. L. Smith-Hicks, X. de la Cruz, A. M. Holtz, H. Z. Elloumi, M. J. Hajianpour, C. Rieubland, D. Braun, S. Banka, D. L. French, E. A. Heller, M. Saade, H. Song, G. Ming, F. S. Alkuraya, P. B. Agrawal, D. Reinberg, E. J. Bhoj, M. A. Martínez-Balbás, N. Akizu, Gain and loss of function variants in EZH1 disrupt neurogenesis and cause dominant and recessive neurodevelopmental disorders. *Nat. Commun.* **14**, 4109 (2023).

55. J. T. den Dunnen, R. Dalgleish, D. R. Maglott, R. K. Hart, M. S. Greenblatt, J. McGowan-Jordan, A.-F. Roux, T. Smith, S. E. Antonarakis, P. E. M. Taschner, HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum. Mutat.* **37**, 564–569 (2016).

56. L. Wiel, C. Baakman, D. Gilissen, J. A. Veltman, G. Vriend, C. Gilissen, MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum. Mutat.* **40**, 1030–1038 (2019).

57. N. M. Ioannidis, J. H. Rothstein, V. Pejaver, S. Middha, S. K. McDonnell, S. Baheti, A. Musolf, Q. Li, E. Holzinger, D. Karyadi, L. A. Cannon-Albright, C. C. Teerlink, J. L. Stanford, W. B. Isaacs, J. Xu, K. A. Cooney, E. M. Lange, J. Schleutker, J. D. Carpten, I. J. Powell, O. Cussenot, G. Cancel-Tassin, G. G. Giles, R. J. MacInnis, C. Maier, C.-L. Hsieh, F. Wiklund, W. J. Catalona, W. D. Foulkes, D. Mandal, R. A. Eeles, Z. Kote-Jarai, C. D. Bustamante, D. J. Schaid, T. Hastie, E. A. Ostrander, J. E. Bailey-Wilson, P. Radivojac, S. N. Thibodeau, A. S. Whittemore, W. Sieh, REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).

58. J. Cheng, G. Novati, J. Pan, C. Bycroft, A. Žemgulytė, T. Applebaum, A. Pritzel, L. H. Wong, M. Zielinski, T. Sargeant, R. G. Schneider, A. W. Senior, J. Jumper, D. Hassabis, P. Kohli, Ž. Avsec, Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).

59. M. I. Valencia-Sánchez, P. De Ioannes, M. Wang, N. Vasilyev, R. Chen, E. Nudler, J.-P. Armache, K.-J. Armache, Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination. *Mol. Cell* **74**, 1010-1019.e6 (2019).

60. K.-J. Armache, J. D. Garlick, D. Canzio, G. J. Narlikar, R. E. Kingston, Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* **334**, 977–982 (2011).

61. P. N. Dyer, R. S. Edayathumangalam, C. L. White, Y. Bao, S. Chakravarthy, U. M. Muthurajan, K. Luger, Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2004).

62. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

63. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

64. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

65. The 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**, iyad031 (2023).

66. F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800 (2011).

67. A Simple Protocol for Informative Visualization of Enriched Gene Ontology Terms. https://bio-protocol.org/exchange/protocoldetail?id=3429&type=1.

68. G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov, A. Sergushichev, Fast gene set enrichment analysis. bioRxiv [Preprint] (2021). https://doi.org/10.1101/060012.

69. V. K. Mootha, C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstråle, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, L. C. Groop, PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

70. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).

71. C. Arshadi, U. Günther, M. Eddison, K. I. S. Harrington, T. A. Ferreira, SNT: a unifying toolbox for quantification of neuronal anatomy. *Nat. Methods* **18**, 374–377 (2021).

72. T. A. Ferreira, A. V. Blackman, J. Oyrer, S. Jayabal, A. J. Chung, A. J. Watt, P. J. Sjöström, D. J. van Meyel, Neuronal morphometry directly from bitmap images. *Nat. Methods* **11**, 982–984 (2014).

73. M. D. Wilson, S. Sethi, P. J. Lein, K. P. Keil, Valid statistical approaches for analyzing sholl data: Mixed effects versus simple linear models. *J. Neurosci. Methods* **279**, 33–43 (2017).

74. L. Gu, S. Kleiber, L. Schmid, F. Nebeling, M. Chamoun, J. Steffen, J. Wagner, M. Fuhrmann, Long-Term In Vivo Imaging of Dendritic Spines in the Hippocampus Reveals Structural Plasticity. *J. Neurosci.* **34**, 13948–13953 (2014).

75. B. Ruszczycki, Z. Szepesi, G. M. Wilczynski, M. Bijata, K. Kalita, L. Kaczmarek, J. Wlodarczyk, Sampling issues in quantitative analysis of dendritic spines morphology. *BMC Bioinformatics* **13**, 213 (2012).

76. A. Mo, E. A. Mukamel, F. P. Davis, C. Luo, G. L. Henry, S. Picard, M. A. Urich, J. R. Nery, T. J. Sejnowski, R. Lister, S. R. Eddy, J. R. Ecker, J. Nathans, Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**, 1369–1384 (2015).

77. M. D. Carpenter, D. K. Fischer, S. Zhang, A. M. Bond, K. S. Czarnecki, M. T. Woolf, H. Song, E. A. Heller, Cell-type specific profiling of histone post-translational modifications in the adult mouse striatum. *Nat. Commun.* **13**, 7720 (2022).

78. A.-H. Pool, H. Poldsam, S. Chen, M. Thomson, Y. Oka, Recovery of missing single-cell RNA-sequencing data with optimized transcriptomic references. *Nat. Methods* **20**, 1506–1515 (2023).

79. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* **9**, giaa151 (2020).

80. S. Goebbels, I. Bormuth, U. Bode, O. Hermanson, M. H. Schwab, K.-A. Nave, Genetic targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice. *genesis* **44**, 611–621 (2006).

81. A. H. Fonseca, G. M. Santana, G. M. Bosque Ortiz, S. Bampi, M. O. Dietrich, Analysis of ultrasonic vocalizations from mice using computer vision and machine learning. *eLife* **10**, e59161 (2021).

Acknowledgments

We thank the patients and their families for sharing data and samples. The authors wish to acknowledge the resources of MSSNG (www.mss.ng), Autism Speaks and The Centre for Applied Genomics at The Hospital for Sick Children, Toronto, Canada. We also thank the participating families and clinicians for their time and contributions to this database, as well as the generosity of the donors who supported this program. We thank Dr. Tanja Vogel for sharing DOT1L;Sun1-sfGFP mouse line and Dr. Andrea Stout for microscopy support. Behavioral procedures were performed at the Neurobehavior Testing Core at the University of Pennsylvania. Single nucleus RNA-sequencing library preparation was performed by the Single Cell Core at Children's Hospital of Pennsylvania. We thank the Bhoj lab for providing fibroblasts.

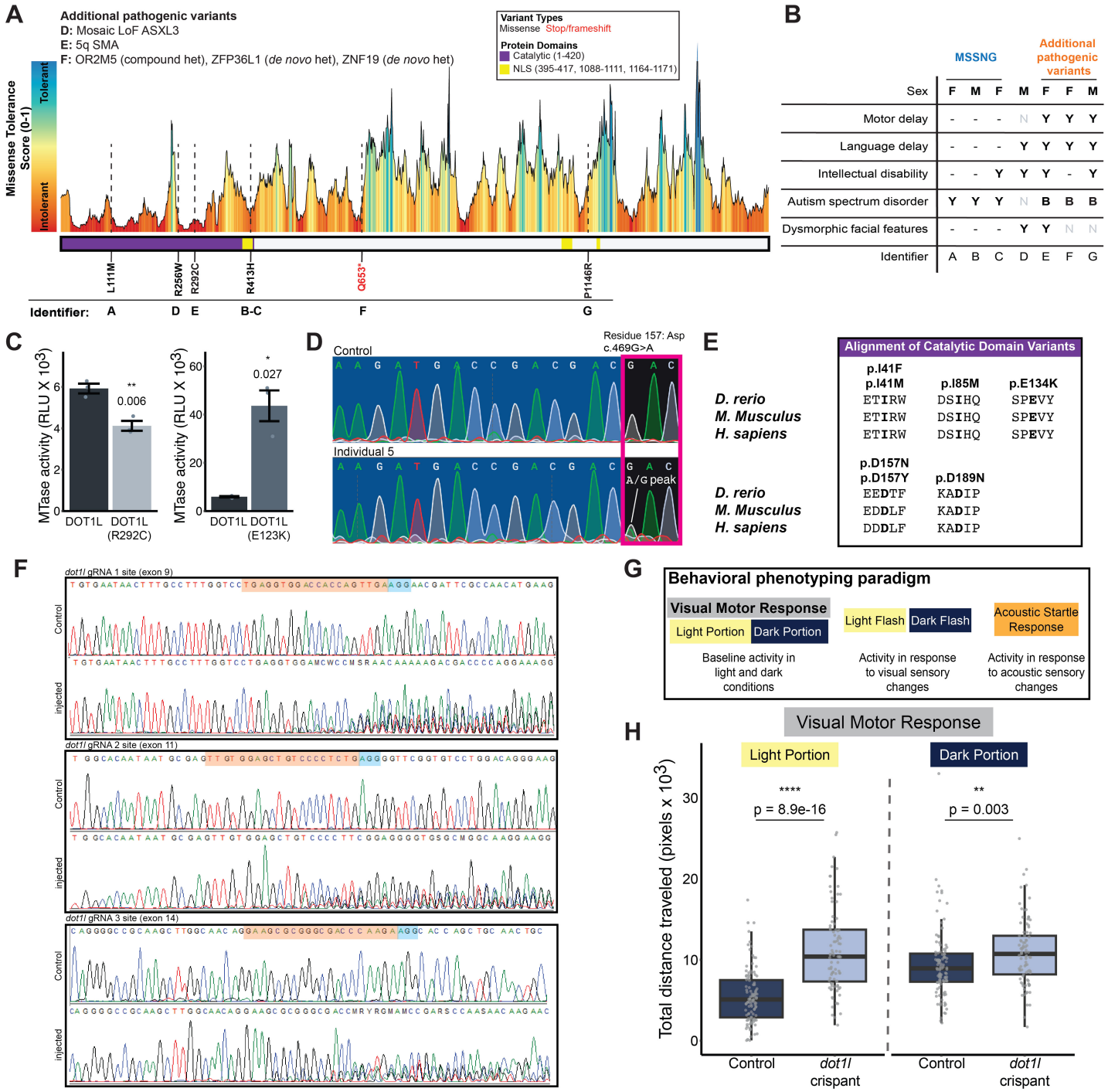
Funding: NIH NINDS grant 1F31NS129242 (MM). NIH/NICHD grant P50 HD105354, Intellectual and Developmental Disabilities Research Center (research support). NIH NINDS grant 1R01NS134755 (EK). NIH NIMH grant 1DP2MH129985 (MB, EK). NIH NIMH grant R00MH111836 (MB, EK). Klingenstein-Simons Fellowship from the Esther A. & Joseph Klingenstein Fund (MB, EK). Simons Foundation (MB, EK). Alfred P. Sloan Foundation Research Fellowship FG-2020-13529 (MB, EK). Brain and Behavior Research Foundation NARSAD Young Investigator Award (MB, EK). SickKids Research Institute (ARD). Azrieli Precision Child Health Platform (ARD). NIHR Manchester Biomedical Research Centre NIHR203308 (SB). MRC Epigenomics of Rare Diseases Node MR/Y008170/1 (SB). Miguel Servet program from Instituto de Salud Carlos III, Spain CP22/00141 (DNB). Netherlands Organisation for Scientific Research ZonMw Vidi, grant 09150172110002 (TSB). EpilepsieNL (TSB). CURE Epilepsy (TSB). NIH NINDS grant K08NS135125 (PDC). University of Pennsylvania Autism Spectrum Program of Excellence (ASPE) (JM, PDC). ANID-Chile Fondecyt grant #1211411 (GMR, VF). "Joan Oró" of the Secretary of Universities and Research of the Department of Research and Universities of the Government of Catalonia with code 2024 FI-1 00075 (BEA). European Union (BEA).

Author contributions: MM designed, performed, and analyzed most experiments. MB performed in vitro RNA-sequencing and immunocytochemistry. KL supported mouse work and analyzed behavioral tests and immunocytochemistry data. ARD gathered clinical information. PC and JM performed zebrafish experiments. RL performed methyltransferase activity experiments. AC performed p.D157N overexpression experiment. AP supported single nucleus RNA-sequencing work. VF, GMR, CM, ALS, CP, GMSM, RS, TSB, CMR, JL, IA, DNB, CO, BEA, FL, KC, AG, JL, XL, AV, AMI, XY, SB, KV, MJ, MK, PS, CIGM, SB, and JLM provided variant clinical information. MG led zebrafish experiments. KA led methyltransferase activity experiments. GC led clinical information compilation. EK led the project.

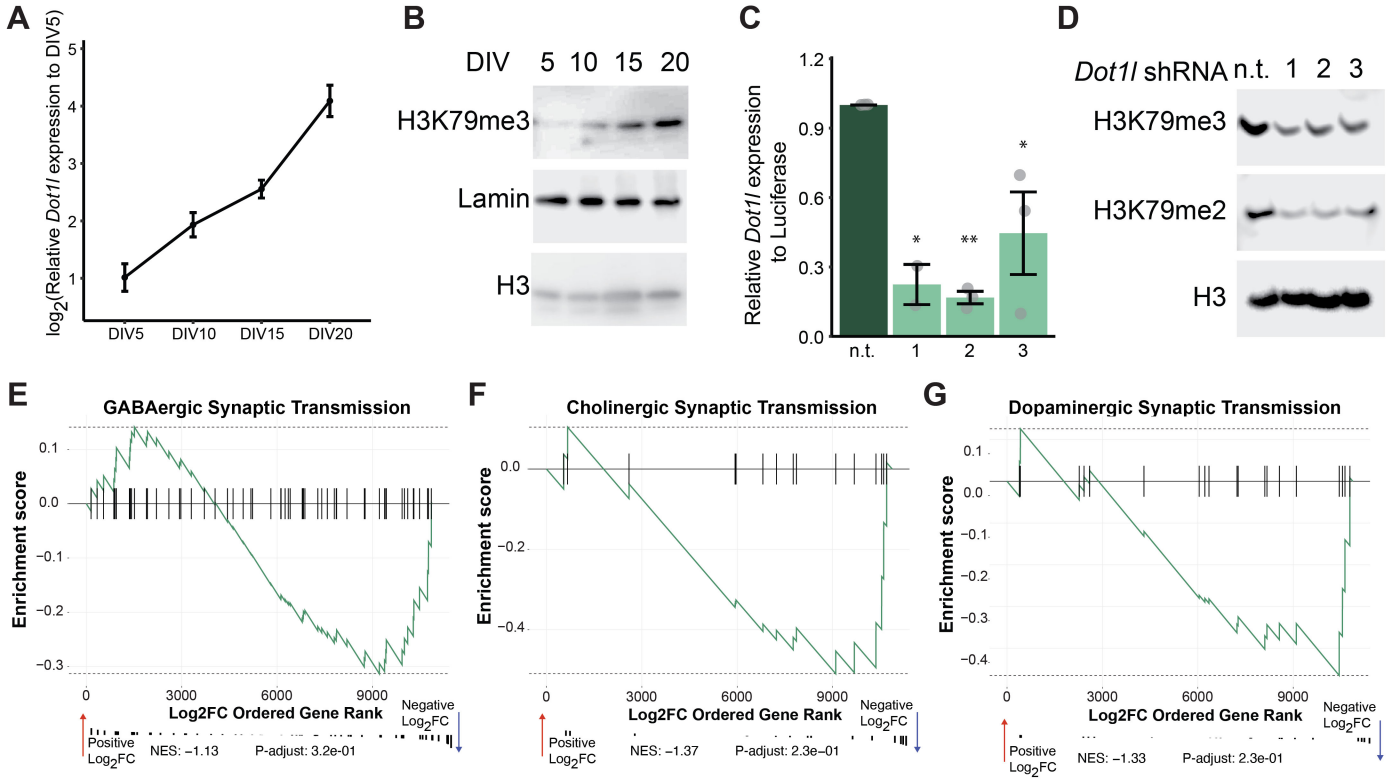
Competing interests:
KV has received honoraria as an advisory board member, travel expenses and speaker fees from Biogen, Santhera, Orchard, ITF and Novartis, outside the submitted work. JLM is an employee of and may own stock in GeneDx, LLC. All other authors declare they have no competing interests.

Data and materials availability:

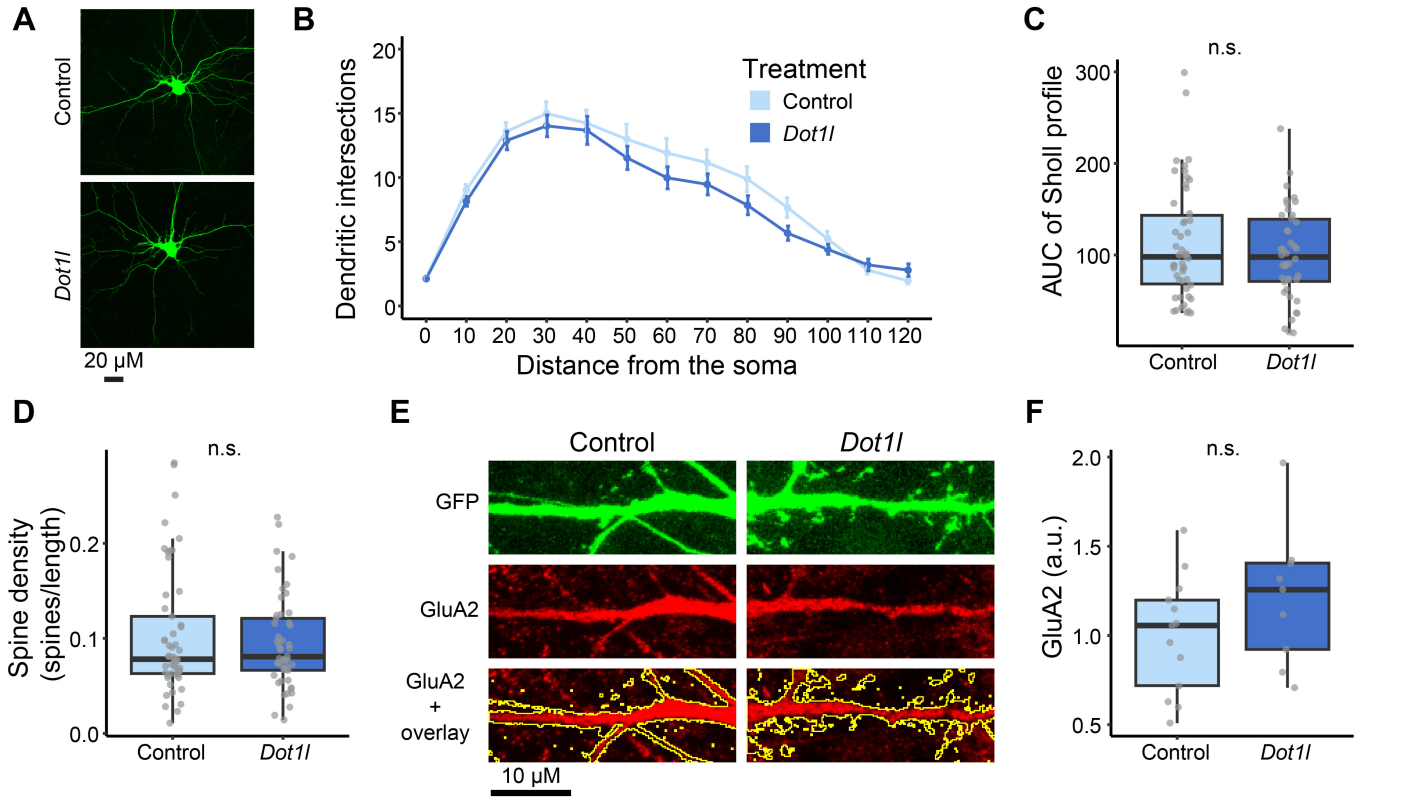
The exome/genome sequencing will be made available upon request provided that privacy and consent criteria are preserved. RNA-sequencing and single nucleus RNA-sequencing data generated in this study can be accessed under the following GEO accession number GSE279978. Variants p.I85M (SCV004169212), E134K (SCV004169195), and Gln598* (SCV003804054) are available on ClinVar. All data are available in the main text or the supplementary materials. Any additional data will be made available within two weeks upon request to corresponding author.



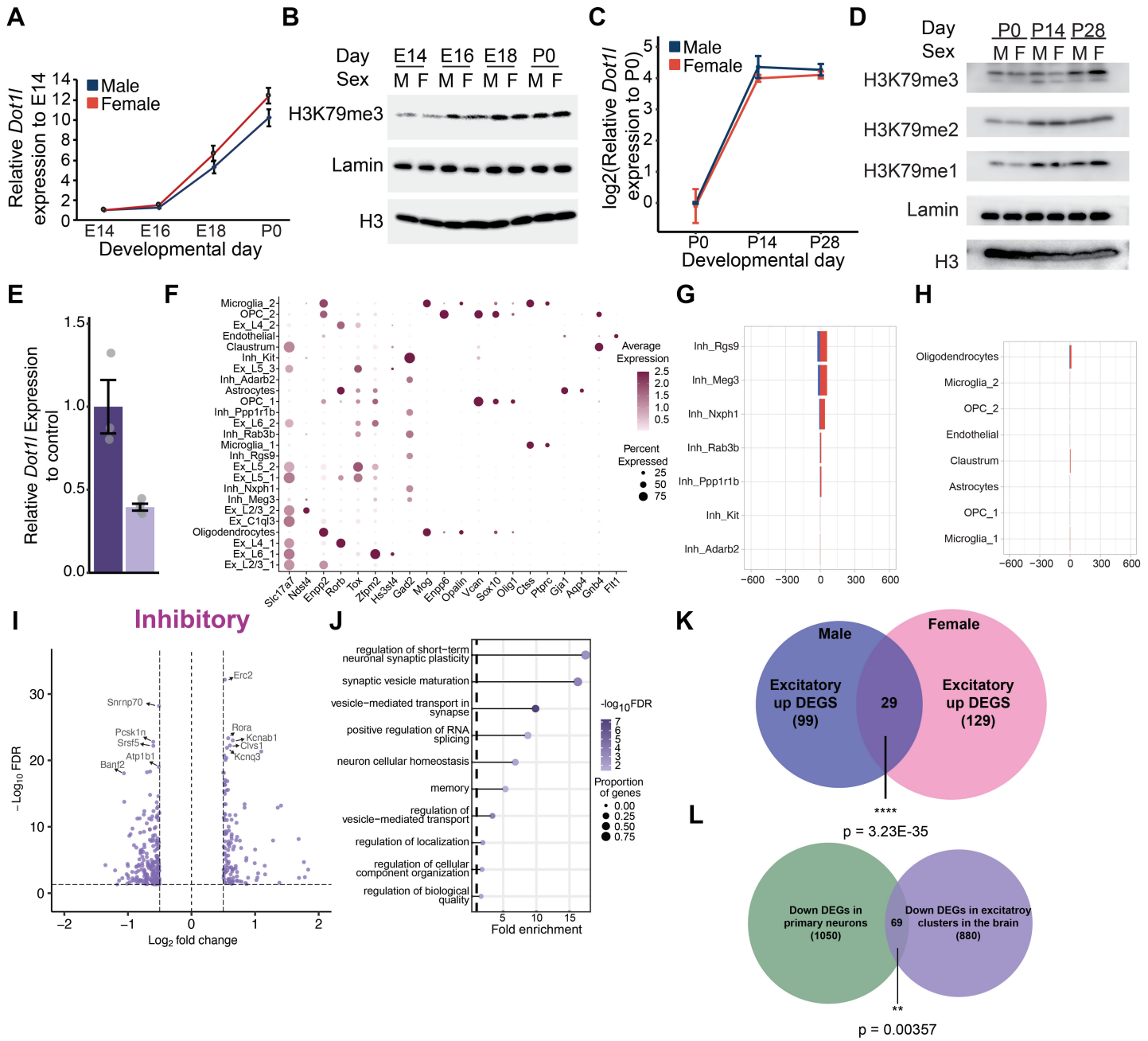
Supplementary Figure 1. (A) The landscape of missense tolerance of DOT1L from Metadome for individuals with ASD from MSSNG database, individuals with variants in the general population, or individuals harboring additional pathogenic variants. Schematic of DOT1L protein domains (purple = catalytic and yellow = nuclear localization signal) and locations of DOT1L variants (black = missense variant, red = stop/frameshift variant). (B) Clinical phenotypes of individuals with *DOT1L* variants from MSSNG, individuals with variants in the general population, or additional pathogenic variants (Y = present, B = borderline, N = not present, - = not reported). (C) Methyltransferase activity of human DOT1L or variants of DOT1L (R292C or E123K) on unmodified nucleosomes. Graph shows mean ± SE (n = 3/condition, unpaired two-tailed t-test). RLU = relative light units. (D) Sanger sequencing of human fibroblasts from individual 5 with variant *DOT1L* (D157N) and from age- and sex-matched control with wildtype *DOT1L*. (E) Alignment of catalytic domain variants to zebrafish Dot1l, mouse DOT1L, and human DOT1L using DIOPT. Individual variants are highlighted in bold. (F) Sanger sequence traces from injected animals with the three dot1l gRNAs used to generate crispants. (G) Schematic behavioral phenotyping paradigm in zebrafish *dot1l* crispants. (H) Total distance traveled with lights on and off during the visual motor response assay in *dot1l* crispant and control zebrafish (control: n = 107, *dot1l* crispant: n = 89; 3 independent experiments, Kruskal-Wallis Test). All box plot bounds indicate the 25th and 75th percentiles, the black line shows the median, and whiskers extend to the minimum and maximum value that are no further than 1.5 * interquartile range. *p<0.05, **p<0.01, ****p<0.0001.



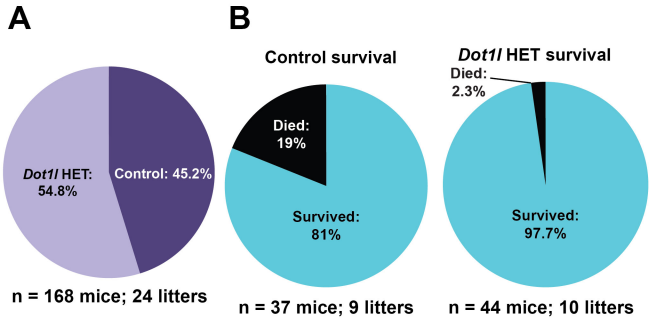
Supplementary Figure 2. (A) *Dot1l* expression in primary cortical neurons from 5-20 days in vitro (DIV) Graph shows mean ± SE. (n=6/DIV). (B) Representative western blot of H3K79me3 in primary cortical neurons from 5-20 DIV. (C) *Dot1l* expression in primary cortical neurons infected with 3 different *Dot1l* shRNAs or non-targeting control (n.t.). Graph shows mean ± SE. (n.t.: n = 3, *Dot1l* shRNA 1: n = 2, *Dot1l* shRNA 2: n = 3, *Dot1l* shRNA 3: n = 3, ANOVA with post-hoc pairwise t-test with Bonferroni correction). (D) Representative western blot of H3K79me2/3 in primary cortical neurons infected with *Dot1l* shRNA or n.t. control. (E-G) Gene set enrichment analysis of genes involved in (E) GABAergic, (F) cholinergic, (G) or dopaminergic synaptic transmission. NES indicates normalized enrichment score. *p<0.05, **p<0.01.



Supplementary Figure 3. (A) Representative images of primary cortical neurons transfected with *Dot1l* or empty vector control. Scale bar = 20 µm. (B) Number of branch intersections per radius of transfected primary cortical neurons (control: n = 50 neurons from 8 biological replicates, *Dot1l*: n = 42 neurons from 8 biological replicates). (C) Area under the curve (AUC) quantification of (B) (control: n = 50 neurons from 8 biological replicates, *Dot1l*: n = 42 neurons from 8 biological replicates, mixed effect model). (D) Spine density of transfected primary cortical neurons (control: n = 50 neurons from 8 biological replicates, *Dot1l*: n = 42 neurons from 8 biological replicates, Kruskal-Wallis Test). (E) Representative images of GluA2 staining in shRNA-transfected primary cortical neurons. (F) Quantification of GluA2 staining in shRNA-transfected primary cortical neurons. (control: n = 13 neurons from 2 biological replicates, *Dot1l* shRNA: n = 9 neurons from 2 biological replicates, Kruskal-Wallis Test). All box plot bounds indicate the 25th and 75th percentiles, the black line shows the median, and whiskers extend to the minimum and maximum value that are no further than 1.5 * interquartile range. n.s. = not significant.

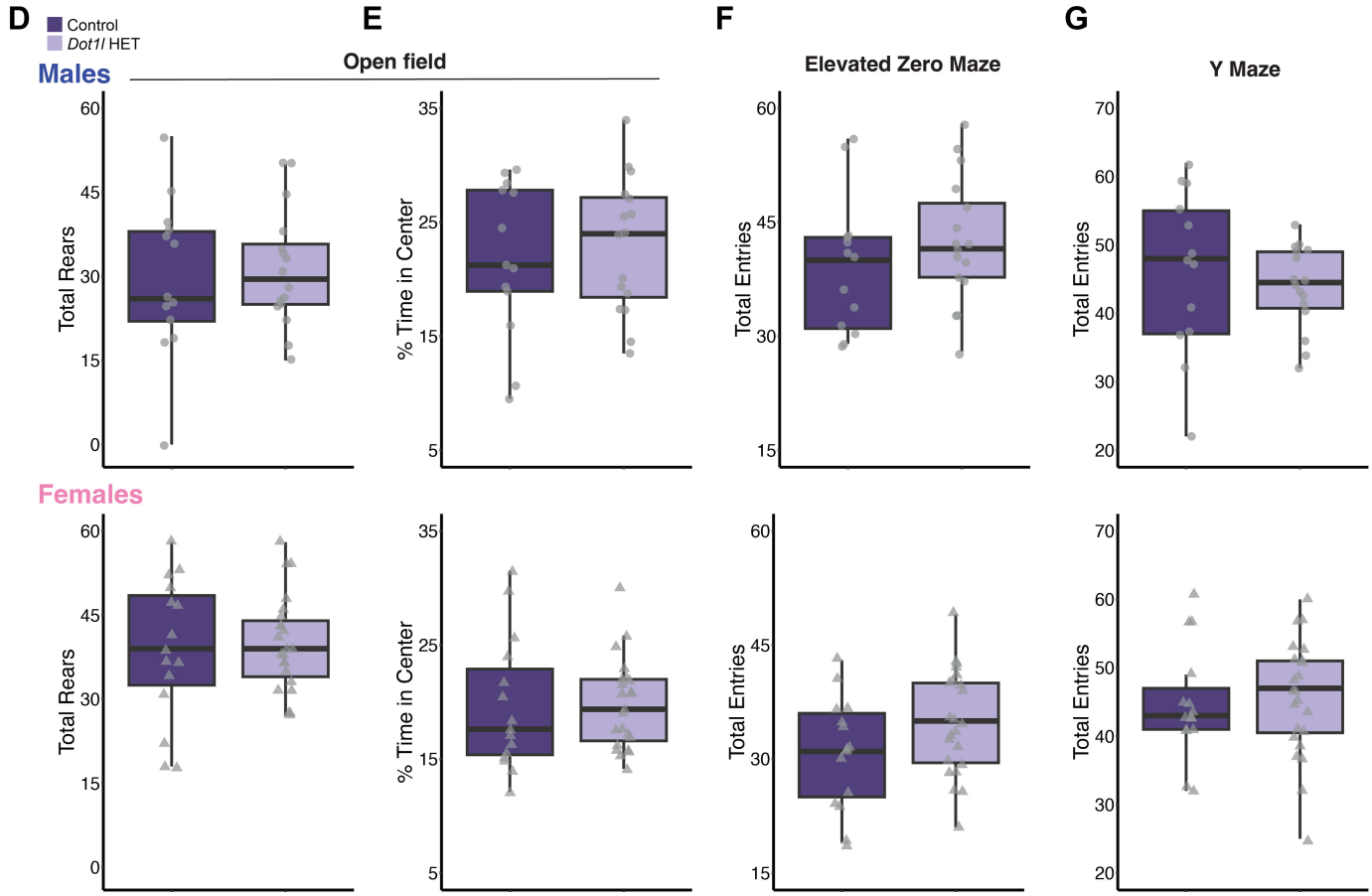


Supplementary Figure 4. (A) *Dot1l* expression in whole cortex from embryonic day 14 (E14) to postnatal day 0 (P0) in male and female mice. Graph shows mean ± SE (E14[male: n = 5, female n= 3], E16[male: n = 4, female: n = 4], E18[male: n = 3, female: n = 6], P0[male: n = 4, female: n = 5]). (B) Representative western of H3K79me3 levels in whole cortex from E14-P0 in male and female mice. (C) *Dot1l* expression in cortical nuclei from P0-P28 in male and female mice (P0[female: n = 2], all remaining groups n = 3/timepoint/sex). (D) Representative western of H3K79me1/2/3 in cortical nuclei from P0-P28 in male and female mice. (E) *Dot1l* expression in control and *Dot1l* HET whole cortex. Graph shows mean ± SE (control: n = 3 [1 male, 2 females]; *Dot1l* HET: n = 4 [2 males, 2 females]). (F) Dot plot showing expression of cell type-specific marker genes for each cluster. (G-H) Number of up- and down-regulated differentially expressed genes (DEGs) within inhibitory clusters (G) and glia, endothelial, and claustrum (H) clusters. (I) Volcano plot showing DEGs from *Dot1l* HET and control combined inhibitory clusters. (E) Biological process gene ontology analysis down-regulated DEGs from combined excitatory clusters. (J) Biological process gene ontology analysis down-regulated DEGs from combined inhibitory clusters. (K) Overlap of male and female up DEGs from combined excitatory clusters (hypergeometric test). (L) Overlap of down-regulated DEGs from primary neurons infected with *Dot1l* shRNA and down-regulated DEGs from combined excitatory clusters from *Dot1l* HET cortical nuclei (hypergeometric test). ** $p < 0.01$, **** $p < 0.0001$.

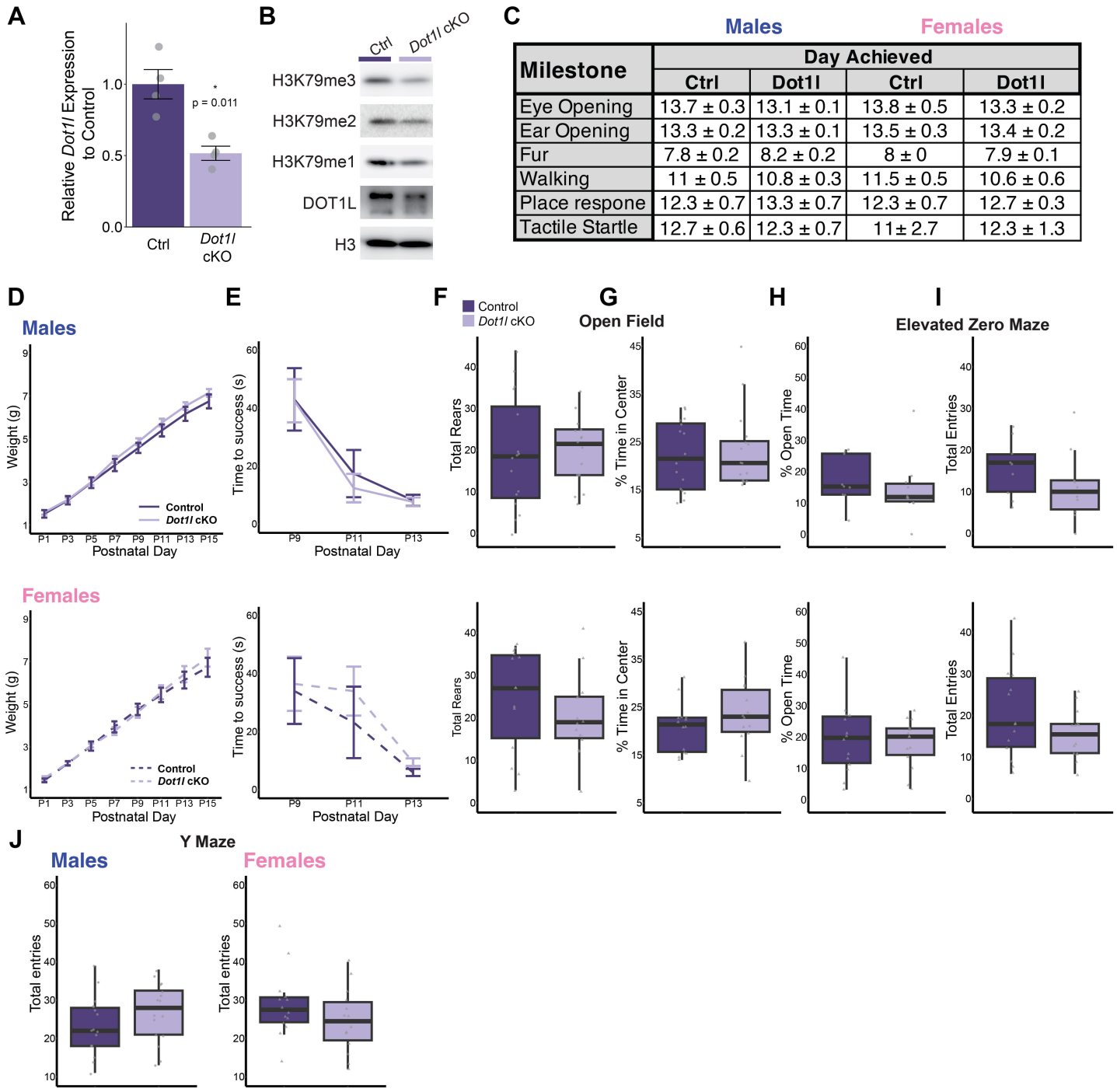


C

Milestone	Males: Day Achieved		Females: Day Achieved	
	Ctrl	Dot1l	Ctrl	Dot1l
Eye Opening	14.1 ± 0.3	14.2 ± 0.2	14.3 ± 0.2	14.2 ± 0.1
Ear Opening	15.1 ± 0.3	15.5 ± 0.4	14.8 ± 0.1	15 ± 0.2
Fur	13.3 ± 0.3	13.7 ± 0.2	13.5 ± 0.2	13.3 ± 0.2
Incisor Eruption	9.5 ± 0.2	10.5 ± 0.2	9.7 ± 0.3	9.5 ± 0.2
Walking	11.2 ± 0.2	11.3 ± 0.2	10.4 ± 0.3	10.9 ± 0.2
Grasp Reflex	6 ± 0.3	6.4 ± 0.3	6.5 ± 0.2	5.9 ± 0.2
Cliff avoidance	9.5 ± 0.5	10.5 ± 0.5	9.7 ± 0.6	9.5 ± 0.5
Place response	9.5 ± 0.3 *	10.5 ± 0.3	9.7 ± 0.4	9.5 ± 0.3
Tactile Startle	11.2 ± 0.4	11.4 ± 0.6	11.4 ± 0.6	11.8 ± 0.4



Supplementary Figure 5. (A) Pie chart of progeny from Dot1l$^{floxed/+}$ x CMV-Cre$^{+/+}$ crosses (n = 168 mice from 24 litters). (B) Survival rate at weaning age (approximately 4 weeks) in *Dot1l* HET and control mice (control: n = 37 mice from 9 litters, *Dot1l* HET: n = 44 mice from 10 litters). (C) Developmental milestone achievements in male and female mice (male[control: n = 13, *Dot1l* HET: n = 16]; female[control: n = 15, *Dot1l* HET: n = 23], unpaired two-tailed t-test). (D) Total number of rears during the open field assay in male and female mice. (males[control: n = 13, *Dot1l* HET: n = 16]; female[-control: n = 15, *Dot1l* HET: n = 23], unpaired two-tailed t-test). (E) Percent of time spent in the center of the open field arena in male and female mice (Kruskal-Wallis Test). (F) Total entries during the elevated zero maze in male and female mice (unpaired two-tailed t-test). (G) Total arm entries during the Y maze in male and female mice (unpaired two-tailed t-test). All box plot bounds indicate the 25th and 75th percentiles, the black line shows the median, and whiskers extend to the minimum and maximum value that are no further than 1.5 * interquartile range.



Milestone	Day Achieved			
	Ctrl	Dot1l	Ctrl	Dot1l
Eye Opening	13.7 ± 0.3	13.1 ± 0.1	13.8 ± 0.5	13.3 ± 0.2
Ear Opening	13.3 ± 0.2	13.3 ± 0.1	13.5 ± 0.3	13.4 ± 0.2
Fur	7.8 ± 0.2	8.2 ± 0.2	8 ± 0	7.9 ± 0.1
Walking	11 ± 0.5	10.8 ± 0.3	11.5 ± 0.5	10.6 ± 0.6
Place respone	12.3 ± 0.7	13.3 ± 0.7	12.3 ± 0.7	12.7 ± 0.3
Tactile Startle	12.7 ± 0.6	12.3 ± 0.7	11± 2.7	12.3 ± 1.3

Supplementary Figure 6. (A) *Dot1l* expression in whole cortex from embryonic day 14 (E14) to postnatal day 0 (P0) in male and female mice. Graph shows mean ± SE (E14[male: n = 5, female n= 3], E16[male: n = 4, female: n = 4], E18[male: n = 3, female: n = 6], P0[male: n = 4, female: n = 5]). (B) Representative western of H3K79me3 levels in whole cortex from E14-P0 in male and female mice. (C) Developmental milestone achievements in male and female mice (male[control: n = 6, *Dot1l* cKO: n = 11]; female[control: n = 4, *Dot1l* cKO: n = 9], unpaired two-tailed t-test). (D) Weight during first two weeks post-birth in male and female mice (repeated measures ANOVA). (E) Time to right self in negative geotaxis assay during development in male and female pups (repeated measures ANOVA). (F) Total number of rears during the open field assay in male and female mice (male[control: n = 16, *Dot1l* cKO: n = 16]; female[control: n = 14, *Dot1l* cKO: n = 14], unpaired two-tailed t-test). (G) Percent of time spent in the center of the open field arena in male and female mice (Kruskal-Wallis Test). (H) Percent of time spent in open arms of the elevated zero maze in male and female mice. (male[control: n = 9, *Dot1l* cKO: n = 10; female[control: n = 14, *Dot1l* cKO: n = 14], unpaired two-tailed t-test). (I) Total entries during the elevated zero maze in male and female mice (unpaired two-tailed t-test). (J) Total arm entries during the Y maze in male and female mice (unpaired two-tailed t-test). All box plot bounds indicate the 25th and 75th percentiles, the black line shows the median, and whiskers extend to the minimum and maximum value. * $p < 0.05$