

Title

NeuroTri2-VISDOT: An open-access tool to harness the power of second trimester human single cell data to inform models of Mendelian neurodevelopmental disorders

Authors

Kelly J. Clark^{1,2*}, Emily E. Lubin^{1,2*}, Elizabeth M. Gonzalez^{1,2}, Annabel K. Sangree^{1,2}, Dana E. Layo-Carris², Emily L. Durham², Rebecca C. Ahrens-Nicklas^{2,3}, Tomoki T. Nomakuchi²⁺, Elizabeth J. Bhoj^{2,3+}

Author Affiliations

¹ Biomedical Graduate School, University of Pennsylvania, Perelman School of Medicine

² Children's Hospital of Philadelphia

³ Department of Pediatrics, University of Pennsylvania, Perelman School of Medicine

* co-first authors

⁺ co-corresponding senior authors

Tomoki Nomakuchi – nomakuchit@chop.edu

Elizabeth Bhoj – bhoje@chop.edu

Abstract

Whole exome and genome sequencing, coupled with refined bioinformatic pipelines, have enabled improved diagnostic yields for individuals with Mendelian conditions and have led to the rapid identification of novel syndromes. For many Mendelian neurodevelopmental disorders (NDDs), there is a lack of pre-existing model systems for mechanistic work. Thus, it is critical for translational researchers to have an accessible phenotype- and genotype-informed approach for model system selection. Single-cell RNA sequencing data can be informative in such an approach, as it can indicate which cell types express a gene of interest at the highest levels across time. For Mendelian NDDs, such data for the developing human brain is especially useful. A valuable single-cell RNA sequencing dataset of the second trimester

developing human brain was produced by Bhaduri et al in 2021, but access to these data can be limited by computing power and the learning curve of single-cell data analysis. To reduce these barriers for translational research on Mendelian NDDs, we have built the web-based tool, Neurodevelopment in Trimester 2 - Visualization of Single cell Data Online Tool (NeuroTri2-VISDOT), for exploring this single-cell dataset, and we have employed it in several different settings to demonstrate its utility for the translational research community.

Intro

Mendelian neurodevelopmental disorders (NDDs) are single-gene conditions that impact the development and function of the brain¹. Forty percent of the more than 5,000 currently classified Mendelian disorders involve the brain or nervous system², and over 1,500 genes have been implicated in these conditions³. Collectively, an estimated 9-18% of children are impacted by NDDs globally³, with approximately 3% of children diagnosed with a Mendelian NDD¹. Every year more precise diagnoses are provided to families as about 300 new genes are identified as harboring causative variants underlying Mendelian NDDs⁴. Often, these novel NDDs fall into the classification of ultra-rare, impacting fewer than 1 in 50,000 individuals around the world⁵.

Putative disease candidate genes can be identified either through individual cases and small cohorts, or through the leveraging of large-scale publicly available datasets, such as gnomAD and GTEx⁶⁻⁸ (Figure 1A). Clinicians who provide care to these individuals can then connect with each other via resources like GeneMatcher and Matchmaker Exchange⁹⁻¹¹. Through this approach, teams of clinicians and researchers can identify larger cohorts who share causative variants in the same genes and build more robust phenotypic profiles to inform care. Model systems are often not readily available to interrogate the pathogenic mechanism of these conditions and, even when they exist, these models can be financially prohibitive to obtain and space-prohibitive to maintain. Further, clinically accessible tissue (CATs) from affected individuals, like blood and skin, are useful for mechanistic studies but their availability is limited by research participation, logistical barriers and cost¹² (Figure 1A). This lack of preexisting functional

work and limited access to CATs for targeted assays can be a barrier to providing prognostic information to families.

Thus, it is imperative for translational geneticists to have access to a systematic, phenotype-informed, genotype-first approach to select a model system. Human induced pluripotent stem cells (hiPSC) are powerful models, but it can be challenging to determine the most appropriate terminal lineage differentiation that recapitulates the pathogenesis of a particular ultra-rare Mendelian syndrome, especially if little to no functional work has been published for a gene of interest. Further, while model organisms like mice and zebrafish are incredibly powerful, especially for interrogating systemic effects of variants, obtaining and maintaining colonies of these organisms can be cost and time prohibitive. Additionally, there are fundamental differences in neurodevelopment between these organisms and humans that must be considered. Nonetheless, targeted exploration of specific cell types using *in vivo* systems can fill existing gaps in our current ability to differentiate hiPSCs to all the cell types that make up the human brain.

With the explosion of large-scale sequencing projects, an overwhelming number of datasets have been generated, and resources like the Chan Zuckerberg Initiative's CELLxGENE platform help make these data more accessible¹³. Identifying the dataset that best aligns with a specific research question can still be a challenge. To specifically bridge the computational chasm for translational research questions motivated by individuals with ultra-rare Mendelian NDDs, we have built the tool Neurodevelopment in Trimester 2 - Visualization of Single cell Data Online Tool (NeuroTri2-VISDOT). NeuroTri2-VISDOT references the powerful second trimester human fetal brain single-cell RNA-sequencing dataset generated by Bhaduri et al¹⁴, which is an important dataset for many reasons. First, these data cannot be widely generated, as access to fetal human brain tissue samples is limited. Additionally, generating and utilizing single cell RNA-sequencing data is a complex and computationally demanding process. The teams who produced these data have already overcome these two significant barriers to access. We sought to make these data more accessible so that the field of translational genetics can use them to responsibly inform selection of models of neurodevelopment, even without a robust computational background.

These second trimester data are an incredibly powerful window into human neurodevelopment (Figure 1B). The development of the human central nervous system (CNS) begins after gestational week 3 (GW3), following the closure of the neural tube, a process called neurulation (Figure 1B)¹⁵. During the first trimester, neural immune cells like native cells and microglia colonize the CNS and begin to proliferate prior to the formation of the blood brain barrier^{16–19} (Figure 1B, Table 1). Neurogenesis and gliogenesis also begin in the first trimester but, at this stage, a limited number of cell types are identifiable based on transcriptomic profiling, including radial glia (RG), neural progenitor cells (NPCs), and neurons²⁰. That stands in contrast to what can be delineated in the second trimester, where transient prenatal cell types, such as Cajal-Retzius cells, as well as many of the CNS cell populations that persist into post-natal life are captured (Figure 1B, gold box; Table 1). Processes essential to neurodevelopment also occur during the second trimester. For instance, RG undergo the human-specific process of becoming a physically discontinuous scaffold between GW16.5-17, which is not captured in commonly employed model systems of neurodevelopment, such as mice or zebrafish^{21,22}.

There are thousands of unique cell types in the human brain. However, identifying all such cell types using single-cell transcriptomic data remains a difficult task. Single cell RNA-sequencing has a high degree of uncertainty, likely due to sequence-specific RNA degradation and random sampling of lowly expressed genes. This results in many expression values being “zero”²³. Further, identification of cell types relies on unsupervised clustering followed by time-consuming manual annotation²³. Thus, rare cell types may be lumped into other clusters, cell types with variable expression patterns may be misidentified, and differentiating continuous cell types must be put into discrete categories. For example, at some point NPCs become neurons. This is a continuous process, but a line must be drawn between the cell types at a discrete point. Cell atlases detailing the expression profiles of different tissues and cell types will be a key resource for this part of single cell data analysis for all model organisms, but they are still a work in progress for human cell types.

The dataset harnessed in NeuroTri2-VISDOT profiles 10 brain structures and 6 neocortical areas. These data are resolved to 10 unique cell types by clustering each sample by cell type and sub-clustering

to identify more granular cell types¹⁴. We provide a primer on those cell populations to facilitate interpretation of the visualized data using NeuroTri2-VISDOT (Table 1), but these cells have been reviewed in more detail elsewhere^{16–19,21,22,24–35,36}. We provide NeuroTri2-VISDOT as a free web-based interface (<https://kellyclark.shinyapps.io/NeuroTri2-VISDOT/>) and as an open-access R Shiny application via GitHub (<https://github.com/kellyclark1/VISDOT>) to visualize the expression patterns of genes across cell types throughout the second trimester in the developing human brain.

In this paper, we employ NeuroTri2-VISDOT in several different settings to evaluate its utility for the translational research community that focuses on Mendelian NDDs. First, we interrogate the expression patterns of the causative genes linked to the subset of chromatinopathies associated with Rubinstein-Taybi Syndrome (RSTS) (OMIM #613684; 180849). We then apply NeuroTri2-VISDOT to the spectrum of syndromes categorized as Noonan Syndrome-like RASopathies, including Noonan Syndrome (NS) (OMIM #163950; #610733; #616559; #613224; #616564; #605275; #619745; #615355; #618624; #618499; #611553); NS with multiple lentigines (OMIM #151100); NS-like disorder with loose anagen hair (OMIM #607721, 617506); Cardiofaciocutaneous Syndrome (OMIM #115150, 615279, 615280); Costello Syndrome (OMIM #218040); CBL-related RASopathy (OMIM #613563); wide spectrum RASopathy (OMIM #615278, 609942); MAPK1-related RASopathy (OMIM # 619087); and newly described NS-like RASopathies caused by variants in CDC42 and YWHAZ^{37,38}. Finally, we employ NeuroTri2-VISDOT to an emerging class of NDDs caused by germline variants in histone genes^{6,39}. Using NeuroTri2-VISDOT, we demonstrate that single cell RNA sequencing data from the developing human brain can be used to motivate the evidence-based selection of relevant cell populations to explore the pathogenic mechanism underlying Mendelian NDDs. Additionally, we show how systematic visualization of the temporal gene expression profiles with NeuroTri2-VISDOT may help elucidate the neurobiology underlying the neuro-phenotypic heterogeneity of historically grouped Mendelian NDDs.

Results

Using NeuroTri2-VISDOT

The NeuroTri2-VISDOT web-based interface has been designed to maximize the utility of the Bhaduri et al dataset for translational researchers (Supplemental Video 1). In the search box, an investigator can input the gene names for as many genes as they are interested in comparing using the notation from Ensembl GRCh38 Release 110 (Supplemental Figure 1A). Multiple genes must be space-separated, and gene names are case-sensitive. Users can opt to select or deselect any combination of cell types they are interested in exploring. All 10 cell types will always be represented in dot plots, but investigators may wish to deselect a cell type that has a higher expression than the other cell types for the line plots of percent and average expression. For example, the gene *PAFAH1B1* (*LISI*) is associated with the structural Mendelian NDD classical lissencephaly, a disorder primarily associated with cytoskeletal reorganization during neuron migration⁴⁰. A large percent of second trimester Cajal-Retzius cells (~80%) express *PAFAH1B1* (Supplemental Figure 1B, top), which can obscure the percent of other cell types expressing the gene. By removing Cajal-Retzius cells from the line plots, an investigator could more closely interrogate the pattern of *PAFAH1B1* expression in the nine other cell types (Supplemental Figure 1B, bottom).

To enhance customizability, investigators can also specify the scale of the plots, which enables the standardization of axes across plots. The two scales that can be customized are those for average expression scaled and percent expression. Average expression scaled is the normalized expression values for that gene in that cell type, scaled to a mean of zero. Percent expression is the percentage of a particular cell type with at least one read mapping to the gene of interest⁴¹. Because genes can be expressed at a low level but still have large impacts on cell function and identity, we chose to focus on percent of cells expressing genes of interest rather than average expression of the gene in each cell in subsequent analyses but the ability to visualize both dimensions of expression is retained in NeuroTri2-VISDOT. Investigators are also able to download not only the individual plots but also the data visualized in the plots if they are interested in performing additional analyses.

Applying NeuroTri2-VISDOT to established NDD families

Chromatinopathies: A focus on Rubinstein-Taybi Syndrome (RSTS)

Chromatinopathies are caused by germline variants in genes that encode for epigenetic machinery: the readers, writers, erasers, and remodelers of the epigenome⁴². This group of Mendelian disorders is highly variable at both the genotypic and phenotypic level. Here, we focus on *CREBBP* and *EP300*, genes which encode histone acetyltransferases in which germline variants are causative for RSTS. Post-natal therapeutic intervention with histone deacetylase inhibitors rescues the markers of neurological dysfunction in mouse models of RSTS, emphasizing the value of more deeply understanding the pathogenesis of this disorder to better understand all chromatinopathies⁴². RSTS is a particularly elegant example of locus heterogeneity underlying phenotypic variability. Individuals who harbor variants in *EP300* tend to have less severe clinical features compared to those with *CREBBP* variants^{43–45}. For example, intellectual disability is typically more severe in individuals with causative *CREBBP* variants, as is autism or autistic behaviors and epilepsy^{43–46}.

This phenotypic severity seems to correlate with the second trimester single cell expression signatures (Figure 2A). Both genes show a peak in expression at GW18 in endothelial cells, reaching a percent expression of about 35% (Figure 2A, right, unfilled boxes). Interestingly, *CREBBP* is expressed in ~48% of Cajal-Retzius cells while *EP300* is only expressed in about 27% of this cell type. This difference in percent expression, specifically in Cajal-Retzius cells, may contribute to the more severe neurological phenotypes of individuals with *CREBBP*-related RSTS compared to individuals with *EP300*-related RSTS. Notably, neuroimaging findings in individuals with RSTS include partial/total agenesis or hypodysgenesis of the corpus callosum, and straight gyrus hypoplasia have also been observed, though the genic contributions have been incompletely delineated⁴⁷. The transcript patterns coupled with the neuroradiologic phenotypes suggest that taking a gene- and cell type-informed approach to interrogate the pathogenic mechanism underlying RSTS may be an appropriate choice to disentangle clinical heterogeneity.

RASopathies: A focus on Noonan Syndrome and Noonan Syndrome-like (NS/NS-like) RASopathies

RASopathies, like chromatinopathies, are a genotypically and phenotypically heterogeneous group of clinically overlapping Mendelian NDDs caused by variants affecting members and modulators of the RAS-MAPK cascade. Here, we focus on NS and NS-like RASopathies because of their universal classification as multisystem developmental disorders that often includes the CNS, which is not the case for non-NS-like RASopathies³⁸. The characteristic inclusion of a neurologic phenotype in NS and NS-like RASopathies makes this variable group of syndromes amenable to interrogation with this neurodevelopmental dataset.

In concordance with previously published work, we define NS-associated genes as *PTPN11*, *SOS1*, *SOS2*, *NRAS*, *LZTR1*, *SPRED2*, *RIT1*, *RRAS2*, *MRAS* and *RAF1*^{37,38} (Figure 2B, left). Of the NS-like RASopathies, we define the associated genes as: *PTPN11* associated with NS with multiple lentigines; *SHOC2* and *PPP1CB* associated with NS-like disorder with loose anagen hair; *BRAF*, *MAP2K1* and *MAP2K2* associated with cardiofaciocutaneous syndrome; *HRAS* associated with Costello Syndrome; and *CBL*, *KRAS*, *MAPK1*, *CDC42*, and *YWHAZ* associated with other RASopathies^{37,38} (Figure 2B, right). When we plot all NS/NS-like RASopathies together, it is challenging to identify any discernable patterns (Supplemental Figure 2). However, when we stratify based on the association of genes with NS compared to NS-like RASopathies, trends begin to emerge (Figure 2C, 2D). Nine of the 10 NS genes could be plotted (*RRAS2* was not able to be queried). Seven of these 9 genes have a peak at GW18 that transcends cell type (Figure 2C, left, unfilled boxes). Additionally, these same genes exhibit an upward trend in expression in both endothelial cells and Cajal-Retzius cells, which is consistent with what we observe when looking at the percent expression line plots (Figure 2C, right, pink and green). Taken together, there is a fairly consistent expression signature shared by genes that harbor causative variants in NS.

Compared to the NS genes, we observe heterogeneity in the second trimester brain expression of the NS-like RASopathy genes (Figure 2D). The first observation is that both the dot plots and the line plots for the newly described NS-like RASopathies associated with *CDC42* and *YWHAZ* do not follow the same trends as other NS-like RASopathy causative genes (Figure 2D, unfilled arrows). In general, the

average and percent expression of these two genes throughout the second trimester is consistently greater than what we observe for any other gene queried in this NS-like RASopathy analysis (Figure 2D). The expression profiles visualized through the percent expression line plots is distinct in comparison to the other NS/NS-like RASopathy genes (Figure 2D, right, unfilled arrows). It has been previously noted that the assignment of the conditions caused by variants in these genes to the RASopathy family is under debate³⁸. These data, along with other disparate features of *CDC42* and *YWHAZ*- related disorders, may support the reevaluation of the assignment of these conditions to the RASopathy family.

The value of looking at both the dot plots and the percent expression line plots is exemplified when probing the genes that underly NS-like disorder with loose anagen hair: *SHOC2* and *PPP1CB*. The dot plots for these genes look similar to each other, but distinct from other NS-like RASopathy genes (Figure 2D, left). However, the percent expression plots of these genes in different cell types across the second trimester are profoundly different. *SHOC2* has a bimodal peak in percent expression in Cajal-Retzius cells at GW18 and GW20 (Figure 2D, right, gray arrows). Conversely, *PPP1CB* has a single peak in percent expression at GW20 in Cajal-Retzius cells, as well as a peak in expression at GW16 in forebrain RG cells. As with the case of *CREBBP* or *EP300*-derived RSTS, these data may support the use of different model systems, for example Cajal-Retzius cells and/or forebrain RG, to study NS-like disorder with loose anagen hair driven by causative variants in different genes. Investigating pathogenic mechanisms through a bifurcated approach that centers on multiple cell types could enable the identification of a shared, downstream perturbation that could be targeted through therapeutic intervention.

Strikingly, the peak in expression at GW16 in forebrain RG cells that we observe in *PPP1CB*-related NS-like disorder with loose anagen hair is also present in percent expression plots for *MAP2K2*-related cardiofaciocutaneous syndrome, *HRAS*-related Costello Syndrome and *KRAS*-related wide spectrum RASopathy (Figure 2D, right, black arrows). These bimodal peaks in expression transcend NS-like RASopathy delineations; gene specification subgroups defined by the ClinGen RASopathy Expert Panel, in which genes with similar function and/or structure have been grouped⁴⁸ (Supplemental Figure

3); and genes that share a gain-of-function disease mechanism (Supplemental Figure 4). Analysis of these expression signatures may prove useful for geneticists when making classification decisions about whether multiple syndromes should be grouped together in one family or into distinct entities.

Applying NeuroTri2-VISDOT to an emerging class of NDDs

Germline variants in genes encoding histones are an emerging class of Mendelian NDDs³⁹, which have recently been classified by OMIM. HIST1H1E syndrome/Rahman Syndrome (OMIM #617537) is caused by germline variants in the gene *HIST1H1E/H1-4* which encodes the histone H1 linker protein. Bryant-Li-Bhoj Syndrome (OMIM #619721, #619720) is caused by germline variants in *H3-3A* and *H3-3B*, the genes that encode histone H3.3. Tessadori-Bicknell-van Haaften NDD (OMIM #619758, #619759, #619950, #619551) is caused by germline variants in the genes *H4C3*, *H4C11*, *H4C5*, and *H4C9*, the genes that encode H4. Additional histone-encoding genes, including *MACROH2A1*, *H2AZ1*, *MACROH2A2*, *H2AZ2*, *H2AX*, and *H1-0*, are predicted to be putative disease candidates⁶.

The genes linked to HIST1H1E syndrome and Tessadori-Bicknell-van Haaften NDD are classified as replication-coupled (RC) histones, which are associated with unique structural and functional features. RC histone genes encode the only known cellular mRNA transcripts that are not poly-adenylated⁶. This lack of a polyA tail on RC histone transcripts renders them undetectable in the most common type of library preparation method used for RNA-sequencing. This means that in most publicly available datasets, including the one employed here, we are unable to explore their expression. Nonetheless, we are able to interrogate the expression of replication-independent (RI) histones that do have polyA tails. Based on prior work, we can stratify these RI histones into known NDD-causing, predicted NDD-causing, and non-disease-causing groups⁶ (Figure 3A).

When we plot the expression of these RI histones, all the genes known or predicted to be associated with Mendelian NDDs show increases in expression across the second trimester, compared to genes not known or predicted to be associated with NDDs, with the exception of *H1-10* (Figures 3B-D). Additionally, these genes seem to have a peak in expression at GW16 irrespective of cell type (Figure 3B-

D, unfilled boxes). Interestingly, this is the window during which RG become physically discontinuous^{21,22}. This expression pattern suggests an importance for RI histones at this very early point in brain development, and that RG would be an effective model to study the phenotypes associated with these variants. Neuroimaging results from individuals with HIST1H1E Syndrome, Bryant-Li-Bhoj Syndrome and Tessadori-Bicknell-van Haaften NDD do not show salient dysregulation of neural migration or cortical development. Thus, the role of RG in the neuropathology of these syndromes has not been explored. Through the visualization of these data using NeuroTri2-VISDOT, we identify a novel terminal lineage to which hiPSCs can be differentiated for subsequent functional work. More broadly, this demonstrates the utility of single cell expression signatures in development for identifying and supporting NDD candidate genes.

Discussion

Clinical genetics is rooted in a rich history of clinical phenotyping that long predates the field's ability to perform diagnostic genetic testing. In some cases, this has led to phenotypically similar syndromes being grouped together under large umbrella characterizations, such as NDDs or leukodystrophies, that may not reflect the distinct genetic processes contributing to the pathophysiology¹. Now, the field of translational genetics is reckoning with this same question: to lump disorders together or split them into functional groups⁴⁹. With NeuroTri2-VISDOT, we introduce a tool that may prove a valuable resource in this pursuit, with utility demonstrated for both established and emerging NDDs.

Single-cell RNA sequencing is an incredibly powerful method for exploring tissue heterogeneity and gene expression across cell types. However, generating a single-cell dataset is expensive and availability of human tissues, especially human fetal tissue, is limited. Further, data analysis requires computational skills and high-performance computing due to the size and complexity of the multidimensional data produced in these experiments. NeuroTri2-VISDOT seeks to address the computational barriers to use of this valuable tool, particularly the size of single-cell data and the computational skills needed to explore it. First, the complete dataset utilized here is available as a 50GB file, which is too large for many

researchers to store and use on their own computers necessitating the use of high-performance computing. NeuroTri2-VISDOT stores a summary of this dataset online, with average scaled expression and percent expression for each gene by cell type rather than read counts for each cell. No downloads are necessary to use the web app, and the raw data for genes of interest can be downloaded as a small file of comma-separated values, reducing the need to download the entire dataset. Second, analysis of single cell data using the R package, Seurat, requires many researchers to dedicate significant time to learning how to navigate R and Seurat objects. NeuroTri2-VISDOT reduces the need for coding expertise by allowing exploration of single cell data with a user-friendly interface that requires only a gene or list of genes as input.

We recognize some limitations associated with our approach. First, this dataset was generated from single-cell transcriptomics performed on microdissected regions of the brain that included 10 major forebrain, midbrain and hindbrain regions in addition to 6 neocortical areas, meaning that only cell types identified in the neocortical samples were powered to be reported in the final dataset¹⁴. Thus, some cellular populations important to neurodevelopment, such as cells of the cerebellum, are not represented. Astrocytes, which differentiate from RG after the peak of neurogenesis at the end of the second trimester, are also not represented. These cells play an important homeostatic role in the brain parenchyma and form the neural component of the blood brain barrier. However, there are thousands of unique cell types in the human brain, and it is not feasible to capture them all. Further, while these data provide a powerful window into neurodevelopment by quantifying gene expression in the second trimester, earlier and later stages of prenatal brain development are not captured here. Nonetheless, NeuroTri2-VISDOT is intended not as an atlas of all cells in all stages of the developing brain, but rather as a tool to efficiently leverage this powerful dataset in translational genetic research.

Another limitation of this analysis is the interdependence and interconnectivity of different cell types during neurodevelopment. For instance, GABAergic interneurons and oligodendrocyte precursor cells require the endothelial-lined vasculature to appropriately migrate. Additionally, NPCs not only arise from but also migrate along RG before eventually differentiating into glutamatergic cortical neurons. This

complexity, in part, motivates some investigators to opt for model organisms, such as mice or zebrafish as opposed to hiPSCs. The expression visualization made possible by NeuroTri2-VISDOT captures this complexity, but also enables investigators to make evidence-based decisions about model selection. For instance, if RG are nominated as a compelling model to interrogate, it is crucial to consider that this cell type undergoes the process of becoming a physically discontinuous scaffold during the middle of the second trimester, which may support the use of hiPSCs over mice and zebrafish in specific situations. These data may also indicate that different cell populations are worth interrogating based on which gene harbors variants driving the phenotype, as in the case of *CREBBP*- versus *EP300*-driven RSTS.

A biological consideration before employing these data is whether the RNA transcript is the appropriate read out for a given gene of interest. Transcript and protein levels are discordant throughout neurodevelopment, most prominently in post-natal life⁵⁰. When exploring post-natal neurodevelopment, protein or phosphorylated protein may be a more precise metric to inform model selection. In cases where transcript level expression is an appropriate biologic read-out, such as in pre-natal neurodevelopment, we propose that there are several ways in which NeuroTri2-VISDOT could be applied in the future (Figure 4). At the level of an individual translational research lab, we envision NeuroTri2-VISDOT as a tool to enable investigators to systemically inform the selection of their model systems. We see this approach being adapted by bioinformaticians to create similar tools for the increasingly available datasets designed to ask similarly targeted questions. For example, a similar tool to visualize neural gene expression across cell types across the lifespan could be employed by biobanks to identify and cluster gene expression signatures of putative disease candidate genes.

In summary, the numerous applications of NeuroTri2-VISDOT, including model selection and interrogation of gene signatures across time and cell types, allow for translational genetics researchers to harness the power of a valuable single-cell second trimester brain dataset without requiring programming experience or access to high performance computing resources. The approach employed to generate and evaluate NeuroTri2-VISDOT also provides a framework for developing tools for similar datasets that will be beneficial to the wider translational research community.

Methods

Data extraction

The Seurat object for the data generated by Bhaduri et al. is publicly available^{14,41}. This object was split into separate objects for each cell type, and plots were generated for genes of interest for each cell type using the Seurat DotPlot function. The data in these dot plots were combined to create a large table containing each gene with the percent expression and average expression scaled at each time point in each cell type. This data is available to download from the NeuroTri2-VISDOT web app. Additionally, the code used to generate NeuroTri2-VISDOT can be found at <https://github.com/kellyclark1/VISDOT>.

NeuroTri2-VISDOT Web app development

The NeuroTri2-VISDOT web app was written in R v4.2.2⁵¹ using the R package shiny v1.7.3⁵² and is available at <https://kellyclark.shinyapps.io/NeuroTri2-VISDOT/>. Plots displayed on the web interface were designed using ggplot2 v3.4.0⁵³.

The web app requires as input at least one human Ensembl GRCh38 Release 110 gene name and will accept as many gene names as a user would like to compare. Users may also select cell types of interest for line plots using the checkboxes in the left panel. The web app allows custom scaling for plots, which should be used with caution. The intent of custom scaling is to allow multiple genes to be plotted separately but at the same scale. However, setting limits that do not include the data will exclude data points outside of those limits.

REFERENCES

1. Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends in Neurosciences*. 2020;43(8):608-621. doi:10.1016/j.tins.2020.05.004
2. Rexach J, Lee H, Martinez-Agosto JA, Németh AH, Fogel BL. Clinical application of next-generation sequencing to the practice of neurology. *The Lancet Neurology*. 2019;18(5):492-503. doi:10.1016/S1474-4422(19)30033-X
3. Leblond CS, Le TL, Malesys S, et al. Operative list of genes associated with autism and neurodevelopmental disorders based on database review. *Molecular and Cellular Neuroscience*. 2021;113:103623. doi:10.1016/j.mcn.2021.103623
4. Bamshad MJ, Nickerson DA, Chong JX. Mendelian Gene Discovery: Fast and Furious with No End in Sight. *The American Journal of Human Genetics*. 2019;105(3):448-455. doi:10.1016/j.ajhg.2019.07.011
5. Smith CIE, Bergman P, Hagey DW. Estimating the number of diseases – the concept of rare, ultra-rare, and hyper-rare. *iScience*. 2022;25(8):104698. doi:10.1016/j.isci.2022.104698
6. Lubin E, Bryant L, Aicher J, Li D, Bhoj E. Analysis of histone variant constraint and tissue expression suggests five potential novel human disease genes: H2AFY2, H2AFZ, H2AFY, H2AFV, H1F0. *Hum Genet*. 2022;141(8):1409-1421. doi:10.1007/s00439-022-02432-1
7. Gudmundsson S, Singer-Berk M, Watts NA, et al. Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation*. 2022;43(8):1012-1030. doi:10.1002/humu.24309
8. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653

9. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Human Mutation*. 2015;36(10):928-930. doi:10.1002/humu.22844
10. Hamosh A, Wohler E, Martin R, et al. The impact of GeneMatcher on international data sharing and collaboration. *Human Mutation*. Published online March 28, 2022:humu.24350. doi:10.1002/humu.24350
11. Azzariti DR, Hamosh A. Genomic Data Sharing for Novel Mendelian Disease Gene Discovery: The Matchmaker Exchange. *Annu Rev Genom Hum Genet*. 2020;21(1):305-326. doi:10.1146/annurev-genom-083118-014915
12. Aicher JK, Jewell P, Vaquero-Garcia J, Barash Y, Bhoj EJ. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genetics in Medicine*. 2020;22(7):1181-1190. doi:10.1038/s41436-020-0780-y
13. CZI Single-Cell Biology Program, Abdulla S, Aevertmann B, et al. *CZ CELL×GENE Discover: A Single-Cell Data Platform for Scalable Exploration, Analysis and Modeling of Aggregated Data*. Cell Biology; 2023. doi:10.1101/2023.10.30.563174
14. Bhaduri A, Sandoval-Espinosa C, Otero-Garcia M, et al. An atlas of cortical arealization identifies dynamic molecular signatures. *Nature*. 2021;598(7879):200-204. doi:10.1038/s41586-021-03910-8
15. Greene NDE, Copp AJ. Neural Tube Defects. *Annu Rev Neurosci*. 2014;37(1):221-242. doi:10.1146/annurev-neuro-062012-170354
16. Prinz M, Erny D, Hagemeyer N. Ontogeny and homeostasis of CNS myeloid cells. *Nat Immunol*. 2017;18(4):385-392. doi:10.1038/ni.3703

17. Herz J, Filiano AJ, Wiltbank AT, Yogev N, Kipnis J. Myeloid Cells in the Central Nervous System. *Immunity*. 2017;46(6):943-956. doi:10.1016/j.immuni.2017.06.007
18. McNamara NB, Munro DAD, Bestard-Cuche N, et al. Microglia regulate central nervous system myelin growth and integrity. *Nature*. 2023;613(7942):120-129. doi:10.1038/s41586-022-05534-y
19. Menassa DA, Gomez-Nicola D. Microglial Dynamics During Human Brain Development. *Front Immunol*. 2018;9:1014. doi:10.3389/fimmu.2018.01014
20. Eze UC, Bhaduri A, Haeussler M, Nowakowski TJ, Kriegstein AR. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat Neurosci*. 2021;24(4):584-594. doi:10.1038/s41593-020-00794-1
21. Nowakowski TJ, Pollen AA, Sandoval-Espinosa C, Kriegstein AR. Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. *Neuron*. 2016;91(6):1219-1227. doi:10.1016/j.neuron.2016.09.005
22. LaMonica BE, Lui JH, Wang X, Kriegstein AR. OSVZ progenitors in the human cortex: an updated perspective on neurodevelopmental disease. *Current Opinion in Neurobiology*. 2012;22(5):747-753. doi:10.1016/j.conb.2012.03.006
23. Lähnemann D, Köster J, Szczurek E, et al. Eleven grand challenges in single-cell data science. *Genome Biol*. 2020;21(1):31. doi:10.1186/s13059-020-1926-6
24. Daneman R, Prat A. The Blood–Brain Barrier. *Cold Spring Harb Perspect Biol*. 2015;7(1):a020412. doi:10.1101/cshperspect.a020412
25. Obermeier B, Daneman R, Ransohoff RM. Development, maintenance and disruption of the blood-brain barrier. *Nat Med*. 2013;19(12):1584-1596. doi:10.1038/nm.3407

26. Crouch EE, Bhaduri A, Andrews MG, et al. Ensembles of endothelial and mural cells promote angiogenesis in prenatal human brain. *Cell*. 2022;185(20):3753-3769.e18. doi:10.1016/j.cell.2022.09.004
27. Ihara M, Yamamoto Y. Transcriptomic mapping of the human cerebrovasculature. *Nat Rev Neurol*. 2022;18(6):319-320. doi:10.1038/s41582-022-00650-9
28. Fujioka T, Kaneko N, Sawamoto K. Blood vessels as a scaffold for neuronal migration. *Neurochemistry International*. 2019;126:69-73. doi:10.1016/j.neuint.2019.03.001
29. Tsai HH, Niu J, Munji R, et al. Oligodendrocyte precursors migrate along vasculature in the developing nervous system. *Science*. 2016;351(6271):379-384. doi:10.1126/science.aad3839
30. Bergles DE, Richardson WD. Oligodendrocyte Development and Plasticity. *Cold Spring Harb Perspect Biol*. 2016;8(2):a020453. doi:10.1101/cshperspect.a020453
31. Barber M, Pierani A. Tangential migration of glutamatergic neurons and cortical patterning during development: Lessons from Cajal-Retzius cells. *Developmental Neurobiology*. 2016;76(8):847-881. doi:10.1002/dneu.22363
32. Elorriaga V, Pierani A, Causeret F. Cajal-retzius cells: Recent advances in identity and function. *Current Opinion in Neurobiology*. 2023;79:102686. doi:10.1016/j.conb.2023.102686
33. Causeret F, Moreau MX, Pierani A, Blanquie O. The multiple facets of Cajal-Retzius neurons. *Development*. 2021;148(11):dev199409. doi:10.1242/dev.199409
34. Olynik BM, Rastegar M. The Genetic and Epigenetic Journey of Embryonic Stem Cells into Mature Neural Cells. *Front Gene*. 2012;3. doi:10.3389/fgene.2012.00081

35. Moretto E, Murru L, Martano G, Sassone J, Passafaro M. Glutamatergic synapses in neurodevelopmental disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2018;84:328-342. doi:10.1016/j.pnpbp.2017.09.014
36. Tang X, Jaenisch R, Sur M. The role of GABAergic signalling in neurodevelopmental disorders. *Nat Rev Neurosci*. 2021;22(5):290-307. doi:10.1038/s41583-021-00443-x
37. Tartaglia M, Aoki Y, Gelb BD. The molecular genetics of RASOPATHIES : An update on novel disease genes and new disorders. *American J of Med Genetics Pt C*. 2022;190(4):425-439. doi:10.1002/ajmg.c.32012
38. Zenker M. Clinical overview on RASOPATHIES. *American J of Med Genetics Pt C*. 2022;190(4):414-424. doi:10.1002/ajmg.c.32015
39. Knapp K, Naik N, Ray S, Van Haaften G, Bicknell LS. Histones: coming of age in Mendelian genetic disorders. *J Med Genet*. 2023;60(3):1-10. doi:10.1136/jmg-2022-109085
40. Dixon-Salazar TJ, Gleeson JG. Genetic regulation of human brain development: lessons from Mendelian diseases. *Annals of the New York Academy of Sciences*. 2010;1214(1):156-167. doi:10.1111/j.1749-6632.2010.05819.x
41. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-3587.e29. doi:10.1016/j.cell.2021.04.048
42. Bjornsson HT. The Mendelian disorders of the epigenetic machinery. *Genome Res*. 2015;25(10):1473-1481. doi:10.1101/gr.190629.115
43. Hamilton MJ, Newbury-Ecob R, Holder-Espinasse M, et al. Rubinstein–Taybi syndrome type 2: report of nine new cases that extend the phenotypic and genotypic spectrum. *Clinical Dysmorphology*. 2016;25(4):135-145. doi:10.1097/MCD.0000000000000143

44. López M, García-Oguiza A, Armstrong J, et al. Rubinstein-Taybi 2 associated to novel EP300 mutations: deepening the clinical and genetic spectrum. *BMC Med Genet*. 2018;19(1):36. doi:10.1186/s12881-018-0548-2
45. Van Gils J, Magdinier F, Fergelot P, Lacombe D. Rubinstein-Taybi Syndrome: A Model of Epigenetic Disorder. *Genes*. 2021;12(7):968. doi:10.3390/genes12070968
46. Fergelot P, Van Belzen M, Van Gils J, et al. Phenotype and genotype in 52 patients with Rubinstein-Taybi syndrome caused by *EP300* mutations. *American J of Med Genetics Pt A*. 2016;170(12):3069-3082. doi:10.1002/ajmg.a.37940
47. Geraldo AF, Rossi A, Severino M. Malformations of Cortical Development. In: *Pediatric Neuroradiology*. Springer, Berlin, Heidelberg; 2021.
48. Gelb BD, Cavé H, Dillon MW, et al. ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. *Genetics in Medicine*. 2018;20(11):1334-1345. doi:10.1038/gim.2018.3
49. Thaxton C, Goldstein J, DiStefano M, et al. Lumping versus splitting: How to approach defining a disease to enable accurate genomic curation. *Cell Genomics*. 2022;2(5):100131. doi:10.1016/j.xgen.2022.100131
50. Breen MS, Ozcan S, Ramsey JM, et al. Temporal proteomic profiling of postnatal human cortical development. *Transl Psychiatry*. 2018;8(1):267. doi:10.1038/s41398-018-0306-4
51. R Core Team. R: A language and environment for statistical computing. Published online 2021. <https://www.R-project.org/>
52. Chang W, Cheng J, Allaire J, et al. shiny: Web Application Framework for R. Published online November 17, 2023. <https://github.com/rstudio/shiny> ; <https://shiny.posit.co/>

484 53. Wickham H, Chang W, Henry L, et al. ggplot2: Elegant Graphics for Data Analysis. Published online
485 2016. <https://ggplot2.tidyverse.org>

486

487

Figure Legends

Figure 1 NeuroTri2-VISDOT fills key gap to inform selection of Mendelian NDD models. A)

Proposed systematic approach to go from identification of putative Mendelian NDD gene to cohort identification to selection of the most useful model for downstream mechanistic work. Black box represents the gap filled by NeuroTri2-VISDOT. B) Timeline of human prenatal neurodevelopment including first- (gestational weeks (GW) 1-13; light gray), second- (GW 14-27; gold), and third-trimester (GW 28-40; gray), as well as the beginning of postnatal neurodevelopment (charcoal). All boxes/lines that extend past GW40 indicate continuance into postnatal life. Unfilled gold box highlights the second trimester of neurodevelopment, including the cell types present and the ongoing neurodevelopmental processes.

Table 1 Primer on transcriptomically-delineated cell types in Bhaduri et al dataset.

Figure 2 NeuroTri2-VISDOT captures genotypic and phenotypic heterogeneity of established

NDDs. (A,C-D) Dot plots (left) – the average expression scaled reflects the normalized expression value for a gene, scaled to zero, and depicted by color gradient; the percent expression reflects the percentage of a particular cell type with at least one read mapping to the gene of interest, depicted by size gradient. Line plots (right) – percent expression by cell type. A) Dot plots (left) and percent expression line plots (right) for *CREBBP* and *EP300* across 10 neural cell types and 8 second-trimester time points. Unfilled boxes highlight GW18, emphasized in text. B) Table of genes associated with NS (left) and NS-like RASopathies (right). C) Dot plots (left) and percent expression line plots (right) for NS genes. Unfilled boxes highlight GW18, emphasized in text. D) Dot plots (left) and percent expression line plots (right) for NS-like RASopathy genes. Unfilled arrows highlight plots for the recently identified RASopathy genes, *CDC42* and *YWHAZ*. Gray arrows highlight plots for *SHOC* and *PPP1CB*, the genes underlying NSLAH. Black arrows highlight plots for *HRAS*, *KRAS* and *MAP2K2*.

Abbreviations: NDDs = neurodevelopmental disorders; NS = Noonan Syndrome; NSML = NS with multiple lentigines; NSLAH = NS-like disorder with loose anagen hair; CFCS = Cardiofaciocutaneous Syndrome; CS = Costello Syndrome; GW = gestational week.

Figure 3 Novel application of NeuroTri2-VISDOT to histone-associated NDDs. A) Table of RI histone genes delineating which are known or predicted to be associated with NDDs. (B-D) Dot plots (left) – the average expression scaled reflects the normalized expression value for a gene, scaled to zero, and depicted by color gradient; the percent expression reflects the percentage of a particular cell type with at least one read mapping to the gene of interest, depicted by size gradient. Line plots (right) – percent expression by cell type. Unfilled boxes highlight GW16, emphasized in text. B) Dot plots (left) and percent expression line plots (right) for RI histone genes with known Mendelian NDD association. C) Dot plots (left) and percent expression plots (right) for RI histone genes with predicted Mendelian NDD association. D) Dot plots (left) and percent expression plots (right) for RI histone genes without known or predicted Mendelian NDD association.

Abbreviations: NDDs = neurodevelopmental disorders; RI = replication-independent; GW = gestational week.

Figure 4 Uses and future applications of NeuroTri2-VISDOT. NeuroTri2-VISDOT can be used, as demonstrated, at the level of a translational genetic research lab to inform model selection. More broadly, similar tools can be built to address the same questions in other tissues and at other developmental stages using different data sets. Expression signatures can also be compared across multiple genes in BioBanks to nominate candidate causative genes.

Data Availability

Data and code used to generate NeuroTri2-VISDOT can be found at <https://github.com/kellyclark1/VISDOT>.

Acknowledgements

We would like to thank Dr. Christopher Sifuentes of the Chan Zuckerberg Initiative not only for his invaluable insights pertaining to data wrangling visualization but also his tireless support of trainees.

Funding Statement

The Chan-Zuckerberg Initiative Neurodegeneration Challenge Network (RAN, EJB) provided the main source of funding for this work. Other sources of funding include: NICHD F30 1F30HD112125-01A1 (EEL); NHGRI T32 5T32HG009495-05 and the Eagles Autism Foundation (DLC); and NIGMS T32 5T32GM008638-27 (ELD, TTN).

Author Contributions

CRedit has 14 categories:

1. Conceptualization – KJC, EEL, EMG, TTN, EJB
2. Data curation – EEL, KJC, EMG
3. Formal analysis – KJC, EEL, EMG
4. Funding acquisition – RAN, EJB
5. Investigation – KJC, EEL, EMG
6. Methodology – KJC, EEL, EMG, TTN, EJB
7. Project administration – TTN, EJB
8. Resources – RAN, EJB
9. Software - KJC
10. Supervision – TTN, EJB
11. Validation – KJC, EEL, EMG, AKS, DLC, ELD, RAN, TTN, EJB
12. Visualization – KJC, EEL
13. Writing-original draft – EEL, KJC
14. Writing-review & editing – EEL, KJC, EMG, AKS, DLC, ELD, RAN, TTN, EJB

Ethics Declaration

This study does not involve human subjects or live invertebrate and/or higher invertebrates.

Conflict of Interest

The authors declare no conflicts of interest.

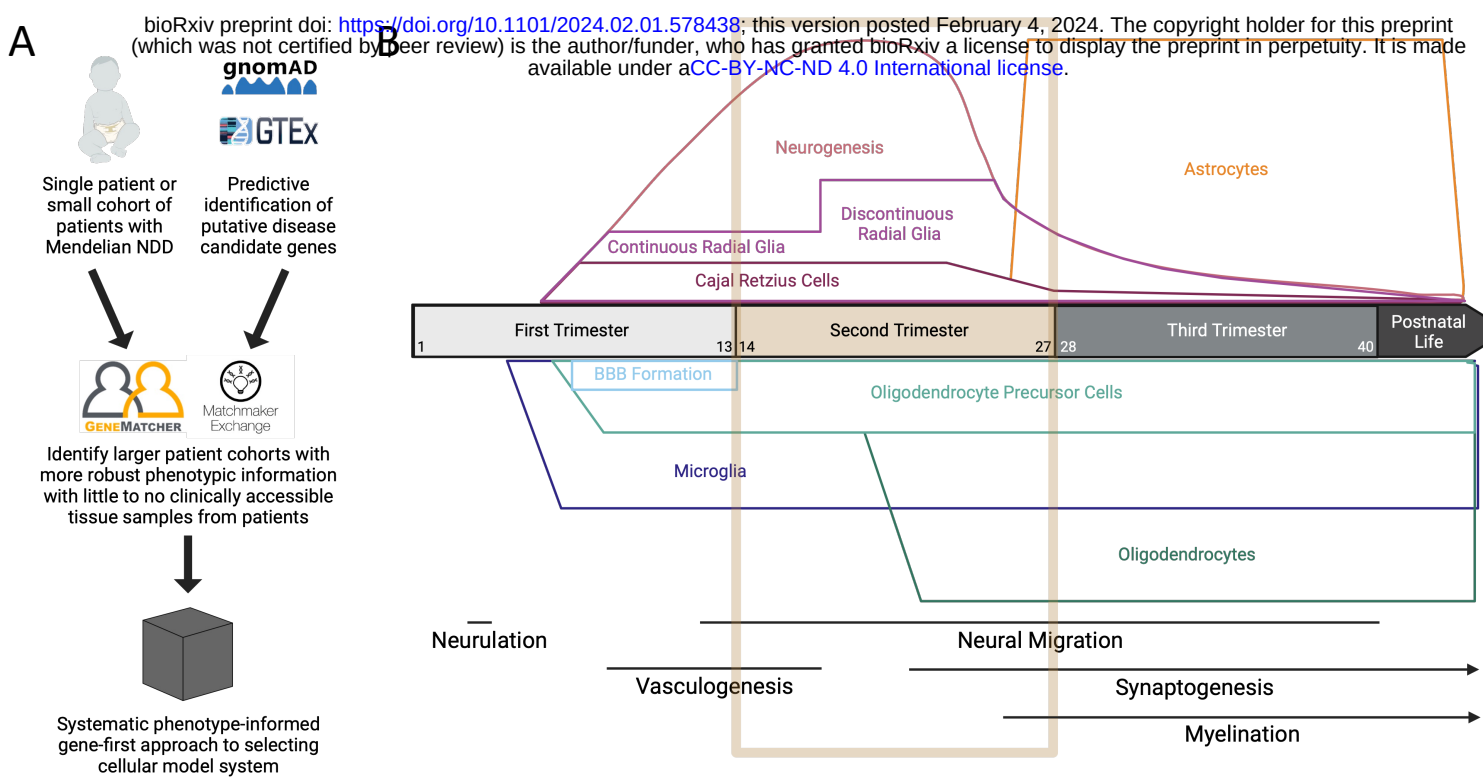
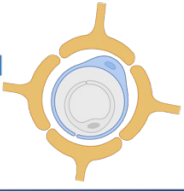
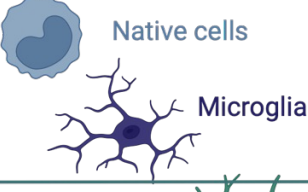
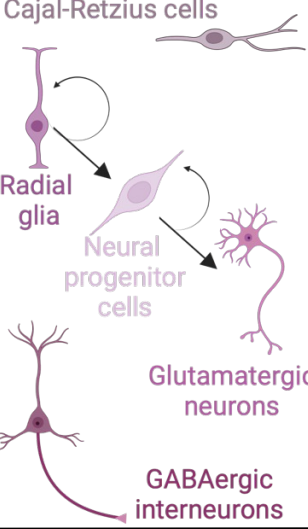
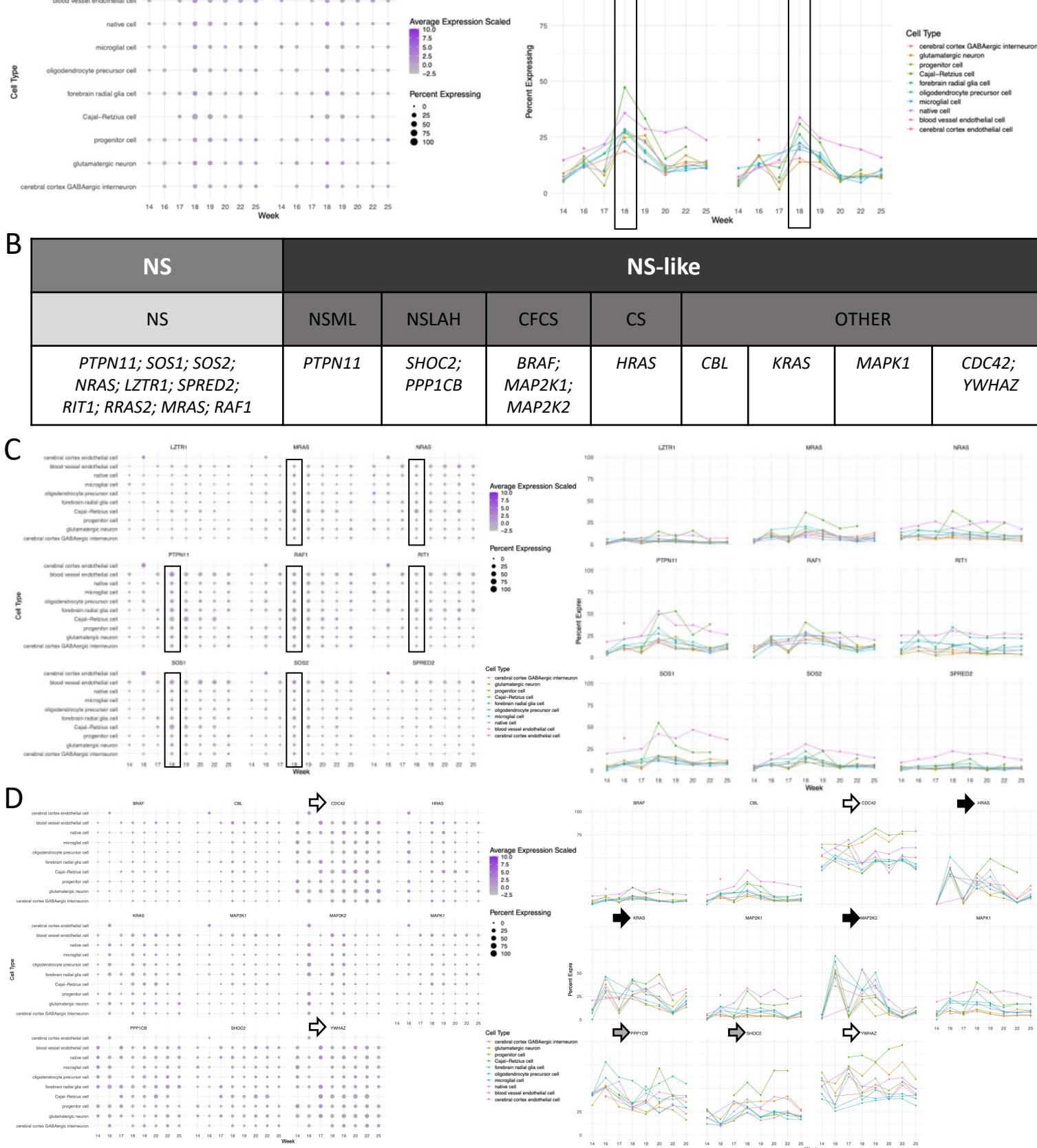


Figure 1 NeuroTri2-VISDOT fills key gap to inform selection of Mendelian NDD models.

Table 1 Primer on transcriptomically-delineated cell types in the Anaduri et al dataset

		Origin	Function	Refs
Endothelial Cells ¹		Mesodermally-derived, modified squamous epithelial cells	- Form the walls of the blood vessels that regulate the interactions with different vascular, immune and neural cells as well as the movement of ions, nutrients, hormones, gases, metabolites, drugs and cells into the brain - One of the two key cell types (in addition to mural cells) that constitute the main structural and functional vascular elements - Vascular maturation and angiogenesis play critical roles in human neurodevelopment, including in vascular-guided migration of GABAergic interneurons	24, 25, 26, 27, 28
Immune Cells ¹		Myeloid cells derived from embryonic hematopoietic progenitors of the yolk sac and fetal liver	- Colonize the brain parenchyma beginning in GW4 - Play an increasingly recognized role in neurodevelopment, and a potentially pathophysiologic role in the pathogenesis of Mendelian and multi-gene neurodegenerative disorders Native cells: non-parenchymal macrophages including perivascular, choroid plexus and meningeal macrophages Microglia: predominant immune cells of the brain that serve as tissue-resident macrophages	16, 17, 18, 19
Glia ¹		Originate from the ventral ventricular zone of the medial and lateral GEs	- Migrate along the vasculature from the medial and lateral GEs - Only glial cells that form direct synapses with neurons - Differentiate into oligodendrocytes, which myelinate neuronal axons in the central nervous system	29, 30
Neuronal Populations ¹		Extraneocortically derived	- Pioneer neurons of the human cerebral cortex that are required for cortical layering through the secretion of reelin; the radial migration of glutamatergic neurons, GABAergic neurons and OPCs; functional area formation; as well as dendritogenesis and synaptogenesis - Largely transient population that apoptose in humans between GW23-28	31, 32, 33
		Neuroepithelially-derived	Radial glia: stem cells of the subventricular zone that begin to divide asymmetrically at GW 7 until, in a human-specific process, they become a physically discontinuous scaffold during GW 16-17 - Contribute to the formation of glutamatergic neurons and also provide a physical scaffold for their radial migration - After the peak of neurogenesis, radial glia differentiate into astrocytes - Asymmetric division of outer radial glia generate neural progenitor cells which then differentiate into glutamatergic neurons of the cerebral cortex Neural progenitor cells are an intermediary precursor population between radial glia and more differentiated neurons Glutamatergic neurons are a class of excitatory neurons that release glutamate and play essential roles in synaptic transmission, plasticity, and long-term potentiation	21, 22, 34, 35
		Derived from progenitors within the ventricular-subventricular zones and GEs	- Inhibitory interneurons that migrate tangentially towards the cortex along blood vessels - Shape the connectivity of neural networks; regulate the excitability of circuits; and modulate the activity and plasticity of the brain	19, 28, 31, 36

¹ The color-coding aligns with the colors of the prenatal to early postnatal human neurodevelopment timeline from Figure 1B

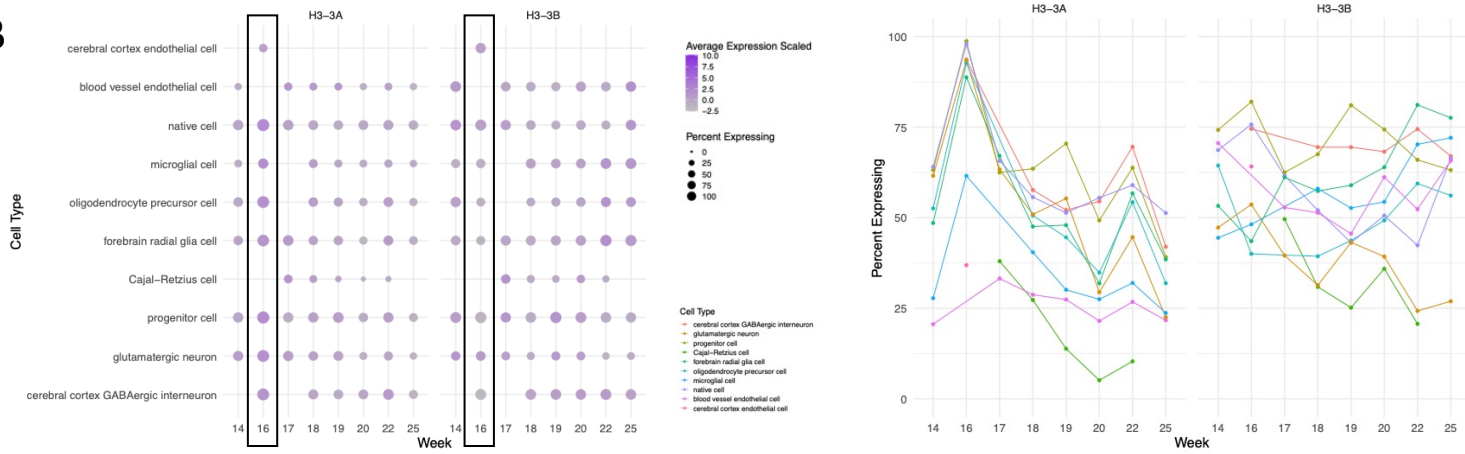


A

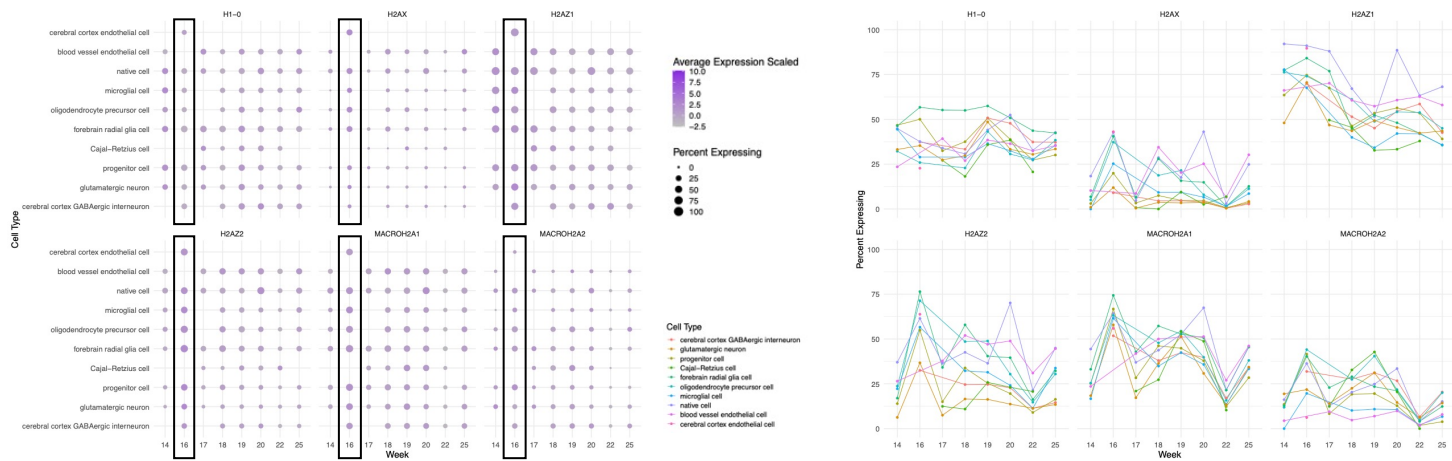
bioRxiv preprint doi: <https://doi.org/10.1101/2024.02.01.578438>; this version posted February 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RI histones with known Mendelian NDD association	RI histones with predicted Mendelian NDD association	RI histones without known or predicted Mendelian NDD association
<i>H3-3A; H3-3B</i>	<i>H1-0; H2AZ1; H2AZ2; H2AX; MACROH2A1; MACROH2A2</i>	<i>H1-7; H1-8; H1-10; H2AJ; H2AP; H2BW1; H2BW2; H3-5; CENPA</i>

B



C



D

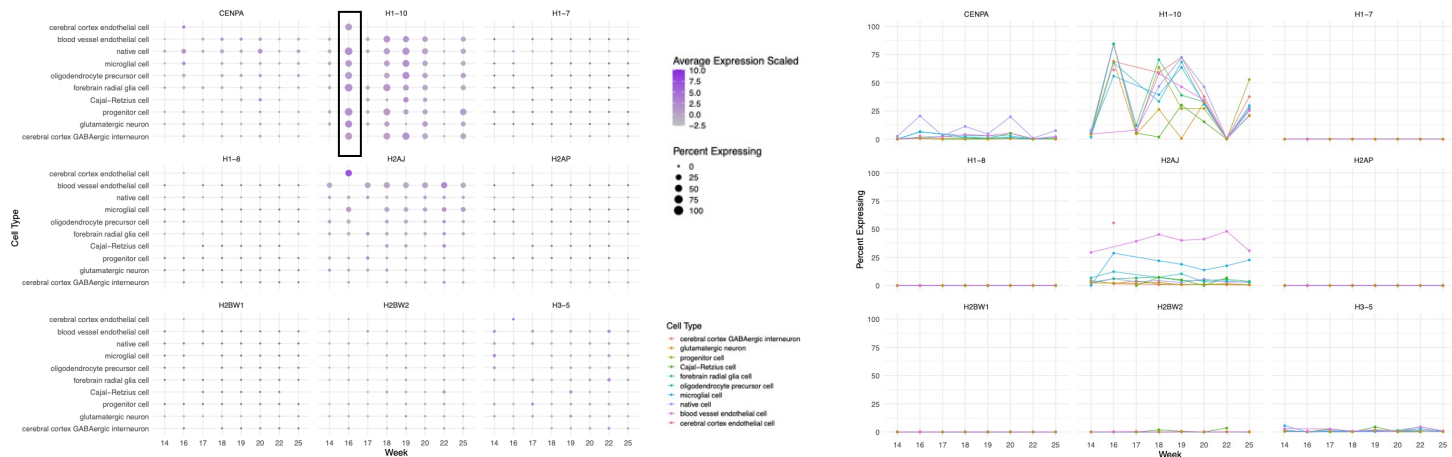


Figure 3 Novel application of NeuroTri2-VISDOT to histone-associated NDDs.

NeuroTri2-VISDOT

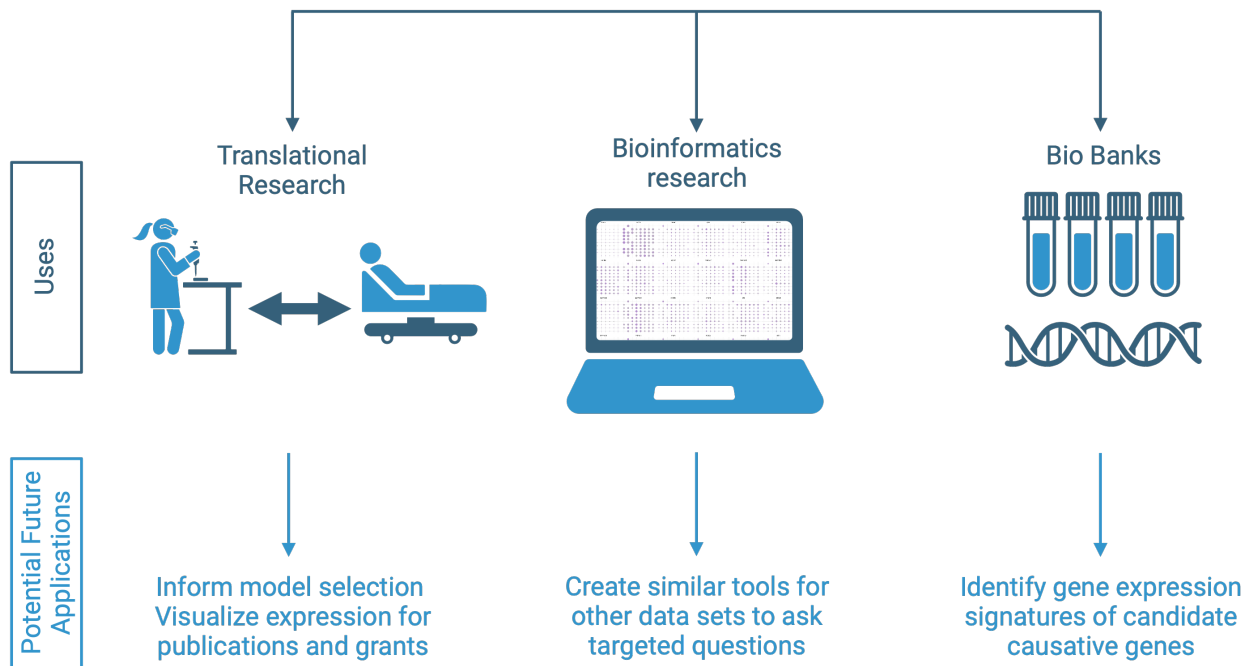


Figure 4 Uses and future applications of NeuroTri2-VISDOT.