Single-Cell Transcriptomics of the Mouse Thalamic Reticular Nucleus

by

Taibo Li

S.B., Massachusetts Institute of Technology (2015)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Author .... Signature redacted .....
Department of Electrical Engineering and Computer Science
May 25, 2017

Certified by.. Signature redacted .....
Guoping Feng
Poitras Professor of Neuroscience, MIT
Thesis Supervisor

Certified by... Signature redacted .....
Kasper Lage
Assistant Professor, Harvard Medical School
Thesis Supervisor

Accepted by.. Signature redacted .....
Christopher Terman
Chairman, Masters of Engineering Thesis Committee

Single-Cell Transcriptomics of the Mouse Thalamic Reticular Nucleus

by

Taibo Li

Submitted to the Department of Electrical Engineering and Computer Science
on May 25, 2017, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

The thalamic reticular nucleus (TRN) is strategically located at the interface between the cortex and the thalamus, and plays a key role in regulating thalamo-cortical interactions. Current understanding of TRN neurobiology has been limited due to the lack of a comprehensive survey of TRN heterogeneity. In this thesis, I developed an integrative computational framework to analyze the single-nucleus RNA sequencing data of mouse TRN in a data-driven manner. By combining transcriptomic, genetic, and functional proteomic data, I discovered novel insights into the molecular mechanisms through which TRN regulates sensory gating, and suggested targeted follow-up experiments to validate these findings.

Thesis Supervisor: Guoping Feng
Title: Poitras Professor of Neuroscience, MIT

Thesis Supervisor: Kasper Lage
Title: Assistant Professor, Harvard Medical School

Acknowledgments

I am deeply grateful for the mentorship of Zhanyan Fu and Joshua Levin, who have taught me how to approach complicated problems in neuroscience from multiple complementary perspectives, and their guidance has been central in my thesis work. Because of the multidisciplinary nature of this project, I tremendously appreciate insightful discussions with Yinqing Li, Mario Arias Garca, Violeta Gisselle Lopez Huerta, April Kim, Edyta Malolepsza, Sasha Krol, Tim Blosser, Kirsten Levandowski, Heiko Horn, Sandrine Muller, Jakob Jespersen, Jonathan Mercer, and Katie Quast, who are experts in their own fields and have helped me formulate various aspects of this thesis. I would also like to thank the many amazing scientists with whom I am fortunate to have collaborated, including Rob Sladek (McGill), Chu Pham Dang (McGill), David An (Broad), Andrew Zimmer (Broad), Liraz Greenfeld (Broad), Arthur Liberzon (Broad), Xian Adiconis (Broad), Lingling Yang (Broad), Ayan Ghoshal (Broad), and Jen Pan (Broad).

I would love to acknowledge my very special gratitude to my thesis supervisors Kasper Lage and Guoping Feng, who have created an unparalleled environment for my thesis research, both socially and intellectually. They have inspired me to pursue a career as a physician-scientist and engage in cutting-edge research to improve peoples lives.

Lastly, I want to thank my parents for their selfless love and support.

BACKGROUND

Autism spectrum disorders (ASDs) are a class of neurodevelopmental disorders affecting approximately one in every 68 children in the United States[1]. Patients with ASD typically experience social interaction deficits and are prone to repetitive, restrictive behaviors and interests[2–4]. Noticeably, there are significant phenotypic variations among ASD patients, suggesting that it is a highly polygenic disease where each genetic risk variant confers only a small incremental risk[5]. The complex genetic landscape and the intricate connectivity pattern of diverse cell types in the brain have been the major challenges in understanding ASD as well as other neuropsychiatric disorders. Importantly, there are currently no cure or effective treatment for ASD core symptoms[6,7].

Among the diverse symptoms with which ASD patients present, atypical sensory-based behaviors are ubiquitously reported[8]. In fact, abnormal regulation of the flow of sensory input has also been observed in other psychiatric disorders including attention deficit/hyperactivity disorder (ADHD)[9] and schizophrenia[10], and recent evidence showed that sensory processing dysfunction can lead to attentional deficiencies with implications for overall cognitive performance[11]. Given the convergence of clinical presentations on sensory overload from distinct diagnoses, sensory gating may be a common circuit perturbed in many of these diseases and therefore present a promising target for therapeutic interventions. Unfortunately, our understanding of the molecular and cellular mechanism which regulates sensory gating has been lacking.

The thalamus plays a key role in relaying sensory inputs to the neocortex for high level sensory processing dictated by behavioral states[12]. Thalamo-cortical interactions are regulated by the thalamic reticular nucleus (TRN), a thin layer of GABAergic neurons strategically located between the neocortex and the thalamus[13,14]. TRN provides strong inhibitory projections to the thalamus, thereby controlling thalamic sensory input and, in addition to sensory gating, regulates other important physiological processes including sleep rhythms, attention selection, and higher cognitive functions[15,16]. It has been suggested that a "leaky thalamus" as a consequence of TRN dysfunction could produce overwhelming cortical sensory input as to disrupt attention and sleep, which has been observed in patients with neurodevelopmental disorders, particularly early-course schizophrenia[17] and ASD[18].

Although TRN consists mainly of GABAergic neurons, recent studies have uncovered considerable diversity in terms of topographic connections, electrophysiological properties, and functions among TRN cells[19–22]. TRN neurons can be further partitioned into functionally divisible modules which control sensory processing in distinct modalities[21,23]. It is therefore likely that different sectors of the TRN play distinct roles in modulating thalamo-cortical interactions, but to date there has been no comprehensive survey of the molecular heterogeneity of TRN neurons.

In this thesis, I aim to leverage the powerful emerging single-cell RNA sequencing technologies to dissect the diversity of TRN cellular identity in detail. Single-cell transcriptomics studies have

been successfully applied in characterizing immune cells[24], tumor cells[25], stem cells[26], and more recently, neurons[27], and are therefore particularly suitable for application to the TRN. In the current study, we obtained mouse TRN samples using PV-tdTomato labeling[28], enriched for neurons using Neun antibody selection, isolated individual nuclei by FACS sorting to 96-well plates, and sequenced single-nucleus full-length transcriptomes using optimized Smart-Seq2 and Nextera protocols (sNuc-Seq)[29]. This method could avoid transcriptional degradation and mRNA level changes caused by neuronal dissociation, and was successful in detecting rare hippocampal cell types and GABAergic neurons in the adult spinal cord[29].

This thesis is organized as follows. In **Section I**, I provide detailed computational analysis of the TRN sNuc-Seq data to dissect the molecular basis of TRN cellular diversity. In **Section II**, I present a TRN-specific co-expression network (TRNNet) and demonstrate its potential to annotate ASD risk genes. In **Section III**, I highlight ongoing computational and experimental follow-up analyses to complement current sNuc-Seq data. I conclude this thesis with a perspective on the extension of this work in **Section IV**.

SECTION I: MOLECULAR BASIS OF TRN HETEROGENEITY

Our TRN dissection yielded single-nucleus RNA sequencing data for 1,687 cells in total. After dimensionality reduction and clustering using a standard pipeline as previously described[29], we identified 694 Gad2^{+}Pvalb^{+} neurons mainly from the TRN (<5% are external globus pallidus [GPe] cells). We found two anti-correlated transcriptional programs among the 694 cells, marked by *Ecel1* and *Spp1* expression respectively. Cells with high levels of *Ecel1* expression are segregated from cells with high levels of *Spp1* expression as observed in the two-dimensional *t*-SNE[30] embedding (**Figure 1**). From the RNA fluorescent *in situ* hybridization (FISH) detection of *Ecel1* and *Spp1* in Pvalb+ TRN neurons, we found that Ecel1+ cells primarily occupy the edge of the TRN while Spp1+ cells reside in the core region (data not shown). Further, using electrophysiological recording, we found that Spp1+ neurons present robust rebound bursting elicited by hyperpolarization with high firing frequencies within a burst, whereas Ecel1+ neurons present only single bursts (data not shown). This suggests that Spp1+ cells might play a key role in sleep spindle generation and sensory filtering, which I will investigate further in this thesis. Additionally, we found that retrograde tracing from first-order thalamo-cortical relay nuclei mainly labeled Spp1+ cells in the center of TRN, while those from high-order nuclei labeled Ecel1+ cells in the peripheral (data not shown). Together, these data revealed remarkable molecular, cellular, functional, and connectivity heterogeneity of TRN neurons (manuscript in preparation).

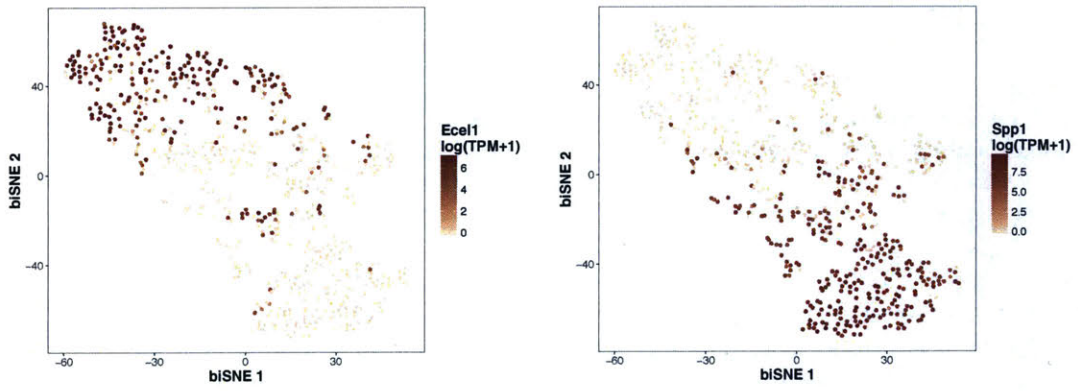

Figure 1 | *Ecel1* and *Spp1* mark two transcriptional programs in TRN inhibitory neurons. Here I plotted the log-transformed expression levels of *Ecel1* (left) and *Spp1* (right) in the 694 TRN Gad2+Pvalb+ neurons, where the coordinates correspond to the *t*-SNE embedding of the high-dimensional gene expression data. We adapted the original *t*-SNE algorithm to be more sensitive for detecting rare cell types (bi-SNE)[29].

Ecel1+ and Spp1+ cells are identified based on gene expression profiles

To investigate the molecular basis of the observed heterogeneity, I first set out to define Ecel1+ and Spp1+ cells based on their gene expression profiles. Both *Ecel1* and *Spp1* expressions exhibit bimodal distributions among the 694 TRN cells (**Figure 2**). To capture the majority of cells in the second peak, I defined the threshold values such that Ecel1+ cells have *Ecel1* expression log(TPM + 1) > 5, resulting in 148 cells, and Spp1+ cells have *Spp1* expression log(TPM + 1) > 6, resulting in 243 cells.

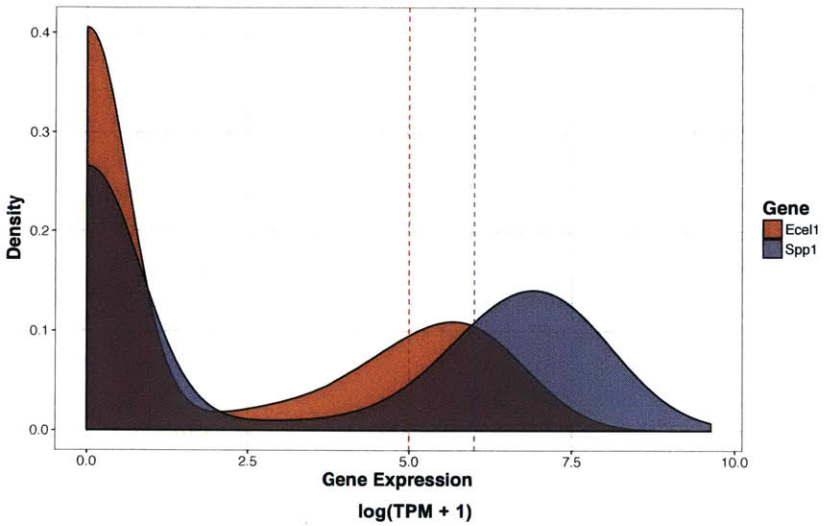

Figure 2 | Log-transformed *Ecel1* and *Spp1* expression profiles both conform to a bimodal distribution among 694 Gad2$^+$Pvalb$^+$ (primarily) TRN neurons. The dotted lines are the threshold values used to determine Ecel1+ and Spp1+ cells. The density distribution was calculated using the default Gaussian kernel estimation from the base package of the R software version 3.2.

To assess the robustness of differential gene expression to the choice of threshold values which define Ecel1+ and Spp1+ cells, I varied threshold around the aforementioned values (5 for *Ecel1* and 6 for *Spp1* log-transformed expression levels), and calculated the one-sided differential expression p-values for each gene in the entire genome between the resulting Ecel1+ and Spp1+ cells. I then measured the concordance of differential gene expression by the rank-based Kendall Tau coefficient for the p-values of each gene under each choice of the threshold values. I noted that p-values are well correlated (above 0.8) in the range of threshold values tested (**Figure 3**), suggesting that the differentially expressed genes are robust to my choices of threshold values. Not surprisingly, the chosen cells are confirmed by the two-dimensional *t*-SNE embedding in that Ecel1+ cells, which reside mainly in the second (II) quadrant, are well segregated from Spp1+ cells, which mainly occupy the fourth (IV) quadrant (**Figure 4**).

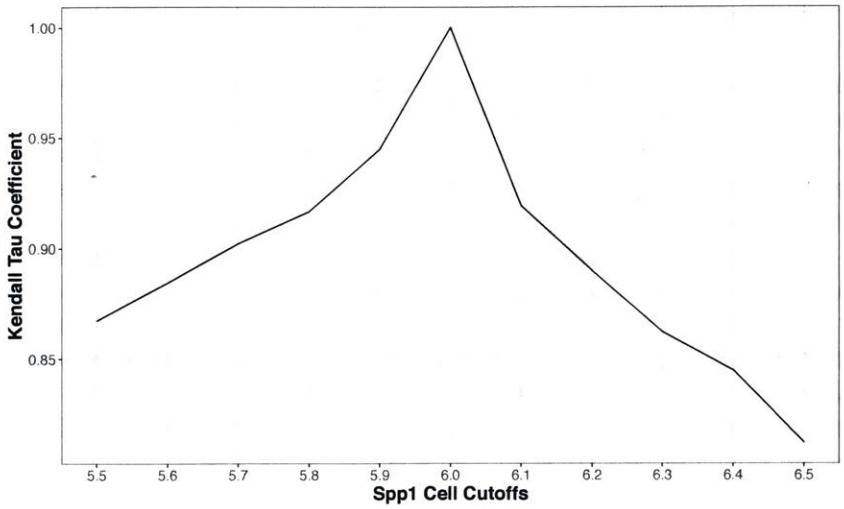

Figure 3 | Differentially expressed genes are robust to the choices of threshold values. I varied the threshold value to define Spp1+ cells within the window [5.5, 6.5] (*x*-axis), and for each population of Spp1+ cells such defined, computed the p-value associated with differential expression between Ecel1+ cells (fixed) and Spp1+ cells for each gene. I computed the Kendall Tau coefficient for the p-values in each case with the real p-values obtained in my choice of threshold 6 (so that the threshold of 6 will have the perfect Kendall Tau coefficient 1), plotted as the *y*-axis.

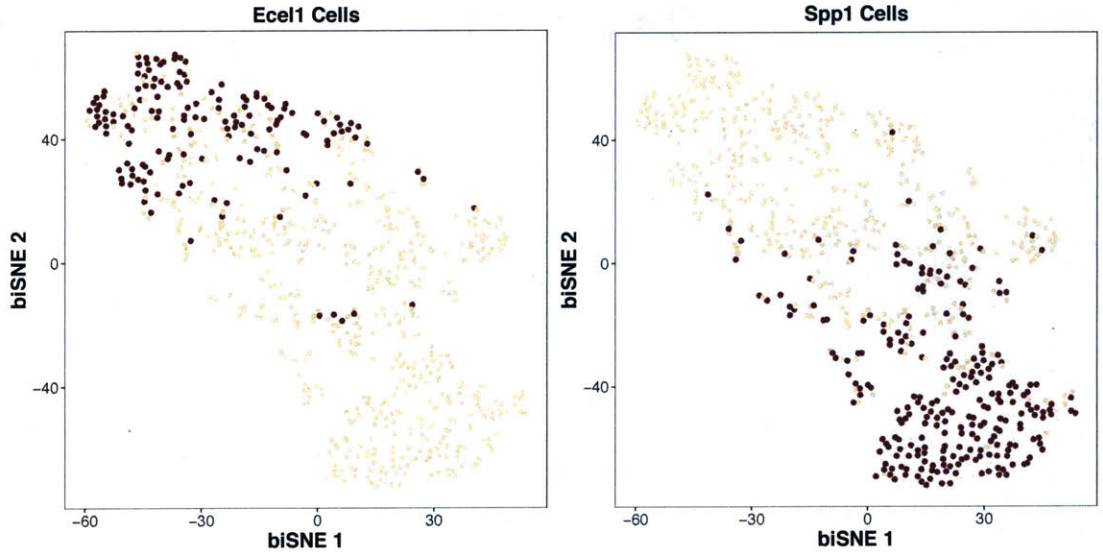

Figure 4 | *t*-SNE embedding of chosen Ecel1+ and Spp1+ cells. Here I colored the chosen cells dark red and the rest of the unlabeled cells light yellow. Ecel1+ cells reside mainly in the second (II) quadrant, while Spp1+ cells reside mainly in the fourth (IV) quadrant.

Identification of additional gene markers for Ecel1+ and Spp1+ cells revealed a continuum of cellular identity

To identify potential gene markers for Ecel1+ and Spp1+ neurons, I looked for genes whose expression are highly consistently correlated with *Ecel1* and *Spp1* across TRN Gad2^{+}Pvalb^{+} neurons. Pearson correlation measures the overall concordance of gene expression using all TRN cells, but it could be sensitive to extensive technical noise owing to the low starting material of the single-cell RNA sequencing experiments, which results in zero-inflated distribution of gene expression values[31]. I therefore devised a novel measure of gene-gene expression similarity, called IoU (intersection over the union). For each gene, I looked for the cells in which the gene has the highest expressions ("signature cells" for each gene), and defined the IoU between a gene pair as the Jaccard index between their signature cells. IoU therefore ranges from 0 to 1, where an IoU of 1 means the gene pair has exactly the same set of cells in which they are highly expressed (thus have very similar expression profiles), and an IoU of 0 means the signature cells between the gene pair do not overlap. IoU is robust against dropout noise because it only considers the cells where our genes of interest are highly expressed. I noted that IoU is in general highly correlated with Pearson correlation except for genes with extensive dropout noise.

I leveraged the fact that our single-nucleus RNA sequencing data come from two batches of mice sacrificed on different dates (January 20, 2016 and October 26-28, 2016), and treated the two batches as two independent samples. I then looked for genes highly correlated with *Ecel1*/*Spp1* by both Pearson correlation and IoU in both batches, thus obtaining genes which are consistently co-regulated with *Ecel1*/*Spp1*. In each case, I used a threshold for Pearson Correlation and IoU to define highly correlated genes based on the empirical distribution (**Figure 5** for *Spp1* as an example).

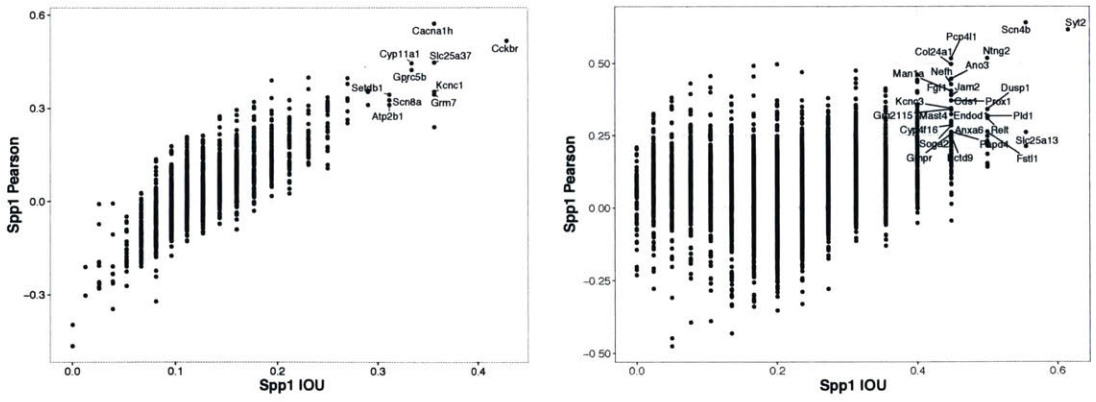

Figure 5 | Additional markers for Spp1+ cells are identified as genes highly correlated with *Spp1* by both Pearson correlation and IoU in both batches. For each gene except *Spp1*, O plotted its Pearson correlation (*y*-axis) and IoU (*x*-axis) with *Spp1* in both batch 1 (left, N = 177 cells) and batch 2 (right, N = 457 cells). Labeled genes were chosen to be highly correlated with *Spp1* according to both measures based on threshold values determined by the empirical distribution.

Using this stringent scheme, I found nine genes for each of Ecel1+ and Spp1+ cells which could serve as additional markers for these populations (**Figure 6**). Interestingly, when viewed through two-dimensional *t*-SNE embedding, the aggregate expression profiles of these marker genes (10 each for Ecel1+ and Spp1+ neurons as defined above) revealed a continuum of cellular identify extending from Ecel1+ to Spp1+ cells (**Figure 7**).

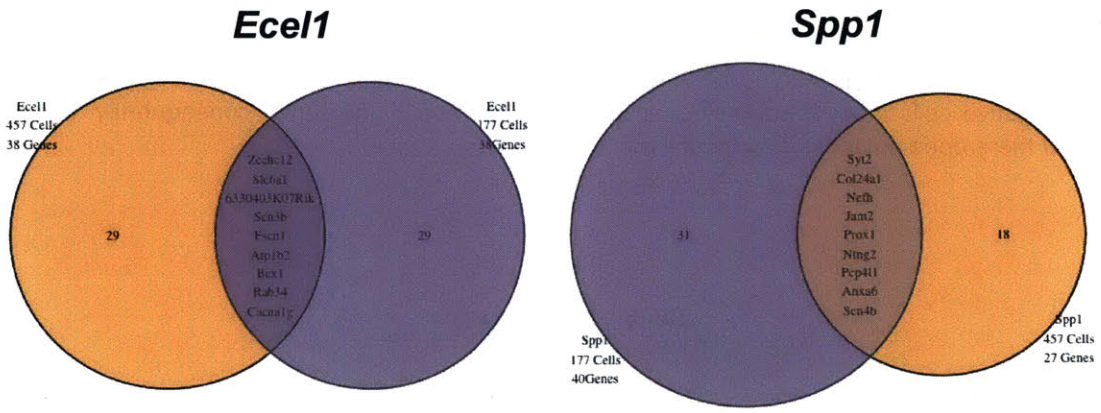

Figure 6 | Genes consistently correlated with *Ecel1* (left) and *Spp1* (right) expression by both Pearson correlation and IoU in both batches. Genes in light blue are from batch 1 (N = 177 cells) and genes in light yellow are from batch 2 (N = 457 cells). Genes identified from both batches are labeled in the overlap region.

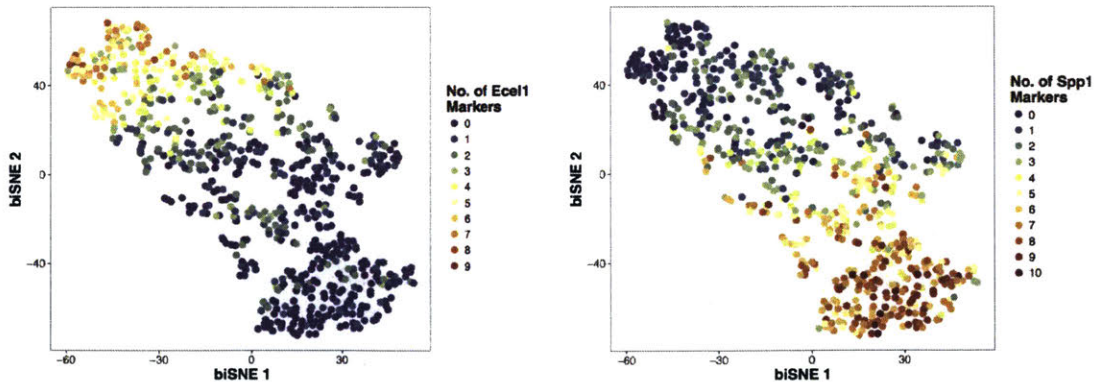

Figure 7 | A continuum of cellular identity is revealed by gene markers. For each Gad2^{+}Pvalb^{+} neuron, I counted the number of marker genes expressed and plotted the count for Ecel1+ cell markers (left) and Spp1+ cell markers (right).

Ecel1^{-}Spp1^{-} (double negative) cells are unlikely to be a distinct cell population

The observation of the continuum prompted me to investigate whether a third intermediate cell type exist in the TRN. I sought to characterize cells with no *Ecel1* or *Spp1* expression ("double negative cells", N = 162, **Figure 8**). Genome-wide scan of the differential gene expression revealed six genes which are statistically enriched in the double negative cells compared to Ecel1+ and Spp1+ cells. Five of them are either predicted genes, cDNA, or pseudo genes (*Gm10471*, *5031410I06Rik*, *Speer7-ps1*, *Gm10220*, and *Speer8-ps1*). The other gene, *Tle2*, is a transcriptional co-repressor which is also enriched in Ecel1+ cells (but not Spp1+ cells). Overall, I could not find gene markers with clear functional meaning for these double negative cells, and thus they are unlikely to be a distinct neuronal population. These data suggest that double negative cells mark the transition in the Ecel1+ – Spp1+ spectrum; consequently, it is likely that there does not exist a third cell type in the TRN.

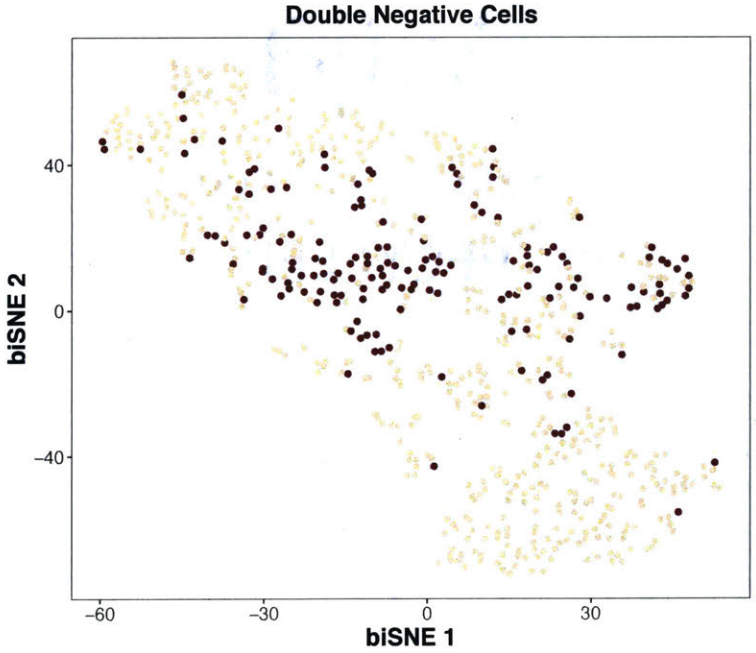

Figure 8 | Cells with no *Ecel1* or *Spp1* expression. Highlighted in dark red, these cells (N = 162) mainly occupy the central region of the *t*-SNE map connecting Ecel1+ cells on the second (II) quadrant and Spp1+ cells on the fourth (IV).

Synaptotagmin 2 is enriched in Spp1+ cells and physically interacts with $Na_V1.7$, which in turn interacts with $Ca_V1.2$ and HCN1

I next sought to investigate the potential functional roles of the marker genes as they could contribute to the observed differences between Ecel1+ and Spp1+ neurons. I noted that the alpha-1G subunit of the T-type voltage-dependent calcium channel ($Ca_V3.1$, encoded by *Cacna1g*) is enriched in Ecel1+ cells, while the alpha-1I subunit ($Ca_V3.3$, encoded by *Cacna1i*) is enriched in Spp1+ cells. This suggests subunit-specific regulation of T-type calcium channels which could be relevant for the differential electrophysiological properties between the two cell populations. I found that *Slc6a1*, the gene encoding a principal GABA transporter that mediates rapid removal of GABA and termination of GABAergic transmission[32], was enriched in Ecel1+ cells.

Furthermore, I also found that synaptotagmin 2 (encoded by *Syt2*) is enriched in Spp1+ cells (**Figure 9**), but not its close paralogs *Syt1* and *Syt9* (data not shown). Synaptotagmin 2 functions as a fast calcium sensor triggering synchronous neurotransmitter release[33], and compared to *Syt1*, it is the primary Ca^{2+} sensor in inhibitory neurons which ensures fast and efficient feedforward inhibition in the central nervous system[34]. Based on these data, I therefore hypothesized that first-order peripheral-thalamic-cortical pathway, which mainly involve Spp1+ cells, may operate at a shorter timescale than higher-order pathways.

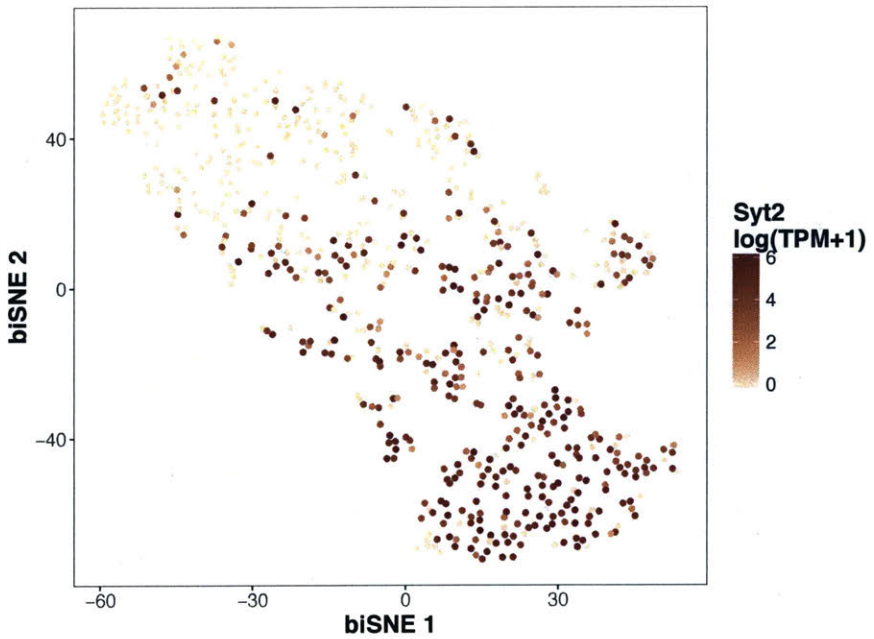

Figure 9 | *Syt2* is enriched in Spp1+ cells. In the *t*-SNE embedding of TRN Gad2+Pvalb+ neurons, cells with high level of *Syt2* expression co-localizes with Spp1+ cells in the fourth (IV) quadrant.

Interestingly, recent experimental evidence suggests that Syt2 physically binds to $Na_V1.7$[35] (encoded by *Scn9a*), a voltage-gated sodium channel. $Na_V1.7$ plays a major role in pain signaling[36] and physically interacts with both $Ca_V1.2$ (alpha-1C subunit of the L-type voltage-dependent calcium channel) and HCN1 (hyperpolarization-activated cyclic nucleotide-gated potassium channel 1) in Ngn-induced human neurons (**Figure 10**). Notably, both *CACNA1C* and *HCN1* genes are located in a locus with an index SNP genome-wide significantly associated with schizophrenia[37]. Because *HCN1* is not highly expressed in the TRN (data not shown) and Syt2 is located at axonal terminal, it is likely that the Syt2–$Na_V1.7$–$Ca_V1.2$ complex mediates presynaptic neurotransmitter release in Spp1+ cells, and the disruption of this pathway could lead to increased risk for psychiatric disorders.

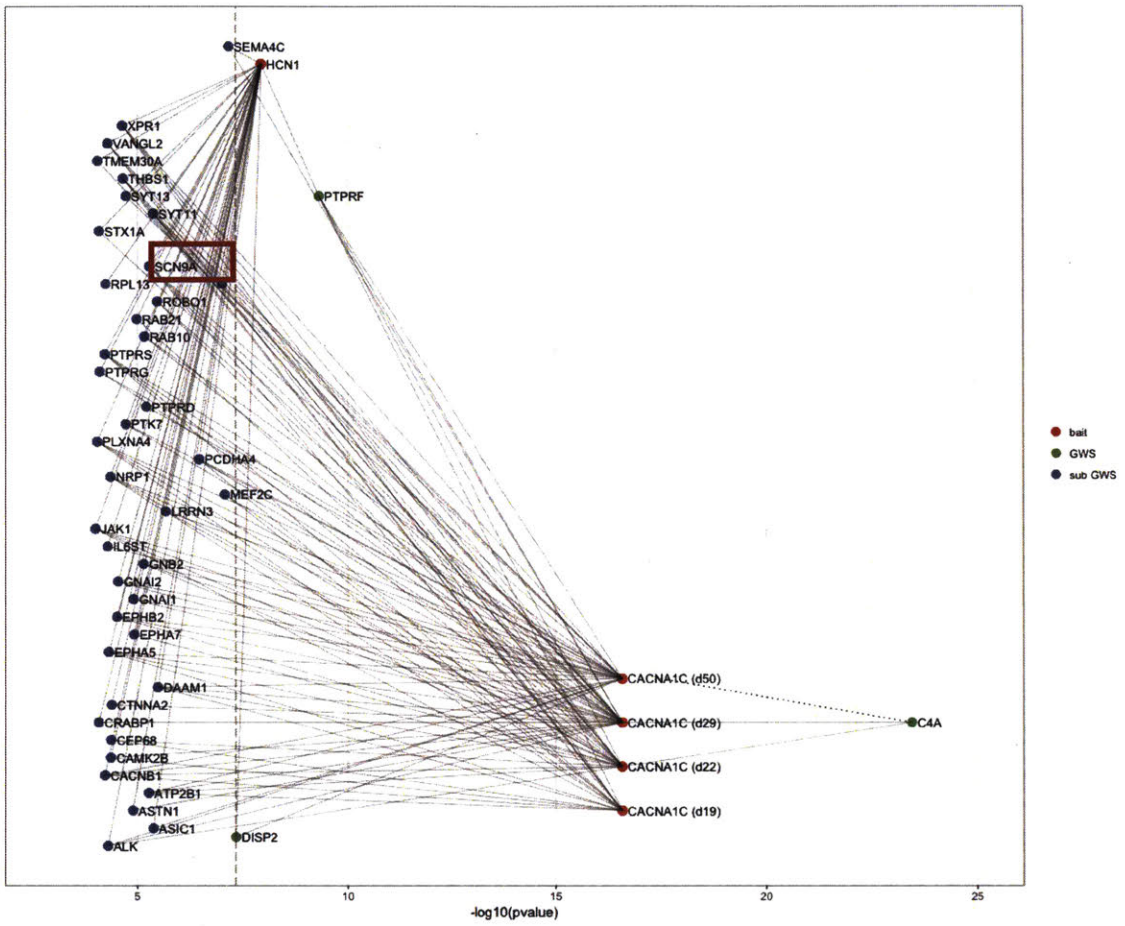

Figure 10 | $Na_V1.7$ (*SCN9A*) physically interacts with $Ca_V1.2$ (*CACNA1C*) and HCN1 in Ngn-induced human cortical neurons. We performed immunoprecipitation followed by tandem mass spectrometry (IP-MS) in Ngn-induced human neurons, using $Ca_V1.2$ (harvested at days 19, 22, 29, and 50 post-differentiation) and HCN1 (harvested at day 50 post-differentiation) as baits. We found that $Na_V1.7$ interacts with $Ca_V1.2$ at all time points and with HCN1. In this plot, the lines connecting gene pairs represent detected physical interactions with false discovery rate less than 0.1. For each gene, the *x*-axis denotes the negative logarithm of the p-values of association with schizophrenia[37] using the nearest SNP. Genes in red are used as baits, genes in green (blue) are genome-wide (sub-genome-wide) significantly associated with schizophrenia. The dotted vertical line represents the genome-wide significance threshold 5E-8. SCN9A is highlighted. GWS – genome-wide significant.

Multiple psychiatric disease risk genes are differentially expressed between Ecel1+ and Spp1+ neurons

Next, given the important role of TRN in sensory gating, attention selection, and sleep rhythm[15], I looked at genes differentially expressed between Ecel1+ and Spp1+ neuron which have been implicated in psychiatric disorders. I curated five datasets from recent studies based on common variants from genome-wide association studies (GWAS) and rare variants from exome-sequencing and copy number variant (CNV) analysis. Given the ongoing debate about ASD risk genes[38], I included all candidate genes from three large-scale sequencing studies (Sanders *et al*, 2015[39], N = 65 genes; Stessman *et al*, 2017[40], N = 87 genes; and Yuen *et al*, 2017[41], N = 61 genes). Not surprisingly, only 26 genes (20.2% of genes from all studies) were found in all three studies, and only 49 genes (38.1%) were found in at least two studies. I also included 93 genes enriched for damaging *de novo* mutations in individuals with neurodevelopmental disorders[42]. Lastly, I included genes found in the loci with an index SNP gemone-wide significantly associated with schizophrenia[37]. Based on the CEU population from the 1000 Genomes project[43], I defined each locus to be delineated by SNPs with r^2 at least 0.6 with the index SNP, plus 50kb up- and downstream, thus partitioning schizophrenia risk genes into two groups: 37 genes which reside in a single-gene locus and 417 genes which reside in a locus with more than one gene (multi-gene locus).

I found 23 genes which are both enriched for psychiatric disease risks from at least one of the above studies, and enriched in either Ecel1+ or Spp1+ neurons (**Figure 11**). Intriguingly, *Slc6a1* is enriched in Ecel1+ cells and implicated in both ASD and neurodevelopmental delay, and *Aph1a* is enriched in Spp1+ cells and implicated in both ASD and schizophrenia. Some genes encoding important channel proteins are also differentially expressed, including *Kcnb1*, *Kcnq2*, *Cacna1i*, *Scn1a*, *Grin2a*, *Scn8a*, and *Grm3*. Of particular interest is ***Scn8a*** (which encodes sodium voltage-gated channel alpha subunit 8, $Na_V1.6$) because *Scn8a* deficiency in the TRN impairs intra-TRN synaptic inhibition and tonic firing output, resulting in absence seizures[44], a common comorbidity with cognitive and developmental deficits[45]. I found that *Scn8a* is enriched in Spp1+ neurons, suggesting that these cells could be important in regulating thalamo-cortical network synchrony.

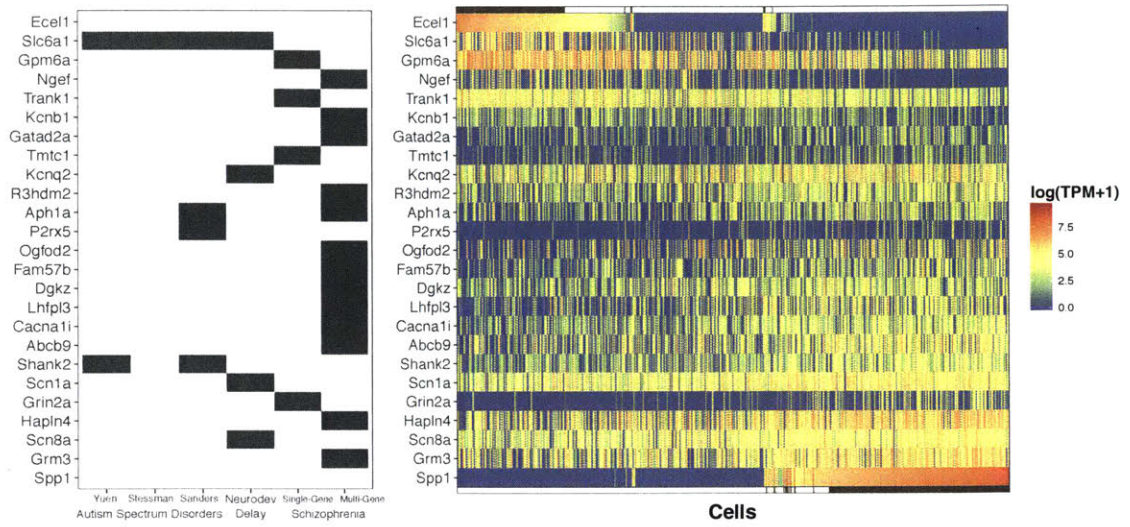

Figure 11 | Differential expression of psychiatric disease risk genes in the TRN. I included disease risk genes differentially expressed between Ecel1+ and Spp1+ cells, plus *Ecel1* and *Spp1*. The panel on the left indicates the source of each gene (i.e. the study in which each gene is reported). The heatmap on the right shows the log-transformed expression level of each gene. The cells (*x*-axis) are ranked based on *Ecel1* expression minus *Spp1* expression. The genes (*y*-axis) are ranked based on the correlation with *Ecel1*. The grey bar on the top indicates Ecel1+ cells, and the grey bar on the bottom indicates Spp1+ cells.

Integrative analysis using proteomic datasets suggests pivotal role of cholinergic transmission in the TRN

To better dissect the molecular basis of the observed differences between Ecel1+ and Spp1+ cells, I sought to investigate the potential functions of *Ecel1* and *Spp1* based on proteomics data. *Ecel1* encodes a member of the M13 family of endopeptidases which is predominantly expressed in the central nervous system[46], and is present in both the plasma membrane and the endoplasmic reticulum[47]. Currently, no natural substrate of Ecel1 has been found[48]. Simulation studies have reported Ecel1 substrate specificity based on its secondary structure[49], suggesting that Ecel1 may have fewer substrates than its paralog Ece1. Although Ecel1 might target neurotransmitters from other nerve terminals, I set out to identify possible Ecel1 substrates in the TRN based on our single-nucleus RNA sequencing data. I collected all 82 neuropeptides from the Neuropeptide Database (http://www.neuropeptides.nl/) and found 10 which are highly expressed in the TRN, among which 4 are statistically enriched in Ecel1+ cells compared to Spp1+ cells: prepronociceptin (*Pnoc*), somatostatin (*Sst*), Secretogranin II (*Scg2*), and thyrotropin releasing hormone (*Trh*). To yield a mechanistic understanding of the role *Ecel1* plays in the TRN, a follow-up experiment using immunoprecipitation followed by tandem mass spectrometry (IP-MS) to probe the interaction partners of Ecel1 in human neurons would be helpful.

Spp1 encodes osteopontin, an extracellular matrix protein with diverse functions[50], most notably cell-mediated immune responses[51,52] and cancer progression and prognosis[53]. Contrary to Ecel1, there is abundant literature on the interaction partners of Spp1. I sought to identify secreted proteins and plasma membrane proteins which physically interact with Spp1 and are relevant to the TRN using an unbiased approach. First, I obtained the human "secretome" from the Human Protein Atlas[54] which contains 2,916 genes predicted to have at least one secreted protein product. I then curated physical interaction partners of Spp1 from six major human protein-protein interaction databases, including InWeb[55,56], IID[57], Mentha[58], HINT[59], PINA[60], and iRefIndex[61,62], and obtained 157 proteins with varying degree of experimental support. Finally, I took the top 10% of genes with the highest mean level of expression across the entire TRN. The rationale of using all 694 TRN neurons rather than only Spp1+ neurons is that osteopontin is a secreted protein and could potentially act on all TRN cells. I found three genes in the overlap of above three datasets (**Figure 12**). Two of them, ***Adam22*** and ***Adam15***, encode members of the ADAM (a disintegrin and metalloprotease domain) family which function in cell-cell and cell-matrix interactions[63]. Of note, Adam22 plays an important role in the correct myelination in the peripheral nervous system[64] and maturation of excitatory synapses[65], making it an interesting candidate with which we could investigate the potential functional role of *Spp1* in regulating TRN cells.

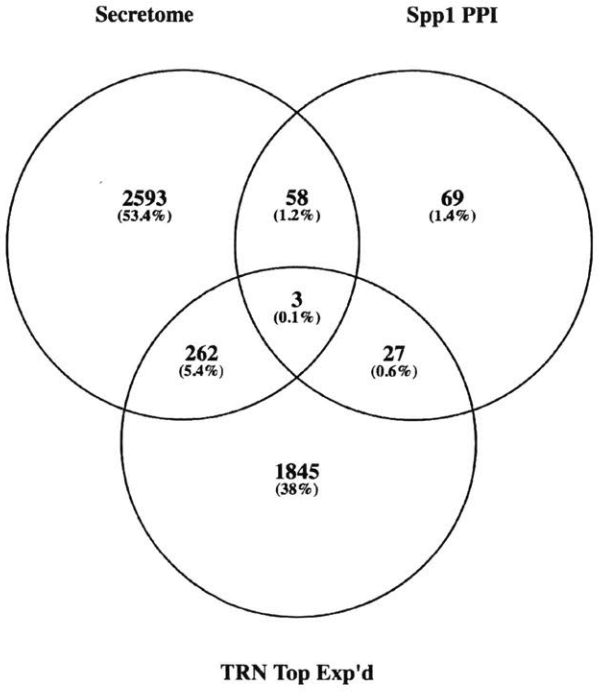

Figure 12 | Data-driven approach identifies *Ache*, *Adam22*, and *Adam15* as interesting candidates. This Venn diagram shows the overlap between genes with secreted protein products ("secretome"[54], 2,916 genes), physical interaction partners of Spp1 in six protein-protein interaction databases[55–62] (157 genes), and genes highly expressed in the entire TRN (2,137 genes) from our single-nucleus RNA sequencing data.

Additionally, I identified AChE (acetylcholinesterase), encoded by ***Ache***, which hydrolyzes acetylcholine (ACh) and thus terminates cholinergic synaptic transmission[66]. *Ache* expression is enriched in the TRN (**Figure 13**), and mutations of *Ache* are associated with ASD risk[39], prompting me to further interrogate functional roles of cholinergic transmission in the TRN. In fact, I found a converging line of evidence when investigating *Grm3*, a schizophrenia risk gene also selectively enriched in the TRN (**Figure 14**). *Grm3* encodes metabotropic glutamate receptor 3 which has 210 known interaction partners in the aforementioned six protein-protein interaction databases[55–62]. I compared our TRN single-nucleus RNA sequencing data with three other single-cell datasets, including Zeisel *et al*, 2015[27] on mouse somatosensory cortex and hippocampal CA1, Tasic *et al*, 2016[67] on mouse V1 cortex, and Habib *et al*, 2016[29] on mouse hippocampus. Importantly, I found that ***Hrh3*** (histamine receptor H3) and ***Chrm2*** (cholinergic receptor muscarinic 2) are both physical interaction partners of mGluR3[68,69] and selectively enriched in the TRN (**Table 1**). The histamine receptor H3 functions as a modulator of the release of neurotransmitters including ACh, whose antagonists exhibit a wide range of cognitive-enhancing effects[70].

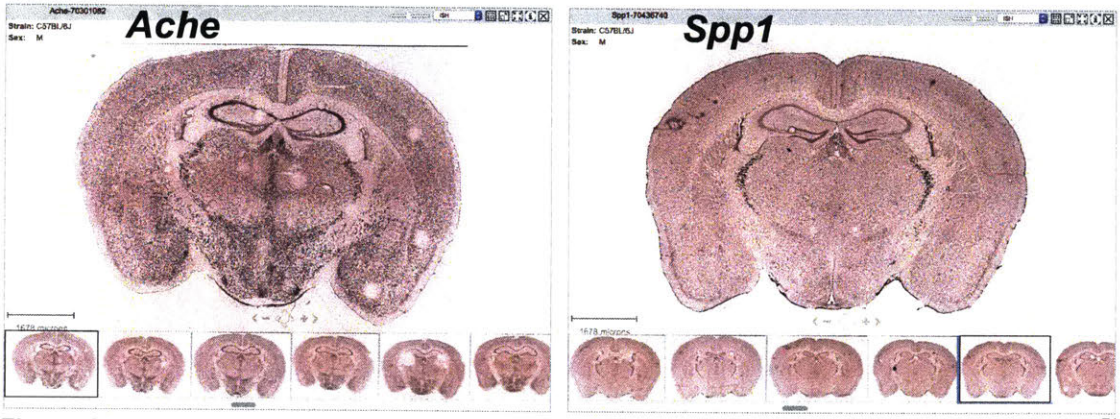

Figure 13 | *Ache* expression is enriched in the TRN. *In situ* hybridization from Allen Brain Atlas[71] shows that expression of *Ache* is enriched in the mouse TRN (left), with *Spp1* expression shown as reference (right). Both slices are from adult mice (P56).

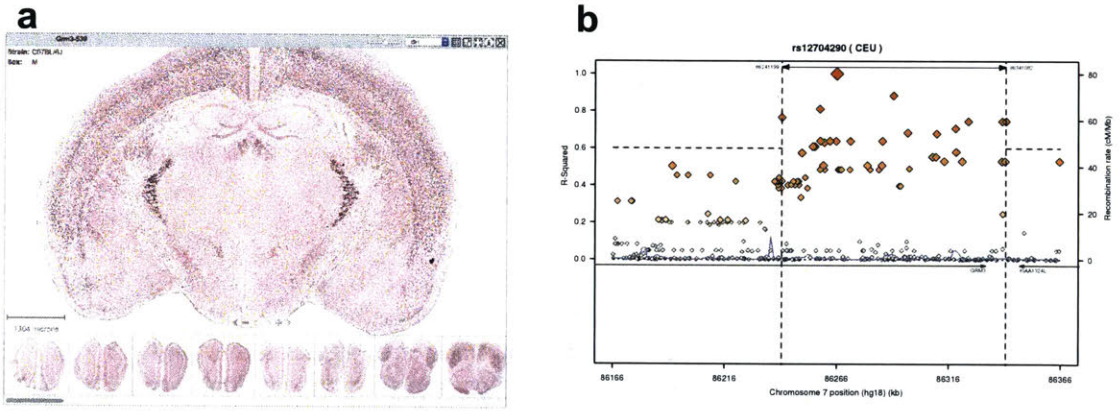

Figure 14 | Metabotropic glutamate receptor 3 (mGluR3), encoded by *Grm3* gene, is enriched in the TRN and also a schizophrenia risk gene. a) *In situ* hybridization from Allen Brain Atlas[71] of *Grm3* expression in adult mouse brain (P56). **b)** rs12704290 is associated with schizophrenia, which lies in the intron of *Grm3* on chromosome 7. The linkage disequilibrium block is defined by a r^2 with the index SNP of at least 0.6 in the CEU population from the 1000 Genomes Project[43]. The *x*-axis represents genomic location and gene annotations, and the *y*-axis represents r^2 with the index SNP. Plot generated by SNP Annotation and Proxy Search tool (http://archive.broadinstitute.org/mpg/snap/ldplot.php).

Gene	TRN	Tasic[67] V1 Cortex		Zeisel[27] Somatosensory cortex & hippocampal CA1		Habib & Li[29] Hippocampus	
	% Cells	% Cells	Adj. P-Value	% Cells	Adj. P-Value	% Cells	Adj. P-Value
Grm3	**84%**	**66%**	**1.42E-18**	**64%**	**4.59E-24**	**15%**	**8.41E-188**
Hrh3	**70%**	**34%**	**2.52E-57**	**33%**	**1.48E-70**	**14%**	**9.91E-133**
Cep164	64%	32%	1.59E-44	29%	2.46E-66	43%	4.70E-17
Chrm2	**49%**	**21%**	**3.33E-43**	**7%**	**3.89E-189**	**2%**	**5.19E-139**
Sst	60%	31%	7.91E-40	54%	0.015	7%	2.10E-136
Pick1	65%	40%	2.02E-27	35%	4.44E-46	56%	0.001
Grm7	83%	60%	1.39E-26	50%	5.92E-56	75%	0.0004
Oprk1	12%	2%	3.28E-21	6%	4.26E-07	2%	9.32E-21
Htr2b	15%	5%	7.20E-16	2%	9.21E-62	11%	0.0174
Npy	53%	40%	6.96E-08	42%	8.36E-07	8%	3.45E-107
Grip1	48%	38%	1.84E-05	24%	4.78E-36	35%	1.56E-07
Grm8	41%	34%	0.0028	13%	1.38E-72	10%	5.75E-55

Table 1 | Interaction partners of mGluR3 which are enriched in the TRN. Genes are sorted by their adjusted p-value (false discovery rate) between Tasic *et al*, 2016[67] and TRN datasets. For each gene in each dataset, I showed the percentage of cells expressing that particular gene and the adjusted p-value of the one-sided proportion test[72]. There are 11 genes that achieve genome-wide significance, among which *Hrh3* and *Chrm2* rank on top. *Grm3* is highlighted in the top row.

These findings complement the emerging literature on the cholinergic afferents to the TRN from the brainstem and basal forebrain. Specifically, the pedunculopontine tegmental nucleus (PPN) has many long-range projections on many thalamic nuclei, including the TRN[73]. Although both GABAergic and glutamatergic neurons in the PPN have distinct roles in modulating cortical activity and sleep/wake states, cholinergic neurons in the PPN specifically suppress slow cortical rhythms[74] and induce REM sleep[75]. In the TRN, optical stimulation of cholinergic fibers induces the generation of sleep spindles regardless of sleep/wake states and promotes sleep[76]. Consistent with my finding of the enrichment of *Chrm2* in the TRN, it was reported that ACh release leads to fast and precise biphasic excitatory-inhibitory synaptic signaling mediated by both nicotinic ACh receptor α4β2 and muscarinic ACh receptor M2[77]. Therefore, cholinergic afferents to the TRN have a major modulatory role in updating behavioral states based on sensory inputs[78].

Recently, synergistic activation of mGluR1 and mGluR5 has been shown to be critical in the regulation of cholinergic synaptic transmission in the TRN[79]. My observations above suggest a potentially novel mechanism in which mGluR3 interfaces with cholinergic signaling such that the mGluR3 – Chrm2/Hrh3 complex could potentially regulate TRN neuronal firing properties. Follow-up experiments are therefore warranted to validate the interaction of Chrm2 and Hrh3 with mGluR3 in the TRN, and to compare TRN neuronal firing activity changes in the presence of mGluR3 agonists with or without Chrm2/Hrh3 antagonists.

Osteopontin interaction partners enriched in Spp1+ cells could potentially explain their unique firing properties

Lastly, I sought to find differentially expressed genes which could explain the unique firing patterns of Spp1+ cells. I started with 157 known interaction partners of Spp1 and looked for those that are enriched in Spp1+ cells. After adjusting for false discovery rate (FDR), I found 16 genes which are statistically enriched, among which ***Kcnip1*** and ***Cacng4*** ranked among the most differentially expressed (**Table 2**). Although both enriched in Spp1+ cells, *Kcnip1* is highly expressed across the TRN while the expression level of *Cacng4* is relatively low (**Figure 15** and **Figure 16**).

Gene	P-Value	FDR
Col11a1	9.15E-15	1.52E-12
Kcnip1	**6.46E-10**	**5.36E-08**
Cc2d1b	2.69E-05	1.20E-03
Map1a	2.88E-05	1.20E-03
Itgav	4.62E-05	1.53E-03
Col1a2	9.43E-05	2.24E-03
Cacng4	**8.82E-05**	**2.24E-03**
Col4a5	2.27E-04	4.71E-03
Fam20c	2.75E-04	5.07E-03
Bag6	8.28E-04	1.38E-02
Fth1	1.07E-03	1.62E-02
Itga4	1.71E-03	2.37E-02
Itgb3	2.88E-03	3.67E-02
Stk39	3.34E-03	3.97E-02
C1rl	4.05E-03	4.36E-02
Atp2a2	4.20E-03	4.36E-02

Table 2 | Interaction partners of Spp1 which are enriched in Spp1+ cells. I examined all known interaction partners of Spp1 and calculated the p-value associated with the one-sided *t*-test of expression levels in Spp1+ cells compared to Ecel1+ cells for each gene, and adjusted for false discovery rate (FDR). There are 16 genes which passed the FDR threshold of 0.05.

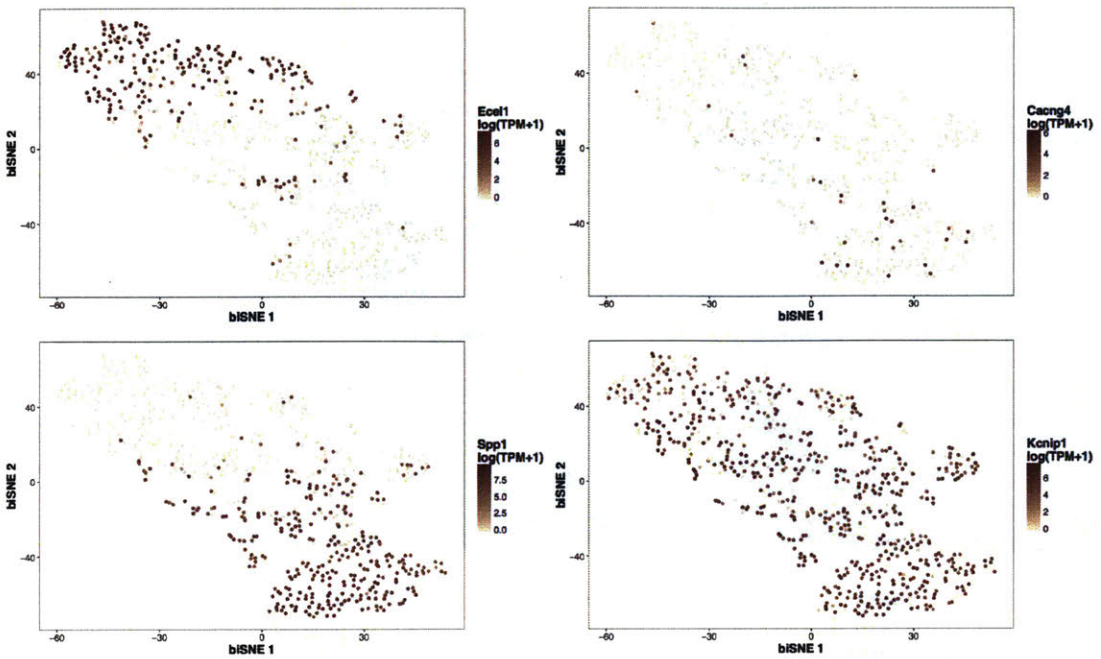

Figure 15 | *Cacng4* and *Kcnip1* are statistically enriched in Spp1+ cells. In the *t*-SNE embedding of TRN Gad2+Pvalb+ neurons, Spp1+ cells in the fourth (IV) quadrant have higher levels of *Cacng4* and *Kcnip1* expression compared to Ecel1+ cells in the second (II) quadrat, although *Kcnip1* is highly expressed across the entire TRN while *Cacng4* expression is restricted to a few cells.

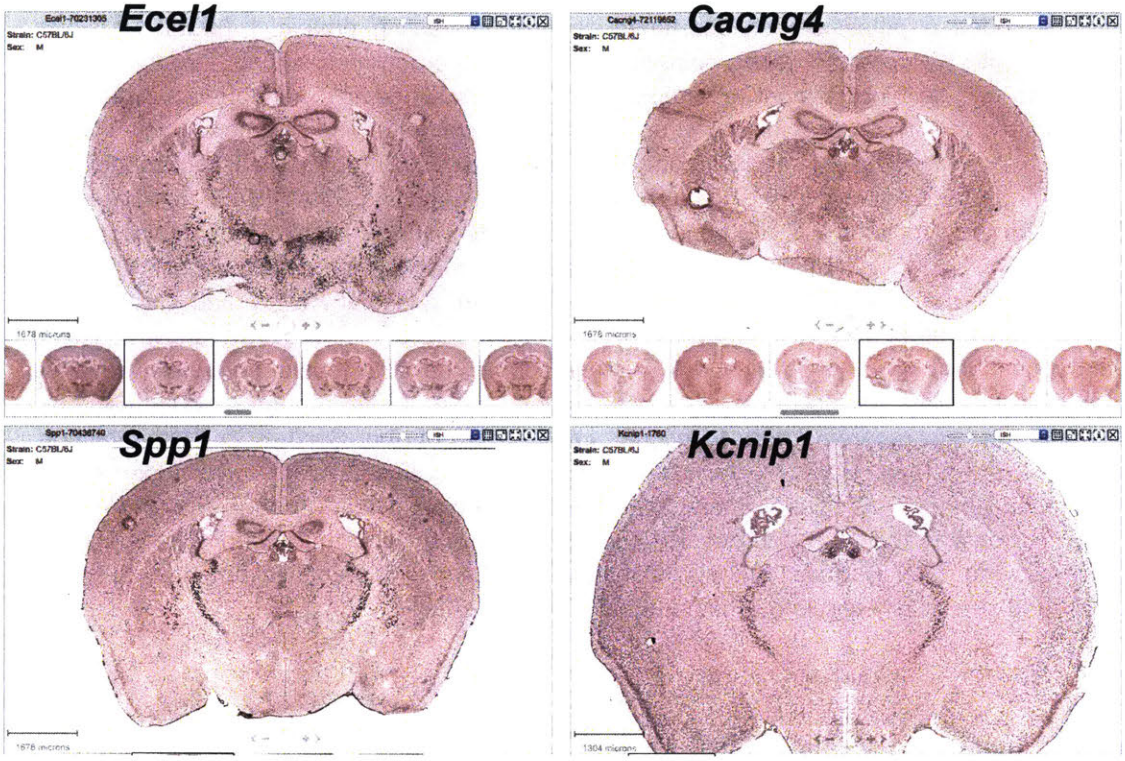

Figure 16 | *Cacng4* and *Kcnip1* are enriched in Spp1+ cells in the center of the TRN. *In situ* hybridization from Allen Brain Atlas[71] confirms that *Cacng4* expression is restricted to a few cells while *Kcnip1* is highly expressed in the TRN. *Ecel1* and *Spp1* expressions are shown as references. All slices are from adult mice (P56).

Cacng4 encodes transmembrane AMPAR regulatory protein (TARP) gamma 4. Members of the TARP gamma family associate with AMPAR on the plasma membrane and can profoundly influence gating. Studies of another gamma subunit, TARP gamma 2, suggested that AMPAR-TARP interaction destabilizes the closed state of AMPAR through electrostatic interactions between the extracellular domains[80,81], which are conserved across TARPs, including TARP gamma 4 (*Cacng4*[82]). In addition, an earlier GWAS identified rs17645023, which lies in the intergenic region upstream of *Cacng4*, to be sub-genome-wide significantly associated with schizophrenia and bipolar disorder[83] (p-value = 6.0E-7). These data suggest that *Cacng4* could be an interesting gene for follow-up studies. Given that glutamate receptor 3 (*Gria3*) are highly expressed in the TRN and enriched in Spp1+ cells (adjusted p-value [FDR] = 3.3E-4), it is therefore possible that *Cacng4* regulates Spp1+ neuron activity through modulating fast glutamatergic synaptic transmissions.

Kcnip1 encodes a potassium voltage-gated channel interacting protein 1 (KChIP1), which associates with Kv4 channel alpha subunit[84] and transduces calcium signals through the EF hand domains[85]. In hippocampal interneurons, KChIP1-Kv4.3 complex enables fast recovery from inhibition of A-type currents and stronger inhibitory control of firing[86]. Interestingly, T-type calcium channels bind to Kv4–KChIP complexes, and this coupling allows efficient modulation of Kv4 activity by calcium currents[87]. Moreover, a long non-coding RNA *NEAT1* directly binds KChIP1 and is associated with neuronal hyperexcitability states[88]. I found that *Kcnip1* (which encodes KChIP1), *kcnd2* (which encodes Kv4.2), and *Cacna1i* (which encodes $Ca_V3.3$) are all highly expressed in the TRN (**Figure 11** and **Figure 16**), and that *Kcnip1* and *Cacna1i* are selectively enriched in the TRN compared to other brain regions[27,29,67] (data not shown). These data suggest that Kv4.2–KChIP1–$Ca_V3.3$ interaction in the TRN could potentially regulate neuronal excitability and sleep spindle generation, most likely through Spp1+ neurons. Questions remain regarding the functional implications of the Spp1–KChIP1 interaction, particularly given that Spp1 is secreted and KChIP1 is intracellular[84]. We have begun investigating potential functional interactions between A-type potassium currents and T-type calcium currents in the TRN. Moreover, based on the observation that KChIP1 is selectively enriched in the TRN (**Figure 16**), we have started exploring therapeutic potential to target KChIP1 and specifically regulate TRN neuronal firing properties.

Section I Summary

Through computational analysis of the single-nucleus RNA sequencing data of 694 TRN $Gad2^+Pvalb^+$ neurons, and integrating genetics and proteomics data, I have found a gradient of cellular identity marked by *Ecel1* and *Spp1* expression, identified several gene candidates which could explain the electrophysiological and functional differences between Ecel1+ and Spp1+ cells, and formed new hypotheses on TRN neurobiology, particularly with respect to cholinergic transmission and sleep spindle generation.

SECTION II: TRN-SPECIFIC CO-EXPRESSION NETWORK

Background

Recent technological advances in mapping genomic, transcriptomic, and proteomic data have made it possible for us to generate unprecedented amounts of functional genomic datasets, which can be conveniently summarized as biological networks, where nodes represent genes and edges represent functional association between gene pairs. These networks have emerged as a powerful tool to dissect complex biological processes that would otherwise have been missed without a comprehensive genome-wide view of the functional associations[89,90].

It has been suggested that diverse functional genomics networks converge on a set of core features, such as

- Scale-free: most genes have few connections and only a handful of genes have many connections[91].
- Small-world: shortest paths between any pair of genes are usually small[92].
- Modular: some genes connect more strongly to each other than to the rest of the network[93].

These features correspond to the fundamental premises of molecular biology in that most genes collaborate with other genes to exert their functions concertedly[94], and that some essential genes represent hubs in the network with many functional connections[95].

Given the biological basis of functional networks, one of their key usage is to interpret large-scale genomic sequencing data. Combining functional genomics networks with exome-sequencing or genome-wide association studies (GWAS) is a cost-efficient and scalable method to uncover candidate cellular circuits enriched for genetic risk in a particular disease, which can be followed up in a targeted manner both computationally and experimentally to dissect disease mechanism and identify novel drug targets.

As a concrete example, in the context of psychiatric disorders, common variants from GWAS and rare variants from exome-sequencing and copy number variant (CNV) analysis have converged on protein complexes including chromatin remodeling complex, glutamate receptor, hyperpolarization-activated cyclic nucleotide-gated (HCN) channel, and the L-type calcium channel[37,39,96–98]. Therefore, using a protein-protein interaction network we developed earlier (InWeb3[56]), I previously identified a novel gene *NAGA* based on its topological similarity to a set of 65 ASD risk genes[39]. Interestingly, *NAGA* has a brain-specific expression quantitative trait loci (eQTL) from the GTEx portal (https://www.gtexportal.org/home/) and is also implicated in schizophrenia GWAS[37]. Mutations of *NAGA* have been implicated in Schindler disease[99], which has overlapping symptoms with ASDs[100]. These data suggest *NAGA* could be an interesting autism candidate gene for targeted follow-up studies (manuscript in preparation), and highlight the utility of network-based methods to interpret and augment psychiatric disease risk gene sets.

Constructing and benchmarking TRN-specific co-expression network (TRNNet)

To further exploit the unbiased nature of genomic networks to aid discovery of novel gene associations in psychiatric disorders, I first developed a TRN-specific co-expression network from the single-nucleus RNA-Sequencing data, treating each cell as independent samples. To construct this network, I applied the Weighted Gene Co-Expression Network (WGCNA) toolkit[101], chose the soft-power coefficient $\beta = 3$ to simulate the scale-free property of the co-expression network, and used a cutoff topological overlap matrix (TOM) coefficient 0.05, resulting in a network with 229,205 edges spanning 11,934 genes (called TRNNet henceforth). In TRNNet, an edge between two genes represents robust co-regulation between those genes in the 694 $Gad2^{+}Pvalb^{+}$ (primarily) TRN cells based on our single-nucleus RNA sequencing data. I used the Vertebrate Homology database from the Mouse Gene Informatics Portal (http://www.informatics.jax.org/homology.shtml, downloaded on March 16, 2017) and the HUGO Gene Nomenclature Committee (http://www.genenames.org/, downloaded on March 15, 2017) to convert mouse genes to human genes.

To establish the uniquely enabling features of TRNNet, we constructed a control co-expression network from the Gene Expression Omnibus[102]. Briefly, we chose the threshold of the expression matrix from Affymetrix arrays to exclude low quality samples by the following criteria:

1) We excluded samples where the ratio of signals coming from 3' end to that coming from 5' end is above 2 for *ACTB* and *GAPDH* gene transcripts. These two genes are standard control which reflect the extent of mRNA degradation and labeling accuracy.
2) We excluded samples where the average signal across all genes is below 150 based on the empirical distribution.
3) We excluded samples where the number of detected genes is less than 30.
4) We excluded samples where the goodness of fit to the scale-free (power-law) regression is less than 4 based on the empirical distribution.

We then constructed the gene correlation matrix from the 19,019 filtered samples, and applied global silencing[103] and network deconvolution[104] to remove indirect effects between pairs of genes. The resulting network (called GEONet henceforth) has 500,000 edges spanning 12,390 genes. Importantly, because the original samples of GEONet come from a variety of tissue types, it is a good control to compare with TRNNet which should possess tissue specificity relevant to studying neurodevelopmental disorders.

To examine potential neuronal specificity of TRNNet, I applied the random forest classifier ("Quack", manuscript in preparation, please see http://apps.broadinstitute.org/genets for details) to assess whether topological features in a particular network can be leveraged to identify neuronal pathway relationships. First, we curated a set of pathways that are likely involved in neurodevelopmental processes, using a text-mining approach on C2:CP (canonical pathways) and C5:BP (GO biological processes) gene sets from the Molecular Signature Database (MSigDB, http://software.broadinstitute.org/gsea/msigdb/). We found 306 such manually-curated pathways which have evidence to be involved in neuronal functions.

Next, I computed 18 topological properties for each gene in each pathway, including node degree, centrality, and local clustering coefficient metrics, and repeated the same calculations for context genes, which are defined as neighboring genes in the network but not pathway members, down-sampled to match the size of pathway genes (**Figure 17a**). Notably, pathway genes and context genes exhibit differential distributions of these topological features (data not shown). I then trained a random forest classifier with an ensemble of 500 decision trees which can optimally distinguish pathway genes from context genes based on the 18 topological features, and calculated the probability for each gene to belong to a particular pathway (**Figure 17b**).

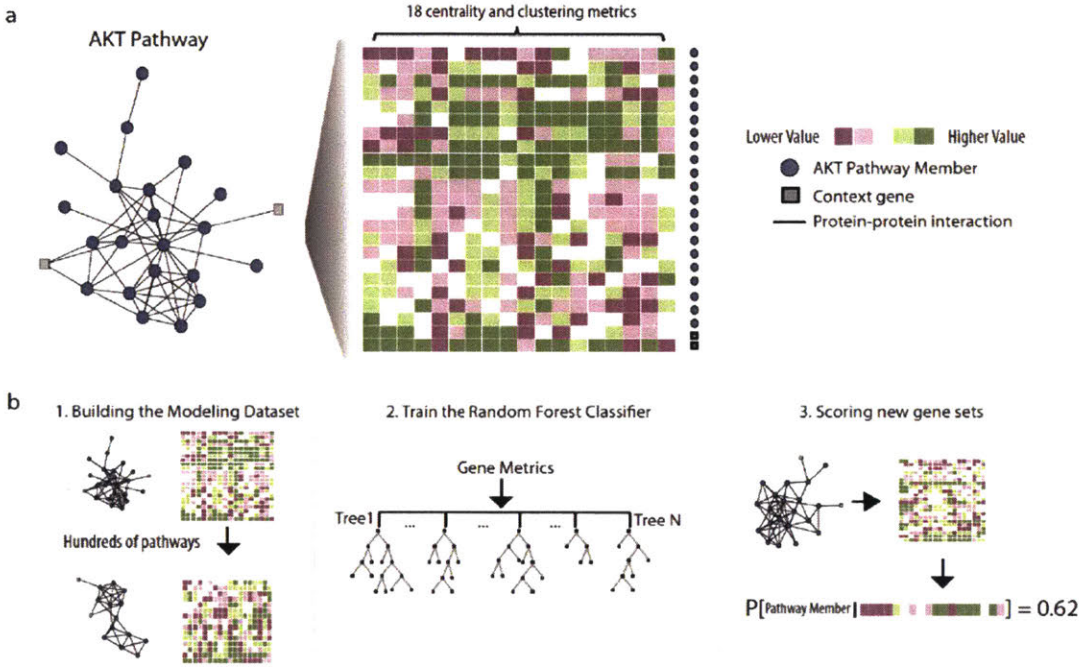

Figure 17 | Building a general classifier to predict pathway membership from networks. **a**) For a given pathway, we measured its topological properties exemplified here with the 21 genes of the AKT pathway in the InWeb3 protein-protein interaction network[56]. In the matrix, there are 18 topological properties shown as columns and the corresponding values for each of the 21 genes in the AKT pathway (black circles) as rows (metric values correspond to colors as indicated in the figure legend). One row in this matrix corresponds to one row in the final modeling dataset. We made the same measurements for genes in the context of the AKT pathway (white squares); only 2 of 2,449 context genes shown in the illustration. **b**) This procedure was repeated for 306 neuronal pathways from which the modeling dataset used to train the classifier is derived. For any candidate gene in a network, the classifier can assign a probability that it belongs to a pathway (e.g., the AKT pathway) as defined by the candidates' topological properties in the overall network and in relation to a specific set of genes (e.g., the 21 AKT genes).

For a subset of the 306 neuronal pathways ("training pathways"), I randomly masked 30% of the pathway members and asked Quack classifier to distinguish these held-out pathway genes from the context genes based on the topological features of the 70% of the remaining pathway genes. I then assessed the ability of a network-specific Quack model to make such predictions by computing the area under the receiver operating characteristic curves (AUCs) using the yet-unseen pathways ("validation pathways"). I found that even though TRNNet contains less data (694 single cells compared to 19,019 samples), it outperforms GEONet in learning the topological features of the neuronal pathways, with an AUC of 0.80 compared to 0.78 of GEONet (**Figure 18**). Further analyses have shown that co-expression networks constructed using tumor samples[105,106] performed worse in this task with AUCs <0.70 (data not shown). These results suggest that TRNNet is particularly suited for studying the architecture of pathways related to neurodevelopmental processes. Given that TRNNet is constructed using only TRN inhibitory neurons, it is highly likely that the increase in predictive power originates from its tissue specificity compared to the generic co-expression networks.

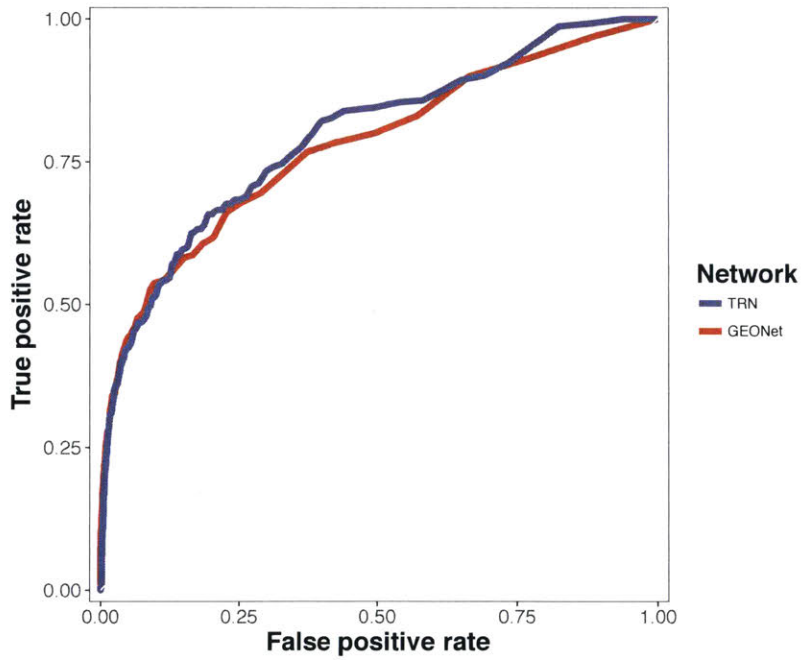

Figure 18 | Comparison of area under the receiver operating characteristic curve (AUC) between TRNNet and GEONet. I applied Quack classifier to assess the ability of TRNNet and GEONet to recapitulate pathway relationships using 306 curated neuronal gene sets. For each network, I trained the Quack random forest classifier using 70% of the pathways ("training pathways"), and validated the model using the 30% validation pathways. In the model training phase, I masked a random 30% of each training pathways asked each model to maximally segregate the held-out pathway members from the context genes, based on the 18 topological features of the 70% of the pathway genes. Among the validation pathways, the Quack model trained on the TRNNet has an AUC of 0.80 compared to that trained on GEONet which has an AUC of 0.78.

TRNNet predicts *ARL6IP4* as a potential novel ASD gene

I next applied the TRNNet-specific Quack model to analyze 65 genes implicated through genetics in ASD[39]. I found 226 genes with a Quack Probability (QuackP) larger than 0.3, which are likely to be functionally related to the 65 ASD seed genes because of their similar topological features in TRNNet. Among these 226 genes, I found ***FAM47A*** (QuackP = 0.54) which emerged as a new ASD candidate gene in a recent whole genome sequencing study (Yuen *et al*, 2017[41]), and ***DOCK8*** (QuackP = 0.40) which was reported in another recent candidate gene sequencing study (Stessman *et al*, 2017[40]). Further, I found a novel gene ***ARL6IP4*** (QuackP = 0.36). *ARL6IP4* has a brain-specific eQTL based on the GTEx portal (https://www.gtexportal.org/home/). Additionally, *ARL6IP4* resides in a locus defined by the linkage disequilibrium with SNP rs2851447 on chromosome 12, which is genome-wide significantly associated with risk of schizophrenia (p-value = 2.19E-14[37], **Figure 19**). Together, these independent genetic data and eQTL data suggests that *ARL6IP4* could be an interesting autism candidate gene that can be followed up with targeted functional experiments. Gene ontology annotations of *ARL6IP4* includes poly(A) RNA binding, which may reflect a novel aspect of ASD pathophysiology.

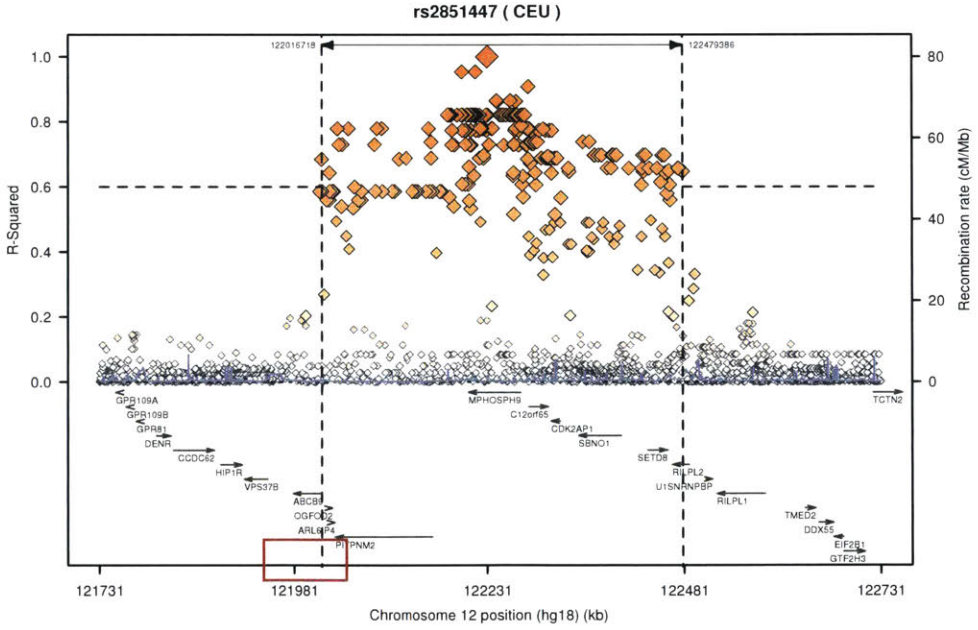

Figure 19 | TRNNet-specific Quack model predicts *ARL6IP4* as a novel ASD gene. *ARL6IP4* resides in a 14-gene locus indexed by SNP rs2851447 on chromosome 12. The linkage disequilibrium block is defined by a r^2 with the index SNP of at least 0.6 in the CEU population from the 1000 Genomes Project[43]. The *x*-axis represents genomic location and gene annotations, and the y-axis represents r^2 with the index SNP. *ARL6IP4* is highlighted in the red box. Plot generated by SNP Annotation and Proxy Search tool (http://archive.broadinstitute.org/mpg/snap/ldplot.php).

Section II Summary

I constructed a TRN-specific co-expression network based on single-nucleus RNA-Sequencing of 694 $Gad2^{+}Pvalb^{+}$ (primarily) TRN neurons (TRNNet). TRNNet exhibits tissue specificity and outperforms generic co-expression networks in predicting neuronal pathway structures. TRNNet-specific Quack model can replicate ASD risk genes emerging from recent sequencing studies, and predict novel ASD candidate genes when complemented by genetic and eQTL data. We have enabled the workflow of network analysis illustrated above on the GeNets web platform (http://apps.broadinstitute.org/genets) where users can easily leverage functional networks like TRNNet to interpret large-scale genomic data (manuscript in preparation).

SECTION III: ONGOING WORK

To complement TRN single-nucleus RNA sequencing data, I am working with various collaborators to perform follow-up experiments and computational analyses to further dissect interesting aspects of TRN neurobiology.

Part I: Causal Gene-Marker Pairs

My computational analysis of the 694 $Gad2^{+}Pvalb^{+}$ (primarily) TRN neurons are mainly based on the assumption that *Ecel1* and *Spp1* are not only marker genes, but also play functional roles in their respective cell populations. However, it could be the case that both genes are only associated with the phenotypic differences between Ecel1+ and Spp1+ cells, while there may exist other genes that causally explain the molecular mechanisms of such divergence. We have started investigating additional gene-marker pairs which could potentially serve as causal genes underlying the electrophysiological and functional differences between Ecel1+ and Spp1+ cells. One such pair is *Sst-Met*. In the *t*-SNE space, cells with high *Sst* expression roughly co-localize with Ecel1+ cells on the second (II) quadrant, while those with high *Met* expression co-localize with Spp1+ cells on the fourth (IV) quadrant. Out of the 694 neurons, 51 (7%) express both *Sst* and *Met*, and 192 (28%) express neither *Sst* nor *Met* (**Figure 20**). These proportions are similar to the case of *Ecel1-Spp1*, where 34 cells (5%) express both *Ecel1* and *Spp1* while 162 cells (23%) express neither.

The reason why the *Sst-Met* pair is particularly of interest is that both have compelling evidence as regulating important neurodevelopmental processes. *Met* is an established autism genetic risk factor[107] and MET signaling is critical in controlling the timing of neuronal growth, glutamatergic maturation, and cortical circuit function[108]. Somatostatin (*Sst*) is a classic marker for GABAergic interneurons whose expression is regulated by brain-derived neurotrophic factor (BDNF)[109]. Reduction in *Sst* expression has been described in diverse brain disorders, including mood disorders[110,111] and neurodegenerative diseases[112]. Therefore, it is possible that *Sst* and *Met* are functionally causal gene markers for the cell populations labeled by *Ecel1* and *Spp1* in our dataset.

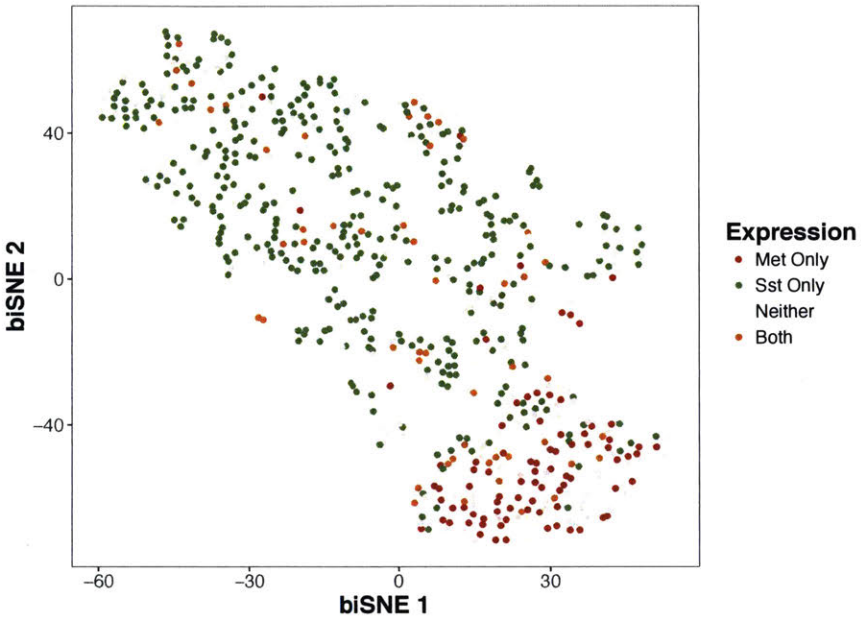

Figure 20 | *Sst* and *Met* mark two anti-correlated transcriptional programs in the TRN. Cells expressing *Sst* but not *Met* (green) roughly co-localizes with Ecel1+ cells on the second (II) quadrant of the *t*-SNE map, while cells expressing *Met* but not *Sst* (red) co-localizes with Spp1+ cells. Cells expressing both *Sst* and *Met* are labeled yellow, while double negative cells are labeled grey.

Part II: Molecular Circuits through which *PTCHD1* Regulates TRN Function

Mutations in the X-linked *PTCHD1* gene have been identified in approximately 1% of patients with intellectual disability and ASD, and these mutations significantly increase the risk of developing ASD-like behaviors[113]. However, the cellular mechanisms associated with *PTCHD1* is poorly understood. Recent evidence suggests that PTCHD1 protein binds with post-synaptic proteins PSD95 and SAB102, whose deficiency could cause excitatory synaptic dysfunction[114]. In the TRN, the expression of *PTCHD1* is highly enriched during early postnatal development. In our single-nucleus RNA sequencing dataset, *Ptchd1* is expressed in more than 80% of TRN neurons, although it is not differentially expressed between Ecel1+ and Spp1+ cells. Functionally, *PTCHD1* deletion attenuates TRN activity by reducing calcium-dependent potassium currents, thereby directly regulating attention, activity, and sleep[115]. These data suggest that studying *PTCHD1* function in the TRN could afford valuable insights into how TRN influences both normal and pathological neurodevelopmental processes.

We are performing two experiments towards this aim.

Comparing TRN Single-Cell RNA Sequencing between Wild-Type and *Ptchd1*$^{y/-}$ Mice

As a pilot experiment, we collected 16,952 cells covering the entire TRN area from four mice (P10), following the standard single cell protocol from 10X Genomics (10X Genomics, Pleasanton, CA). The four mice consist of two pairs of littermates with one *Ptchd1* knock-out (KO, mouse 2 and 4) and one wild type (WT, mouse 1 and 3) each, to ensure comparable environment and genetic background[116]. We used the default `cellranger aggr` from the Cell Ranger software v1.3 to aggregate data from all four mice, by downsampling and normalizing to have the same sequencing depth for each sample.

We applied the Seurat package for data processing[117] and identified 297 TRN neurons (**Figure 21a**). We did not observe significant batch effect, although there are many more cells from Mice 1 and 2 compared to those from 3 and 4 (**Figure 21b**). We have identified several robustly differentially expressed (DE) genes between WT and KO; however, the yield of TRN Gad2+Pvalb+ cells are fairly low (1.8%). We are working to improve the dissociation protocol to obtain more cells from more mice, which will increase power for DE gene detection.

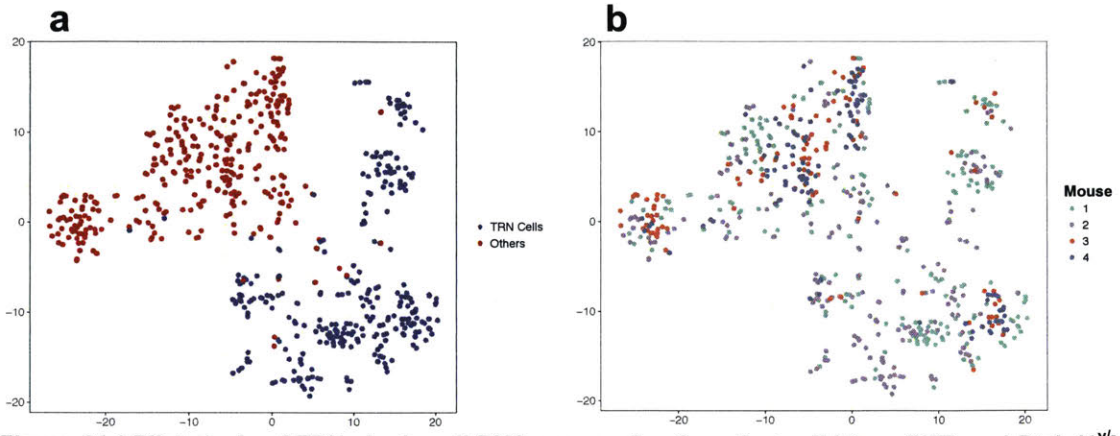

Figure 21 | Pilot study of TRN single-cell RNA sequencing from four wild-type (WT) and Ptchd1$^{y/-}$ (KO) mice. **a**) *t*-SNE embedding of TRN Gad2^{+}Pvalb^{+} neurons (blue, N = 297 cells) combined with excitatory neurons (red, N = 326 cells). **b**) Cells labeled by their mouse of origin. Mice 1 and 3 are WT, and mice 2 and 4 are KO. Among the 297 TRN neurons, there are 130 cells from mouse 1, 115 cells from mouse 2, 24 cells from mouse 3, and 28 cells from mouse 4.

Identifying the interaction partners of PTCHD1 protein using BioID

BioID is a new technique which can screen for protein interactions in living mammalian cells based on proximity[118]. By fusing a promiscuous biotin ligase to the bait protein, BioID identifies physiologically relevant interacting proteins in native cellular environment, including those with weak or transient interactions[119-120]. Roughly half of BioID-detected proteins likely reside within 20-30nm of the bait[119].

To identify interaction partners of PTCHD1 using BioID, we have transfected N-terminus-labeled *PTCHD1* construct into HEK293 cells and selected for stable clones. We have also transfected *TCF4*, *CACNB2*, *CUL3*, and *C4A* constructs. We plan to repeat these experiments in human glioblastoma cells (LN-229 from ATCC, Manassas, VA), and compare the interaction partners identified from two cell lines to assess cell-type specificity (or lack thereof).

Once we obtain proximity-based interaction partners of these genes, as quality control, we will cross-reference these interaction partners with those identified using traditional IP-MS approaches to evaluate the concordance between different experimental. Based on the important roles of *PTCHD1* in regulating TRN functions through calcium-dependent potassium currents[115], we plan to screen for evidence of interaction between PTCHD1 and SK channel proteins. We also plan to search among PTCHD1 interaction partners for enrichment of genes with rare deleterious variants and *de novo* variants in ASD and schizophrenia patients, and enrichment of genes with depletion of missense mutations in the normal population using the ExAC database[121]. Finally, we will overlay genes differentially-expressed in *Ptchd1*$^{y/-}$ mice with genes encoding proteins interacting with PTCHD1. All of these filtering analyses will facilitate prioritization of candidate genes for targeted follow-up experiments.

Part III: Elucidating Brain-Specific Pathways Using Single-Cell-Type Proteomics

As part of the ongoing Brain Interaction Network (BINe) project, we have performed systematic quantitative interaction proteomics followed by tandem mass spectrometry targeting genes implicated in psychiatric disorders using Ngn-induced human neurons. Such single-cell-type proteomic experiments have yielded compelling insights on human neurobiology.

For instance, one of such genes we have studied is *CACNA1C*, which encodes the alpha-1C subunit of L-type calcium channel ($Ca_V1.2$). *CACNA1C* is genome-wide significantly associated with schizophrenia[37]. We found 109 proteins interacting with $Ca_V1.2$ at a false discovery rate (FDR) threshold of 0.1, in which all parts of the L-type calcium channel machinery were identified (α, β, and $\alpha_2\delta$ subunits). We also detected an enrichment of known $Ca_V1.2$ interaction partners based on our comprehensive catalog of the human interactome (InWeb_IM)[55] (**Figure 22a**). Notably, compared to the proteomics data obtained in heart tissues, $Ca_V1.2$ interaction partners in the brain are expressed at much higher levels in the prenatal period based on the BrainSpan Atlas of the developing human brain[122] (**Figure 22b**), suggesting their important roles during development. Interestingly, we have also found evidence that $Ca_V1.2$ interacts with complement C4A (**Figure 10**). C4A is a member of the classical complement pathway and has recently been found to mediate synaptic pruning during postnatal development and established as a schizophrenia risk gene[123]. This observation has prompted us to study the potential functional role of the $Ca_V1.2$–C4A complex in regulating synapse elimination. Overall, $Ca_V1.2$ pulldown data are reproducible and of high quality, exhibit neuronal tissue-specificity, and have provided us with novel hypotheses which can direct future experiments.

We plan to screen for the interaction partners using this pipeline for the gene candidates we identified from TRN single-nucleus RNA sequencing data. Currently, we have planned for the pulldown experiments of one particular gene *GRIN2A*, which is enriched in Spp1+ cells and genome-wide significantly associated with schizophrenia (**Figure 11**).

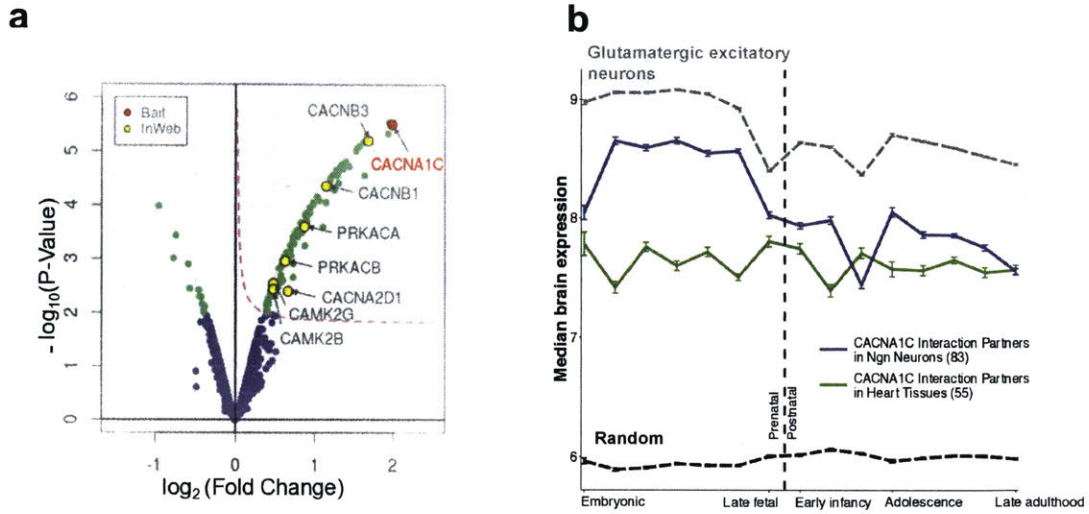

Figure 22 | CACNA1C proteomics on induced Ngn-human neurons. a) Volcano plot of experimental data. Green points represent identified proteins achieving FDR $q \leq 0.1$. $Ca_V1.2$ is the bait and most enriched protein (red), and known interaction partners of $Ca_V1.2$ in InWeb_IM are highlighted in yellow. *x*-axis represents the log fold change between antibody and IgG control, *y*-axis represents negative log p-value of the enrichment based on two replciates. **b**) BrainSpan[122] expression profile of $Ca_V1.2$ interaction partners in human neurons (blue, N = 83 proteins) compared to those in homogenized heart tissues[124] (green, N = 55 proteins). Grey dotted line represents genes highly expressed in glutamatergic excitatory neurons, and black dotted line represents random genes.

Part IV: Annotating Recent Psychiatric Genetics Datasets Using TRNNet

We have recently developed an efficient and powerful method (called network mutation burden, or NMB) to computationally assess the mutation burden of genes by aggregating the mutation frequencies of its neighboring genes in the network, effectively increasing power to detect genes with low mutation burden[125]. Using this method, we were able to annotate cancer driver genes and ASD risk genes with good accuracy[55]. We have also curated genes with multiple *de novo* protein-truncating mutations from 3,982 ASD trios[126], several cohorts of patients with intellectual disability[127–129], and 3,954 individuals with neurodevelopmental delay[42]. Given TRN's important role in sensory gating, we plan to apply NMB on TRNNet to uncover novel gene associations and identify pathways disturbed by pathogenic mutations in neurodevelopmental delay and ASD.

Section III Summary

We have initiated both targeted functional experiments based on interesting candidate genes, and computational analyses using orthogonal datasets, to elucidate the molecular circuits through which TRN regulates sensory gating. Such efforts have resulted in deep collaborations with laboratories not only at MIT and Harvard but also at McGill University in Canada.

SECTION IV: OVERALL SUMMARY

Single-cell technologies have enabled us to dissect the molecular basis of cellular diversity in complex tissues with unprecedented resolution. By applying state-of-the-art computational methods on our single-nucleus RNA sequencing data, I discovered many novel insights into the mechanisms through which TRN regulates sensory gating, and formulated actionable hypotheses to further interrogate TRN neurobiology. Importantly, many of these findings have profound implications for the diagnosis and treatment of psychiatric disorders like autism spectrum disorders, and I have initiated follow-up experiments through collaborations to identify promising therapeutic targets based on these data.

More generally, my thesis established a comprehensive analytical pipeline where I complemented high-throughput single-cell transcriptomics data with unbiased human genetics and functional proteomics data. I have demonstrated that such deep integration of diverse data types can yield important biological insights. As large-scale data-driven research become increasingly common in biomedicine, this work could serve as a useful exemplar for future investigations.

REFERENCES

1. Christensen, D. L. *et al.* Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years--Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR. Surveill. Summ.* **65,** 1–23 (2016).
2. Lord, C. & Bishop, S. L. Recent advances in autism research as reflected in DSM-5 criteria for autism spectrum disorder. *Annu. Rev. Clin. Psychol.* **11,** 53–70 (2015).
3. Frazier, T. W. *et al.* Validation of proposed DSM-5 criteria for autism spectrum disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **51,** 28–40.e3 (2012).
4. Mandy, W., Charman, T., Puura, K. & Skuse, D. Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: evidence from Finnish and UK samples. *Autism* **18,** 45–54 (2014).
5. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48,** 552–5 (2016).
6. Robinson, E. B., Neale, B. M. & Hyman, S. E. Genetic research in autism spectrum disorders. *Curr. Opin. Pediatr.* **27,** 685–91 (2015).
7. Pankevich, D. E., Altevogt, B. M., Dunlop, J., Gage, F. H. & Hyman, S. E. Improving and accelerating drug development for nervous system disorders. *Neuron* **84,** 546–553 (2014).
8. Marco, E. J., Hinkley, L. B. N., Hill, S. S. & Nagarajan, S. S. Sensory processing in autism: a review of neurophysiologic findings. *Pediatr. Res.* **69,** 48R–54R (2011).
9. Pfeiffer, B., Daly, B. P., Nicholls, E. G. & Gullo, D. F. Assessing sensory processing problems in children with and without attention deficit hyperactivity disorder. *Phys. Occup. Ther. Pediatr.* **35,** 1–12 (2015).
10. Javitt, D. C. & Freedman, R. Sensory processing dysfunction in the personal experience and neuronal machinery of schizophrenia. *Am. J. Psychiatry* **172,** 17–31 (2015).
11. Anguera, J. A. *et al.* A pilot study to determine the feasibility of enhancing cognitive abilities in children with sensory processing dysfunction. *PLoS One* **12,** e0172616 (2017).
12. Aguilar, J. R. & Castro-Alamancos, M. A. Spatiotemporal gating of sensory inputs in thalamus during quiescent and activated states. *J. Neurosci.* **25,** 10990–1002 (2005).
13. Pinault, D. The thalamic reticular nucleus: structure, function and concept. *Brain Res. Brain Res. Rev.* **46,** 1–31 (2004).
14. Huguenard, J. R. & McCormick, D. A. Thalamic synchrony and dynamic regulation of global forebrain oscillations. *Trends Neurosci.* **30,** 350–6 (2007).
15. Bolkan, S. & Gordon, J. A. Neuroscience: Untangling autism. *Nature* 5–6 (2016). doi:10.1038/nature17311
16. Neyer, C. *et al.* mGluR-mediated calcium signalling in the thalamic reticular nucleus. *Cell Calcium* **59,** 312–23 (2016).
17. Manoach, D. S. *et al.* Sleep spindle deficits in antipsychotic-naïve early course schizophrenia and in non-psychotic first-degree relatives. *Front. Hum. Neurosci.* **8,** 762 (2014).
18. Limoges, E., Mottron, L., Bolduc, C., Berthiaume, C. & Godbout, R. Atypical sleep architecture and the autism phenotype. *Brain* **128,** 1049–61 (2005).
19. Lee, S.-H., Govindaiah, G. & Cox, C. L. Heterogeneity of firing properties among rat thalamic reticular nucleus neurons. *J. Physiol.* **582,** 195–208 (2007).
20. Deleuze, C. & Huguenard, J. R. Distinct electrical and chemical connectivity maps in the thalamic reticular nucleus: potential roles in synchronization and sensation. *J. Neurosci.* **26,** 8633–45 (2006).
21. Lam, Y.-W. & Sherman, S. M. Functional organization of the thalamic input to the thalamic reticular nucleus. *J. Neurosci.* **31,** 6791–9 (2011).
22. Lam, Y.-W. & Sherman, S. M. Functional topographic organization of the motor

reticulothalamic pathway. *J. Neurophysiol.* **113,** 3090–7 (2015).
23. Halassa, M. M. *et al.* State-dependent architecture of thalamic reticular subnetworks. *Cell* **158,** 808–21 (2014).
24. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498,** 236–40 (2013).
25. Lohr, J. G. *et al.* Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat. Biotechnol.* **32,** 479–84 (2014).
26. Kumar, R. M. *et al.* Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516,** 56–61 (2014).
27. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347,** 1138–42 (2015).
28. Kaiser, T., Ting, J. T., Monteiro, P. & Feng, G. Transgenic labeling of parvalbumin-expressing neurons with tdTomato. *Neuroscience* **321,** 236–45 (2016).
29. Habib, N. *et al.* Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353,** 925–8 (2016).
30. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-sne. *J. Mach. Learn. Res.* **9,** 2579–2605 (2008).
31. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10,** 1093–5 (2013).
32. Gonzalez-Burgos, G. GABA transporter GAT1: a crucial determinant of GABAB receptor activation in cortical circuits? *Adv. Pharmacol.* **58,** 175–204 (2010).
33. Kochubey, O., Lou, X. & Schneggenburger, R. Regulation of transmitter release by Ca(2+) and synaptotagmin: insights from a large CNS synapse. *Trends Neurosci.* **34,** 237–46 (2011).
34. Chen, C., Arai, I., Satterfield, R., Young, S. M. & Jonas, P. Synaptotagmin 2 Is the Fast Ca(2+) Sensor at a Central Inhibitory Synapse. *Cell Rep.* **18,** 723–736 (2017).
35. Millet, Q., Lolignier, S., Morohashi, T. & Gossage, S. J. Mapping protein interactions of sodium channel Na V 1 . 7 using epitope-tagged gene targeted mice. *bioRxiv* (2017). doi:https://doi.org/10.1101/118497
36. Dib-Hajj, S. D., Yang, Y., Black, J. A. & Waxman, S. G. The Na(V)1.7 sodium channel: from molecule to man. *Nat. Rev. Neurosci.* **14,** 49–62 (2013).
37. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–7 (2014).
38. Barrett, J. C. *et al.* New mutations, old statistical challenges. *bioRxiv* (2017).
39. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87,** 1215–33 (2015).
40. Stessman, H. A. F. *et al.* Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49,** 515–526 (2017).
41. C Yuen, R. K. *et al.* Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20,** 602–611 (2017).
42. McRae, J. F. *et al.* Prevalence and architecture of de novo mutations in developmental disorders. *Nature* (2017). doi:10.1038/nature21062
43. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
44. Makinson, C. D. *et al.* Regulation of Thalamic and Cortical Network Synchrony by Scn8a. *Neuron* **93,** 1165–1179.e6 (2017).
45. Nickels, K. C., Zaccariello, M. J., Hamiwka, L. D. & Wirrell, E. C. Cognitive and neurodevelopmental comorbidities in paediatric epilepsy. *Nat. Rev. Neurol.* **12,** 465–76 (2016).
46. Valdenaire, O., Richards, J. G., Faull, R. L. & Schweizer, A. XCE, a new member of the

endothelin-converting enzyme and neutral endopeptidase family, is preferentially expressed in the CNS. *Brain Res. Mol. Brain Res.* **64,** 211–21 (1999).
47. Benoit, A., Vargas, M. A., Desgroseillers, L. & Boileau, G. Endothelin-converting enzyme-like 1 (ECEL1) is present both in the plasma membrane and in the endoplasmic reticulum. *Biochem. J.* **380,** 881–8 (2004).
48. Dieterich, K. *et al.* The neuronal endopeptidase ECEL1 is associated with a distinct form of recessive distal arthrogryposis. *Hum. Mol. Genet.* **22,** 1483–92 (2013).
49. Ul-Haq, Z., Iqbal, S. & Moin, S. T. Dynamic changes in the secondary structure of ECE-1 and XCE account for their different substrate specificities. *BMC Bioinformatics* **13,** 285 (2012).
50. Denhardt, D. T. & Guo, X. Osteopontin: a protein with diverse functions. *FASEB J.* **7,** 1475–82 (1993).
51. Ashkar, S. *et al.* Eta-1 (osteopontin): an early component of type-1 (cell-mediated) immunity. *Science* **287,** 860–4 (2000).
52. Wang, K. X. & Denhardt, D. T. Osteopontin: role in immune regulation and stress responses. *Cytokine Growth Factor Rev.* **19,** 333–45
53. Briones-Orta, M. A. *et al.* Osteopontin splice variants and polymorphisms in cancer progression and prognosis. *Biochim. Biophys. Acta* **1868,** 93–108 (2017).
54. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).
55. Li, T. *et al.* A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods* **14,** 61–64 (2017).
56. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25,** 309–16 (2007).
57. Kotlyar, M., Pastrello, C., Sheahan, N. & Jurisica, I. Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic Acids Res.* **44,** D536-41 (2016).
58. Calderone, A., Castagnoli, L. & Cesareni, G. mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods* **10,** 690–1 (2013).
59. Das, J. & Yu, H. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst. Biol.* **6,** 92 (2012).
60. Cowley, M. J. *et al.* PINA v2.0: mining interactome modules. *Nucleic Acids Res.* **40,** D862-5 (2012).
61. Razick, S., Magklaras, G. & Donaldson, I. M. iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics* **9,** 405 (2008).
62. Turner, B. *et al.* iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database (Oxford).* **2010,** baq023 (2010).
63. Edwards, D. R., Handsley, M. M. & Pennington, C. J. The ADAM metalloproteinases. *Mol. Aspects Med.* **29,** 258–89 (2008).
64. Ozkaynak, E. *et al.* Adam22 is a major neuronal receptor for Lgi4-mediated Schwann cell signaling. *J. Neurosci.* **30,** 3857–64 (2010).
65. Lovero, K. L., Fukata, Y., Granger, A. J., Fukata, M. & Nicoll, R. A. The LGI1-ADAM22 protein complex directs synapse maturation through regulation of PSD-95 function. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E4129-37 (2015).
66. McHardy, S. F., Wang, H.-Y. L., McCowen, S. V & Valdez, M. C. Recent advances in acetylcholinesterase Inhibitors and Reactivators: an update on the patent literature (2012-2015). *Expert Opin. Ther. Pat.* **27,** 455–476 (2017).
67. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19,** 335–46 (2016).
68. Lerea, C. L., Bunt-Milam, A. H. & Hurley, J. B. Alpha transducin is present in blue-, green-, and red-sensitive cone photoreceptors in the human retina. *Neuron* **3,** 367–76 (1989).

69. Vastrik, I. *et al.* Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.* **8,** R39 (2007).
70. Esbenshade, T. A. *et al.* The histamine H3 receptor: an attractive target for the treatment of cognitive disorders. *Br. J. Pharmacol.* **154,** 1166–81 (2008).
71. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445,** 168–76 (2007).
72. Newcombe, R. G. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat. Med.* **17,** 873–90 (1998).
73. Kolmac, C. I. & Mitrofanis, J. Patterns of brainstem projection to the thalamic reticular nucleus. *J. Comp. Neurol.* **396,** 531–43 (1998).
74. Kroeger, D. *et al.* Cholinergic, Glutamatergic, and GABAergic Neurons of the Pedunculopontine Tegmental Nucleus Have Distinct Effects on Sleep/Wake Behavior in Mice. *J. Neurosci.* **37,** 1352–1366 (2017).
75. Van Dort, C. J. *et al.* Optogenetic activation of cholinergic neurons in the PPT or LDT induces REM sleep. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 584–9 (2015).
76. Ni, K. M. *et al.* Selectively driving cholinergic fibers optically in the thalamic reticular nucleus promotes sleep. *Elife* **5,** 1–17 (2016).
77. Sun, Y.-G. *et al.* Biphasic cholinergic synaptic transmission controls action potential activity in thalamic reticular nucleus neurons. *J. Neurosci.* **33,** 2048–59 (2013).
78. Mena-Segovia, J. & Bolam, J. P. Rethinking the Pedunculopontine Nucleus: From Cellular Organization to Function. *Neuron* **94,** 7–18 (2017).
79. Sun, Y.-G. *et al.* mGluR1 and mGluR5 Synergistically Control Cholinergic Synaptic Transmission in the Thalamic Reticular Nucleus. *J. Neurosci.* **36,** 7886–96 (2016).
80. Ben-Yaacov, A. *et al.* Molecular Mechanism of AMPA Receptor Modulation by TARP/Stargazin. *Neuron* **93,** 1126–1137.e4 (2017).
81. Twomey, E. C., Yelshanskaya, M. V, Grassucci, R. A., Frank, J. & Sobolevsky, A. I. Elucidation of AMPA receptor-stargazin complexes by cryo-electron microscopy. *Science* **353,** 83–6 (2016).
82. Cais, O. *et al.* Mapping the interaction sites between AMPA receptors and TARPs reveals a role for the receptor N-terminal domain in channel gating. *Cell Rep.* **9,** 728–40 (2014).
83. Curtis, D. *et al.* Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes. *Psychiatr. Genet.* **21,** 1–4 (2011).
84. An, W. F. *et al.* Modulation of A-type potassium channels by a family of calcium sensors. *Nature* **403,** 553–6 (2000).
85. Ames, J. B. & Lim, S. Molecular structure and target recognition of neuronal calcium sensor proteins. *Biochim. Biophys. Acta* **1820,** 1205–13 (2012).
86. Bourdeau, M. L., Laplante, I., Laurent, C. E. & Lacaille, J.-C. KChIP1 modulation of Kv4.3-mediated A-type K(+) currents and repetitive firing in hippocampal interneurons. *Neuroscience* **176,** 173–87 (2011).
87. Anderson, D. *et al.* Regulation of neuronal activity by Cav3-Kv4 channel signaling complexes. *Nat. Neurosci.* **13,** 333–7 (2010).
88. Barry, G. *et al.* The long non-coding RNA NEAT1 is responsive to neuronal activity and is associated with hyperexcitability states. *Sci. Rep.* **7,** 40127 (2017).
89. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12,** 56–68 (2011).
90. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5,** 101–13 (2004).
91. Barabási, A.-L. Scale-free networks: a decade and beyond. *Science* **325,** 412–3 (2009).
92. Grabow, C., Grosskinsky, S., Kurths, J. & Timme, M. Collective relaxation dynamics of small-world networks. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **91,** 52815 (2015).

93. Mitra, K., Carvunis, A.-R., Ramesh, S. K. & Ideker, T. Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet.* **14,** 719–32 (2013).
94. Lage, K. Protein-protein interactions and genetic diseases: The interactome. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842,** 1971–1980 (2014).
95. Goh, K.-I. *et al.* The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 8685–90 (2007).
96. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485,** 242–5 (2012).
97. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet (London, England)* **381,** 1371–9 (2013).
98. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485,** 246–50 (2012).
99. Clark, N. E. & Garman, S. C. The 1.9 a structure of human alpha-N-acetylgalactosaminidase: The molecular basis of Schindler and Kanzaki diseases. *J. Mol. Biol.* **393,** 435–47 (2009).
100. Bakker, H. D. *et al.* Human alpha-N-acetylgalactosaminidase (alpha-NAGA) deficiency: no association with neuroaxonal dystrophy? *Eur. J. Hum. Genet.* **9,** 91–6 (2001).
101. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).
102. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30,** 207–10 (2002).
103. Barzel, B. & Barabási, A.-L. Network link prediction by global silencing of indirect correlations. *Nat. Biotechnol.* **31,** 720–5 (2013).
104. Feizi, S., Marbach, D., Médard, M. & Kellis, M. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat. Biotechnol.* **31,** 726–33 (2013).
105. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313,** 1929–35 (2006).
106. Cowley, G. S. *et al.* Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci. data* **1,** 140035 (2014).
107. Peng, Y., Huentelman, M., Smith, C. & Qiu, S. MET receptor tyrosine kinase as an autism genetic risk factor. *Int. Rev. Neurobiol.* **113,** 135–65 (2013).
108. Peng, Y. *et al.* The autism-associated MET receptor tyrosine kinase engages early neuronal growth mechanism and controls glutamatergic circuits development in the forebrain. *Mol. Psychiatry* **21,** 925–35 (2016).
109. McKinney, B. C. *et al.* Hypermethylation of BDNF and SST Genes in the Orbital Frontal Cortex of Older Individuals: A Putative Mechanism for Declining Gene Expression with Age. *Neuropsychopharmacology* **40,** 2604–13 (2015).
110. Lin, L.-C. & Sibille, E. Reduced brain somatostatin in mood disorders: a common pathophysiological substrate and drug target? *Front. Pharmacol.* **4,** 110 (2013).
111. Tripp, A., Kota, R. S., Lewis, D. A. & Sibille, E. Reduced somatostatin in subgenual anterior cingulate cortex in major depression. *Neurobiol. Dis.* **42,** 116–24 (2011).
112. Grosser, C., Neumann, L., Horsthemke, B., Zeschnigk, M. & van de Nes, J. Methylation analysis of SST and SSTR4 promoters in the neocortex of Alzheimer's disease patients. *Neurosci. Lett.* **566,** 241–6 (2014).
113. Chaudhry, A. *et al.* Phenotypic spectrum associated with PTCHD1 deletions and truncating mutations includes intellectual disability and autism spectrum disorder. *Clin. Genet.* **88,** 224–33 (2015).
114. Ung, D. C. *et al.* Ptchd1 deficiency induces excitatory synaptic and cognitive dysfunctions

in mouse. *Mol. Psychiatry* 1–12 (2017). doi:10.1038/mp.2017.39
115. Wells, M. F., Wimmer, R. D., Schmitt, L. I., Feng, G. & Halassa, M. M. Thalamic reticular impairment underlies attention deficit in Ptchd1(Y/-) mice. *Nature* **532,** 58–63 (2016).
116. Holmdahl, R. & Malissen, B. The need for littermate controls. *Eur. J. Immunol.* **42,** 45–7 (2012).
117. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33,** 495–502 (2015).
118. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196,** 801–10 (2012).
119. Roux, K. J., Kim, D. I. & Burke, B. BioID: a screen for protein-protein interactions. *Curr. Protoc. protein Sci.* **74,** Unit 19.23. (2013).
120. Mehus, A. A., Anderson, R. H. & Roux, K. J. BioID Identification of Lamin-Associated Proteins. *Methods Enzymol.* **569,** 3–22 (2016).
121. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–91 (2016).
122. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508,** 199–206 (2014).
123. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530,** 177–83 (2016).
124. Lundby, A. *et al.* Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. *Nat. Methods* **11,** 868–74 (2014).
125. Horn, H. *et al.* A comparative analysis of network mutation burdens across 21 tumor types augments discovery from cancer genomes. *bioRxiv* (2015). doi:10.1101/025445
126. Kosmicki, J. A. *et al.* Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49,** 504–510 (2017).
127. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet (London, England)* **380,** 1674–82 (2012).
128. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367,** 1921–9 (2012).
129. Lelieveld, S. H. *et al.* Meta-analysis of 2,104 trios provides support for 10 novel candidate genes for intellectual disability. *Nat. Neurosci.* **19,** 1194–1196 (2016).