TechSights

Brain Transcriptome Databases: A User's Guide

Jason M. Keil,[1,2,3] **Adel Qalieh,**[1,2] and **Kenneth Y. Kwan**[1,2]

[1]Molecular and Behavioral Neuroscience Institute, [2]Department of Human Genetics, and [3]Medical Scientist Training Program, University of Michigan, Ann Arbor, Michigan 48109

Transcriptional programs instruct the generation and maintenance of diverse subtypes of neural cells, establishment of distinct brain regions, formation and function of neural circuits, and ultimately behavior. Spatiotemporal and cell type-specific analyses of the transcriptome, the sum total of all RNA transcripts in a cell or an organ, can provide insights into the role of genes in brain development and function, and their potential contribution to disorders of the brain. In the previous decade, advances in sequencing technology and funding from the National Institutes of Health and private foundations for large-scale genomics projects have led to a growing collection of brain transcriptome databases. These valuable resources provide rich and high-quality datasets with spatiotemporal, cell type-specific, and single-cell precision. Most importantly, many of these databases are publicly available via user-friendly web interface, making the information accessible to individual scientists without the need for advanced computational expertise. Here, we highlight key publicly available brain transcriptome databases, summarize the tissue sources and methods used to generate the data, and discuss their utility for neuroscience research.

Key words: transcriptome; database; RNA-seq

Introduction

The exquisite control of spatiotemporal gene expression enables the specification of diverse neural cell types, the development of distinct brain regions, the wiring and function of neural circuits, and ultimately controls behavior. For decades, individual genes have been studied using lower-throughput methods, such as *in situ* hybridization and RT-PCR, to query their expression levels and spatiotemporal patterns. The advents of microarray technology and RNA sequencing (RNA-seq) made possible genome-wide, unbiased interrogation of the transcriptome, the sum total of all RNA transcripts in a cell or an organ. Recent advances in sequencing technology, cell isolation techniques, genetic access to specific cell types, and data analysis have enabled transcriptomic studies with increasingly greater precision and granularity, giving new insights into gene expression in specific organs, cell types, and single cells. For neuroscience, large-scale transcriptomic data hold tremendous potential to inform molecular and cellular brain studies, the neural substrates and biomarkers of brain disorders, the validity of *in vitro* and *in vivo* models, and potential therapeutic strategies for neurological and psychiatric disease.

Recognizing the importance of gene expression data for basic and translational research, the National Institutes of Health

and private foundations, notably the Allen Institute for Brain Science, have prioritized funding for large-scale, often collaborative efforts to catalog and analyze the transcriptomes of cells and tissues in humans, nonhuman primates, and model organisms. Importantly, data sharing of the resulting transcriptome datasets has become common. Journal publishers and funders have put in place policies for deposition of transcriptome data into open repositories such as Gene Expression Omnibus and Sequence Read Archive (SRA) to drive further analyses by other groups and enable across group comparisons. Importantly, many datasets are housed in user-friendly databases, where individual scientists without advanced data analysis expertise can query and access the data via web interface. These databases have tremendous additional value. They condense what could otherwise be an overwhelming amount of data into a format that is easily accessible to the research community and thus can propel basic and translational research in individual laboratories.

In this review, we highlight publicly available brain transcriptome databases that can be accessed without specialized computational expertise (Table 1), focusing on where to access the data, what types of data are available, how they may be used for research, and what the considerations are for the use of these resources. We organize these databases based on the type of transcriptome analysis: spatiotemporal, cell type-specific, single-cell, and integrative.

Highlighted databases

Spatiotemporal analysis

The brain is functionally organized into regions, which are distinguished by distinct compositions of molecularly defined cell types and region-dependent patterns of long- and short-range connectivities. These anatomical, circuit, and functional

Table 1. Highlighted brain transcriptome databases [a]

Analysis	Web Interface	Reference	Species	Age	Sample	Method	Isoform	Accession
Spatiotemporal	http://hbatlas.org	Johnson et al., 2009	Human	Lifespan	Multi, macrodissection	Microarray	—	GSE13344
		Kang et al., 2011						GSE25219
	http://hbatlas.org/mouseNCXtranscriptome	Fertuzinhos et al., 2014	Mouse	Postnatal	Ctx layer, microdissection	RNA-seq	—	SRP031888
	http://www.blueprintnhpatlas.org	Bakken et al., 2016	Macaque	Lifespan	Multi, macrodissection, and LMD	Microarray	—	At database
Spatial	http://human.brain-map.org	Hawrylycz et al., 2012	Human	Adult	Multi, macrodissection, and LMD	Microarray	—	At database
	http://genserv.anat.ox.ac.uk/layers	Belgard et al., 2011	Mouse	Adult	Ctx layer, microdissection	RNA-seq	+	GSE27243
	http://rakiclab.med.yale.edu/transcriptome	Ayoub et al., 2011	Mouse	Embryonic	Ctx embryonic layer, LMD	RNA-seq	+	GSE30765
	http://www.brainspan.org/lcm	Miller et al., 2014	Human	Midfetal	Multi, LMD	Microarray	—	At database
	https://www.gtexportal.org	GTEx Consortium, 2015	Human	Adult	Many tissues and cell lines	RNA-seq	+	At database
Temporal	http://braincloud.jhmi.edu	Colantuoni et al., 2011	Human	Lifespan	Prefrontal Ctx, macrodissection	Microarray	—	GSE30272
Cell type- specific	http://brainrnaseq.org	Zhang et al., 2014	Mouse	Adult	Ctx, genetic labeling, immunopanning	RNA-seq	+	GSE52564
		Zhang et al., 2016	Human	Fetal/adult	Ctx, Hp, immunopanning	RNA-seq	+	GSE73721
	http://genetics.wustl.edu/jdlab/csea-tool-2	Doyle et al., 2008	Mouse	Adult	Multi, genetic labeling, ribosome affinity purification	Microarray	—	GSE13379
		Xu et al., 2014						
	http://decon.fas.harvard.edu	Molyneaux et al., 2015	Mouse	Embryonic	Ctx, transcription factor FACS	RNA-seq	+	GSE63482
	http://hipposeq.janelia.org	Cembrowski et al., 2016	Mouse	Adult	Hp, genetic labeling, manual selection	RNA-seq	—	GSE74985
	http://neuroseq.janelia.org	Sugino et al., 2017	Mouse	Adult	Multi, genetic labeling, manual selection	RNA-seq	+	GSE79238
Single-cell	http://linnarssonlab.org/cortex	Zeisel et al., 2015	Mouse	Adult	Ctx, Fluidigm	RNA-seq	—	GSE60361
	http://genebrowser.unige.ch/science2016	Telley et al., 2016	Mouse	Embryonic	Ctx, ventricle dye, FACS, Fluidigm	RNA-seq	—	NA
	https://portals.broadinstitute.org/single_cell	Shekhar et al., 2016	Mouse	Adult	Retina, genetic labeling, Drop-seq	RNA-seq	—	GSE81905
	https://portals.broadinstitute.org/single_cell	Habib et al., 2016	Mouse	Adult	Hp, single nuclei, FACS, sNuc-seq	RNA-seq	—	GSE84371
	https://bit.ly/cortexSingleCell	Nowakowski et al., 2017	Human	Fetal	Ctx, ganglionic eminence, Fluidigm	RNA-seq	—	PRJNA295469
	http://gbmseq.org	Darmanis et al., 2017	Human	Adult	Ctx tumor, immunopanning, FACS	RNA-seq	—	GSE84465
Integrative	https://www.encodeproject.org	ENCODE Project Consortium, 2012	Many	Many	Many tissues and cell lines	Multiomics	+	Many
	http://celltypes.brain-map.org	Tasic et al., 2016	Mouse	Adult	Ctx, genetic labeling, FACS	RNA-seq	—	GSE71585

[a]Ctx, Cortex; Hp, hippocampus; multi, multiple brain regions. Isoform column indicates availability of isoform information via web interface.

differences are reflected in the transcriptome. Several transcriptomic studies have therefore sought to capture region-specific differences in gene expression profiles, leveraging a variety of dissection techniques ranging from manual macrodissection to laser microdissection. Furthermore, the brain undergoes protracted periods of development, refinement, and maturation, spanning the early fetal periods to adolescence. The brain can also undergo aging and degeneration, especially in humans. To capture the temporal transcriptomic differences in the brain, several transcriptomic studies have analyzed brain samples from distinct stages of development and aging.

In one of the largest human brain transcriptome studies to date, Kang et al. (2011) analyzed 16 human brain regions (11 neocortical areas, cerebellar cortex, mediodorsal nucleus of the thalamus, striatum, amygdala, and hippocampus) from 57 postmortem human brains. The tissue samples were collected from early fetal development (5.7 weeks after conception) through aging (82 years), essentially covering the entire human lifespan, making this a comprehensive database of the spatiotemporal transcriptome of the human brain. This dataset was generated from 1340 individual tissue samples collected by macrodissection and incorporated data from an earlier study from the same group (Johnson et al., 2009). Using Affymetrix exon microarrays, the transcriptome data interrogated exon usage and thus enabled isoform differences to be inferred. This work showed that the majority of brain-expressed genes are spatiotemporally regulated and that the transcriptome is organized into coexpression networks. Notably, these data showed that the vast majority (90%) of genes expressed in the brain were spatiotemporally regulated and that the prenatal brain transcriptome is highly dynamic. The transcriptome data and the developmental trajectories of genes and pathways are publicly available in a user-friendly format from HB Atlas at http://hbatlas.org/, where individual genes can be queried for their spatiotemporal patterns. Of note is that all samples from this study were also analyzed using RNA-seq. The

RNA-seq data are available from the BrainSpan Consortium website http://www.brainspan.org/rnaseq/search/index.html, where the reads per kilobase transcript per million mapped reads (RPKMs) of individual genes in each sample can be queried and the RNA-seq and microarray data can be downloaded for further analysis.

The human brain has undergone molecular, connectional, and structural changes during recent evolution. To capture human-specific patterns of spatiotemporal gene expression in the brain, Sousa et al. (2017) recently analyzed by RNA-seq the transcriptome of 16 adult brain regions in chimpanzee and rhesus macaque for direct comparison with orthologous regions of the adult human brain (Kang et al., 2011). Although the majority of genes showed conserved spatiotemporal patterns, substantial species differences were found: approximately one-fourth of protein-coding genes were differentially expressed between at least two species in one or more brain regions. The data from this study can be accessed through SRA.

In a more focused study, Colantuoni et al. (2011) analyzed temporal-specific changes in one important region for human brain function and disease: the dorsolateral prefrontal cortex (BA46/9). Focusing on one region, this study was able to analyze a very large number of human brain samples (269 in total) spanning gestational week 14 in fetal development to aging (>80 years) using two-color custom-spotted arrays. Interestingly, this study identified gene expression dynamics occurring during fetal development that were reversed in early postnatal development and that these reversals were mirrored late in life during aging and neurodegeneration. The data can be queried to access the developmental trajectory of individual genes through Brain-Cloud at http://braincloud.jhmi.edu/plots/.

Beyond macroscopic regionalization, the brain is also functionally organized by subregional compartmentalization. To capture with high spatial resolution the subregional transcriptomes of the human brain, Hawrylycz et al. (2012) used both macrodissection and laser microdissection (LMD) to profile ~900 neuro-

anatomically precise subdivisions from two high-quality human brains using Agilent $8 \times 60K$ custom-design arrays. With this high-resolution dataset, one of the most anatomically comprehensive to date, it was shown that the spatial topography of the neocortex is reflected in its transcriptomic topography; physically closer regions have more similar gene expression patterns. The data from this study can be queried and visualized through the Allen Brain Map portal http://human.brain-map.org/.

In addition to areas, the neocortex is also organized into horizontal layers (layers 1–6) composed of distinct subsets of neurons exhibiting layer-dependent connections and patterns of gene expression. To capture these differences, several groups have performed microdissection of cortical layers for transcriptome analysis. In an early transcriptomic study targeting neocortical layers, J. G. Chen et al. (2005) microdissected upper layers (L2–L4) and deep layers (L5 and L6) from early postnatal mice (P7) for microarray analysis. This work provided the pioneering data supporting a number of gene-specific functional studies (Britanova et al., 2008; Kwan et al., 2008; Han et al., 2011; Shim et al., 2012). In a more comprehensive study of the mature mouse brain, Belgard et al. (2011) microdissected cortical layers 1–3, 4, upper 5, lower 5, 6, and 6b from adult (P56) mouse brain slices and performed RNA-seq on the Illumina GA IIx platform. This high-precision study led to the identification of 5835 protein-encoding genes and 66 noncoding RNAs as differentially expressed between cortical layers. The RNA-seq data are available in a user-friendly format from http://genserv.anat.ox.ac.uk/layers, where the fragments per kilobase transcript per million mapped reads (FPKMs) of individual genes and their layer-specific enrichment probabilities are provided.

In a more recent study, Fertuzinhos et al. (2014) leveraged a genetically encoded layer-specific fluorescent reporter mouse line (*Dcdc2a-Gfp*, GENSAT) (Gong et al., 2003) to perform precise fluorescence-guided microdissection of infragranular layers, granular layer, and supragranular layers. Cortical layer gene expression was analyzed at postnatal day (P) P4, P6, P8, P10, P14, and P180 (adult), covering key periods of synaptogenesis and neural circuit refinement in mouse cortical development. By using RNA-seq on the Illumina Hi-Seq platform, this group identified spatiotemporally regulated splicing events, as well as potential microRNA (miRNA)-mRNA interactions during circuit development in the mouse cortex. Coding mRNA and miRNA data are available for query through http://hbatlas.org/mouseNCXtranscriptome/. In a recent study, He et al. (2017) analyzed the layer-dependent transcriptomes of human, chimpanzee, and rhesus macaque prefrontal cortex, using horizontal sections grouped into bins. The high-resolution, multispecies study revealed a number of genes with human-specific layer patterns that are driven by neuronal and non-neuronal expression. The data can be downloaded through SRA.

During brain development, anatomical regionalization is dynamic. For example, distinct from the neocortical layer organization of the postnatal brain, the fetal neocortex is characterized by developmental layers that are only transiently present during prenatal periods. These include the germinal layers (ventricular zone and subventricular zone), which contain the proliferative neural stem cells, and the cortical plate, which contains postmigration neurons. To capture the transcriptomic differences in these developmental cortical layers, Ayoub et al. (2011) leveraged LMD to finely dissect the ventricular zone, subventricular zone/intermediate zone, and cortical plate from embryonic (E14.5) mouse cortex and performed RNA-seq on Illumina GA IIx. This work identified a number of zone-specific transcrip-

tional programs. The layer-specific RPKMs can be queried at http://rakiclab.med.yale.edu/transcriptome/index.aspx.

In a more recent study, Miller et al. (2014) very comprehensively profiled layer-dependent transcriptomes in the human fetal cortex, which contains cytoarchitecturally distinct layers and sublayers not present in the mouse. Using LMD, 9 fetal layers were delineated in mid-fetal human brain samples from ∼25 areas of the developing neocortex and profiled on custom 64K Agilent microarrays. With high layer and area resolution, this study revealed molecular gradients present in the germinal and postmitotic zones and patterned expression in genes associated with human brain disorders or evolution. The data are available for individual gene query and correlated gene search from BrainSpan at http://www.brainspan.org/lcm/search/index.html, where a supporting fetal human brain anatomical atlas is also available.

Human tissue samples are precious, and even the most extensive brain collections have gaps in coverage of specific brain regions or developmental stages. Bakken et al. (2016), therefore, undertook a highly precise spatiotemporal study extensively covering prenatal and postnatal development of the macaque monkey cortex. This work analyzed LMD samples of cortical layers from multiple brain regions at 10 stages of development (embryonic days 40–48 months postnatal) using Affymetrix Rhesus Macaque GeneChip microarrays. This comprehensive, high-resolution spatiotemporal study of the macaque brain revealed rapid prenatal expression changes in progenitors and neurons, disease-specific spatiotemporal enrichment of genes associated with human neurodevelopmental disorders, and evidence of human-specific gene regulation. The data from this study are available through the NIH Blueprint NHP Atlas at http://www.blueprintnhpatlas.org.

In addition to expression data from within the brain, several databases are available with transcriptomes from many organs and systems. A key example is the Genotype-Tissue Expression project (GTEx), which is a multisite consortium funded by the National Institutes of Health to generate a large-scale dataset linking genetic variation to gene expression in multiple tissues of the human body, including several regions of the brain (GTEx Consortium, 2015, 2017). Profiling >50 tissues across the body from >600 donors, GTEx has generated a rich dataset of tissue-specific transcriptomes, providing insights into exon usage, splicing, and the tissue specificity of these events. Importantly, each donor is genotyped for common single-nucleotide polymorphisms, thus enabling expression quantitative trait loci (eQTL) studies. eQTLs can reveal the contribution of an individual variant to expression of local (*cis*-eQTLs) and distant (*trans*-eQTLs) genes and, with wide sampling of many tissue types, whether such effects may be tissue-specific. These data are available to investigators through https://www.GTExportal.org, offering an intuitive and rich resource to query individual genes and the eQTLs linked to the expression of these genes.

In addition to transcriptome databases, spatial gene expression data can also be found in several databases housing large-scale *in situ* hybridization datasets. Many of these resources preceded the transcriptome era but remain important as they provide single-cell gene expression data in a precise anatomical context. Briefly, these *in situ* databases include the developing mouse brain (http://developingmouse.brain-map.org/, http://www.eurexpress.org/ee/), adult mouse brain (http://mouse.brain-map.org/), and adult human brain (http://human.brain-map.org/ish/search) (Lein et al., 2007; Diez-Roux et al., 2011; Hawrylycz et al., 2011; Zeng et al., 2012).

Cell type-specific analysis

The brain is a highly heterogeneous tissue composed of diverse cell types characterized by distinct patterns of gene expression. In transcriptome analyses of whole tissues, RNAs from all cell types are analyzed *en masse*. Cell type-specific patterns of gene expression and regulation, therefore, would be diluted and may be missed altogether. Several strategies have been used to physically isolate particular cell types so that they can be more specifically analyzed. In the 2000s, genetic access to specific cells in the brain became possible with the availability of transgenic mice expressing a reporter gene (e.g., EGFP) in specific cell types by random genomic integration or promoter-driven gene expression (Feng et al., 2000; Gong et al., 2003). In a pioneering study, Sugino et al. (2006) leveraged genetic access to specific cell types and FACS to isolate glutamatergic and GABAergic neurons in the cortex, hippocampus, amygdala, and thalamus. The resulting microarray data revealed the transcriptomic profiles of 12 distinct adult neuronal populations in the mouse brain.

With advances in recombineering and transgenesis, more precise genetic targeting of cell types became possible. The GENSAT consortium, for example, was funded by National Institutes of Health to generate BAC transgenic EGFP reporter mice targeting specific cell types in the brain (Gong et al., 2003). Using this targeting strategy, Doyle et al. (2008) and Heiman et al. (2008) pioneered BACarray or bacTRAP, which enabled cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). By targeting the expression of an EGFP-tagged ribosomal subunit (RPL10a) to specific cell types by BAC transgenesis, polysomal mRNAs from genetically defined cell populations can be isolated by affinity purification. The bacTRAP project generated a large collection of bacTRAP lines for cell type-specific transcriptomic analysis, at that time by microarrays. Although a web interface for these data was not provided at the time, Xu et al. (2014) generated a convenient webtool, Specific Expression Analysis, at http://genetics.wustl.edu/jdlab/csea-tool-2/ to query and visualize these data, which contain transcriptomic data from 27 genetically targeted cell types in the brain. This and a similar Cre-dependent strategy for genetic ribosomal tagging (Sanz et al., 2009) are now being used in many fields of biology. Other compartments of the cell can be targeted to facilitate cell purification. In a more recent study, Mo et al. (2015) genetically targeted the expression of a nuclear tag using INTACT (isolation of nuclei tagged in specific cell types) for combined transcriptomic and epigenomic profiling in specific cell types in the mouse brain (Mo et al., 2015). The integrated transcriptomic, DNA methylation, and DNA accessibility (ATAC-seq) data for excitatory pyramidal neurons, parvalbumin-expressing interneurons, and VIP-expressing interneurons in the mouse cortex can be accessed at http://neomorph.salk.edu/mm_intact/.

In addition to genetic targeting, which requires the generation of transgenic or knock-in mice, cell surface markers can also be used to purify specific cell types. In a comprehensive cell type-specific analysis of the brain transcriptome, Cahoy et al. (2008) and Zhang et al. (2014, 2016) used cell surface antigen immunopanning and FACS to isolate known major cell types in the adult brain (neurons, astrocytes, oligodendrocytes, microglia, and endothelial cells) from human and mouse. RNA-seq was performed on the Illumina HiSeq and NextSeq platforms. This species- and cell type-dependent analysis enabled, among other insights, the identification of transcriptomic differences in mouse and human astrocytes, as well as microglia (Bennett et al., 2016). The resulting RNA-seq data are available in a user-friendly format from the Brain RNA-seq database at http://brainrnaseq.org/.

Neurons are distinguished by their connectivities. The projection neurons of the cerebral cortex, for example, can be broadly classified into callosal, corticothalamic, corticotectal, and corticospinal. In a pioneering study of connection-specific transcriptomic analysis, Arlotta et al. (2005) combined retrograde axon tracing and FACS to isolate corticospinal, callosal, and corticotectal neurons for microarray analysis, which identified key genes for follow-up functional studies (Arlotta et al., 2005; Molyneaux et al., 2005; Lai et al., 2008). More recently, Molyneaux et al. (2015) isolated callosal, subcerebral, and corticothalamic projection neurons by nuclear marker labeling and FACS using E15.5, E16.5, E18.5, and P1 mouse neocortex. The RNA-seq data revealed spatiotemporal usage of alternative promoters and exons during fate specification, as well as numerous noncoding RNAs with dynamic expression. These data are available for web query from DeCoN at http://decon.fas.harvard.edu/. In another recent study, Ekstrand et al. (2014) used retrograde adeno-associated virus to tag ribosomes in long range projection neurons based on their projection target. The technique was used to profile midbrain dopamine neurons projecting to the NAc.

The hippocampus is a well-studied brain region with important subregional specializations. Cembrowski et al. (2016) analyzed with high precision all hippocampal excitatory neuronal classes (dentate gyrus granule and mossy cells, and CA1, CA2, CA3 pyramidal cells) by combining genetic targeting of cell types, microdissection, and manual sorting. The high-resolution RNA-seq data revealed the expression profiles of lesser-known cell classes and unexpected spatial differences in expression across the dorsal-ventral axis. The data are available as a public resource through HippoSeq at http://hipposeq.janelia.org/.

A new database (NeuroSeq) that has recently become available houses the transcriptomes of, to date, the most comprehensive collection of genetically and anatomically identified neuronal classes from many regions of the adult mouse brain (Sugino et al., 2017). By combining genetic access, achieved through a large collection of cell type-specific Cre driver lines, brain region-specific microdissection, and manual selection of cell pools, Sugino et al. (2017) performed deep RNA-seq on 181 molecularly and spatially precise cell types for NeuroSeq. Using a strategy that is distinct from the single-cell sequencing techniques discussed below, NeuroSeq can achieve significantly higher sequencing depth from genetically labeled and manually separated cell types. The resulting data showed that homeobox transcription factors contain the highest information content for distinguishing cell types and the lowest noise. Therefore, a cell type classifier using a decision tree of gene expression levels can be based on a small, informative set of transcription factors. The comprehensive transcriptome dataset can be queried and explored at http://neuroseq.janelia.org/ through user-selected tracks in an interactive browser. In addition, a high-resolution mouse brain atlas of Cre driver and reporter expression is also available to explore the cell type and location specificity of the transgenic mice used to build the transcriptomic data.

Single-cell analysis

With genetic access to specific cell types and advanced cell purification and microdissection techniques, the transcriptomes of specific groups of brain cells can be assessed with high precision. However, these techniques are unable to reveal the molecular taxonomy of the brain at its fundamental unit, the cell. The brain is a highly heterogeneous tissue composed of single cells, each proposed to be characterized by a unique transcriptome. Cell-to-cell transcriptomic heterogeneity is a fundamental property of any multicellular system and is likely to be especially important

for the assembly and function of neural circuits in the brain. Recent advances in Next Generation Sequencing (NGS), single-cell isolation, and molecular barcoding techniques have enabled the generation and sequencing of cDNA libraries from a single cell; thus, cell-to-cell heterogeneity at the level of the transcriptome can be assessed. This technique, single-cell RNA-seq (scRNA-seq), is emerging as a powerful tool to classify new cell populations, characterize rare cell types, and track the longitudinal development of cells at single-cell resolution. In the few years since technology has enabled high-throughput scRNA-seq, several variations of the technique have already been applied to the brain, including adult human brain (Darmanis et al., 2015; Lake et al., 2016), fetal human brain (Pollen et al., 2014; Johnson et al., 2015; S. J. Liu et al., 2016), human brain organoids (Quadrato et al., 2017), adult mouse dentate gyrus (Luo et al., 2015; Shin et al., 2015; Lacar et al., 2016; Dulken et al., 2017), adult mouse striatum (Gokce et al., 2016), adult mouse hypothalamus (Campbell et al., 2017), adult mouse DRG (Li et al., 2016), and fetal mouse ganglionic eminences (Y. J. Chen et al., 2017).

In one of the first studies to provide a database of brain single-cell transcriptomes, Zeisel et al. (2015) analyzed 3005 single cells from the adult mouse somatosensory cortex and hippocampal CA1, using the Fluidigm C1 instrument to capture single cells without selection and generate scRNA-seq libraries. Clustering of single cells based on their transcriptomes revealed nine major classes, which were identified by the expression of known marker genes as pyramidal neurons, interneurons, oligodendrocytes, astrocytes, microglia, vascular endothelial cells, mural cells, and ependymal cells. Further clustering revealed 47 molecularly distinct subclasses (e.g., layer-specific pyramidal neurons) and identified new molecularly defined subpopulations of layer 1 interneurons and postmitotic oligodendrocytes, highlighting the ability of scRNA-seq to characterize new or rare cell populations. The single-cell transcriptome data can be interactively accessed at http://linnarssonlab.org/cortex/.

In the developing brain, diverse subtypes of neurons are born from neural progenitor cells. During the acquisition of neuronal identity, the cell is thought to undergo rapid transcriptomic changes as it transitions from proliferation to neuronal specification and differentiation. To capture these early waves of transcriptomic changes, Telley et al. (2016) used a pulsed intraventricular dye to label isochronic cohorts of apical progenitors, basal progenitors, and neurons at early and later stages of differentiation from the embryonic mouse cortex for scRNA-seq. Labeled cells were isolated by FACS, and single cells were captured using the Fluidigm C1 system. The resulting dataset of 272 single-cell transcriptomes revealed dynamic, neuron-specific transcriptional waves that instruct the sequence of specification and differentiation. This high temporal resolution scRNA-seq dataset can be queried and visualized at http://genebrowser.unige.ch/science2016/.

In a comprehensive new study, Nowakowski et al. (2017) performed scRNA-seq of human fetal cortex and medial ganglionic eminence across key stages of prenatal neurogenesis (from 6 to 37 weeks after conception) using the Fluidigm C1 instrument. Analysis and clustering of 4261 cells revealed lineage-dependent trajectories of transcriptional regulators, and that modest transcriptional differences in cortical radial glial stem cells cascade into robust cell-type-dependent differences in neurons. The single-cell transcriptome data can be queried and cell clustering can be visualized through a web interface at http://bit.ly/cortexSingleCell.

With advances in microfluidics, several higher-throughput single-cell isolation methods have become available. Shekhar et al. (2016) used Drop-seq to profile the single-cell transcriptomes of 25,000 mouse retinal bipolar cells. Drop-seq is a technique that encapsulates, in a nanoliter volume droplet, a single cell together with a barcoded bead, thus barcoding individual cells for highly parallel scRNA analysis (Macosko et al., 2015). As proof of concept for Drop-seq, Macosko et al. (2015) first analyzed the single-cell transcriptomes of 44,808 mouse retinal cells. More recently, Shekhar et al. (2016) focused on the bipolar cells, which comprise only ~7% of retinal cells in mice. The data led to a molecular classification with 15 bipolar cell types that corresponded to morphologic classification, including all known cell types and two new cell types. The resulting data can be accessed and visualized through the Broad Single Cell Portal at https://portals.broadinstitute.org/single_cell.

A critical step in scRNA-seq is the isolation of intact single cells from tissues. This can be particularly challenging for neurons, which elaborate long processes that can increase the likelihood of cell membrane rupture during dissociation. In addition, typical scRNA-seq protocols require the use of fresh tissues because freezing disrupts the cell membrane, making the isolation of intact single cells from frozen samples impractical. To address these issues, techniques have been developed to isolate single nuclei from fresh or frozen (archival) tissues for single-nucleus RNA-Seq (sNuc-Seq), which applies scRNA-seq to single nuclei isolated using fluorescence activated nuclear sorting (Lake et al., 2016). Habib et al. (2016) recently extended on the sNuc-Seq technique to develop Div-Seq, which combines sNuc-Seq with pulse labeling of S-phase cells by deoxythymidine analog EdU to profile single proliferating cells in the adult hippocampal neurogenic niche. Clustering analysis of 1402 single-nuclear transcriptomes identified closely related cell types and enabled tracking of transcriptional trajectories as cells mature from progenitors to neurons. These data can be accessed and visualized through the Single Cell Portal at https://portals.broadinstitute.org/single_cell. The same group has also developed DroNc-seq, which leverages the Drop-seq strategy to profile single nuclei at high throughput, and applied the technique to archived human and mouse tissues (Habib et al., 2017). This and other techniques have now made possible the profiling of single cells from archival human pathological tissues, including those from individuals with neurological and psychiatric disorders, which holds great potential to reveal mechanistic insights into brain disorders.

In addition to postmortem samples, human pathology tissues can also be collected from surgical resections. Recently, Darmanis et al. (2017) collected surgically excised glioblastoma tumors for scRNA-seq. Data from 3589 cells from four patients showed clustering of major cell types (neoplastic, vascular, myeloid, neurons, oligodendrocytes, astrocytes) and revealed that, despite significant heterogeneity, infiltrating glioblastoma cells are characterized by a consistent transcriptomic signature, thus providing insights into shared mechanisms of infiltration. These data can be accessed and visualized through GBMseq at http://gbmseq.org/.

In addition to the cerebral cortex, scRNA-seq databases with data from other brain regions are available. Kee et al. (2017) recently used scRNA-seq to analyze the progenitors of mesencephalic dopamine neurons in embryonic mice by using FACS to isolate single cells that were labeled and unlabeled by the *Lmx1a-EGFP* transgene. The resulting data revealed closely related, but distinct, transcriptomic profiles between mesencephalic dopamine neurons and subthalamic nucleus neurons. The data can be

visualized through http://rshiny.nbis.se/shiny-server-apps/shiny-apps-scrnaseq/Kee_2016/.

Integrative analysis

Upstream of the transcriptome are exquisite gene regulatory mechanisms that precisely control spatiotemporal gene expression, whereas downstream of the transcriptome is the execution of essentially all aspects of cellular function. Integrative transcriptomic databases facilitate the covisualization of transcriptomic and other forms of genomic and cellular data, thus enabling users to correlate gene expression with upstream regulatory processes or downstream cellular phenotypes.

Essential to a full understanding of gene regulation is the functional annotation of genomic regulatory elements. The multisite Encyclopedia of DNA Elements (ENCODE) Consortium is an international collaboration of research groups funded by the National Institutes of Health to comprehensively map coding and noncoding functional elements in the human, mouse, fly, and worm genomes, including regulatory elements that act at the DNA, RNA, and protein levels and the tissue- and cell type-dependent contexts of their function (ENCODE Project Consortium, 2012). To date, the multiphase ENCODE project has generated large-scale datasets for nearly 9000 projects. The multiomics data generated include transcriptome (e.g., polyA RNA-seq, miRNA-seq), DNA methylation (e.g., WGBS, DNAme array), DNA accessibility (e.g., DNase-seq, FAIRE-seq, ATAC-seq), chromatin interaction (e.g., ChIA-PET, 5-C, Hi-C), RNA-binding protein interaction (e.g., RIP-seq, CLIP-seq), and chromatin modification (e.g., H3K27me3 ChIP, H3K4me3 ChIP) from a variety of cell lines (e.g., ESCs, HeLa) and tissues (e.g., adult brain, fetal heart) at different stages of differentiation or development. Importantly, this comprehensive and highly informative dataset can be explored in a user-friendly way through data tracks on the widely used UCSC genome browser (https://genome.ucsc.edu/encode/). Users can search for and select multiple data tracks of interest and simultaneously visualize the data on the well-annotated UCSC genome browser together with existing UCSC tracks (e.g., conservation, single-nucleotide polymorphisms, RepeatMasker). The ENCODE project is ongoing and the updated availability of data can be found at the project portal (https://www.encodeproject.org/). Additionally, pertinent to brain research is the Psychiatric Encyclopedia of DNA Elements (PsychENCODE) consortium (Akbarian et al., 2015), a recently launched multisite project funded by the National Institute of Mental Health that aims to comprehensively catalog regulatory elements, epigenetic modifications, and noncoding RNAs in ∼1000 phenotypically well-characterized, high-quality healthy and disease-affected human brains. These collaborative efforts are now underway and expected to generate large-scale data that provide direct insights into the molecular pathology underpinning brain disorders, and how genetic variants contribute to disease. These data are available at https://www.synapse.org//#!Synapse:syn4921369/wiki/235539.

Downstream of the transcriptome is the regulation of nearly all aspects of cellular phenotypes. In neurons, these phenotypes include morphological and electrophysiological properties. The Allen Brain Atlas Cell Types database is an ongoing survey of electrophysiological, morphological, and transcriptomic data derived from individual mouse and human brain cells. The current data include a comprehensive analysis of single cells from one mouse cortical region, primary visual cortex, by scRNA-seq. For scRNA-seq, Tasic et al. (2016) captured molecularly distinct single cells using established and new cell type-specific Cre and reporter transgenic mouse lines, which is a strategy distinct from the unselected analysis of Zeisel et al. (2015). Based on single-cell transcriptomic data, Tasic et al. (2016) identified 49 cell types, including 23 GABAergic, 19 glutamatergic, and 7 non-neuronal cell types. By applying unsupervised clustering to all cells, this work characterized the specificity of genetic access to the identified cell types by transgenic Cre lines. Importantly, Tasic et al. (2016) showed that some of these cell types are characterized by distinguishing electrophysiological or axon projection properties, thus associating single-cell transcriptome signatures with specific cellular phenotypes, and revealing a transcriptomic cell type taxonomy that is supported by genetic, physiological, morphological, and projectional evidence. Integrating these data, the Allen Brain Atlas Cell Types database enables users to browse the electrophysiological response data and reconstructed neuronal morphologies of molecularly defined cell types using the Cell Feature Search tool (http://celltypes.brain-map.org/). The transcriptomic properties of the same molecularly defined cell types can be accessed through the web or by download (http://celltypes.brain-map.org/download).

The field is entering a new era of integrative analysis. Very recently, Paul et al. (2017) combined scRNA-seq with anatomical and electrophysiological data to specifically study subtypes of GABAergic interneurons in the mouse neocortex. They found that the synaptic connectivity patterns of GABAergic neuron types are well delineated by transcriptional architecture into 6 categories. In addition, with improvements in library generation, it is now possible to sequence the cDNA of a single cell derived from aspiration. This the basis of Patch-seq, which enables analysis of electrophysiological characteristics and transcriptome in a single cell, by using a patch-clamp protocol followed by aspiration of the cytoplasm into the recording pipette for sequencing library generation (Fuzik et al., 2016). In a very recent study, Lake et al. (2018) combined single cell transcriptome and epigenome analyses in >60,000 adult human brain cells, revealing regulatory elements and transcriptional programs underlying distinct cell types. With further technical improvements, integrative analysis of cellular characteristics, epigenomes, and transcriptomes in single cells is becoming an exciting possibility.

Considerations for use of data

The databases we highlight house high-quality data that can be a tremendous resource to many fields of neuroscience. These data are, however, not without caveats. Understanding the potential limitations can enable database users to more appropriately and effectively use these resources. Here we discuss key methodologies for transcriptome analysis, what can and cannot be measured, and biological and technical considerations for use of transcriptome databases.

What does the transcriptome measure?

The transcriptome is broadly defined as the sum total of all RNA transcripts in a cell or an organ. When it is measured, the transcriptome represents a snapshot of transcript levels at the time of analysis, which is the combined outcome of the transcriptional activities that produce the RNAs and the post-transcriptional processes that stabilize or degrade them. A number of newer techniques have analyzed nascent RNA to distinguish transcriptional from post-transcriptional activity (Core et al., 2008; Paulsen et al., 2014). Interestingly, the correlation between nascent and static RNA-seq can differ on a gene-by-gene basis, suggesting gene-dependent differences in RNA stability or degradation. The transcriptome can also provide insights into transcript isoforms. Alternative splicing is emerging as an important

mechanism to generate protein diversity and sites for differential post-transcriptional regulation. Exon microarrays and short-read RNA-seq can provide insights into alternative exon usage. Indeed, several of the highlighted databases provide transcript isoform information (Zhang et al., 2014; GTEx Consortium, 2015; Sugino et al., 2017). It is important to note, however, that short-read RNA-seq or microarrays based on short DNA probes can inform the usage of an exon but cannot reliably measure the upstream or downstream exons to which it is spliced. Direct analysis of splice junctions throughout the transcript requires multijunction-spanning reads, which is possible with longer read sequencing technology. With emerging techniques in long-read sequencing and direct RNA sequencing (e.g., nanopore) (Garalde et al., 2018), such analyses are becoming a possibility. Transcriptomic analyses can also be used to study species differences. These comparisons can be challenging, especially for phylogenetically distant species because of differences in genome annotation, tissue processing, and analysis standardization. For the brain, the high degree of regionalization presents an additional challenge in defining orthologous regions between species. For more closely related species (e.g., human, chimp, and rhesus macaque), anatomical identification is feasible (Sousa et al., 2017). Although challenging, these comparisons can provide important insights into the molecular bases of brain evolution.

For many researchers, transcriptomic data are used to infer protein expression, largely because of the relative ease and lower cost of quantifying mRNA levels compared with their protein products. Protein levels, however, are regulated through a combination of transcriptional, post-transcriptional, and post-translational mechanisms. Transcript and protein level correlation has long been a subject of study and debate (Schwanhäusser et al., 2011; Edfors et al., 2016; Y. Liu et al., 2016). A recent statistical analysis of mRNA and protein levels across human tissues found that, for a given gene, protein levels across tissues are only poorly predicted by mRNA levels (i.e., the protein-to-mRNA ratio for an individual gene varies greatly between tissues), likely reflecting the effects of tissue-dependent post-transcriptional regulation. However, and quite importantly, for between-gene comparisons, protein levels are well predicted by mRNA levels (e.g., highly abundant proteins tend to be encoded by highly abundant mRNAs and vice versa). This phenomenon, termed Simpson's paradox, is discussed in detail previously (Franks et al., 2017). Recently, Carlyle et al. (2017) compared mRNA and protein levels from multiple human brain regions from the BrainSpan project. Their analysis demonstrated a modest median correlation ($r = 0.32$) between mRNA and protein, providing support for use of the transcriptome as an informative tool, while highlighting the imperfect prediction of protein levels based on mRNA. Given the potential for gene-specific differences in RNA-protein correlation, it may be important, depending on the ultimate goal of the user, to validate RNA or protein levels in additional tissues or using orthogonal methods, which we further discuss below in Validation.

RNA-seq versus microarray

Early studies of brain transcriptomes were performed using microarray, a hybridization-based method for measuring gene expression (Hitzemann et al., 2014). A microarray is a collection of DNA probes immobilized on a solid surface to which fluorescently labeled target cDNA hybridize. Probe-target hybridization is then quantified by measuring fluorescence to determine the relative abundance of target cDNA, which is identified by the coordinates of the probe spot. Microarrays

have been reliable and accurate and, with more advanced oligonucleotide arrays targeting individual exons, have provided biological insights at the level of transcript isoforms. In the past several years, RNA-seq has emerged as the dominant method for transcriptome profiling. RNA-seq is a sequencing-based method that leverages the massively parallel throughput of NGS to simultaneously quantify RNA species in an organ or a cell, thus enabling unbiased profiling of the transcriptome (Mortazavi et al., 2008). Briefly, RNA-seq starts with isolation of the desired species of RNA, such as mRNA, by selection or depletion methods. The RNA is then converted into cDNA by reverse transcription, often after an RNA fragmentation step. For NGS, an RNA-seq library, which contains the end sequences compatible with sequencing-by-synthesis for a particular NGS platform, is generated by attaching adapters to the cDNA via ligation or PCR. Individual libraries from different RNA samples can be uniquely barcoded so that they can be mixed and sequenced together to reduce cost. The transcriptome is then assembled by assigning sequencing reads to particular features (e.g., a specific gene), usually by alignment to the genome or the transcriptome. Quantification can be achieved based on sequence read counts after normalization, which must take into account transcript length and total library size.

Although the brain transcriptome data generated by microarray remain valuable and reliable resources, analyses by RNA-seq have several advantages. First, RNA-seq directly provides sequence reads and thus can detect splice variants and novel transcripts without prior knowledge or probe selection bias. Second, RNA-seq offers a much wider dynamic range, more accurately quantifying high and low expressors. Third, the increased sensitivity of RNA-seq enables analysis from a small amount of starting material, including from single cells. RNA-seq analyses, however, are not without limitations. We discuss some of these considerations below.

Quantifying gene expression

A key goal of expression analysis is quantification of transcript abundance. Essentially all transcriptome data represent relative, not absolute, measurements that require the data to be normalized. Microarrays provide analog results with values that are inherently arbitrary and normalized within the experiment, usually using reference probes. RNA-seq produces shot-gun sequencing reads that need to be processed and mapped. The choice of reference, mapping method, and quantification strategy can have a significant impact on the result (Williams et al., 2017). Normalization is generally implemented as a unit to account for variations in total read count, which can vary between samples, even within an experimental batch. Normalization also needs to account for gene length, as longer transcripts are more likely to be sequenced and would have a higher read count at a given expression level. Importantly, there are a number of conventions for normalizing RNA-seq data, which lead the data to be reported in different units between groups. These units include RPKMs, FPKMs, and transcripts per million (TPM). The most commonly used unit is RPKM. Introduced by one of the first RNA-seq papers (Mortazavi et al., 2008), RPKM normalizes for total read count and gene length but produces units that are not comparable across samples due to inconsistencies in the final normalized counts across samples. FPKM is calculated by an analogous method for paired-end read sequencing libraries (Trapnell et al., 2010). In contrast, TPM, which also normalizes for total read count and gene length, quantifies transcript abundance by a clever reordering of the terms so that the final values are all nor-

malized equally, making it much more feasible to compare values across samples (Li et al., 2010). Although TPM has not yet been adopted widely, recent computational and statistical packages have transitioned to using TPM as the unit of choice (Wagner et al., 2012).

Technical factors
Transcriptome analysis is a technically challenging endeavor, and many methodological factors can affect the accuracy and reproducibility of the data. Postmortem interval, tissue quality, tissue dissection, and RNA extraction methods can influence transcriptome analysis. Notably, for analysis of human brain samples, antemortem and postmortem factors can skew results. Most of the highlighted human transcriptome databases therefore report postmortem interval, brain pH (a measure of tissue quality), and RNA integrity numbers (a measure of RNA degradation) for each specimen. Recently, it was shown that correcting for RNA quality biases can significantly improve replication rates (Jaffe et al., 2017).

For RNA-seq, library generation is a critical step and potential source of technical variability. Library generation requires PCR amplification, which can be nonuniform and thus introduce bias. Bias can also be introduced during polyA selection or rRNA depletion, techniques that are commonly used to enrich for mRNAs. In addition, library generation requires a long, multistep protocol that can introduce batch-to-batch variations that are well documented. These potential issues, however, can be mitigated. For example, by using RNA standard spike-in (e.g., ERCC), batch-to-batch variations in library quality can be normalized. Recently, unique molecular identifiers have been used to tag individual starting cDNAs before amplification, thus reducing the impact of amplification bias. These quality control techniques have already been used in some of the databases and are expected to become more widely adopted. In addition, computational tools can be used to assess library quality based on batch effects, overamplification, strand directionality, signal-to-noise, and other metrics (Islam et al., 2014; Patro et al., 2017; Pimentel et al., 2017). However, computational pipelines themselves can differ in methodology (Baruzzo et al., 2017; Everaert et al., 2017) and be an additional source of technical variability. Although computational tools are rapidly improving, together, these technical factors emphasize the importance of validation experiments, which we discuss in Validation.

It is also important to note that transcriptomic analysis does not measure all genes with equal accuracy. For RNA-seq, the confidence of transcript quantification increases with the number of reads representing the transcript. Transcripts that are very low in abundance can be difficult to quantify due to the limited number of representative reads. Therefore, measurements, such as fold change, can be skewed for low expressors. For example, two genes that are measured at 0.01 FPKM and 0.05 FPKM, respectively, are likely not indeed fivefold different due to uncertainty in RNA-seq measurements and quantification. Computational algorithms do take this into account and can determine statistical confidence for each gene. Nonetheless, this highlights the need for validation experiments.

Single-cell transcriptomics represent an area of tremendous promise. Investments into large, collaborative efforts, such as the Human Cell Atlas (Regev et al., 2017), will continue to accelerate the discovery and molecular definition of cell types in the coming years. It should be noted, however, that current scRNA-seq techniques lack sufficient depth for full analysis of the transcriptome on a per-cell basis. As a single cell contains only a very small amount of starting material, uneven loss of RNA can lead to spurious gene drop-out, the absence of reads representing an expressed gene, and affect quantitative analysis (M. Chen and Zhou, 2017). In particular for scRNA-seq, genes with short transcript length or low transcript abundance are susceptible to drop-out and can be difficult to measure accurately. As methodologies in single-cell capture, barcoding, and library generation improve, deep analysis of single-cell transcriptomes is expected to become feasible in the near future.

Validation
Validation is essential to discovery science. For transcriptomic data, biological and technical factors can contribute to accuracy and reproducibility. Importantly, transcriptome analyses, by definition, simultaneously probe all possible transcripts, each of which has the potential to produce a statistically significant discovery. With more transcripts being compared, the likelihood that two groups will appear to differ on at least one transcript increases due to random sampling error (Fang and Cui, 2011). This phenomenon, termed multiple testing, can be accounted for by calculating the false discovery rate (Robinson et al., 2010; Love et al., 2014). Although multiple testing has been taken into account in many of the highlighted databases, the possibility of false discovery underscores the importance of independent validation of transcriptomic data. Depending on the specific needs of the end user, validation experiments may be necessary to ascertain transcript or protein levels in the tissues or cells of interest. At the transcript level, quantitative RT-PCR or digital droplet RT-PCR can provide highly accurate quantification, whereas *in situ* hybridization can provide spatial information. At the protein level, Western blotting, ELISA, and immunostaining can be used. For some of the databases highlighted here, the data are derived from tissues that are not readily accessible to many researchers, such as fetal human and macaque tissue. In these cases, independent validation may still be possible by using data from multiple databases or sources, including other published studies.

Utility of brain transcriptomic databases
The collection of transcriptomic resources highlighted here has tremendous potential to inform neuroscience research very broadly. Here we briefly discuss how these data can be used to propel studies of gene function and regulation as well as genetic studies of human brain disorders, using several examples to illustrate the potential utility of transcriptome data.

Studies of gene function and regulatory mechanisms
A key pursuit in molecular neuroscience is to understand the roles of individual genes in brain development and function. The spatiotemporal and cell type specificity of gene expression can provide important clues into potential function and inform the tissues and developmental time points most appropriate for experimental study. The transcriptomic databases highlighted here represent valuable sources of curated, high-quality data accessible without the need for specialized expertise in informatics or data analysis, enabling individual scientists to easily query the genes they study or those of potential interest. These combined resources can rapidly provide candidate gene expression profiles with spatiotemporal, cell type, and single-cell specificity, giving a high-level overview that would otherwise be too costly or time-consuming for individual laboratories to generate. These data can be used to identify candidate genes for further study

and support or refine working hypotheses by enabling a more unbiased approach based on cell type specificity and developmental trajectory of gene expression. In addition, clues into potential mechanisms and gene–gene interactions may be gained by examining expression overlap or mutual exclusivity. Here we highlight the potential utility of these transcriptomic resources with a small sampling of the individual gene function studies they have enabled or facilitated.

Even from very early studies, it became clear that transcriptome data can provide candidate genes for functional studies and help generate specific hypotheses. For example, some of the earliest transcriptomic studies of mouse cortical layers led to the identification of candidate transcription factors with layer- or cell type-dependent expression (Arlotta et al., 2005; J. G. Chen et al., 2005). Specific hypotheses informed by expression were formed and tested, ultimately revealing genes encoding transcription factors, such as *Bcl11b*, *Fezf2*, and *Sox5*, as critical determinants of cortical projection neuron identity and corticofugal axon pathfinding (Arlotta et al., 2005; J. G. Chen et al., 2005; Molyneaux et al., 2005; Kwan et al., 2008; Lai et al., 2008).

Analysis of gene coexpression or mutual exclusion in a variety of contexts can help inform potential regulatory mechanisms, gene–gene interactions, or codependence. This endeavor, however, requires a large dataset. The HB Atlas (Kang et al., 2011), which characterized a large collection of human brain tissue at high spatial and temporal resolution, enabled more precise gene coexpression studies. In one example illustrating the use of these data, Shim et al. (2012) sought to identify the transcriptional regulators of the *Fezf2* E4 enhancer. Coexpression analysis between *FEZF2* and members of the SOX transcription factor family was undertaken, leading to the identification of SOX4 and SOX11 as candidate regulators. Furthermore, the spatiotemporal usage of these data is exemplified by Bae et al. (2014), who sought to determine how a mutation upstream of *GPR56* led to perisylvian polymicrogyria, a disruption of gyri formation spatially restricted to an area around the sylvian fissure. Because this mutation specifically affected a particular region of the cortex, Bae et al. (2014) used the HB Atlas data to assess candidate gene expression in the ventrolateral prefrontal cortex and during fetal development, which facilitated the identification of specific members of the RFX transcription factor family.

The HB Atlas data (Kang et al., 2011) have also been used to inform species-dependent studies. For example, Kwan et al. (2012) used HB Atlas data to refine a list of candidate RNA-binding proteins interacting with *NOS1* mRNA in human fetal cortex, an interaction that is absent from mouse brain. This led to the identification of FMRP as a post-transcriptional regulator of human NOS1 expression. In another study of species differences, Ataman et al. (2016) used these data to investigate the function of Osteocrin, which they identified as an activity-dependent secreted factor in human fetal brain cultures. In mouse, Osteocrin is expressed selectively in skeletal muscle and bone. The HB Atlas data, however, showed that Osteocrin is highly expressed in human cortex during critical periods for synapse development and plasticity. Subsequent analyses identified species-dependent upstream elements that drive activity-dependent neural expression of Osteocrin via the transcription factor MEF2.

In addition to the spatiotemporal data, cell type-specific transcriptome data have also been shown to be tremendously informative for functional studies. For example, Lui et al. (2016) derived support for the hypothesis that progranulin suppresses inappropriate microglial activation via regulation of complements production from the Brain RNA-seq cell type-specific database, which showed significant enrichment of *progranulin* and multiple complement genes in microglia. Further transcriptome analysis in *progranulin*-deficient mice revealed disruption of complement cascade members and illuminated C1qa as a key player in microglial activation. The Brain RNA-seq data were also used to inform the hypothesis that microglia induce reactive (A1) astrocytes (Liddelow et al., 2017). By showing that LPS activation of A1 requires TLR4 signaling expressed specifically by microglia, these data helped move forward the experiments that ultimately showed reactive astrocyte induction by activated microglia.

These examples of functional and mechanistic studies illustrate how spatiotemporal and cell type-specific transcriptome data can be harnessed to interrogate molecular and cellular processes. The impact of the myriad new resources, including the single-cell transcriptome databases, will certainly become evident over the next few years. Undoubtedly, these emerging resources will facilitate new and innovative avenues to ask critical questions about brain development and function.

Genetic studies of human brain disorders

In addition to the basic molecular and cellular mechanisms underpinning normal brain function, a key effort of the field is to gain a mechanistic understanding of brain disorders. With recent advances in sequencing technologies, the number of genes and loci carrying risk for disease, including many psychiatric and neurodevelopmental disorders, are being identified at an accelerating pace. How mutations in these genes contribute to brain disorders, and the spatiotemporal and cell type specificity with which they function, are incompletely understood. This knowledge gap remains an obstacle on the path toward molecular medicine. Here, we highlight some examples of how transcriptomic data can be leveraged to gain biological insights into human genetic data, by providing clues into the molecular pathways and cellular substrates underpinning disease etiology.

To illustrate the utility of transcriptomic data to inform human genetic studies, we highlight recent work from the field of autism spectrum disorder (ASD). For decades, the genetics of ASD have been a challenge for the field because of phenotypic and genotypic heterogeneity. In the past few years, next generation sequencing and analysis of hundreds of affected trio families have led to the identification of scores of ASD risk genes with high statistical confidence, in particular, genes carrying *de novo* loss-of-function variants. From these studies, it has become clear that no single locus accounts for >1% of ASD cases and the number of contributing genes is likely to be in the hundreds (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012a, b; Sanders et al., 2012; Talkowski et al., 2012; Jiang et al., 2013; De Rubeis et al., 2014; Iossifov et al., 2014; L. Liu et al., 2014; Sanders et al., 2015). Ironically, the genetic heterogeneity that had been a challenge for the field can now be leveraged to provide clues into etiology: by analyzing these genes as a group to identify convergent molecular pathways and cellular processes. Importantly, understanding the spatiotemporal convergence of ASD-related mutations will not only inform the potential neural substrates of the disorder but also enable the design of productive functional studies (State and Šestan, 2012).

Several of the large transcriptome resources we highlight provide a sufficiently rich dataset for gene coexpression analyses, which offer insights into potential relationships between genes by assessing correlation across multiple samples, thus revealing nuances that may not be apparent in binary analyses of differentially expressed genes. Numerous studies have harnessed the power of

network analysis, which identifies networks of genes that covary across multiple samples (Langfelder and Horvath, 2008). This allows for identification of "modules" of connected genes with "hub" genes representing the most central and representative members of the group. In pioneering studies, Willsey et al. (2013) and Parikshak et al. (2013) applied network analysis to ASD-associated genes to gain insights into pathogenesis. Willsey et al. (2013) hypothesized that analysis of the normal expression patterns of high confidence ASD genes can inform the brain region, developmental timing, and cell type underpinning ASD pathogenesis. Using the spatially rich and developmental stage-precise mRNA expression data of the HB Atlas (Kang et al., 2011), they used the nine genes with the strongest genetic evidence for association with ASD as seed genes to construct gene coexpression networks. Further intersection with layer-specific expression data (Fertuzinhos et al., 2014; Miller et al., 2014) led to the unbiased identification of a key point of convergence in midfetal deep-layer cortical projection neurons. Parikshak et al. (2013) used a slightly different strategy, leveraging coexpression modules built from analysis of gene expression in ASD brains (Voineagu et al., 2011) and using a longer list of ASD-associated genes. By intersecting with coexpression modules generated by Kang et al. (2011) and layer-specific transcriptome data from adult macaque brain (Bernard et al., 2012), Parikshak et al. (2013) similarly implicated the midfetal developing cortex, and additionally identified upper-layer cortical neurons as a potential site of pathogenesis. Together, these studies implicate midfetal cortical development as an important nexus of convergence for ASD etiology. Of note, a similar study of schizophrenia by Gulsuner et al. (2013) using the same Kang et al. (2011) data, also identified the fetal prefrontal cortex as a likely site of pathological development for schizophrenia.

In a very recent study, Turner et al. (2017) used a different strategy, investigating the expression of candidate ASD genes in adult brain cell types. After identifying de novo SNVs and CNVs in ASD simplex families, they analyzed the expression of these genes using cell type-specific data generated in adult mice by bacTRAP (Doyle et al., 2008; Heiman et al., 2008). They concluded that potentially pathogenic de novo mutations are enriched for adult expression in striatal neurons, adding to the literature of ASD pathogenicity in striatal circuitry and identifying new mutations that may contribute to the disease (Turner et al., 2017). Interestingly, beyond informing the roles of identified candidate genes, transcriptomic data may be leveraged to estimate the potential risk of genes not yet directly associated with disease. This is based on the idea that genes covarying in expression level throughout the developmental trajectory are likely to serve a shared purpose or may be regulated in a similar fashion. For example, the DAWN algorithm, by modeling rare variations and gene coexpression data from HB Atlas (Kang et al., 2011), identified 127 genes that plausibly affect ASD risk, as well as a set of likely ASD subnetworks (L. Liu et al., 2014).

The trio study design that has seen success in the past few years has identified rare or private variants contributing a large effect to disease. Although rare variants can inform mechanisms of disease pathogenesis, emerging genetic data suggest that, for many neuropsychiatric disorders, polygenic contributions of common variants are responsible for the majority of disease risk (Craddock and Sklar, 2013; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Sanders et al., 2015), and their study will be critical to therapeutic development (Birnbaum and Weinberger, 2017). It remains to be seen whether transcriptome databases will be useful in dissecting

the individually smaller contribution of common variations to brain disorders. A few studies, however, have used these resources to gain biological insights into genetic data. For example, a landmark study identifying risk for schizophrenia at 108 commonly variant loci (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) used bacTRAP data (Doyle et al., 2008; Heiman et al., 2008) to interrogate the expression patterns of genes near identified risk loci. This showed an enrichment of disease loci near transcriptional start sites of genes expressed in multiple neuronal lineages, but not in glial populations.

In the future, critically important to a full understanding of common genetic variation will be resources, such as the GTEx database, which will uncover links between genetic variants and expression changes. Furthermore, large-scale genomics and transcriptomics of human brains from neurotypical and diseased donors, such as the PsychENCODE project, are also expected to reveal molecular insights into disease. As we enter a new era in disease genomics, transcriptome data will become an increasingly important resource for dissecting genetic risk, disease pathogenesis, and therapeutic strategies.

References

Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, Jaffe AE, Pinto D, Dracheva S, Geschwind DH, Mill J, Nairn AC, Abyzov A, Pochareddy S, Prabhakar S, Weissman S, Sullivan PF, State MW, Weng Z, Peters MA, et al. (2015) The PsychENCODE project. Nat Neurosci 18:1707–1712. CrossRef Medline

Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, Macklis JD (2005) Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. Neuron 45:207–221. CrossRef Medline

Ataman B, Boulting GL, Harmin DA, Yang MG, Baker-Salisbury M, Yap EL, Malik AN, Mei K, Rubin AA, Spiegel I, Durresi E, Sharma N, Hu LS, Pletikos M, Griffith EC, Partlow JN, Stevens CR, Adli M, Chahrour M, Sestan N, et al. (2016) Evolution of Osteocrin as an activity-regulated factor in the primate brain. Nature 539:242–247. CrossRef Medline

Ayoub AE, Oh S, Xie Y, Leng J, Cotney J, Dominguez MH, Noonan JP, Rakic P (2011) Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. Proc Natl Acad Sci U S A 108:14950–14955. CrossRef Medline

Bae BI, Tietjen I, Atabay KD, Evrony GD, Johnson MB, Asare E, Wang PP, Murayama AY, Im K, Lisgo SN, Overman L, Šestan N, Chang BS, Barkovich AJ, Grant PE, Topçu M, Politsky J, Okano H, Piao X, Walsh CA (2014) Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. Science 343:764–768. CrossRef Medline

Bakken TE, Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, Dalley RA, Royall JJ, Lemon T, Shapouri S, Aiona K, Arnold J, Bennett JL, Bertagnolli D, Bickley K, Boe A, Brouner K, Butler S, Byrnes E, et al. (2016) A comprehensive transcriptional map of primate brain development. Nature 535:367–375. CrossRef Medline

Baruzzo G, Hayer KE, Kim EJ, Di Camillo B, FitzGerald GA, Grant GR (2017) Simulation-based comprehensive benchmarking of RNA-seq aligners. Nat Methods 14:135–139. CrossRef Medline

Belgard TG, Marques AC, Oliver PL, Abaan HO, Sirey TM, Hoerder-Suabedissen A, García-Moreno F, Molnár Z, Margulies EH, Ponting CP (2011) A transcriptomic atlas of mouse neocortical layers. Neuron 71: 605–616. CrossRef Medline

Bennett ML, Bennett FC, Liddelow SA, Ajami B, Zamanian JL, Fernhoff NB, Mulinyawe SB, Bohlen CJ, Adil A, Tucker A, Weissman IL, Chang EF, Li G, Grant GA, Hayden Gephart MG, Barres BA (2016) New tools for studying microglia in the mouse and human CNS. Proc Natl Acad Sci U S A 113:E1738 –E1746. CrossRef Medline

Bernard A, Lubbers LS, Tanis KQ, Luo R, Podtelezhnikov AA, Finney EM, McWhorter MM, Serikawa K, Lemon T, Morgan R, Copeland C, Smith K, Cullen V, Davis-Turak J, Lee CK, Sunkin SM, Loboda AP, Levine DM, Stone DJ, Hawrylycz MJ, et al. (2012) Transcriptional architecture of the primate neocortex. Neuron 73:1083–1099. CrossRef Medline

Birnbaum R, Weinberger DR (2017) Genetic insights into the neurodevel-

opmental origins of schizophrenia. Nat Rev Neurosci 18:727–740. CrossRef Medline

Britanova O, de Juan Romero C, Cheung A, Kwan KY, Schwark M, Gyorgy A, Vogel T, Akopov S, Mitkovski M, Agoston D, Sestan N, Molnár Z, Tarabykin V (2008) Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. Neuron 57:378–392. CrossRef Medline

Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA, Thompson WJ, Barres BA (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci 28:264–278. CrossRef Medline

Campbell JN, Macosko EZ, Fenselau H, Pers TH, Lyubetskaya A, Tenen D, Goldman M, Verstegen AM, Resch JM, McCarroll SA, Rosen ED, Lowell BB, Tsai LT (2017) A molecular census of arcuate hypothalamus and median eminence cell types. Nat Neurosci 20:484–496. CrossRef Medline

Carlyle BC, Kitchen RR, Kanyo JE, Voss EZ, Pletikos M, Sousa AM, Lam TT, Gerstein MB, Sestan N, Nairn AC (2017) A multiregional proteomic survey of the postnatal human brain. Nat Neurosci 20:1787–1795. CrossRef Medline

Cembrowski MS, Wang L, Sugino K, Shields BC, Spruston N (2016) Hipposeq: a comprehensive RNA-seq database of gene expression in hippocampal principal neurons. Elife 5:e14997. CrossRef Medline

Chen JG, Rasin MR, Kwan KY, Sestan N (2005) Zfp312 is required for subcortical axonal projections and dendritic morphology of deep-layer pyramidal neurons of the cerebral cortex. Proc Natl Acad Sci U S A 102: 17792–17797. CrossRef Medline

Chen M, Zhou X (2017) Controlling for confounding effects in single cell RNA sequencing studies using both control and target genes. Sci Rep 7:13587. CrossRef Medline

Chen YJ, Friedman BA, Ha C, Durinck S, Liu J, Rubenstein JL, Seshagiri S, Modrusan Z (2017) Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types. Sci Rep 7:45656. CrossRef Medline

Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, Colantuoni EA, Elkahloun AG, Herman MM, Weinberger DR, Kleinman JE (2011) Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature 478:519–523. CrossRef Medline

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322:1845–1848. CrossRef Medline

Craddock N, Sklar P (2013) Genetics of bipolar disorder. Lancet 381:1654–1662. CrossRef Medline

Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, Quake SR (2015) A survey of human brain transcriptome diversity at the single cell level. Proc Natl Acad Sci U S A 112:7285–7290. CrossRef Medline

Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, Zhang Y, Neff N, Kowarsky M, Caneda C, Li G, Chang SD, Connolly ID, Li Y, Barres BA, Gephart MH, Quake SR (2017) Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. Cell Rep 21:1399–1410. CrossRef Medline

De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, Singh T, Klei L, Kosmicki J, Shih-Chen F, Aleksic B, Biscaldi M, Bolton PF, Brownfeld JM, Cai J, Campbell NG, et al. (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 515:209–215. CrossRef Medline

Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, Lin-Marq N, Koch M, Bilio M, Cantiello I, Verde R, De Masi C, Bianchi SA, Cicchini J, Perroud E, Mehmeti S, et al. (2011) A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol 9:e1000582. CrossRef Medline

Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, Gong S, Greengard P, Heintz N (2008) Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell 135:749–762. CrossRef Medline

Dulken BW, Leeman DS, Boutet SC, Hebestreit K, Brunet A (2017) Single-cell transcriptomic analysis defines heterogeneity and transcriptional dynamics in the adult neural stem cell lineage. Cell Rep 18:777–790. CrossRef Medline

Edfors F, Danielsson F, Hallström BM, Käll L, Lundberg E, Pontén F,

Forsström B, Uhlén M (2016) Gene-specific correlation of RNA and protein levels in human cells and tissues. Mol Syst Biol 12:883. CrossRef Medline

Ekstrand MI, Nectow AR, Knight ZA, Latcha KN, Pomeranz LE, Friedman JM (2014) Molecular profiling of neurons based on connectivity. Cell 157:1230–1242. CrossRef Medline

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. CrossRef Medline

Everaert C, Luypaert M, Maag JLV, Cheng QX, Dinger ME, Hellemans J, Mestdagh P (2017) Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Sci Rep 7:1559. CrossRef Medline

Fang Z, Cui X (2011) Design and validation issues in RNA-seq experiments. Brief Bioinform 12:280–287. CrossRef Medline

Feng G, Mellor RH, Bernstein M, Keller-Peck C, Nguyen QT, Wallace M, Nerbonne JM, Lichtman JW, Sanes JR (2000) Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. Neuron 28:41–51. CrossRef Medline

Fertuzinhos S, Li M, Kawasawa YI, Ivic V, Franjic D, Singh D, Crair M, Sestan N (2014) Laminar and temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex. Cell Rep 6:938–950. CrossRef Medline

Franks A, Airoldi E, Slavov N (2017) Post-transcriptional regulation across human tissues. PLoS Comput Biol 13:e1005535. CrossRef Medline

Fuzik J, Zeisel A, Máté Z, Calvigioni D, Yanagawa Y, Szabó G, Linnarsson S, Harkany T (2016) Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. Nat Biotechnol 34:175–183. CrossRef Medline

Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Pantic N, Admassu T, James P, Warland A, Jordan M, Ciccone J, Serra S, Keenan J, Martin S, McNeill L, Wallace EJ, Jayasinghe L, Wright C, Blasco J, Young S, et al. (2018) Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods. Advance online publication. Retrieved Jan. 15, 2018. doi: 10.1038/nmeth.4577. CrossRef Medline

Gokce O, Stanley GM, Treutlein B, Neff NF, Camp JG, Malenka RC, Rothwell PE, Fuccillo MV, Südhof TC, Quake SR (2016) Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-seq. Cell Rep 16:1126–1137. CrossRef Medline

Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, Heintz N (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 425:917–925. CrossRef Medline

GTEx Consortium (2015) Human genomics. the genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348:648–660. CrossRef Medline

GTEx Consortium (2017) Genetic effects on gene expression across human tissues. Nature 550:204–213. CrossRef Medline

Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, Rippey C, Shahin H, Nimgaonkar VL, Go RC, Savage RM, Swerdlow NR, Gur RE, Braff DL, King MC, McClellan JM (2013) Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell 154:518–529. CrossRef Medline

Habib N, Li Y, Heidenreich M, Swiech L, Avraham-Davidi I, Trombetta JJ, Hession C, Zhang F, Regev A (2016) Div-seq: single-nucleus RNA-seq reveals dynamics of rare adult newborn neurons. Science 353:925–928. CrossRef Medline

Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, Zhang F, Regev A (2017) Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods 14:955–958. CrossRef Medline

Han W, Kwan KY, Shim S, Lam MM, Shin Y, Xu X, Zhu Y, Li M, Sestan N (2011) TBR1 directly represses Fezf2 to control the laminar origin and development of the corticospinal tract. Proc Natl Acad Sci U S A 108: 3041–3046. CrossRef Medline

Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, Abajian C, Beckmann CF, Bernard A, Bertagnolli D, Boe AF, Cartagena PM, Chakravarty MM, Chapin M, Chong J, Dalley RA, et al. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489:391–399. CrossRef Medline

Hawrylycz M, Ng L, Page D, Morris J, Lau C, Faber S, Faber V, Sunkin S, Menon V, Lein E, Jones A (2011) Multi-scale correlation structure of

gene expression in the brain. Neural Netw 24:933–942. CrossRef Medline

He Z, Han D, Efimova O, Guijarro P, Yu Q, Oleksiak A, Jiang S, Anokhin K, Velichkovsky B, Grünewald S, Khaitovich P (2017) Comprehensive transcriptome analysis of neocortical layers in humans, chimpanzees and macaques. Nat Neurosci 20:886–895. CrossRef Medline

Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, Suárez-Fariñas M, Schwarz C, Stephan DA, Surmeier DJ, Greengard P, Heintz N (2008) A translational profiling approach for the molecular characterization of CNS cell types. Cell 135:738–748. CrossRef Medline

Hitzemann R, Darakjian P, Walter N, Iancu OD, Searles R, McWeeney S (2014) Introduction to sequencing the brain transcriptome. Int Rev Neurobiol 116:1–19. CrossRef Medline

Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekritsky M, Pradhan K, et al. (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74:285–299. CrossRef Medline

Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, Smith JD, Paeper B, Nickerson DA, Dea J, Dong S, Gonzalez LE, Mandell JD, Mane SM, Murtha MT, Sullivan CA, et l. (2014) The contribution of de novo coding mutations to autism spectrum disorder. Nature 515:216–221. CrossRef Medline

Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P, Linnarsson S (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods 11:163–166. CrossRef Medline

Jaffe AE, Tao R, Norris AL, Kealhofer M, Nellore A, Shin JH, Kim D, Jia Y, Hyde TM, Kleinman JE, Straub RE, Leek JT, Weinberger DR (2017) qSVA framework for RNA quality correction in differential expression analysis. Proc Natl Acad Sci U S A 114:7130–7135. CrossRef Medline

Jiang YH, Yuen RK, Jin X, Wang M, Chen N, Wu X, Ju J, Mei J, Shi Y, He M, Wang G, Liang J, Wang Z, Cao D, Carter MT, Chrysler C, Drmic IE, Howe JL, Lau L, Marshall CR, et al. (2013) Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am J Hum Genet 93:249–263. CrossRef Medline

Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanović D, Geschwind DH, Mane SM, State MW, Sestan N (2009) Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron 62:494–509. CrossRef Medline

Johnson MB, Wang PP, Atabay KD, Murphy EA, Doan RN, Hecht JL, Walsh CA (2015) Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. Nat Neurosci 18:637–646. CrossRef Medline

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AM, Pletikos M, Meyer KA, Sedmak G, Guennel T, Shin Y, Johnson MB, Krsnik Z, Mayer S, Fertuzinhos S, Umlauf S, Lisgo SN, Vortmeyer A, Weinberger DR, et al. (2011) Spatio-temporal transcriptome of the human brain. Nature 478: 483–489. CrossRef Medline

Kee N, Volakakis N, Kirkeby A, Dahl L, Storvall H, Nolbrant S, Lahti L, Björklund ÅK, Gillberg L, Joodmardi E, Sandberg R, Parmar M, Perlmann T (2017) Single-cell analysis reveals a close relationship between differentiating dopamine and subthalamic nucleus neuronal lineages. Cell Stem Cell 20:29–40. CrossRef Medline

Kwan KY, Lam MM, Johnson MB, Dube U, Shim S, Rašin MR, Sousa AM, Fertuzinhos S, Chen JG, Arellano JI, Chan DW, Pletikos M, Vasung L, Rowitch DH, Huang EJ, Schwartz ML, Willemsen R, Oostra BA, Rakic P, Heffer M, et al. (2012) Species-dependent posttranscriptional regulation of NOS1 by FMRP in the developing cerebral cortex. Cell 149:899–911. CrossRef Medline

Kwan KY, Lam MM, Krsnik Z, Kawasawa YI, Lefebvre V, Sestan N (2008) SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons. Proc Natl Acad Sci U S A 105:16021–16026. CrossRef Medline

Lacar B, Linker SB, Jaeger BN, Krishnaswami S, Barron J, Kelder M, Parylak S, Paquola A, Venepally P, Novotny M, O'Connor C, Fitzpatrick C, Erwin J, Hsu JY, Husband D, McConnell MJ, Lasken R, Gage FH (2016) Nuclear RNA-seq of single neurons reveals molecular signatures of activation. Nat Commun 7:11022. CrossRef Medline

Lai T, Jabaudon D, Molyneaux BJ, Azim E, Arlotta P, Menezes JR, Macklis JD (2008) SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. Neuron 57:232–247. CrossRef Medline

Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, Wildberg A, Gao D,

Fung HL, Chen S, Vijayaraghavan R, Wong J, Chen A, Sheng X, Kaper F, Shen R, Ronaghi M, Fan JB, Wang W, Chun J, et al. (2016) Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 352:1586–1590. CrossRef Medline

Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, Duong TE, Gao D, Chun J, Kharchenko PV, Zhang K (2018) Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol 36:70–80. CrossRef Medline

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559. CrossRef Medline

Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445:168–176. CrossRef Medline

Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN (2010) RNA-seq gene expression estimation with read mapping uncertainty. Bioinformatics 26: 493–500. CrossRef Medline

Li CL, Li KC, Wu D, Chen Y, Luo H, Zhao JR, Wang SS, Sun MM, Lu YJ, Zhong YQ, Hu XY, Hou R, Zhou BB, Bao L, Xiao HS, Zhang X (2016) Somatosensory neuron types identified by high-coverage single-cell RNA-sequencing and functional heterogeneity. Cell Res 26:83–102. CrossRef Medline

Liddelow SA, Guttenplan KA, Clarke LE, Bennett FC, Bohlen CJ, Schirmer L, Bennett ML, Münch AE, Chung WS, Peterson TC, Wilton DK, Frouin A, Napier BA, Panicker N, Kumar R, Buckwalter MS, Rowitch DH, Dawson VL, Dawson TM, Stevens B, et al. (2017) Neurotoxic reactive astrocytes are induced by activated microglia. Nature 541: 481–487. CrossRef Medline

Liu L, Lei J, Sanders SJ, Willsey AJ, Kou Y, Cicek AE, Klei L, Lu C, He X, Li M, Muhle RA, Ma'ayan A, Noonan JP, Sestan N, McFadden KA, State MW, Buxbaum JD, Devlin B, Roeder K (2014) DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. Mol Autism 5:22. CrossRef Medline

Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ, He D, Weissman JS, Kriegstein AR, Diaz AA, Lim DA (2016a) Single-cell analysis of long non-coding RNAs in the developing human neocortex. Genome Biol 17:67. CrossRef Medline

Liu Y, Beyer A, Aebersold R (2016b) On the dependency of cellular protein levels on mRNA abundance. Cell 165:535–550. CrossRef Medline

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. CrossRef Medline

Lui H, Zhang J, Makinson SR, Cahill MK, Kelley KW, Huang HY, Shang Y, Oldham MC, Martens LH, Gao F, Coppola G, Sloan SA, Hsieh CL, Kim CC, Bigio EH, Weintraub S, Mesulam MM, Rademakers R, Mackenzie IR, Seeley WW, et al. (2016) Progranulin deficiency promotes circuit-specific synaptic pruning by microglia via complement activation. Cell 165:921–935. CrossRef Medline

Luo Y, Coskun V, Liang A, Yu J, Cheng L, Ge W, Shi Z, Zhang K, Li C, Cui Y, Lin H, Luo D, Wang J, Lin C, Dai Z, Zhu H, Zhang J, Liu J, Liu H, deVellis J, et al. (2015) Single-cell transcriptome analyses reveal signals to activate dormant neural stem cells. Cell 161:1175–1186. CrossRef Medline

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214. CrossRef Medline

Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, Arnold JM, Bennet C, Bertagnolli D, Brouner K, Butler S, Caldejon S, Carey A, Cuhaciyan C, Dalley RA, Dee N, et al. (2014) Transcriptional landscape of the prenatal human brain. Nature 508:199–206. CrossRef Medline

Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, Urich MA, Nery JR, Sejnowski TJ, Lister R, Eddy SR, Ecker JR, Nathans J (2015) Epigenomic signatures of neuronal diversity in the mammalian brain. Neuron 86:1369–1384. CrossRef Medline

Molyneaux BJ, Goff LA, Brettler AC, Chen HH, Hrvatin S, Rinn JL, Arlotta P (2015) DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. Neuron 85:275–288. CrossRef Medline

Molyneaux BJ, Arlotta P, Hirata T, Hibi M, Macklis JD (2005) Fezl is required for the birth and specification of corticospinal motor neurons. Neuron 47:817–831. CrossRef Medline

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 5:621–628. CrossRef Medline

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, et al. (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485:242–245. CrossRef Medline

Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, Haeussler M, Sandoval-Espinosa C, Liu SJ, Velmeshev D, Ounadjela JR, Shuga J, Wang X, Lim DA, West JA, Leyrat AA, Kent WJ, Kriegstein AR (2017) Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science 358:1318–1323. CrossRef Medline

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, et al. (2012a) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485:246–250. CrossRef Medline

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, Munson J, Hiatt JB, Turner EH, Levy R, O'Day DR, Krumm N, Coe BP, Martin BK, Borenstein E, Nickerson DA, et al. (2012b) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 338:1619–1622. CrossRef Medline

Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, Horvath S, Geschwind DH (2013) Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155:1008–1021. CrossRef Medline

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419. CrossRef Medline

Paul A, Crow M, Raudales R, He M, Gillis J, Huang ZJ (2017) Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. Cell 171:522–539.e20. CrossRef Medline

Paulsen MT, Veloso A, Prasad J, Bedi K, Ljungman EA, Magnuson B, Wilson TE, Ljungman M (2014) Use of bru-seq and BruChase-seq for genome-wide assessment of the synthesis and stability of RNA. Methods 67:45–54. CrossRef Medline

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L (2017) Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods 14:687–690. CrossRef Medline

Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, Ramalingam N, Sun G, Thu M, Norris M, Lebofsky R, Toppani D, Kemp DW 2nd, Wong M, Clerkson B, Jones BN, et al. (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat Biotechnol 32:1053–1058. CrossRef Medline

Quadrato G, Nguyen T, Macosko EZ, Sherwood JL, Min Yang S, Berger DR, Maria N, Scholvin J, Goldman M, Kinney JP, Boyden ES, Lichtman JW, Williams ZM, McCarroll SA, Arlotta P (2017) Cell diversity and network dynamics in photosensitive human brain organoids. Nature 545:48–53. CrossRef Medline

Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Göttgens B, Hacohen N, et al. (2017) The human cell atlas. Elife 6:e27041. CrossRef Medline

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. CrossRef Medline

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, et al. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485:237–241. CrossRef Medline

Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, Goldberg AP, Jinlu C, Keaney JF 3rd, Klei L, Mandell JD, Moreno-De-Luca D, Poultney CS, Robinson EB, Smith L, Solli-Nowlan T, et al. (2015) Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 87:1215–1233. CrossRef Medline

Sanz E, Yang L, Su T, Morris DR, McKnight GS, Amieux PS (2009) Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proc Natl Acad Sci U S A 106:13939–13944. CrossRef Medline

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

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. Nature 473:337–342. CrossRef Medline

Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin JZ, Nemesh J, Goldman M, McCarroll SA, Cepko CL, Regev A, Sanes JR (2016) Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. Cell 166:1308–1323.e30. CrossRef Medline

Shim S, Kwan KY, Li M, Lefebvre V, Sestan N (2012) Cis-regulatory control of corticospinal development and evolution. Nature 486:74–79. CrossRef Medline

Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, Enikolopov G, Nauen DW, Christian KM, Ming GL, Song H (2015) Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis. Cell Stem Cell 17:360–372. CrossRef Medline

Sousa AM, Zhu Y, Raghanti MA, Kitchen RR, Onorati M, Tebbenkamp AT, Stutz B, Meyer KA, Li M, Kawasawa YI, Liu F, Perez RG, Mele M, Carvalho J, Skarica M, Gulden FO, Pletikos M, Shibata A, Stephenson AR, Edler MK, et al. (2017) Molecular and cellular reorganization of neural circuits in the human lineage. Science 358:1027–1032. CrossRef Medline

State MW, Šestan N (2012) The emerging biology of autism spectrum disorders. Science 337:1301–1303. CrossRef Medline

Sugino K, Hempel CM, Miller MN, Hattox AM, Shapiro P, Wu C, Huang ZJ, Nelson SB (2006) Molecular taxonomy of major neuronal classes in the adult mouse forebrain. Nat Neurosci 9:99–107. CrossRef Medline

Sugino K, Clark E, Schulmann A, Shima Y, Wang L, Hunt DL, Hooks BM, Trankner D, Chandrashekar J, Picard S, Lemire A, Spruston N, Hantman A, Nelson SB (2017) The transcriptional logic of mammalian neuronal diversity. Advance online publication. bioRxiv. CrossRef

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, Pereira S, Ruderfer D, Kirby A, Ripke S, Harris DJ, Lee JH, Ha K, Kim HG, Solomon BD, Gropman AL, et al. (2012) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell 149:525–537. CrossRef Medline

Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, et al. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci 19:335–346. CrossRef Medline

Telley L, Govindan S, Prados J, Stevant I, Nef S, Dermitzakis E, Dayer A, Jabaudon D (2016) Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. Science 351:1443–1446. CrossRef Medline

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515. CrossRef Medline

Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC, Kronenberg ZN, Hormozdiari F, Raja A, Pennacchio LA, Darnell RB, Eichler EE (2017) Genomic patterns of de novo mutation in simplex autism. Cell 171:710–722.e12. CrossRef Medline

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474:380–384. CrossRef Medline

Wagner GP, Kin K, Lynch VJ (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci 131:281–285. CrossRef Medline

Williams CR, Baccarella A, Parrish JZ, Kim CC (2017) Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-seq. BMC Bioinformatics 18:38. CrossRef Medline

Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, Murtha MT, Bichsel C, Niu W, Cotney J, Ercan-Sencicek AG, Gockley J, Gupta AR, Han W, He X, Hoffman EJ, et al. (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155:997–1007. CrossRef Medline

Xu X, Wells AB, O'Brien DR, Nehorai A, Dougherty JD (2014) Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. J Neurosci 34:1420–1431. CrossRef Medline

Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S (2015) Brain structure. cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 347:1138–1142. CrossRef Medline

Zeng H, Shen EH, Hohmann JG, Oh SW, Bernard A, Royall JJ, Glattfelder KJ, Sunkin SM, Morris JA, Guillozet-Bongaarts AL, Smith KA, Ebbert AJ, Swanson B, Kuan L, Page DT, Overly CC, Lein ES, Hawrylycz MJ, Hof PR, Hyde TM, et al. (2012) Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. Cell 149: 483–496. CrossRef Medline

Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N, Deng S, Liddelow SA, Zhang C, Daneman R, Maniatis T, Barres BA, Wu JQ (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci 34:11929–11947. CrossRef Medline

Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, Vogel H, Steinberg GK, Edwards MS, Li G, Duncan JA 3rd, Cheshier SH, Shuer LM, Chang EF, Grant GA, Gephart MG, Barres BA (2016) Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. Neuron 89:37–53. CrossRef Medline