Brain transcriptome atlases

A computational perspective

Mahfouz, Ahmed; Huisman, Sjoerd M.H.; Lelieveldt, Boudewijn P.F.; Reinders, Marcel J.T.

Citation (APA)
Mahfouz, A., Huisman, S. M. H., Lelieveldt, B. P. F., & Reinders, M. J. T. (2017). Brain transcriptome atlases: A computational perspective. *Brain Structure and Function*, 1557-1580. https://doi.org/10.1007/s00429-016-1338-2

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

CrossMark

REVIEW

Brain transcriptome atlases: a computational perspective

Ahmed Mahfouz[1,2] · Sjoerd M. H. Huisman[1,2] · Boudewijn P. F. Lelieveldt[1,2] ·
Marcel J. T. Reinders[2]

Abstract The immense complexity of the mammalian brain is largely reflected in the underlying molecular signatures of its billions of cells. Brain transcriptome atlases provide valuable insights into gene expression patterns across different brain areas throughout the course of development. Such atlases allow researchers to probe the molecular mechanisms which define neuronal identities, neuroanatomy, and patterns of connectivity. Despite the immense effort put into generating such atlases, to answer fundamental questions in neuroscience, an even greater effort is needed to develop methods to probe the resulting high-dimensional multivariate data. We provide a comprehensive overview of the various computational methods used to analyze brain transcriptome atlases.

Keywords Brain atlases · Gene expression · Co-expression · Omics integration · Imaging genetics

Mapping gene expression in the brain

The mammalian brain is a complex system consisting of billions of neuronal and glia cells that can be categorized into hundreds of different subtypes. Understanding the organization of these cells, throughout development, into functional circuits carrying out sophisticated cognitive tasks can help us better characterize disease-associated changes. Advances in technology and automation of laboratory procedures have facilitated high-throughput characterization of functional neuronal circuits and connections at different scales (Pollock et al. 2014). For example, the Human Connectome Project maps the complete wiring of the brain using magnetic resonance imaging (Van Essen and Ugurbil 2012). Despite the importance of these imaging modalities in characterizing brain pathologies and development, it is imperative to analyze the molecular structure to gain a better mechanistic understanding of how the brain works. However, studying the molecular mechanisms of the brain has proved very challenging due to the unknown large number of cell types (Sunkin 2006).

The complexity of the brain is largely reflected in the underlying patterns of gene expression that defines neuronal identities, neuroanatomy, and patterns of connectivity. With 80% of the 20,000 genes in the mammalian genome expressed in the brain (Lein et al. 2007), characterizing spatial and temporal gene expression patterns can provide valuable insights into the relationship between genes and brain function and their role throughout neurodevelopment. Brain transcriptome atlases have proven to be extremely instrumental for this task.

Following earlier progress in other model organisms (Kim et al. 2001; Spencer et al. 2011; Milyaev et al. 2012), several projects have assessed gene expression in the mouse brain with various degrees of coverage for genes, anatomical regions, and developmental time-points (Sunkin 2006; Pollock et al. 2014). In rodents, the Gene Expression Nervous System Atlas (GENSAT) (Gong et al. 2003; Heintz 2004) and GenePaint (Visel et al. 2004) mapped gene expression in both the adult and developing mouse brain, while the EurExpress (Diez-Roux et al. 2011) and the e-Mouse Atlas of Gene Expression (EMAGE) (Richardson et al. 2014) focused on the developing mouse

✉ Ahmed Mahfouz
a.mahfouz@lumc.nl

1 Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands

2 Delft Bioinformatics Laboratory, Delft University of Technology, Delft, The Netherlands

brain. Comparable atlases of gene expression in the human brain are far less abundant due to the challenges posed by difference in size between the human and mouse brain as well as the scarcity of post-mortem tissue. However, several studies have profiled the human brain transcriptome to analyze expression variation across the brain (Lonsdale 2013), expression developmental dynamics (Oldham et al. 2008; Colantuoni et al. 2011; Kang et al. 2011), and differential expression in the autistic brain (Voineagu et al. 2011), albeit in a limited number of coarse brain regions.

The Allen Institute for Brain Science provides the most comprehensive maps of gene expression in the mouse and human brain in terms of the number of genes, the spatial-resolution, and the developmental stages covered (Pollock et al. 2014). Several atlases have been released which map gene expression in the adult and developing mouse brain (Lein et al. 2007; Thompson et al. 2014), the adult and developing human brain (Hawrylycz et al. 2012; Miller et al. 2014a), and the adult and developing non-human primate (NHP) brain (Bernard et al. 2012; Bakken et al. 2016); see Fig. 1. Sunkin et al. (2013) provides a complete review of the Allen Brain Atlas resources.

The availability of genome-wide spatially mapped gene expression data provides a great opportunity to understand the complexity of the mammalian brain. It provides the necessary data to decode the molecular functions of different cell populations and brain nuclei. However, the diversity of cell types and their molecular signatures and the effect of mutations on the brain remain poorly understood. For example, de novo loss-of-function mutations in autistic children have been shown to converge on three distinct pathways: synaptic function, Wnt signaling, and chromatin remodeling (Krumm et al. 2014; De Rubeis et al. 2014). Except for the synaptic role of autism-related genes, it is not clear how alternations in basic cell functions, such as Wnt signaling and chromatin remodeling, can result in the complex phenotype of autism spectrum disorders (ASD). A recent effort to map somatic mutations in cortical neurons using single-cell sequencing has shown that neurons have on average ~ 1500 transcription-associated mutations (Lodato et al. 2015). The significant association of these single-neuron mutations and genes with cortical expression indicates the vulnerability of genes active in human neurons to somatic mutations, even in normal individuals. The difference between these patterns in the normal and diseases brains remains unclear. Efforts to understand genotype-phenotype relationships in the brain face several challenges, including the complexity of the underlying molecular mechanisms and the poor definition of clinically based neurological disorders. In addition, the high-dimensionality of the data makes most studies underpowered to detect any associations. This is especially true in the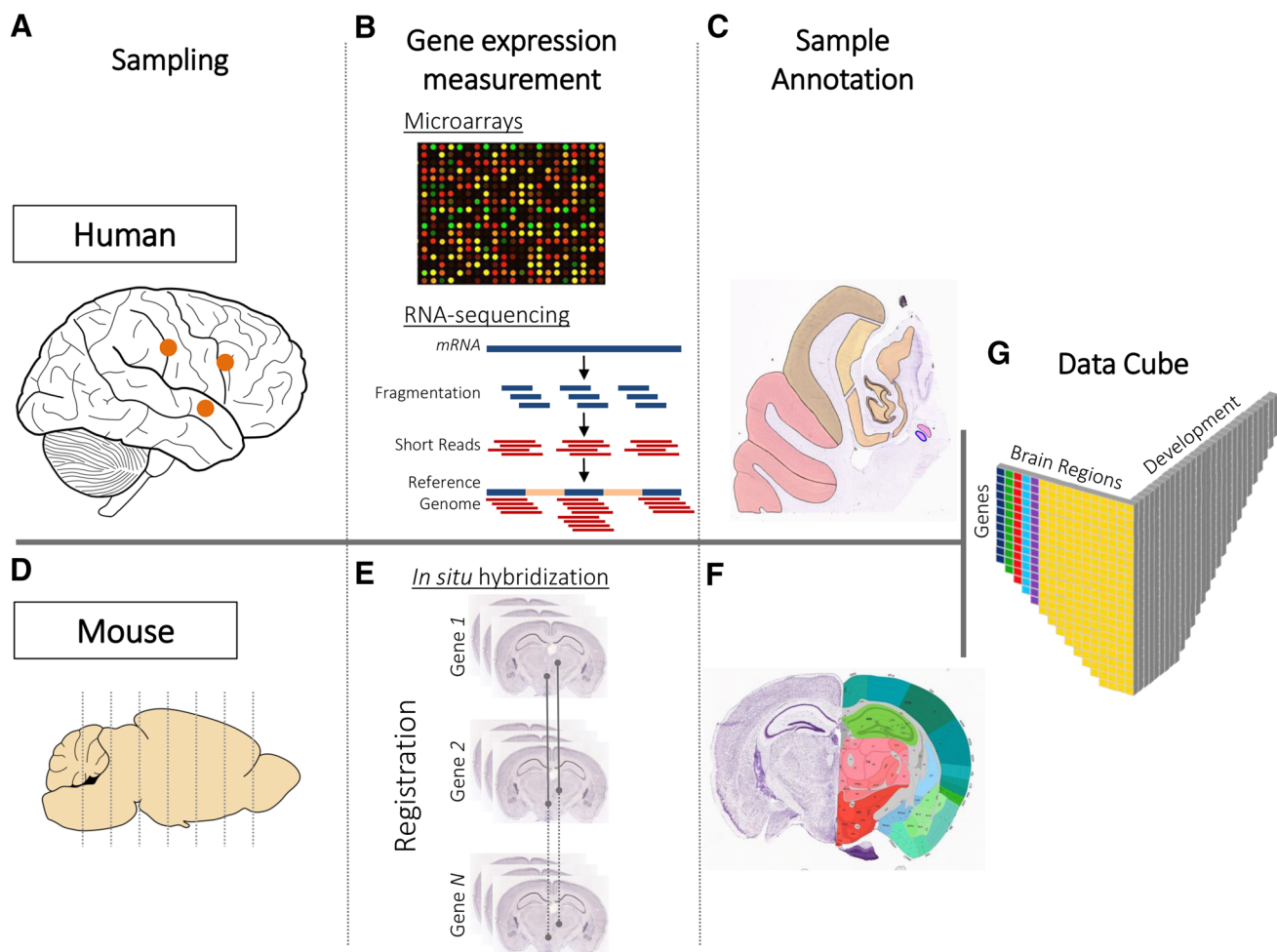 case of testing genetic associations with phenotype markers, such as imaging measurements (Medland et al. 2014). A combination of efforts to map the genomic landscape of the brain and data-driven approaches can add to our understanding of the underlying genetic etiology of neurological processes and how they are altered in neurological disorders.

Several review articles provide extensive insights into the gene expression maps of the brain. French and Pavlidis (2007) provide a global overview of neuroinformatics, including ontology, semantics, databases, connectivity, electrophysiology, and computational neuroscience. Jones et al. (2009) give an overview on developing the mouse atlas, the challenges faced, the community reaction, limitations, and atlas usage examples, as well as the data mining tools provided by the Allen institute. Pollock et al. (2014) provide a detailed review of the technology and tools which are currently advancing the field of molecular neuroanatomy. Recently, Parikshak et al. (2015) illustrated the power of using network approaches to leverage our understanding of the genetic etiology of neurological disorders. Yet, a global overview of the computational methodologies applied to brain transcriptome atlases to increase our understanding of neurological processes and disorders remains missing.

In this review, we provide an overview of the computational approaches used to expand our understanding of the relationship between gene expression on one hand and the anatomical and functional organization of the mammalian brain on the other hand. We focus our discussion on spatial and temporal brain transcriptomes mapped by the Allen Institute for Brain Sciences. Nevertheless, we also discuss how the methods can be extended to epigenomes and proteomes of the brain and other human tissues. We describe the different computational approaches taken to analyze the high-dimensional data and how they have contributed to our understanding of the functional role of genes in the brain, molecular neuroanatomy, and genetic etiology of neurological disorders. Finally, we discuss how these methods can help solve some of the data-specific challenges, and how the integration of several data types can further our understanding of the brain at different scales, ranging from molecular to behavioral.

Computational analysis of spatial and temporal gene expression data in the brain

Spatio-temporal transcriptomes of the brain pose several challenges due to their high-dimensionality. In this section, we identify the different types of approaches taken to analyze the spatially mapped gene expression data. We show the strengths of each approach and demonstrate how it has enriched neuroscience research. We divide the

Fig. 1 Spatially mapped gene expression in the mammalian brain. To map gene expression across the human and mouse brains, the Allen Institute for Brain Sciences followed two different strategies. In the human brain, samples covering all brain regions are extracted (**a**) and gene expression is measured using either microarray or RNA-sequencing (Hawrylycz et al. 2012; Miller et al. 2014b) (**b**). Accompanying histology sections and MRI scans are acquired to localize samples. Manual delineation of anatomical regions on the histology sections allowed for accurate sample annotation (**c**). In the mouse brain, gene expression is measured in coronal and sagittal sections using in situ hybridization (Lein et al. 2007) (**d**). Several slices covering the mouse brain are extracted per gene. Image registration methods are used to align the set of sections acquired for each gene to a common reference atlas (**e**). Anatomical regions are delineated on the reference atlas allowing for sample annotation (**f**). Data from the mouse and human atlases can be represented in a data matrix of three dimensions representing: genes, brain regions, and developmental stages (in case of the developmental atlases) (**g**)
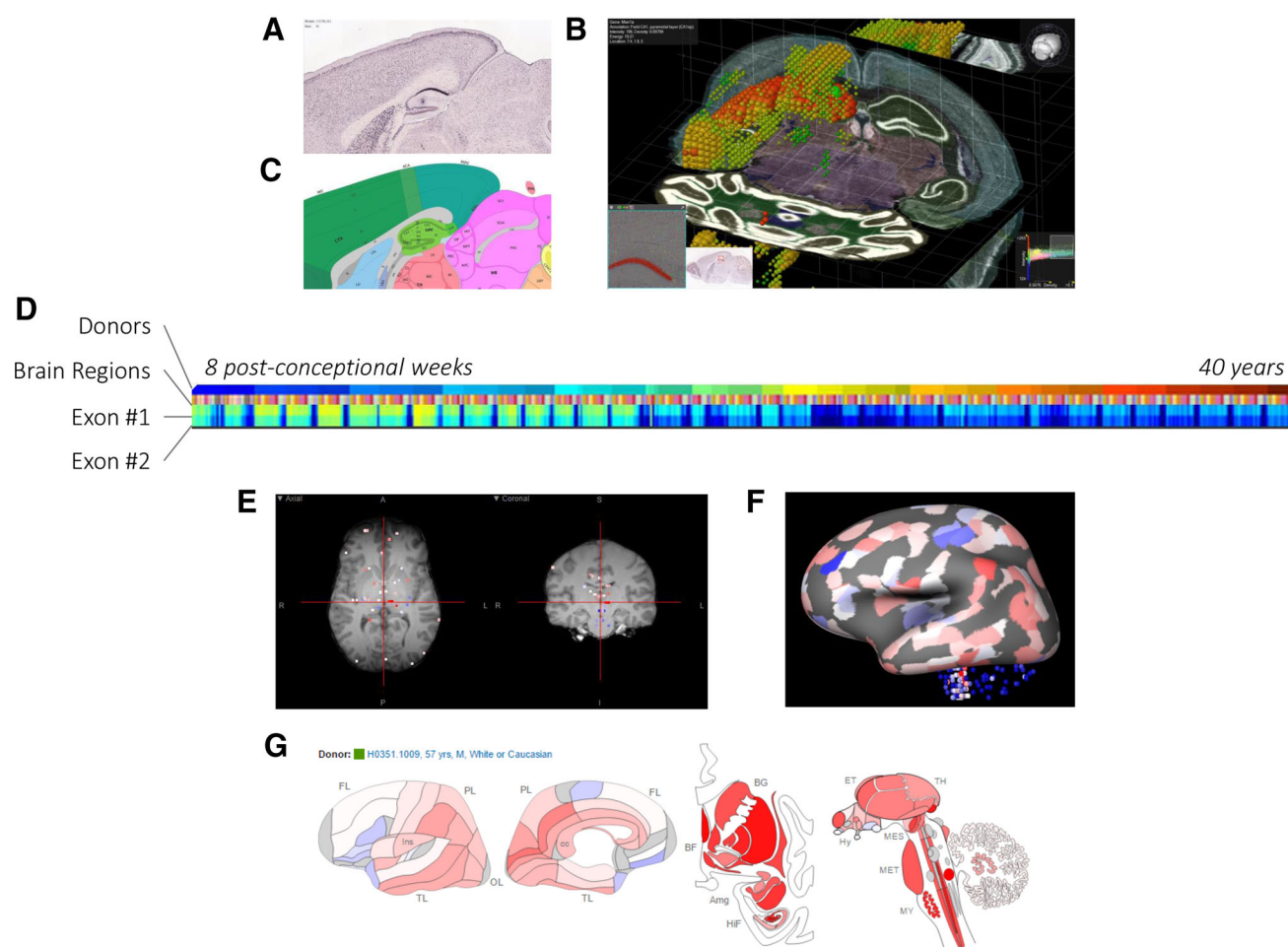
different methods into two categories. First, we describe a class of methods used to analyze the expression profile of gene(s) across different brain regions, cell types, and developmental stages. Second, we discuss methods focusing on the molecular organization and the genetic signature of the brain.

Analyzing the expression patterns of genes in the brain

Mapping gene expression across the brain is very helpful in determining the neural function of a gene of interest by associating it with a specific brain region and/or developmental stage or in identifying genetic markers of those

brain regions and developmental stages. Brain transcriptome atlases, such as the Allen Brain Atlases, provide useful information about the expression of a gene under "normal" conditions. Such information can be used to direct in-depth studies about a specific gene in biologically/clinically relevant cohorts. With the increasing number of genes implicated in neurological diseases as well as the realization that complex phenotypes of the brain likely result from the combined activity of several genes, a number of studies analyze gene sets rather than individual candidate genes. By studying the expression of a gene set rather than a single gene, neuroscientists are faced with a challenge on how to summarize this data to understand the relationship between genes and neuronal phenotypes.

Fig. 2 Gene expression visualization. Gene expression of spatially mapped samples can be visualized using several approaches. **a** Mouse gene expression data of the gene *Man1a* can be investigated using the original ISH sections. **b** BrainExplorer software allows visualization of the 3D expression volume with an overlay of the anatomical atlas and the ability to go back to the original high-resolution ISH section. **c** Simultaneously, viewing the ISH section and the corresponding atlas section helps in localizing gene expression to brain regions. **d** Heatmaps are commonly used to visualize gene expression. Expression of the two exons of the *NEUROD6* gene from the BrainSpan Atlas is visualized using a heatmap in which samples are ordered according to the age of the donor. **e** Samples from the Allen Human Brain Atlas are associated with coordinates of their location in the corresponding brain MRI. **f** Using the BrainExplorer, expression values of *Mecp2* can be mapped to an inflated white matter surface for better visualization of the cortex. **g** Alternatively, expression values can be mapped on an anatomical atlas of the human brain

Gene expression visualization

High-throughput data visualization approaches can facilitate the exploration of complex patterns in multivariate high-dimensional gene expression data sets (Pavlopoulos et al. 2015). For example, heatmaps are commonly used to visualize gene expression levels across a set of samples using a two-dimensional false-color image (Fig. 2f). However, heatmaps are not ideal to represent brain transcriptomes, because they fail to capture the multivariate nature of the data (genes, samples, and time-points) and to represent the inherent spatial and temporal relationships between different brain regions and developmental stages, respectively. To acquire high-resolution gene expression maps, the Allen atlases of the developing and adult mouse brain rely of ISH images (Fig. 2a). The Brain Explorer 3D viewer (Lau et al. 2008) is an interactive desktop application that allows the visualization of the 3D expression of one or more genes with the possibility to link them back to the high-resolution ISH images (Sunkin et al. 2013) (Fig. 2b). ISH images can be synchronized between different genes and also with the anatomical atlas of the mouse brain (Fig. 2c), facilitating the analysis of a group of genes. For the adult and developing human atlases, the gene expression data (microarray or RNA-seq) are mainly visualized using heatmaps (Fig. 2d). In the adult human atlas, the expression data can also be visualized on top of the magnetic resonance images (Fig. 2e). The Brain Explorer 3D viewer is also used to visualize gene expression from cortical samples using an inflated cortical

surface, a surface-based representation of the cortex that allows better representation of the relative locations of laminar, columnar, and areal features (Fig. 2f). In addition, gene expression can be mapped to an anatomical representation of the brain to facilitate interpretation (Fig. 2g). Ng et al. developed a method to construct surface-based flatmaps of the mouse cortex that enables mapping of gene expression data from the Allen Mouse Brain Atlas (Ng et al. 2010). Similarly, French (2015) developed a pipeline to map the expression of any gene from the Allen Human brain atlas to the cortical atlas built into the FreeSurfer software, which shall facilitate integration with medical imaging studies.

Summary statistics and visualization-based methods

The early studies employing the Allen Brain Atlases used a variety of visualization and qualitative measurements to analyze the expression of gene sets associated with dopamine neurotransmission (Björklund and Dunnett 2007), consummatory behavior in the mouse brain (Olszewski et al. 2008), midbrain dopaminergic neurons (Alavian and Simon 2009), and changes in locomotor activity in the mouse brain (Mignogna and Viggiano 2010). Kondapalli et al. (2014) used a similar qualitative approach to analyze the expression of Na^+/H^+ exchangers (NHE6 and NHE9), which are linked to several neuropsychiatric disorders, in the adult and developing mouse brain atlases.

To provide better quantitative representations of the expression of gene sets, several studies relied on basic summary statistics, such as the mean and standard deviation. Zaldivar and Krichmar (2013) used summations to summarize the expression of cholinergic, dopaminergic, noradrenergic, and serotonergic receptors in the amygdala, and in neuromodulatory areas. By plotting the average expression of genes harboring de novo loss-of-function mutations identified by means of exome sequencing across human brain development, Ben-David and Shifman (2012a) identified two clusters with antagonistic expression patterns across development. In addition, spatio-temporal exonic expression in the BrainSpan atlas correlates inversely with the burden of deleterious de novo mutations identified by exome sequencing in autism, schizophrenia, or intellectual disability (Uddin et al. 2014). For genes mutated in autism, the inverse relationship was found to be strongest in prenatal orbital frontal cortex, highlighting the value of the BrainSpan atlas to associate genetic variation with specific brain regions and developmental stages. Dahlin et al. (2009) developed a custom score (expression factor) of gene expression in the mouse brain based on the ISH images of the Allen Mouse Brain Atlas. They computed the mean and the standard deviation of the expression factor to assess the global expression and heterogeneity of solute carrier genes,

respectively. To deal with the qualitative ISH-based expression data from the Allen Mouse Brain Atlas, Roth et al. (2013) used a non-parametric representation of the data (using ranks instead of the raw expression values) to study the relationship between genes associated with grooming behavior in mice and 12 major brain structures.

Most of the studies analyzing gene expression in the brain focused on scores describing the expression of a gene or a gene set within each brain region of interest. Liu et al. (2014) proposed a characterization of the stratified expression pattern of sonic hedgehog (*Shh*), a classical signal molecule required for pattern formation along the dorsal–ventral axis, and its receptor *Ptch1*. Using a combination of differential expression, transcription factor motif analysis, and CHIP-seq, they identified the role of *Gata3*, *Fox2*, and their downstream targets in pattern formation in the early mouse brain. These results illustrate the power of characterizing complex expression patterns across the brain rather than solely summarizing the expression of each gene within individual brain regions.

Box1 | Gene Sets

Complex biological functions and disorders usually involve several rather than a single gene. Gene sets are groups of genes that share common biological functions and that can be defined either based on prior knowledge (e.g. about biochemical pathways or diseases) or experimental data (e.g. transcription factor targets identified using CHIP-seq). Gene set databases organize existing knowledge about these groups of genes by arranging them in sets that are associated with a functional term, such as a pathway name or a transcription factor that regulates the genes. Gene sets can be classified into 5 types:

Gene Ontology (GO)

The Gene Ontology project (Ashburner et al. 2000) developed three hierarchically structured vocabularies (ontologies) that describe gene products in terms of their associated *biological processes*, *cellular components* and *molecular functions*. Genes annotated with the same GO term(s) constitute a gene set.

Biological Pathways

Biological pathways are networks of molecular interactions underlying biological processes. Pathway databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata et al. 1999) and REACTOME (Croft et al. 2014), catalog physical entities (proteins and other macromolecules, small molecules, complexes of these entities and post-translationally modified forms of them), their subcellular locations and the transformations they can undergo (biochemical reaction, association to form a complex and translocation from one cellular compartment to another).

Transcription

Transcription databases include information on regulation of genes by transcription factors (TFs) binding to the DNA, or post-transcriptional regulation by microRNA binding to the mRNA. Determining these physical interactions can be done either *in silico* using computational inference (motif enrichment analysis) or using experimental data (such as CHIP-seq and microRNA binding data). For the motif enrichment analysis, position weight

matrices (PWMs) from databases TRANSFAC (Matys et al. 2006) and JASPER (Portales-Casamar et al. 2010) can be used to scan the promoters of genes in the region around the transcription factor start site (TSS). CHIP-seq data, such as the large collection of experiments from the Encyclopedia of DNA Elements (ENCODE) project (Bernstein et al. 2012b) and the Roadmap Epigenomics consortium (Consortium 2015a), is used to identify genes targeted by the TFs. Similarly, microRNA targets can be extracted from databases such as TargetScan (Lewis et al. 2003).

Cell-type markers

Cell type-specific transcriptional data provide a very rich source of cell type marker genes. Genes are identified as a cell type marker if they are up-regulated in one cell population compared to other cell populations. Several studies have used microarrays and RNA-seq to profile the transcriptome of a number of neuronal cell types (Cahoy et al. 2008; Zhang et al. 2014). Recently, studies are using single-cell sequencing to precisely capture the transcriptome of individual neuronal cells (Darmanis et al. 2015; Zeisel et al. 2015).

Disease

Genes can be grouped into sets based on their association to the same diseases. Public databases, such as OMIM (2015a) and DisGeNet (Pinero et al. 2015), contains curated information from literature and public sources on gene-disease association. Another source to obtain disease-related gene sets is by identifying genes harboring variants identified using GWAS (Simón-Sánchez and Singleton 2008; Welter et al. 2014), exome-sequencing (2015b), or whole-genome sequencing.

Identifying genes with localized expression patterns

The complexity of the brain implies that genes are involved in more than one function and that their function is region- or cell-type-specific. Neuronal cell types have been classically defined using cell morphology, electrophysiological and connectivity properties. Similarly, classical neuroanatomy identifies regions based on their cyto-, myelo-, or chemo-architecture. Genomic transcriptome measurements provide an alternative route to define functional cell types and brain regions based on their genetic makeup.

Several studies have analyzed the ISH-based gene expression images of the Allen Mouse Brain Atlas to identify cell-type-specific genes and genes with localized gene expression. Loerch et al. (2008) studied the localization of age-related gene expression changes in different neuronal cell types in the mouse and human brains. At the brain region level, David and Eddy (2009) developed ALLENMINER, a tool that searches the Allen Mouse Brain Atlas for genes with a specific expression pattern in a user-defined brain region. At a finer scale, Kirsch et al. (2012) described an approach to identify genes with a localized expression pattern in a specific layer of the mouse cerebellum. They represented each ISH image (gene) using a histogram of local binary patterns (LBP) at multiple-scales. Predicting the localization

of gene activity to each of the four cerebellar layers is done using two-level classification. First, they used a support vector machine (SVM) classifier to assign a cerebellar layer to each image and then used multiple-instance learning (MIL) to combine the resulting image classification into gene classification. Similarly, to identify cell-type specific genes, Li et al. (2014) used scale-invariant feature transform (SIFT) features of the ISH images. They further classified genes, using a supervised learning approach (regularized learning), based on their expression in different brain cell types. Zeng et al. (2015) compared two models to extract features from the ISH images of the developing mouse brain atlas to train a classification model to annotate gene expression patterns in brain structures. In one approach, they used SIFT features and the bag-of-words approach to represent the expression of each gene across the entire brain. In addition, they used a transfer learning approach by training a deep convolutional neural network on natural images to extract useful features from the ISH images. Their results show a superior performance for the deep convolutional neural network, indicating the applicability of transfer learning from natural to biological images (Zeng et al. 2015).

Ramsden et al. (2015) studied the molecular components underlying the neural circuits encoding spatial positioning and orientation in the medial entorhinal cortex (MEC). They developed a computational pipeline for automated registration and analysis of ISH images of the Allen Mouse Brain Atlas at laminar resolution. They showed that while very few genes are uniquely expressed in the MEC, differential gene expression defines its borders with neighboring brain structures, and its laminar and dorso-ventral organization. Their analysis identifies ion channel-, cell adhesion- and synapse-related genes as candidates for functional differentiation of MEC layers and for encoding of spatial information at different scales along the dorso-ventral axis of the MEC. Finally, they reveal laminar organization of genes related to disease pathology and suggest that a high metabolic demand predisposes layer II to neurodegenerative pathology.
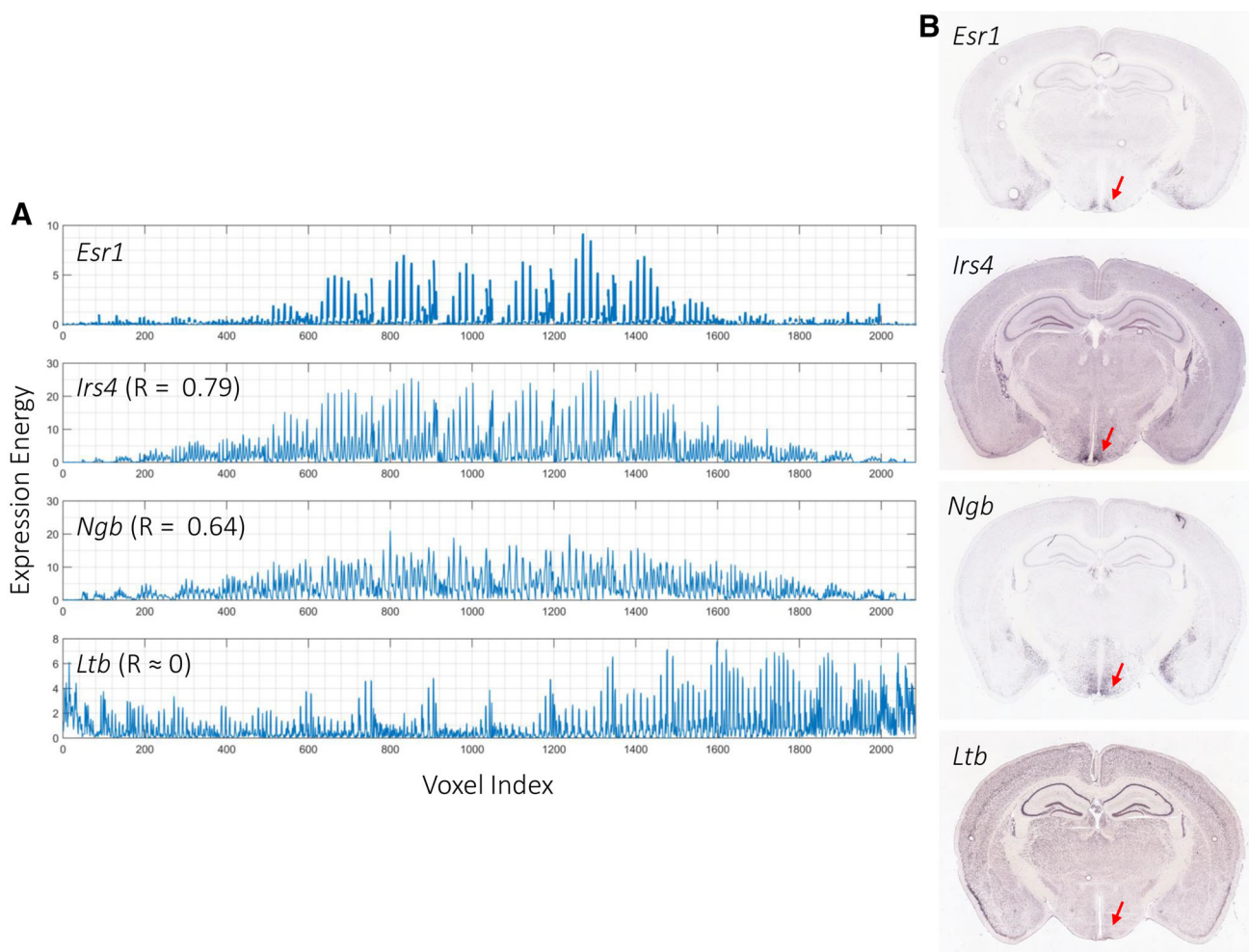
Spatial and temporal gene co-expression

Genes with similar expression patterns over a set of samples are said to be co-expressed and are more likely to be involved in the same biological processes (guilt by association) (Stuart et al. 2003). Applying the same approach to brain transcriptomes can identify co-expressed genes based on their spatial and/or temporal expression across the brain. This can serve as a powerful tool to characterize genes with respect to their context-specific functions. In addition, co-expression has been used to assess the quality of RNA-seq data, such as the BrainSpan atlas, by modeling the effects of noise within observed co-expression (Ballouz and Gillis 2016a).

The high dimensionality of transcriptomes, and other biological data (e.g. proteomes, epigenomes, etc.), provides a challenge for visualization as well as for selecting informative features for clustering and classification. Dimensionality-reduction approaches aim at finding a smaller number of features that can adequately represent the original high dimensional data in a lower dimensional space. The conventional principal component analysis (PCA) is the most commonly used dimensionality reduction method. Despite its utility, PCA can only capture linear rather than non-linear relationships, which are inherent in many biological applications. Several non-linear dimensionality reduction techniques have been proposed (e.g. Isomap (Tenenbaum et al. 2000)), see (Lee and Verleysen 2005) for an extensive review. The t-distributed stochastic neighbor embedding (t- SNE) method (Maaten and Hinton 2008) has been widely used to visualize biological data in two dimensions by preserving both the global and local relationships between the data points in the high-dimensional space (Saadatpour et al. 2015).

Several similarity/distance measurements have been used to characterize the similarity in spatial/temporal expression patterns between a pair of genes. Of these, correlation-based measures are mostly used to assess gene co-expression patterns across the brain. NeuroBlast is a search tool developed by the Allen Institute for Brain Sciences to identify genes with a similar 3D spatial expression to that of a gene of interest in a given anatomical region, based on Pearson correlation (Hawrylycz et al. 2011). Figure 3a shows an example of the obtained correlations of estrogen receptor alpha (*Esr1*) in the mouse hypothalamus. The ISH sections in Fig. 3b show that correlation can effectively be used to identify genes' functional association with *Esr1*. For example, the top correlated gene to *Esr1* in the hypothalamus is insulin receptor substrate 4 (*Irs4*), a target gene of *Esr1* associated with sex-specific behavior (Xu et al. 2012). NeuroBlast was

Fig. 3 Spatial gene co-expression in the mouse brain. **a** Expression energy profiles of voxels in the hypothalamus region of the mouse brain using the same linear ordering. The estrogen receptor alpha (*Esr1*) gene shows high expression in the hypothalamus. The expression patterns of *Irs4* and *Ngb* are highly correlated with that of *Esr1* ($R = 0.79$ and $R = 0.64$, respectively). On the other hand, the expression pattern of *Ltb* is not correlated with that of *Esr1* ($R = 8.01 \times 10^{-4}$). Correlation is calculated using Pearson correlation. **b** *Esr1* and its highly correlated genes (*Irs4* and *Ngb*) are highly expressed in the hypothalamus (*red arrow*), while *Ltb* is not
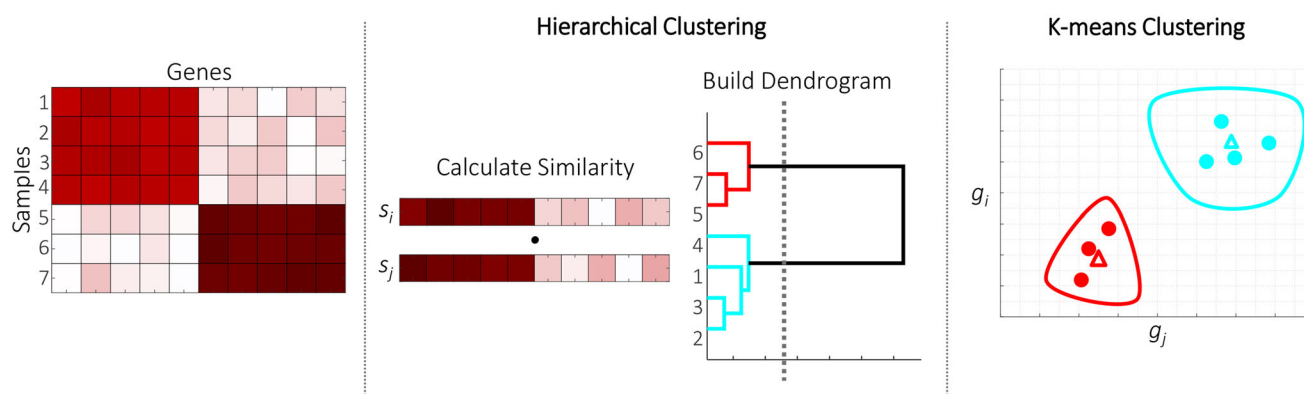
used to identify genes with a similar expression profile to *Wnt3a*, a ligand in the Wnt signaling pathway, in the developing mouse brain and identified eight Wnt signaling genes among the top correlated genes (Thompson et al. 2014). Using Spearman correlation coefficient, French et al. analyzed gene-pairs with positive and negative co-expression in the mouse brain. By focusing on genes with a

associated genes. Their analysis showed that *CHD8* was expressed throughout cortical and sub-cortical structures at the early prenatal ages and that expression decreased through development. In addition, they showed a significant enrichment of autism-candidate genes among genes with correlated temporal patterns to *CHD8* in the BrainSpan atlas.

Box 3 | Clustering

Clustering is the unsupervised learning process of identifying distinct groups of objects (clusters) in a dataset (Duda et al. 2000). There are two main types of clustering: hierarchical and partitional. Hierarchical clustering algorithms start by calculating all the pair-wise similarities between samples and then building a dendrogram by iteratively grouping the most similar sample pairs. By cutting the tree at an appropriate height, the samples are grouped into clusters. On the other hand, partitional clustering optimizes the number of simple models to fit the data. Examples of partitional clustering include k-means, Gaussian mixture models (GMMs), density-based clustering, and graph-based methods.

In order to cluster the samples hierarchically, all the pair-wise similarities between sample S_i and S_j are calculated. Samples are then grouped iteratively based on the calculated similarities (grouping the most similar first). Once the full dendrogram is built, a cut-off (dashed line) is used to group samples into groups. For k-means we set the number of clusters based on the data heatmap. K-means groups samples by minimizing the within-cluster sum of square distances between each point in the cluster and the cluster center.

---

strong negative correlation, they showed that variation in gene expression in the adult normal mouse brain can be explained as reflecting regional variation in glia to neuron ratios, and is correlated with degree of connectivity and location in the brain along the anterior–posterior axis (French et al. 2011). Tan et al. (2013) extended the analysis to the adult human brain and identified conserved co-expression patterns between the mouse and the human brain. To characterize the role of *SNCA*, a gene harboring a causative mutation for Parkinson's disease, Liscovitch and French (2014) analyzed the co-expression relationships of *SNCA* in the adult and developing human brain. They identified a negative spatial co-expression between *SNCA* and interferon-gamma signaling genes in the normal brain and a positive co-expression in post-mortem samples from Parkinson's patients, suggesting an immune-modulatory role of *SNCA* that may provide insight into neurodegeneration. Another example is given by Bernier et al. (2014), in which the developing human, macaque, and mouse brain atlases were used to analyze the expression and co-expression patterns of *CHD8*, one of the key autism-

Gene co-expression can serve as a very powerful tool for in silico prediction and prioritization of disease genes, by identifying genes with similar expression pattern to known disease genes. Piro et al. (2010) described a candidate gene prioritization method using the Allen Mouse Brain Atlas. They showed that the spatial gene-expression patterns can be successfully exploited for the prediction of gene–phenotype associations by applying their method to the case of X-linked mental retardation. By extending their methods to the human brain atlas, they showed that spatially mapped gene expression data from the human brain can be employed to predict candidate genes for Febrile seizures (FEB) and genetic epilepsy with febrile seizures plus (GEFS$^+$) (Piro et al. 2011). Both examples illustrate the power of using computational approaches to prioritize disease genes before carrying out empirical analysis in the lab.

In measuring gene co-expression, correlation-based methods are not specific to spatially mapped expression data and do not fully model the complexity of the brain transcriptomes. To identify gene-pairs with similar
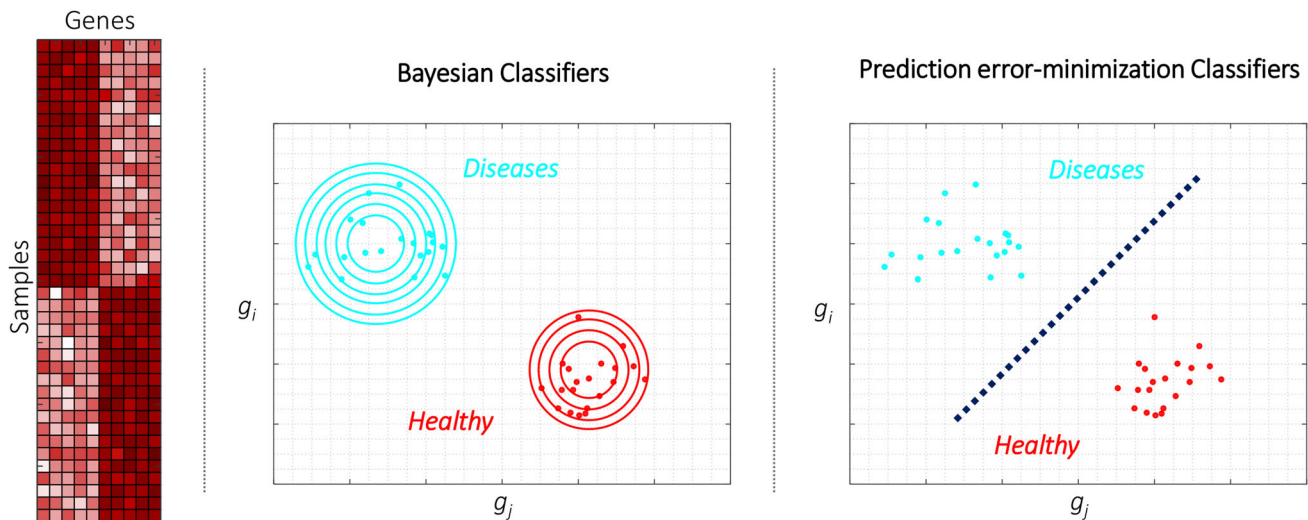
expression patterns in the adult mouse brain based on the ISH images, Liu et al. (2007) compared three image similarity metrics: a naïve pixel-wise metric, an adjusted pixel-wise metric, and a histogram- row-column (HRC) metric. They showed that HRC performs better than voxel-based methods, indicating the superiority of methods that capture the local structure in spatially mapped data. Miazaki and Costa (2012) used Voronoi diagrams to measure the similarity of the density distribution between gene expressions in the adult mouse brain. Inspired by computer vision algorithms, Liscovitch et al. (2013) used the similarity of scale-invariant feature transform (SIFT) descriptors of the ISH images of the mouse brain to predict the gene ontology (GO) labels of genes.

pathways, share common regulatory elements, or are involved in the same biological process. Co-expression networks avoid the problem of relying on prior knowledge, such as protein–protein interactions and pathway information, which are valuable but incomplete. Gene co-expression networks have heavily been used to identify disrupted molecular mechanisms in cancer (Chuang et al. 2007; Yang et al. 2014) and aging (van den Akker et al. 2014).

Hierarchical clustering is a widely used unsupervised approach to identify groups of co-expressed genes across a set of samples. Using hierarchical clustering, Gofflot et al. (2007) identified the functional networks of nuclear receptors based on their global expression across different

---

Box 4 | Classification

Classification is a supervised learning process of labeling unseen objects (test set) given a set of labeled objects (training set) (Duda et al. 2000). Classification approaches can be divided into Bayesian methods and prediction error minimization methods. The former group is based on Bayesian decision theory and uses statistical inference to find the best class for a given object. Bayesian methods can be further divided into parametric classifiers (e.g nearest-mean classifier and Hidden Markov Model) and non-parametric classifiers (e.g. Parzen window or k-nearest neighbor classifier). Alternatively, classifiers can be designed to minimize a measure of the prediction error. Well-known classifiers in this category include regression classifiers (e.g. Lasso regression), support vector machines, decision trees and artificial neural networks. Neural networks (in particular Deep Learning), have become very successful in solving problems in a wide range of applications, including bioinformatics (Xiong et al. 2014; Alipanahi et al. 2015; Engelhardt and Brown 2015).

A low dimensional embedding of the samples is generated using two features (genes). A Baysian Classifier assigns each sample to one of the two classes (Diseases or Healthy) based on statistical inference. A prediction error-minimization classifier updates the classification boundary (dashed line) based on the prediction error and terminates when a certain criterion is met.

Gene co-expression networks

As we have shown, the guilt by association paradigm has been successfully employed to identify pairs of spatially co-expressed genes sharing a neuronal function, based on various similarity measures. To extend the co-expression analysis of gene-pairs, clustering and network-based approaches can be used to identify molecular interaction networks of a group of genes that signal through similar

regions of the mouse brain. By focusing on subsets of brain structures involved in specialized behavioral functions, such as feeding and memory, they elucidated links between nuclear receptors and these specialized brain functions that were initially undetected in a global analysis. Dahlin et al. (2009) used hierarchical clustering to explore potential functional relatedness of the solute carrier genes and anatomic association with brain microstructures.

Another approach to unsupervised clustering is to use gene co-expression relationships to construct a co-expression network where nodes are genes and edges represent the similarity of the expression profile of those genes. Weighted gene co-expression network analysis (WGCNA) (Zhang and Horvath 2005) is a commonly used method to construct modules of co-regulated genes based on the topological overlap between genes in a weighted co-expression network. WGCNA has widely been used to identify transcription networks in the mammalian brain. Oldham et al. (2006) demonstrated the first utility of WGCNA to examine the conservation of co-expression networks between the human and chimpanzee brains. They found that module conservation in cerebral cortex is significantly weaker than module conservation in sub-cortical brain regions, which is in line with evolutionary hierarchies. WGCNA has been applied to identify modules of co-regulated genes in the developing and adult human brain transcriptomes (Kang et al. 2011; Hawrylycz et al. 2012), the developing rhesus monkey brain (Miller et al. 2013), the developing mouse brain (Thompson et al. 2014), and the prenatal human cortex (Miller et al. 2014a), see Fig. 3b. The methods provide a valuable insight into the molecular organization of the brain by identifying modules reflecting primary neural cell types and molecular functions. For example, modules constructed based on the prenatal human cortex correspond to cortical layers and age, while no areal patterning was observed (Miller et al. 2014a). In addition, WGCNA was used to identify a set of 32 functionally and anatomically distinct modules of genes with highly reproducible gene expression patterns across six human brains (Hawrylycz et al. 2015). There are numerous technical considerations to considere while constructing co-expression networks that go beyond the scope of this review (Allen et al. 2012; Ballouz et al. 2015). To analyze regional specificity of co-expression networks in the adult human brain, Myers et al. (2015) analyzed the modularity of a given gene set in region-specific co-expression networks. The developed method was used to compare networks that are constructed using expression data from a large sample size, but coarse neuroanatomical data set (Gibbs et al. 2010) to region-specific networks derived from the Allen Human Brain Atlas.

Box 5 | Co-expression Measurements

Gene co-expression is widely used for functional annotation, pathway analysis, and the reconstruction of gene regulatory networks. Co-expression measurements assess the similarity between a pair of gene expression profiles by detecting bivariate associations between them. These co-expression measurements can be summarized in five categories (Kumari et al. 2012; Allen et al. 2012; Song et al. 2012; Wang et al. 2014):

Correlation

The most widely used co-expression measure is *Pearson correlation*, due to its straightforward conceptual interpretation and computational efficiency. However, Pearson correlation can only capture linear relationships between variables. Alternatively, Spearman correlation is a nonparametric measure of non-linear associations. Other correlation-based methods include Renyi correlation, Kendall rank correlation, and bi-weight mid-correlation.

Partial correlation

Partial correlation is used to measure direct relationships between a pair of variables, excluding indirect relationships. Based on Gaussian graphical models, partial correlations infer conditional dependency as the non-zero entries in the precision matrix (the inverse of the covariance matrix).

Mutual-Information

Mutual information-based methods measure general statistical dependence between two variables. Based on information theory, mutual information does not assume monotonic relationships and hence can capture non-linear dependencies.
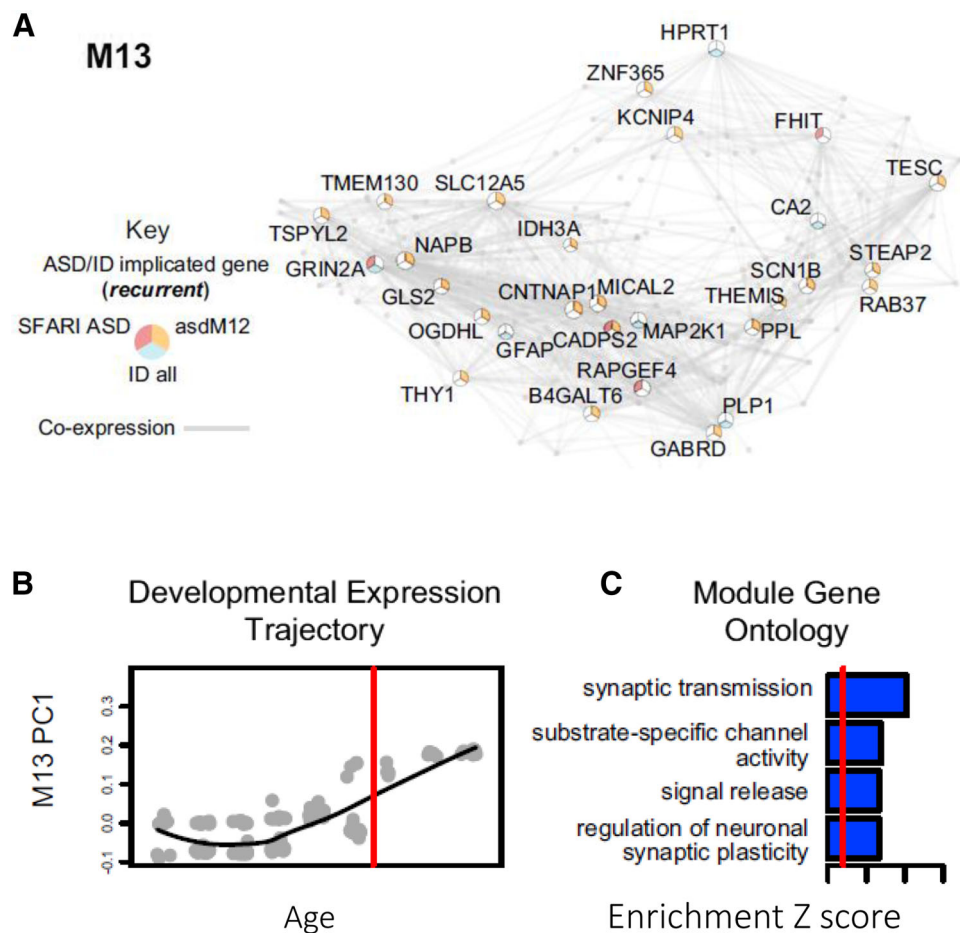
Other measures

Euclidian distance; Cosine similarity; Kullback-Leibler divergence; Hoeffding's D, distance covariance, and probabilistic measures (as used in Baysian networks).

Co-expression of disease-related genes

Complex neuropsychiatric and neurological disorders involve dysregulation of multiple genes, each conferring a small but incremental risk, which potentially converge in deregulated biological pathways or cellular functions. Using genome-wide association studies (GWAS), exome sequencing, and whole-genome sequencing (WGS), hundreds of variants have been linked to complex neurological disorders, such as autism (Iossifov et al. 2012; Neale et al. 2012; O'Roak et al. 2012; Sanders et al. 2012; Dong et al. 2014; De Rubeis et al. 2014), schizophrenia (Fromer et al. 2014; Ripke et al. 2014), Migraine (Freilinger et al. 2012), and Alzheimer's (Bettens et al. 2013; Zhang et al. 2013). With the increasing numbers of samples included in these studies, the number of variants associated to each disease is set to increase (Krumm et al. 2014). Gene co-expression networks provide a framework to identify the underlying molecular mechanisms on which these variants converge. Ben-David and Shifman (2012b) analyzed co-expression networks of genes affected by common and rare variants in autism using WGCNA. Menashe et al. (2013) used the cosine similarity of expression profiles to build a co-expression network of autism-related genes in the mouse brain. Both studies provide an important link between gene networks associated with autism and specific brain regions. However, for neurodevelopmental disorders, such as autism and schizophrenia, it is more beneficial to study when and where implicated genes are expressed during brain development. Gulsuner et al. (2013) studied the

Fig. 4 Gene co-expression networks. **a** Module M13 of co-expressed genes from Parikshak et al. (2013) (reprinted from Parikshak et al. Parikshak et al. 2013, Copyright (2016), with permission from Elsevier.). The shown module is significantly enriched in autism-related genes. The shown network comprises the top 200 connected genes (highest correlation) and their top 1000 connections in the subnetwork (also ordered on correlation). Genes are labeled if they are members of relevant gene sets. **b** Pattern of gene expression of genes in the shown module is summarized using the first principal component (eigengene). The *red line* indicates birth. **c** Gene Ontology terms enriched in the shown module. The *blue bars* indicate relative enrichment compared to all cortex-expressed genes in terms of Z score. The *red line* indicates Z = 2

transcriptional co-expression of genes harboring de novo mutations in schizophrenia patients using the BrainSpan atlas of the Developing Human Brain. Parikshak et al. (2013) used WGCNA to identify modules of co-expressed genes during human brain development using the BrainSpan atlas. They identified modules with significant enrichment in autism-related genes (Fig. 4). Willsey et al. (2013) used the BrainSpan atlas to generate co-expression networks around nine genes harboring recurrent de novo loss-of-function mutations in autism probands. Mahfouz et al. (2015b) used a combination of differential expression and genome-wide co-expression analysis to identify shared pathways among autism-related genes. To assess the functional convergence of distinct sets of genetic variants, Ballouz and Gillis (2016b) analyzed the connectivity of autism-candidate genes within a co-expression network constructed from the BrainSpan atlas. Their results show that gene sets with a higher proportion of burden genes exhibit higher interconnectivity, indicating stronger functional associations.

Using gene co-expression networks to study relationships between disease-related genes is a valuable approach

to understand disease mechanisms. In addition, using networks facilitates the integration of different types of interactions between genes, including but not limited to: co-expression, protein–protein interactions, and literature-based interactions. This can be very useful to our understanding of the etiologies of complex neurological diseases at different levels. In a recent study, Hormozdiari et al. (2015) integrated gene co-expression based on the BrainSpan atlas and PPI networks to identify networks of genes related to autism and intellectual disability. For a review on using gene networks to investigate the molecular mechanisms underlying neurological disorders, we refer to Gaiteri et al. (2014) and Parikshak et al. (2015).

Box 6 | Co-expression Networks

Gene co-expression networks provide a framework to uncover the molecular mechanisms underlying biological processes based on gene expression data. A co-expression network consists of nodes to represent genes and edges to encode the co-expression between two genes. A weighted network is a network in which the edges have continuous values to indicate the strength of co-expression. Networks with binary edges (an edge either exists or

not) are termed binary networks. Analysis of co-expression networks can be summarized in four main steps:

Network Construction

The first step in building a co-expression network is to construct a similarity matrix, by quantifying the similarity between the expression profiles of each pair of genes (i.e. co-expression). Several methods to measure gene co-expression are discussed in Box 5. For non-regularized estimations of co-expression, all off-diagonal elements of this similarity matrix will be nonzero. We can take these similarities as edge weights in the network, but that will give a fully connected network (each gene is connected to each gene). An additional step can be to threshold the similarity matrix, either to prune edges, or to binarize (absent/present) the similarities to obtain an adjacency matrix. In the latter case, pairs of genes with co-expression values above a threshold will be connected in a binary network. In the weighted gene co-expression network analysis (WGCNA) framework the similarity matrix undergoes a power transformation and a weight diffusion step, to optimize the topological properties and stability of the network (Zhang and Horvath 2005).

Network Characterization

The obtained networks can be analyzed in a number of ways. Topological measures characterize the structure of the network, and quantify the importance of genes in their network context. These measures have been extended to weighted networks (Zhang and Horvath 2005), and can capture topology on different levels of scale (Hulsman et al. 2014). Sets of networks can also be aligned and compared (Przulj 2007; Hayashida and Akutsu 2010; Fionda 2011). Network comparison can be used either to assess changes between different conditions, or to replicate a network in an independent dataset for validity assessment.

Module Identification

To interpret a network, it can be divided into sub-networks, or gene modules. To do this, the network edges are often treated as similarities in a clustering approach (see Box 3). Alternatively, graph properties, such as topological overlap or modularity, can be used to divide a network into modules (Blondel et al. 2008).

Module Characterization

Finally, modules can be characterized using a wide range of approaches. The expression profile of genes within the same module can be summarized using the average or the first principle component (also called eigengene (Oldham et al. 2006)). Alternatively, one can characterize a module according to its hub genes: the genes with the largest number of connections within the module. Another option is to assess the association of a module to external data by testing statistical enrichment in various gene sets (see Box 1 for different types of gene sets). In addition, modules can be characterized based on changes between conditions (e.g. health and disease) in their summary statistics (average expression profile), their topological measures (inter-connectivity), or the number of differentially-expressed genes they include.

Analyzing genetic signature of brain regions

Spatially mapped gene expression data allow for the exploration of neuroanatomy from a molecular point of view. Individual genes with spatially differential expression have long been used to define the structural organization of the brain and to break it down into regions and sub-regions. Genes have also been used to identify different classes of neuronal cell types. Studying the "genetic signature" of different brain regions can be useful for a multitude of applications. Spatially mapped gene expression data allow for the analysis of the similarity between brain regions in terms of their expression profiles. Regions sharing an expression profile are likely to be involved in the same neuronal functions or be part of the same neuronal circuit. Moreover, studying the expression profiles of functionally and anatomically connected structures provides valuable insights into the molecular basis of brain connectivity.

Spatial and temporal similarity of regional gene expression patterns

Each of the Allen Brain Atlases assigns a spatial location and a time point to each sample, allowing the exploration of the structural organization of the brain based on spatial and temporal similarities between different brain regions across the expression of thousands of genes. The Anatomic Gene Expression Atlas (AGEA) is a Web-based tool to calculate voxel-wise correlations based on gene expression in the adult and developing mouse brain atlases (Ng et al. 2009). To show the value of using the similarity of gene expression patterns to study anatomical organization, Dong et al. (2009) used AGEA to identify three distinct functional domains in the CA1 region of the mouse hippocampus. Hawrylycz et al. (2010) used AGEA to show that a consistent expression-based organization of areal patterning in the mouse cortex exists when clustered on a laminar basis. Using a combination of voxel–voxel similarities in gene expression (AGEA) and gene–gene similarities in expression patterns (NeuroBlast), Wagner et al. (2016) identified transcriptional markers of the mouse habenula as well as its subnuclear organization. In contrast to methods identifying regional markers by analyzing one gene at a time (Ramsden et al. 2015), using correlations between voxels (AGEA) and genes (NeuroBlast) simultaneously, such as (Dong et al. 2009; Wagner et al. 2016), reveals the transcriptomic–anatomic organization of brain areas.

Voxel correlation maps, such as those obtained by AGEA, can be used to cluster the mouse brain voxels into regions with similar gene expression profile. To analyze whether anatomically delineated regions, as defined classically, can also be distinguished based on their expression profile, Bohland et al. (2010) clustered the adult mouse brain voxels based on the similarity of their expression profiles. Using k-means clustering, they showed that their parcellations are quantitatively similar to the classically defined neuroanatomical atlas. These results show that the

spatially mapped gene expression data can be very valuable in identifying the molecular basis of brain organization. Similarly, Goel et al. (2014) used a combination of dimensionality reduction and spectral clustering to investigate the correspondence between spatial clusters of gene expression and human brain anatomy.

To identify which genes are responsible for brain organization, Ko et al. (2013) used a similar approach to cluster brain voxels based on their expression of gene markers of different cell types. Their results show that the neuroanatomical boundaries within a mouse brain can be defined by the clustering of only 170 neuron-specific genes. To identify the driving mechanism of spatial co-expression of genes in the brain, Grange et al. (2014) modeled co-expression patterns based on the spatial distribution of underlying cell types. Their model can be used to estimate cell-type specific maps of the mouse brain and to identify brain regions based on their genetic signatures. The model proposed in (Grange et al. 2014) was used to estimate the similarity between the expression profiles of two cliques of two cliques of co-expressed autism genes (Menashe et al. 2013) and the spatial distribution of cell types (Grange et al. 2015).

The temporal dynamics of gene expression patterns of brain regions, throughout brain development, have been considered in several studies. To understand gene expression specialization of mouse brain regions during development, Liscovitch and Chechik (2013) assessed the dissimilarities between brain regions based on gene expression and how these changeover time. Their results suggest an hourglass pattern, with high dissimilarity early in development that decreases to reach a minimum at birth after which it increases again. Using differential expression among regions of the human cortex at each development stage, Pletikos et al. (2014) also reported a highly similar temporal hourglass pattern of dissimilarity between brain regions. Another study by Mahfouz et al. (2014) analyzed the similarity between gene expression patterns of brain regions during human development. Using a network-based approach, they characterized the topology of the connectivity network of autism-related genes across development.

Gene expression and brain connectivity

Another way to study brain organization and function is to consider brain connectivity. Brain connectivity has been linked to many neurological disorders, such as ischemic stroke, autism, and schizophrenia. The relationship between gene expression and neuronal connectivity has long been studied in model organisms, such as *Caenorhabditis elegans*, to identify genes involved in synaptogenesis and axon guidance (Varadan et al. 2006; Kaufman et al. 2006; Baruch et al. 2008).

Zaldivar and Krichmar (2013) used the Allen mouse brain atlas to study the expression patterns of neurotransmitters in the brain. Since the expression of a transmitter must be coupled with the expression of appropriate receptors in the postsynaptic target, they have also analyzed the expression of receptors in target regions. This study shows that known neurobiological concepts can be seen back in the Allen brain atlas. To take it one step further, French and Pavlidis (2011) and Wolf et al. (2011) analyzed the relationship between gene expression similarity of brain regions and their connectivity. Both studies used the Allen mouse brain atlas to calculate the similarity in gene expression between different regions and the neural connectivity data of the rat brain from the Brain Architecture Management System (BAMS) (Bota and Swanson 2010). Genes involved in brain development and neurodevelopmental disorders, such as autism, showed strong correlations with anatomical connectivity patterns.

With the recent availability of the Allen mouse connectivity atlas, it has become possible to study the relationship between gene expression and brain connectivity within the same species. Rubinov et al. (2015) used a multivariate dimensionality reduction approach, partial least squares, to explore the association between gene expression and connectivity in the mouse brain. Rather than assessing the correlation between the gene expression similarity and connectivity, Ji et al. (2014) and Fakhry and Ji (2014) set out to predict connectivity based on gene expression patterns. By analyzing highly connected regions (i.e., hubs) in the mouse brain, Fulcher and Fornito (2016) showed that these hubs are more likely to interconnect with each other and are more likely to be transcriptionally similar. More interestingly, the genes with the highest contribution to the transcriptional similarity between hubs are involved in regulating the synthesis and metabolism of ATP, which is the primary energy source for neural activity.

Integrating gene expression and brain imaging data

The anatomical locations of samples in the Allen Human Brain Atlas have been indicated in the MRI scans of each of the six donor brains. These scans have been mapped to the Montreal Neurological Institute (MNI) standardized coordinate space, allowing for easy integration with other imaging studies. Rizzo et al. (2014) tested the predictive power of mRNA transcription maps extracted from the Allen Human Brain Atlas to predict in vivo protein distributions acquired using positron emission tomography (PET) imaging. By analyzing genes involved in two neurotransmission systems with different regulatory mechanisms, they showed that in vivo protein distributions can be predicted from mRNA transcription maps when expression

is being regulated translationally instead of posttranscriptionally. In another study, mRNA data from the Allen Human Brain Atlas were used to estimate the specific and non-displaceable components of PET radioligands for brain receptors, such as Serotonin 5-HT1A receptor; *HTR1A* (Veronese et al. 2016). Because many receptors are expressed across the whole brain, identifying a reference region that is devoid of the receptor requires pharmacological blockade. The method proposed by Veronese et al. estimates the specific and non-displaceable components of radioligand uptake based on the correlation between the abundance of the receptor gene transcript (using data from the Allen Human Brain Atlas) and the PET measurements of the expressed protein, without the need for blocking drugs.

Another promising research direction is the integration of data from the Allen Human Brain Atlas into fMRI studies to better understand the molecular mechanisms underlying functional connectivity in the human brain. One of the earliest efforts to link neuroimaging data and gene expression data in the human brain is presented by Goel et al. (2014). They explored whether structurally connected regions, those connected by white matter tracts determined by MR diffusion tensor imaging, have similar gene expression patterns as observed in rodents (French and Pavlidis 2011; Wolf et al. 2011). Despite finding no significant association between pair-wise connectivity and gene expression similarity, their results indicate that the overall connectivity of the brain is influenced by the underlying gene expression patterns. A large-scale analysis of the association between several cognitive phenomena and their underlying molecular mechanisms has been carried out in Fox et al. (2014). The study makes use of Neurosynth (Yarkoni et al. 2011), a framework to automatically synthesize brain-wide functional activation maps of cognitive tasks and psychological states based on published fMRI studies. By quantifying the spatial similarity between the expression patterns of all genes and several psychological topics, they demonstrated the ability to replicate known gene-cognition associations, such as between dopamine and reward. They further used their analysis to pinpoint previously unknown associations that can serve as a guide for researchers towards testable hypotheses about the genetic etiology of complex cognitive tasks. Cioli et al. (2014) used the Allen Human Brain Atlas to characterize the molecular differences between two sets of cortical functional networks. Using discriminant correspondence analysis, they predicted to which set of functional networks a cortical region belongs based on its gene expression profile. Richardi et al. (2015) showed that functionally connected regions, defined by a synchronized activity as measured by fMRI, are similar in their gene expression patterns compared with disconnected regions.

Furthermore, they identified a set of genes underlying the relationship between correlated gene expression and functional networks, and through GO analysis, they found that these genes are significantly enriched for ion channels. Similarly, Wang et al. (2015) used a region-specific measurement of brain activity based on fMRI to identify genes that correlate with brain activity in the default mode network that is brain regions with coherent fMRI signal fluctuations at the resting state. The correlated genes were enriched in neurons as well as genes down-regulated in autism. By analyzing the relationship between genes with consistent expression patterns across individuals and resting-state functional connectivity data from the Human Connectome Project, Hawrylycz et al. (2015) suggested that functional circuits are linked to conserved gene expression patterns across the cortex. Krienen et al. (2016) analyzed the association between corticocortical functional networks and the co-expression patterns of 19 genes uniquely enriched in the supragranular layers of the human cerebral cortex, in contrast to mice. The resulting strong association of major functional cortical classes (sensory/motor, paralimbic, or associational) supports the hypothesis that this unique molecular signatures of the human upper cortical layers underlie long-distance corticocortical connections, distinguishing humans from rodents. To extend this analysis, Vértes et al. used partial least squares (Rubinov et al. 2015) to identify the transcriptional signatures associated with topological parameters of fMRI networks indicating whether cortical regions are involved in long- or short-distance connections (Vértes et al. 2016). They showed that the transcriptional profiles of hub regions are, indeed, enriched in genes specific to supragranular layers as well as genes involved in oxidative metabolism and mitochondria, supporting the high cost associated with long-distance connections.

In contrast to the aforementioned studies on integrating functional activation maps of the human brain with gene expression patterns, fewer studies analyzed the link between structural changes in MRI scans and patterns of gene expression. Whitaker et al. (2016) used MRI to study maturation of human brain structures by quantifying changes in cortical thickness and myelination throughout adolescence. To understand the molecular mechanisms underlying changes in cortical thickness and myelination at different brain regions, they analyzed the relationship between these MRI markers and gene expression patterns from the Allen Human Brain Atlas. Using a multivariate dimensionality reduction technique (partial least squares), they identified associations between the expression patterns of all genes ($\sim$20,000) and four MRI-based variables. Peng et al. (2016) investigated whether the relationships among cortical regions can be explained from genetic factors using genotype data from twins and unrelated

individuals. In addition, they reported high concordance between inter-regional genetic correlations (based on genotype) and the inter-regional similarity of expression profiles using data from the Allen Human Brain Atlas, further confirming the genetic basis of cortical patterning. With the increasing interest in linking neuroimaging data to gene expression data, Rizzo et al. (2016) developed MENGA (Multimodal Environment for Neuroimaging and Genomic Analysis), which is a framework to integrate neuroimaging data from various modalities, such as PET and MRI, to gene expression patterns from the Allen Human Brain Atlas. MENGA was evaluated by analyzing the correlation between image data from different modalities focusing on the serotonin and the dopamine systems as well as myelin in brain tissue.

Romme et al. (2016) extended the study of associations between brain wiring and the underlying transcriptional signatures of connected regions to examine the role of genes in connectivity disruptions observed in schizophrenia patients. Using cross-correlation analysis of expression profiles of SCZ risk genes, identified using GWAS, and diffusion-weighted MRI, they found a strong association between the expression of the risk genes and regional macroscale dysconnectivity in schizophrenia patients. Valli et al. (2016) used the expression profiles of the glucocorticoid and mineralocorticoid receptors across the human brain to analyze the relationship between cortisol levels and gray-matter volume in individuals with ultra-high risk for psychosis. By assuming that the relationship between gray-matter volume and cortisol levels likely occurs in brain areas with high expression of cortisol receptor genes, they used an adaptive threshold to identify significant associations. These results further highlight the value of studying associations of alternations observed in brain images and the underlying transcriptional profile of the affected areas to uncover disease mechanisms as well as to identify new disease genes.

Studying brain organization using dimensionality reduction methods

An alternative approach to analyze the relationship between gene expression and neuroanatomy is dimensionality reduction (Box 2). Mapping high-dimensional data in two dimensions allows for the exploration of how gene expression patterns relate to brain organization. Ji (2013) used t-distributed stochastic neighborhood embedding (t-SNE) to map the Allen developing mouse brain atlas and showed that t-SNE clearly outperforms PCA. The results show that clustering voxels in the low-dimensional space is more consistent with neuroanatomy than in the original space. Mahfouz et al. (2015a) used a computationally efficient implementation of t-SNE, named Barnes-

Hut-SNE, to map the sagittal and coronal adult mouse atlas and the brain transcriptome of the six human donors (Fig. 5). They quantitatively showed that BH-SNE maps are superior in their separation of neuroanatomical regions in comparison to PCA and MDS. Similarly, dimensionality reduction approaches can be used to analyze the gene–gene relationships. A low-dimensional embedding of genes in which distances represent similarity of the spatial and/or temporal expression profile of genes across the brain can be very informative.
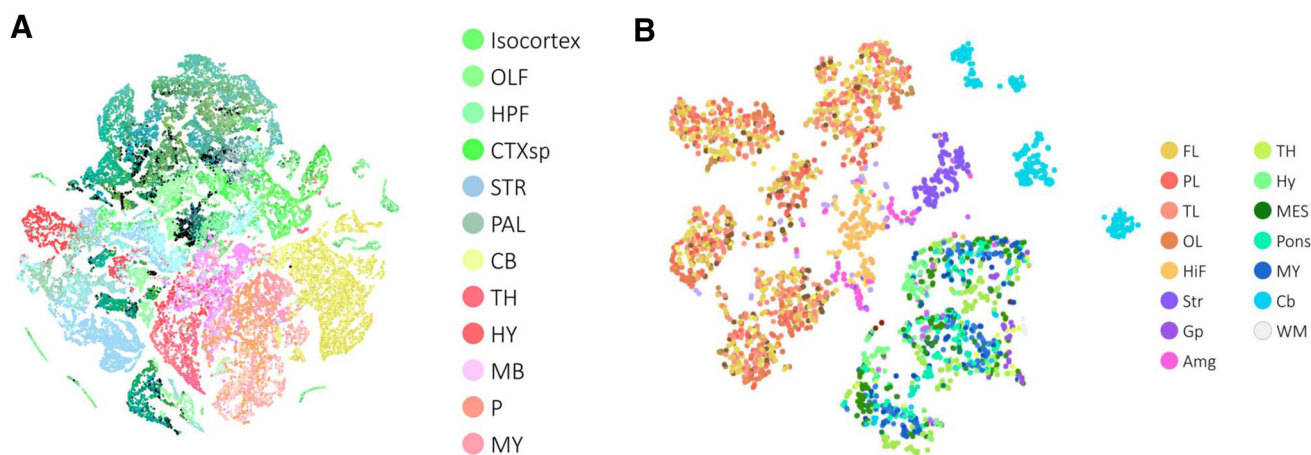
Perspective on the future of computational analysis of brain transcriptomes

Brain transcriptome atlases are no cell-type-specific

The identification of the molecular profile of the different cell types in the brain, their connectivity patterns, and their electrophysiological properties is crucial to our understanding of the functional organization of the brain. Despite the valuable information provided by the brain transcriptomes, these resources remain limited in their ability to quantify cell-type-specific expression of genes. New technologies targeting specific cell populations, such as viral, optogenetic and single-cell sequencing approaches, will allow us to better characterize cell types and their role in brain function. So far, these techniques are limited in their scalability and computational methods still provide a feasible alternative approach. Using spatial clustering of gene expression patterns of cell-type-specific genes in the adult mouse, Ko et al. (2013) showed that astrocytes and oligodendrocytes differ between brain regions, but these regional differences in expression are less pronounced than differences in neuronal composition. Similarly, Grange et al. (2014) proposed a model to estimate cell-type-specific maps of the mouse brain. Kuhn et al. (2011) developed a method to analyze brain samples of varying cellular composition. Their method detected myelin-related abnormalities in brain samples from Huntington's disease patients, which was not detected using standard differential expression. These examples illustrate the power of computational models in untangling the complex composition of the different cell types in the brain.

With the recent advances in single-cell mRNA sequencing, it has become feasible to measure the expression of thousands of genes and their variability between different cell types (Shapiro et al. 2013). In addition, single-cell sequencing has indicated that neurons from small cortical regions come from different clones with distinct somatic mutations (Lodato et al. 2015). Understanding how these different clones of neurons contribute to the aggregated gene expression from a specific

Fig. 5 Dimensionality reduction of brain transcriptomes. Samples from brain transcriptomes can be embedded in a low-dimensional space by means of dimensionality reduction methods. **a** 2D embedding of ∼60,000 voxels from the Allen Mouse Brain Atlas. **b** 2D embedding of ∼3700 samples from the six donors in the Allen Human Brain Atlas. Both embeddings were generated using Barnes-Hut t-SNE. In both maps, *colors* correspond to anatomical regions of the mouse and human brain. Data from Mahfouz et al. (2015a)

brain region will be of great interest to understand the role of mutations in neurological disorders. The vast amount of data generated by these projects illustrates the importance of computational methods that can identify distinct groups of cells with a common functional role (Pettit et al. 2014; Grün et al. 2015).

Limited resolution of brain transcriptomes

There are several limitations associated with the current spatial and temporal brain transcriptomes. Despite their unprecedented spatial and temporal resolution, human brain transcriptomes are still of low resolution with ∼1000 samples per brain. This relatively low resolution presents a fundamental limitation, especially when integration with imaging-based data (e.g., MRI or PET) is considered. The ISH-based mouse transcriptomes offer a much higher resolution. Although the original ISH data provide a near-cellular resolution (∼1 μm), the genome-wide data registered to the common 3D space offer a much lower resolution (∼200 μm). Several studies used re-registration of a limited set of the high-resolution ISH images from the Allen Mouse Brain atlas to acquire genome-wide data at a higher resolution. The aforementioned study by (Ko et al. 2013) found more transcriptionally distinct brain regions than a previous study (Bohland et al. 2010), mainly due to the usage of cell-type specific genes. However, Ko et al. have also realigned the ISH images of the mouse brain atlas and performed their analysis on a higher resolution grid (100 μm). Ramsden et al. (2015) used non-linear registration to realign the ISH data of the mouse. By analyzing genome-wide data at a resolution of 10 μm, they were able to identify genes whose expression pattern delineates the borders and layers of the medial entorhinal cortex.

There is still need for more generic approaches to map spatially mapped gene expression data (from ISH experiments) generated at different labs to the standard 3D space of the Allen Reference Atlas. Tools, such as BrainAligner (Peng et al. 2011), are available for analyzing *Drosophila melanogaster* neural expression patterns. The availability of similar tools for the mouse and human brain can enormously enhance our understanding of disease molecular mechanisms by allowing researchers to map their own data to the same space.

Current atlases focus only on protein-coding mRNA

Most of the atlases profiling the mammalian brain transcriptome and its relationship to brain development and function have mainly focused on profiling the expression of protein-coding mRNA. These atlases mostly provided limited or no information about other RNA species, such as non-coding RNA (ncRNA) and microRNA (miRNA), despite their recognized role in brain development and neurological disorders (Ponjavic et al. 2009; Qureshi and Mehler 2012). Using the Allen Mouse Brain Atlas, long ncRNAs showed regionally enriched expression patterns, such as those observed for protein-coding mRNAs (Mercer et al. 2008), further supporting their functional role in the brain. By profiling the developmental transcriptome of the neocortex using deep sequencing, Fertuzinhos et al. defined the dynamics of mRNA, miRNA, and ncRNA across the different layers of the mouse cortex (Fertuzinhos et al. 2014). The BrainSpan atlas provides the most comprehensive map of miRNA expression in the developing human brain. Ziats and Rennert (2013) used the BrainSpan miRNA data to define a pattern of increased inter-regional expression differences of miRNA through development,

potentially driving regional specialization. Moreover, targets of differentially expressed miRNAs were mostly related to transcriptional regulation and neurodevelopmental disorders, highlighting the importance of studying miRNAs as potential biomarkers. Additional measurement of ncRNAs and miRNAs as well as a detailed analysis of their role in gene regulatory networks can help our understanding of their relationship to genes related to neurodevelopmental disorders.

Integrating brain transcriptomes and other neuro-omics data sets

Advances in high-throughput molecular profiling have facilitated acquiring various omics data sets spanning a wide spectrum of cellular processes. For instance, the rapid developments in next-generation sequencing (NGS) technology enabled genome-wide measurement of genomic, transcriptomic, and epigenomic data of brain tissues. While transcriptomes provide detailed information on the abundance of RNA, epigenomic features, such as histone modifications, methylation, and chromatin interactions, describe the underlying mechanisms of distinct cell-specific transcriptomes. Moreover, most disease-related variants are in the non-coding regulatory regions of the genome, making epigenomic studies crucial to uncover a larger proportion of the genetic contribution to complex traits than can be explained by coding variants alone. Increasingly, studies are gathering data across different platforms from a wide range of tissues and cell types to uncover mechanisms underlying complex phenotypes and disease. The Encyclopedia of DNA Elements (ENCODE) (Bernstein et al. 2012a) and the Roadmap Epigenome project (Consortium 2015a) have profiled the epigenome of several tissues and cell types, while the Genotype Tissue Expression project (GTEx) (Lonsdale 2013) is generating genotype and gene expression data from 25 unique human tissues, including 13 brain regions. In addition, The Cancer Genome Atlas project (TCGA) (Weinstein et al. 2013) and the International Cancer Genome Consortium (ICGC) (Hudson et al. 2010) provide comprehensive genomic and transcriptomic and epigenomic data from multiple cancer types. However, most of these studies have profiled samples from cancer cell lines or normal cells from non-brain tissues due to limitations specific to the brain, such as the requirement of large amount of genomic material and the high heterogeneity of cell types within the same sample (Shin et al. 2014). Recently, the isolation of more homogeneous samples from the brain as well as developments in single-cell analysis is greatly advancing the field of neuroepigenomics (Maze et al. 2014; Shin et al. 2014). For example, efforts have been made to map the brain methylome (Illingworth et al. 2015) and to identify *cis*-

regulatory elements across brain regions (Vermunt et al. 2014). The PsychENCODE consortium (Akbarian et al. 2015) is an ongoing project to profile neurobiological epigenetic landscape of the healthy and diseased developing and adult human brains. For large-scale multi-omics data sets, systems genomics approaches, which integrate different genome-wide data types, can minimize false positive discoveries as well as unravel the molecular mechanisms underlying the phenotype or disease of interest. Several approaches have been developed to integrate multi-omics data (Consortium 2015b), clearly illustrating the added value of collecting multiple omics measurements from a large number of samples.

Integrating brain transcriptomes and imaging mass spectroscopy

Over the past few years, imaging mass spectrometry (IMS) (Caprioli et al. 1997) has emerged as a powerful technique to capture the spatial distribution of large biomolecules, such as proteins, peptides, and lipids in biological samples. Similar to ISH, imaging mass spectroscopy holds great potential in studying the chemical organization of complex samples from the brain (Hanrieder et al. 2013). Methods have been developed to align IMS-based sections of the mouse brain to histology-based sections from the Allen Mouse Brain Atlas to anatomically localize biomolecules within the brain (Abdelmoula et al. 2014; Carreira et al. 2015). However, recently, these methods have been extended to link protein expression to the expression of the encoding gene as well as its co-expressed genes based on the Allen Mouse Brain Atlas (Škrášková et al. 2015). There is a great potential for applications based on the integration of ISH-based gene expression and IMS-based protein expression measurements to help our understanding of translational mechanisms in the brain. Yet, more complex modeling of the two data types is needed. Methods developed to integrate spatially mapped gene and protein expression data can also be used to study spatial localization within the cell using data from the Human Protein Atlas (Uhlen et al. 2015).

Imaging genetics

In an attempt to better understand gene-disease associations, researchers are searching for genes that affect intermediate disease biomarkers. Brain imaging studies can be used to reveal genetic effects on brain structure, function, and circuitry, providing valuable mechanistic insights. Imaging genetics have emerged as a field concerned with finding associations between genetic variants (typically SNPs) and imaging-based measurements (Hibar et al. 2011b). Due to the millions of statistical tests that need to

be performed, stringent statistical thresholds are required to limit the false discovery rate (Medland et al. 2014). Recently, the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium (Hibar 2015) analyzed SNPs associations with the volume of sub-cortical structures in ∼30,000 individuals, providing the first large-scale analysis of the genetic causes of human brain variability. Several methods have been developed to limit the number of statistical tests performed in genome-wide, brain-wide analysis by either exploiting the dependency between brain voxels and/or testing for associations with genes or pathways instead of individual variants (Hibar et al. 2011a). In addition, efforts have been made to jointly model imaging and genetic observations from Alzheimer's Disease Neuroimaging Initiative (ADNI) data (adni.loni.ucla.edu), using multivariate statistical methods (Wang et al. 2012; Batmanghelich et al. 2013). These methods remain computational very expensive, limiting the number of variables analyzed. Brain transcriptomes can play an important role in imaging genetics by providing region-specific information about gene expression that can be used to prioritize genes and variants for testing. For example, incorporating spatial gene co-expression of amyloid-related candidate genes from the Allen Human Brain Atlas as prior knowledge to their statistical model significantly improved the prediction of associations between SNPs in the APOE gene and amyloid deposition measures among cortical regions (Yan et al. 2014). There is need for more advanced methods to link genomic measurements which are usually collected from blood samples to intermediate disease phenotypes observed in brain images.

Unexplored computational avenues

The multiple dimensions of the brain transcriptomes (genes, regions, and time) provide a framework to explore spatio-temporal regulation of gene expression during development. Clustering the data along one dimension only yields global patterns of similarity, while in a complex system, such as the brain, it is always more useful to identify more localized patterns of correlation. For example, the effect of steroid hormones on the brain is highly region-specific, depending on the availability of target genes and co-regulators affecting the steroid receptors at the site of action. Analyzing the region-specific co-expression relationships of steroid receptors and their coactivators can be used to predict steroid responsiveness and selective activation of particular circuits with synthetic ligands (Zalachoras et al. 2013).

Biclustering is a type of technique to simultaneously identify a subset of genes associated with a subset of conditions (this can be brain regions and/or time-points), allowing for the identification of local spatial or temporal patterns of co-expression. Biclustering has been particularly effective in analyzing time-series expression data (Goncalves and Madeira 2014). Similarly, applying biclustering to expression data from the Allen Mouse Brain Atlas resulted in more GO-enriched clusters than those obtained by independently clustering genes or regions (Jagalur et al. 2007). Ji et al. (2013) described a co-clustering method based on graph approximation to explore the spatio-temporal regulation of gene expression during the mouse brain development. Yet, they apply biclustering to each developmental stage independently and do not consider the time-varying nature of the developing mouse brain data, due to the lack of correspondence between the voxels across different stages. To fully exploit the multi-dimensionality of the developing brain transcriptomes, triclustering methods provide an interesting approach to identify groups of genes that show spatial and temporal co-expression (Tchagang et al. 2012). Recently, Jung et al. (2015) used three-component analysis to identify genes associated with aging by analyzing longitudinal gene expression, methylation, and histone modification data of human skin fibroblasts. Their three-component analysis represents an integrative approach to jointly model temporal changes in different data types. An extension of their methods to incorporate spatial information available in brain transcriptomes can lead to a complete approach of modeling spatial and temporal changes of different omics data from the brain.

Graphical models (e.g., conditional random fields) are commonly used for data segmentation using local features, especially in computer vision application. The Roadmap Epigenome project has used a Hidden Markov Model to classify the human genome into chromatin states based on epigenetic markers (Consortium 2015a). These models can be used to model the spatial and/or temporal relationships between genes in brain transcriptome atlases.

A greater challenge lies in identifying causal relationships rather than associations in gene–gene interactions and the brain is no exception. Systems biology approaches provide an interesting avenue to explore causal relationships between genes by means of quantitative modeling. The resulting mathematical models enable formal analysis and simulation of complex biological processes (Kolch et al. 2015). However, inferring causal relationships between the different variables requires a vast amount data, limiting their usability to a small number of genes (Lausted et al. 2014). Hwang et al. (2009) presented a system approach to analyze genes differentially expressed in the mouse brain across time in Prion disease. An extension of such a model to include spatial information on gene expression can help refine the model as well as associate disease-related changes to specific brain areas.

References

Abdelmoula WM, Carreira RJ, Shyti R et al (2014) Automatic registration of imaging mass spectrometry data to the Allen Brain Atlas transcriptome. Anal Chem 9034:90343M. doi:10.1117/12.2043653

Akbarian S, Liu C, Knowles JA et al (2015) The PsychENCODE project. Nat Neurosci 18:1707–1712. doi:10.1038/nn.4156

Alavian KN, Simon HH (2009) Linkage of cDNA expression profiles of mesencephalic dopaminergic neurons to a genome-wide in situ hybridization database. Mol Neurodegener 4:6. doi:10.1186/1750-1326-4-6

Alipanahi B, Delong A, Weirauch MT, Frey BJ (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 33:1–9. doi:10.1038/nbt.3300

Allen JDJ, Xie Y, Chen M et al (2012) Comparing statistical methods for constructing large scale gene networks. PLoS One 7:e29348. doi:10.1371/journal.pone.0029348

Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29. doi:10.1038/75556

Bakken TE, Miller JA, Ding S-L et al (2016) Comprehensive transcriptional map of primate brain development. Nature. doi:10.1038/nature18637

Ballouz S, Gillis J (2016a) AuPairWise: a method to estimate RNA-seq replicability through co-expression. PLoS Comput Biol 12:e1004868. doi:10.1371/journal.pcbi.1004868

Ballouz S, Gillis J (2016b) Assessment of functional convergence across study designs in autism. bioRxiv, 1–38. doi:10.1111/jdi.12545

Ballouz S, Verleyen W, Gillis J (2015) Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. Bioinformatics 31:2123–2130. doi:10.1093/bioinformatics/btv118

Baruch L, Itzkovitz S, Golan-Mashiach M et al (2008) Using expression profiles of *Caenorhabditis elegans* neurons to identify genes that mediate synaptic connectivity. PLoS Comput Biol 4:e1000120. doi:10.1371/journal.pcbi.1000120

Batmanghelich NK, Dalca AV, Sabuncu MR, Golland P (2013) Joint modeling of imaging and genetics. Inf Process Med Imaging 7917:766–777. doi:10.1007/978-3-642-38868-2_64

Ben-David E, Shifman S (2012a) Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. Mol Psychiatry 18:1054–1056. doi:10.1038/mp.2012.148

Ben-David E, Shifman S (2012b) Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. PLoS Genet 8:e1002556. doi:10.1371/journal.pgen.1002556

Bernard A, Lubbers LS, Tanis KQ et al (2012) Transcriptional architecture of the primate neocortex. Neuron 73:1083–1099. doi:10.1016/j.neuron.2012.03.002

Bernier R, Golzio C, Xiong B et al (2014) Disruptive CHD8 mutations define a subtype of autism early in development. Cell. doi:10.1016/j.cell.2014.06.017

Bernstein BE, Birney E, Dunham I et al (2012a) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. doi:10.1038/nature11247

Bernstein BE, Birney E, Dunham I et al (2012b) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. doi:10.1038/nature11247

Bettens K, Sleegers K, Van Broeckhoven C (2013) Genetic insights in Alzheimer's disease. Lancet Neurol 12:92–104. doi:10.1016/S1474-4422(12)70259-4

Björklund A, Dunnett SB (2007) Dopamine neuron systems in the brain: an update. Trends Neurosci 30:194–202. doi:10.1016/j.tins.2007.03.006

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 10008:6. doi:10.1088/1742-5468/2008/10/P10008

Bohland JW, Bokil H, Pathak SD et al (2010) Clustering of spatial gene expression patterns in the mouse brain and comparison with classical neuroanatomy. Methods 50:105–112. doi:10.1016/j.ymeth.2009.09.001

Bota M, Swanson LW (2010) Collating and curating neuroanatomical nomenclatures: principles and use of the brain architecture knowledge management system (BAMS). Front Neuroinform 4:3. doi:10.3389/fninf.2010.00003

Cahoy JD, Emery B, Kaushal A et al (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci 28:264–278. doi:10.1523/JNEUROSCI.4178-07.2008

Caprioli RM, Farmer TB, Gile J (1997) Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. Anal Chem 69:4751–4760. doi:10.1021/Ac970888i

Carreira RJ, Shyti R, Balluff B et al (2015) Large-scale mass spectrometry imaging investigation of consequences of cortical spreading depression in a transgenic mouse model of migraine. J Am Soc Mass Spectrom. doi:10.1007/s13361-015-1136-8

Chuang H-Y, Lee E, Liu Y-T et al (2007) Network-based classification of breast cancer metastasis. Mol Syst Biol 3:1–10. doi:10.1038/msb4100180

Cioli C, Abdi H, Beaton D et al (2014) Differences in human cortical gene expression match the temporal properties of large-scale functional networks. PLoS One 9:1–28. doi:10.1371/journal.pone.0115913

Colantuoni C, Lipska BK, Ye T et al (2011) Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature 478:519–523. doi:10.1038/nature10524

Consortium RE (2015a) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330. doi:10.1038/nature14248

Consortium TUP (2015b) The UK10K project identifies rare variants in health and disease. Nature 526:82–90. doi:10.1038/nature14962

Croft D, Mundo AF, Haw R et al (2014) The Reactome pathway knowledgebase. Nucleic Acids Res 42:D472–D477. doi:10.1093/nar/gkt1102

Dahlin A, Royall J, Hohmann JG, Wang J (2009) Expression profiling of the solute carrier gene family in the mouse brain. J Pharmacol Exp Ther 329:558–570. doi:10.1124/jpet.108.149831

Darmanis S, Sloan SA, Zhang Y et al (2015) A survey of human brain transcriptome diversity at the single cell level. Proc Natl Acad Sci 112:201507125. doi:10.1073/pnas.1507125112

Davis FP, Eddy SR (2009) A tool for identification of genes expressed in patterns of interest using the Allen Brain Atlas. Bioinformatics 25:1647–1654. doi:10.1093/bioinformatics/btp288

De Rubeis S, He X, Goldberg AP et al (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. doi:10.1038/nature13772

Diez-Roux G, Banfi S, Sultan M et al (2011) A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol 9:e1000582. doi:10.1371/journal.pbio.1000582

Dong H-W, Swanson LW, Chen L et al (2009) Genomic-anatomic evidence for distinct functional domains in hippocampal field CA1. Proc Natl Acad Sci 106:11794–11799. doi:10.1073/pnas.0812608106

Dong S, Walker MF, Carriero NJ et al (2014) De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. Cell Rep 9:16–23. doi:10.1016/j.celrep.2014.08.068

Duda RO, Hart PE, Stork DG (2000) Pattern classification. Wiley-Interscience, New Jersey

Engelhardt BE, Brown CD (2015) Diving deeper to predict noncoding sequence function. Nat Methods 12:925–926. doi:10.1038/nmeth.3604

Exome Variant Server (2015b). In: NHLBI GO Exome Seq. Proj. (ESP), Seattle. http://evs.gs.washington.edu/EVS/

Fakhry A, Ji S (2014) High-resolution prediction of mouse brain connectivity using gene expression patterns. Methods 73C:71–78. doi:10.1016/j.ymeth.2014.07.011

Fertuzinhos S, Li M, Kawasawa YI et al (2014) Laminar and temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex. Cell Rep 6:938–950. doi:10.1016/j.celrep.2014.01.036

Fionda V (2011) Biological network analysis and comparison: mining new biological knowledge. Cent Eur J Comput Sci 1:185–193. doi:10.2478/s13537-011-0013-1

Fox AS, Chang LJ, Gorgolewski KJ, Yarkoni T (2014) Bridging psychology and genetics using large-scale spatial analysis of neuroimaging and neurogenetic data. bioRxiv. doi:10.1101/012310

Freilinger T, Anttila V, de Vries B et al (2012) Genome-wide association analysis identifies susceptibility loci for migraine without aura. Nat Genet 44:777–782. doi:10.1038/ng.2307

French L (2015) A FreeSurfer view of the cortical transcriptome generated from the Allen Human Brain Atlas. Front Neurosci 9:1–5. doi:10.3389/fnins.2015.00323

French L, Pavlidis P (2007) Informatics in neuroscience. Brief Bioinform 8:446–456. doi:10.1093/bib/bbm047

French L, Pavlidis P (2011) Relationships between gene expression and brain wiring in the adult rodent brain. PLoS Comput Biol 7:e1001049. doi:10.1371/journal.pcbi.1001049

French L, Tan PPC, Pavlidis P (2011) Large-scale analysis of gene expression and connectivity in the rodent brain: insights through data integration. Front Neuroinform 5:12. doi:10.3389/fninf.2011.00012

Fromer M, Pocklington AJ, Kavanagh DH et al (2014) De novo mutations in schizophrenia implicate synaptic networks. Nature 506:179–184. doi:10.1038/nature12929

Fulcher BD, Fornito A (2016) A transcriptional signature of hub connectivity in the mouse connectome. Proc Natl Acad Sci USA. doi:10.1073/pnas.1513302113

Gaiteri C, Ding Y, French B et al (2014) Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. Genes Brain Behav 13:13–24. doi:10.1111/gbb.12106

Gibbs JR, van der Brug MP, Hernandez DG et al (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in Human Brain. PLoS Genet 6:29. doi:10.1371/journal.pgen.1000952

Goel P, Kuceyeski A, Locastro E, Raj A (2014) Spatial patterns of genome-wide expression profiles reflect anatomic and fiber connectivity architecture of healthy human brain. Hum Brain Mapp 35:4204–4218. doi:10.1002/hbm.22471

Gofflot F, Chartoire N, Vasseur L et al (2007) Systematic gene expression mapping clusters nuclear receptors according to their function in the brain. Cell 131:405–418. doi:10.1016/j.cell.2007.09.012

Goncalves J, Madeira S (2014) LateBiclustering: efficient heuristic algorithm for time-lagged bicluster identification. IEEE/ACM Trans Comput Biol Bioinform. doi:10.1109/TCBB.2014.2312007

Gong S, Zheng C, Doughty ML et al (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 425:917–925. doi:10.1038/nature02033

Grange P, Bohland JW, Okaty BW et al (2014) Cell-type-based model explaining coexpression patterns of genes in the brain. Proc Natl Acad Sci 111:5397–5402. doi:10.1073/pnas.1312098111

Grange P, Menashe I, Hawrylycz MJ (2015) Cell-type-specific neuroanatomy of cliques of autism-related genes in the cell-type-specific neuroanatomy of cliques of autism-related genes in the mouse brain. Front Comput Neurosci. doi:10.3389/fncom.2015.00055

Grün D, Lyubimova A, Kester L et al (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature. doi:10.1038/nature14966

Gulsuner S, Walsh T, Watts AC et al (2013) Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell 154:518–529. doi:10.1016/j.cell.2013.06.049

Hanrieder J, Phan NT, Kurczy ME, Ewing AG (2013) Imaging mass spectrometry in neuroscience. ACS Chem Neurosci 4:666–679. doi:10.1021/cn400053c

Hawrylycz MJ, Bernard A, Lau C et al (2010) Areal and laminar differentiation in the mouse neocortex using large scale gene expression data. Methods 50:113–121. doi:10.1016/j.ymeth.2009.09.005

Hawrylycz MJ, Ng L, Page D et al (2011) Multi-scale correlation structure of gene expression in the brain. Neural networks 24:933–942. doi:10.1016/j.neunet.2011.06.012

Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL et al (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489:391–399. doi:10.1038/nature11405

Hawrylycz MJ, Miller JA, Menon V et al (2015) Canonical genetic signatures of the adult human brain. Nat Neurosci. doi:10.1038/nn.4171

Hayashida M, Akutsu T (2010) Comparing biological networks via graph compression. BMC Syst Biol 4(Suppl 2):S13. doi:10.1186/1752-0509-4-S2-S13

Heintz N (2004) Gene expression nervous system atlas (GENSAT). Nat Neurosci 7:483. doi:10.1038/nn0504-483

Hibar DP (2015) Common genetic variants influence human subcortical brain structures. Nature. doi:10.1038/nature14101

Hibar DP, Kohannim O, Stein JL et al (2011a) Multilocus genetic analysis of brain images. Front Genet 2:73. doi:10.3389/fgene.2011.00073

Hibar DP, Stein JL, Kohannim O et al (2011b) Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. Neuroimage 56:1875–1891. doi:10.1016/j.neuroimage.2011.03.077

Hormozdiari F, Penn O, Borenstein E, Eichler EE (2015) The discovery of integrated gene networks for autism and related disorders. Genome Res. doi:10.1101/gr.178855.114.142

Hudson TJ, Anderson W, Aretz A, Barker AD (2010) International network of cancer genome projects. Nature 464:993–998. doi:10.1038/nature08987

Hulsman M, Dimitrakopoulos C, De Ridder J (2014) Scale-space measures for graph topology link protein network architecture to

function. Bioinformatics 30:237–245. doi:10.1093/bioinformatics/btu283

Hwang D, Lee I, Yoo H et al (2009) A systems approach to prion disease. Mol Syst Biol 5:252. doi:10.1038/msb.2009.10

Illingworth RS, Gruenewald-Schneider U, De Sousa D et al (2015) Inter-individual variability contrasts with regional homogeneity in the human brain DNA methylome. Nucleic Acids Res 43:732–744. doi:10.1093/nar/gku1305

Iossifov I, Ronemus M, Levy D et al (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74:285–299. doi:10.1016/j.neuron.2012.04.009

Jagalur M, Pal C, Learned-Miller E et al (2007) Analyzing in situ gene expression in the mouse brain with image registration, feature extraction and block clustering. BMC Bioinform 8(Suppl 10):S5. doi:10.1186/1471-2105-8-S10-S5

Ji S (2013) Computational genetic neuroanatomy of the developing mouse brain: dimensionality reduction, visualization, and clustering. BMC Bioinform 14:222. doi:10.1186/1471-2105-14-222

Ji S, Zhang W, Li R (2013) A probabilistic latent semantic analysis model for coclustering the mouse brain atlas. IEEE ACM Trans Comput Biol Bioinform 10:1460–1468. doi:10.1109/TCBB.2013.135

Ji S, Fakhry A, Deng H (2014) Integrative analysis of the connectivity and gene expression atlases in the mouse brain. Neuroimage 84:245–253. doi:10.1016/j.neuroimage.2013.08.049

Jones AR, Overly CC, Sunkin SM (2009) The Allen Brain Atlas: 5 years and beyond. Nat Rev Neurosci 10:821–828. doi:10.1038/nrn2722

Jung M, Jin S-G, Zhang X et al (2015) Longitudinal epigenetic and gene expression profiles analyzed by three-component analysis reveal down-regulation of genes involved in protein translation in human aging. Nucleic Acids Res 43:1–14. doi:10.1093/nar/gkv473

Kang HJ, Kawasawa YI, Cheng F et al (2011) Spatio-temporal transcriptome of the human brain. Nature 478:483–489. doi:10.1038/nature10523

Kaufman A, Dror G, Meilijson I, Ruppin E (2006) Gene expression of Caenorhabditis elegans neurons carries information on their synaptic connectivity. PLoS Comput Biol 2:e167. doi:10.1371/journal.pcbi.0020167

Kim SK, Lund J, Kiraly M et al (2001) A gene expression map for Caenorhabditis elegans. Science 293:2087–2092. doi:10.1126/science.1061603

Kirsch L, Liscovitch N, Chechik G (2012) Localizing genes to cerebellar layers by classifying ISH images. PLoS Comput Biol 8:e1002790. doi:10.1371/journal.pcbi.1002790

Ko Y, Ament SA, Eddy JA et al (2013) Cell type-specific genes show striking and distinct patterns of spatial expression in the mouse brain. Proc Natl Acad Sci 110:3095–3100. doi:10.1073/pnas.1222897110

Kolch W, Halasz M, Granovskaya M, Kholodenko BN (2015) The dynamic control of signal transduction networks in cancer cells. Nat Rev Cancer 15:515–527. doi:10.1038/nrc3983

Kondapalli KC, Prasad H, Rao R (2014) An inside job: how endosomal Na(+)/H(+) exchangers link to autism and neurological disease. Front Cell Neurosci 8:172. doi:10.3389/fncel.2014.00172

Krienen FM, Yeo BTT, Ge T et al (2016) Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. Proc Natl Acad Sci. doi:10.1073/pnas.1510903113

Krumm N, O'Roak BJ, Shendure J, Eichler EE (2014) A de novo convergence of autism genetics and molecular neuroscience. Trends Neurosci 37:95–105. doi:10.1016/j.tins.2013.11.005

Kuhn A, Thu D, Waldvogel HJ et al (2011) Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. Nat Methods 8:945–947. doi:10.1038/nmeth.1710

Kumari S, Nie J, Chen H-S et al (2012) Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. PLoS One 7:e50411. doi:10.1371/journal.pone.0050411

Lau C, Ng L, Thompson C et al (2008) Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. BMC Bioinform 9:153. doi:10.1186/1471-2105-9-153

Lausted C, Lee I, Zhou Y et al (2014) Systems approach to neurodegenerative disease biomarker discovery. Annu Rev Pharmacol Toxicol 54:457–481. doi:10.1146/annurev-pharmtox-011613-135928

Lee JA, Verleysen M (2005) Nonlinear dimensionality reduction of data manifolds with essential loops. Neurocomputing 67:29–53. doi:10.1016/j.neucom.2004.11.042

Lein ES, Hawrylycz MJ, Ao N et al (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445:168–176. doi:10.1038/nature05453

Lewis BP, Shih I, Jones-Rhoades MW et al (2003) Prediction of mammalian microRNA targets. Cell 115:787–798. doi:10.1016/S0092-8674(03)01018-3

Li R, Zhang W, Ji S (2014) Automated identification of cell-type-specific genes in the mouse brain by image computing of expression patterns. BMC Bioinform 15:209. doi:10.1186/1471-2105-15-209

Liscovitch N, Chechik G (2013) Specialization of gene expression during mouse brain development. PLoS Comput Biol. doi:10.1371/journal.pcbi.1003185

Liscovitch N, French L (2014) Differential co-expression between α-synuclein and IFN-γ signaling genes across development and in Parkinson's disease. PLoS One 9:e115029. doi:10.1371/journal.pone.0115029

Liscovitch N, Shalit U, Chechik G (2013) FuncISH: learning a functional representation of neural ISH images. Bioinformatics 29:i36–i43. doi:10.1093/bioinformatics/btt207

Liu Z, Yan SF, Walker JR et al (2007) Study of gene function based on spatial co-expression in a high-resolution mouse brain atlas. BMC Syst Biol 1:19. doi:10.1186/1752-0509-1-19

Liu J, Wang X, Li J et al (2014) Reconstruction of the gene regulatory network involved in the sonic Hedgehog pathway with a potential role in early development of the mouse brain. PLoS Comput Biol 10:e1003884. doi:10.1371/journal.pcbi.1003884

Lodato MA, Woodworth MB, Lee S et al (2015) Somatic mutation in single human neurons tracks developmental and transcriptional history. Science 350:94–98

Loerch PM, Lu T, Dakin KA et al (2008) Evolution of the aging brain transcriptome and synaptic regulation. PLoS One 3:e3329. doi:10.1371/journal.pone.0003329

Lonsdale J (2013) The genotype-tissue expression (GTEx) project. Nat Genet 45:580–585. doi:10.1038/ng.2653

Mahfouz A, Ziats MN, Rennert OM et al (2014) Genomic connectivity networks based on the BrainSpan atlas of the developing human brain. SPIE Medical Imaging, pp 90344G–90344G

Mahfouz A, van de Giessen M, van der Maaten L et al (2015a) Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings. Methods 73:79–89. doi:10.1016/j.ymeth.2014.10.004

Mahfouz A, Ziats MN, Rennert OM et al (2015b) Shared pathways among autism candidate genes determined by co-expression network analysis of the developing human brain transcriptome. J Mol Neurosci 57:580–594. doi:10.1007/s12031-015-0641-3

Matys V, Kel-Margoulis OV, Fricke E et al (2006) TRANSFAC(R) and its module TRANSCompel(R): transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34:D108–D110. doi:10.1093/nar/gkj143

Maze I, Shen L, Zhang B et al (2014) Analytical tools and current challenges in the modern era of neuroepigenomics. Nat Neurosci. doi:10.1038/nn.3816

Medland SE, Jahanshad N, Neale BM, Thompson PM (2014) Whole-genome analyses of whole-brain data: working within an expanded search space. Nat Neurosci 17:791–800. doi:10.1038/nn.3718

Menashe I, Grange P, Larsen EC et al (2013) Co-expression profiling of autism genes in the mouse brain. PLoS Comput Biol 9:e1003128. doi:10.1371/journal.pcbi.1003128

Mercer TR, Dinger ME, Sunkin SM et al (2008) Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci 105:716–721. doi:10.1073/pnas.0706729105

Miazaki M, Costa LDF (2012) Study of cerebral gene expression densities using Voronoi analysis. J Neurosci Methods 203:212–219. doi:10.1016/j.jneumeth.2011.09.009

Mignogna P, Viggiano D (2010) Brain distribution of genes related to changes in locomotor activity. Physiol Behav 99:618–626. doi:10.1016/j.physbeh.2010.01.026

Miller JA, Nathanson J, Franjic D et al (2013) Conserved molecular signatures of neurogenesis in the hippocampal subgranular zone of rodents and primates. Development 140:4633–4644. doi:10.1242/dev.097212

Miller JA, Ding S-L, Sunkin SM et al (2014a) Transcriptional landscape of the prenatal human brain. Nature 508:199–206. doi:10.1038/nature13185

Miller JA, Menon V, Goldy J et al (2014b) Improving reliability and absolute quantification of human brain microarray data by filtering and scaling probes using RNA-Seq. BMC Genom 15:1–14. doi:10.1186/1471-2164-15-154

Milyaev N, Osumi-sutherland D, Reeve S et al (2012) The virtual fly brain browser and query interface. Bioinformatics 28:411–415. doi:10.1093/bioinformatics/btr677

Myers EM, Bartlett CW, Machiraju R, Bohland JW (2015) An integrative analysis of regional gene expression profiles in the human brain. Methods 73:54–70. doi:10.1016/j.ymeth.2014.12.010

Neale BM, Kou Y, Liu L et al (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485:242–245. doi:10.1038/nature11011

Ng L, Bernard A, Lau C et al (2009) An anatomic gene expression atlas of the adult mouse brain. Nat Neurosci 12:356–362. doi:10.1038/nn.2281

Ng L, Lau C, Sunkin SM et al (2010) Surface-based mapping of gene expression and probabilistic expression maps in the mouse cortex. Methods 50:55–62. doi:10.1016/j.ymeth.2009.10.001

O'Roak BJ, Vives L, Girirajan S et al (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485:246–250. doi:10.1038/nature10989

Ogata H, Goto S, Sato K et al (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27:27–30. doi:10.1093/nar/27.1.29

Oldham MC, Horvath S, Geschwind DH (2006) Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc Natl Acad Sci 103:17973–17978. doi:10.1073/pnas.0605938103

Oldham MC, Konopka G, Iwamoto K et al (2008) Functional organization of the transcriptome in human brain. Nat Neurosci 11:1271–1282. doi:10.1038/nn.2207

Olszewski PK, Cedernaes J, Olsson F et al (2008) Analysis of the network of feeding neuroregulators using the Allen Brain Atlas. Neurosci Biobehav Rev 32:945–956. doi:10.1016/j.neubiorev.2008.01.007

Online Mendelian inheritance in man, OMIM (2015a). In: McKusick-Nathans Inst. Genet. Med. Johns Hopkins Univ., Baltimore. http://omim.org/

Parikshak NN, Luo R, Zhang A et al (2013) Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155:1008–1021. doi:10.1016/j.cell.2013.10.031

Parikshak NN, Gandal MJ, Geschwind DH (2015) Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat Rev Genet 16:441–458. doi:10.1038/nrg3934

Pavlopoulos GA, Malliarakis D, Papanikolaou N et al (2015) Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future. Gigascience. doi:10.1186/s13742-015-0077-2

Peng H, Chung P, Long F et al (2011) BrainAligner: 3D registration atlases of Drosophila brains. Nat Methods 8:493–500. doi:10.1038/nmeth.1602

Peng Q, Schork A, Bartsch H et al (2016) Conservation of distinct genetically-mediated human cortical pattern. PLoS Genet 12:1–18. doi:10.1371/journal.pgen.1006143

Pettit J-B, Tomer R, Achim K et al (2014) Identifying cell types from spatially referenced single-cell expression datasets. PLoS Comput Biol 10:e1003824. doi:10.1371/journal.pcbi.1003824

Pinero J, Queralt-Rosinach N, Bravo A et al (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database 2015:bav028–bav028. doi:10.1093/database/bav028

Piro RM, Molineris I, Ala U et al (2010) Candidate gene prioritization based on spatially mapped gene expression: an application to XLMR. Bioinformatics 26:i618–i624. doi:10.1093/bioinformatics/btq396

Piro RM, Molineris I, Ala U, Di Cunto F (2011) Evaluation of candidate genes from orphan FEB and GEFS$^+$ loci by analysis of human brain gene expression atlases. PLoS One 6:e23149. doi:10.1371/journal.pone.0023149

Pletikos M, Sousa AMM, Sedmak G et al (2014) Temporal specification and bilaterality of human neocortical topographic gene expression. Neuron 81:321–332. doi:10.1016/j.neuron.2013.11.018

Pollock JD, Wu DY, Satterlee JS (2014) Molecular neuroanatomy: a generation of progress. Trends Neurosci 37:106–123. doi:10.1016/j.tins.2013.11.001

Ponjavic J, Oliver PL, Lunter G, Ponting CP (2009) Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet 5:e1000617. doi:10.1371/journal.pgen.1000617

Portales-Casamar E, Thongjuea S, Kwon AT et al (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res 38:D105–D110. doi:10.1093/nar/gkp950

Przulj N (2007) Biological network comparison using graphlet degree distribution. Bioinformatics 23:e177–e183. doi:10.1093/bioinformatics/btl301

Qureshi IA, Mehler MF (2012) Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. Nat Rev Neurosci 13:528–541. doi:10.1038/nrn3234

Ramsden HL, Sürmeli G, McDonagh SG, Nolan MF (2015) Laminar and dorsoventral molecular organization of the medial entorhinal cortex revealed by large-scale anatomical analysis of gene expression. PLoS Comput Biol 11:e1004032. doi:10.1371/journal.pcbi.1004032

Richardson L, Venkataraman S, Stevenson P et al (2014) EMAGE mouse embryo spatial gene expression database: 2014 update. Nucleic Acids Res 42:D703–D709. doi:10.1093/nar/gkt1155

Richiardi J, Altmann A, Jonas R (2015) Correlated gene expression supports synchronous activity in brain networks. Science 348:1241–1244. doi:10.1126/science.1255905

Ripke S, Neale BM, Corvin A et al (2014) Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427. doi:10.1038/nature13595

Rizzo G, Veronese M, Heckemann RA et al (2014) The predictive power of brain mRNA mappings for in vivo protein density: a positron emission tomography correlation study. J Cereb Blood Flow Metab 34:827–835. doi:10.1038/jcbfm.2014.21

Rizzo G, Veronese M, Expert P et al (2016) MENGA: a new comprehensive tool for the integration of neuroimaging data and the allen human brain transcriptome atlas. PLoS One 11:e0148744. doi:10.1371/journal.pone.0148744

Romme IA, de Reus MA, Ophoff RA et al (2016) Connectome disconnectivity and cortical gene expression in schizophrenia. Biol Psychiatry. doi:10.1016/j.biopsych.2016.07.012

Roth A, Kyzar EJ, Cachat J et al (2013) Potential translational targets revealed by linking mouse grooming behavioral phenotypes to gene expression using public databases. Prog Neuro-Psychopharmacol Biol Psychiatry 40:312–325. doi:10.1016/j.pnpbp.2012.10.015

Rubinov M, Ypma RJF, Watson C, Bullmore ET (2015) Wiring cost and topological participation of the mouse brain connectome. Proc Natl Acad Sci 112:201420315. doi:10.1073/pnas.1420315112

Saadatpour A, Lai S, Guo G, Yuan G-C (2015) Single-cell analysis in cancer genomics. Trends Genet 31:576–586. doi:10.1016/j.tig.2015.07.003

Sanders SJ, Murtha MT, Gupta AR et al (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485:237–241. doi:10.1038/nature10945

Shapiro E, Biezuner T, Linnarsson S (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet 14:618–630. doi:10.1038/nrg3542

Shin J, Ming G, Song H (2014) Decoding neural transcriptomes and epigenomes via high-throughput sequencing. Nat Neurosci 17:1463–1475. doi:10.1038/nn.3814

Simón-Sánchez J, Singleton A (2008) Genome-wide association studies in neurological disorders. Lancet Neurol 7:1067–1072. doi:10.1016/S1474-4422(08)70241-2

Škrášková K, Khmelinskii A, Abdelmoula WM et al (2015) Precise anatomic localization of accumulated lipids in Mfp2 deficient murine brains through automated registration of SIMS images to the Allen Brain Atlas. J Am Soc Mass Spectrom. doi:10.1007/s13361-015-1146-6

Song L, Langfelder P, Horvath S (2012) Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC Bioinform 13:328. doi:10.1186/1471-2105-13-328

Spencer WC, Zeller G, Watson JD et al (2011) A spatial and temporal map of C. elegans gene expression. Genome Res 21:325–341. doi:10.1101/gr.114595.110.Freely

Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302:249–255. doi:10.1126/science.1087447

Sunkin SM (2006) Towards the integration of spatially and temporally resolved murine gene expression databases. Trends Genet 22:211–217. doi:10.1016/j.tig.2006.02.006

Sunkin SM, Ng L, Lau C et al (2013) Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Res. doi:10.1093/nar/gks1042

Tan PPC, French L, Pavlidis P (2013) Neuron-enriched gene expression patterns are regionally anti-correlated with oligodendrocyte-enriched patterns in the adult mouse and human brain. Front Neurosci 7:1–12. doi:10.3389/fnins.2013.00005

Tchagang AB, Phan S, Famili F et al (2012) Mining biological information from 3D short time-series gene expression data: the OPTricluster algorithm. BMC Bioinform 13:54. doi:10.1186/1471-2105-13-54

Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323. doi:10.1126/science.290.5500.2319

Thompson CL, Ng L, Menon V et al (2014) A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. Neuron 83:1–15. doi:10.1016/j.neuron.2014.05.033

Uddin M, Tammimies K, Pellecchia G et al (2014) Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. Nat Genet 46:742–747. doi:10.1038/ng.2980

Uhlen M, Fagerberg L, Hallstrom BM et al (2015) Tissue-based map of the human proteome. Science 347:1260419. doi:10.1126/science.1260419

Valli I, Crossley NA, Day F et al (2016) HPA-axis function and grey matter volume reductions: imaging the diathesis-stress model in individuals at ultra-high risk of psychosis. Transl Psychiatry 6:e797. doi:10.1038/tp.2016.68

van den Akker EB, Passtoors WM, Jansen R et al (2014) Meta-analysis on blood transcriptomic studies identifies consistently coexpressed protein-protein interaction modules as robust markers of human aging. Aging Cell 13:216–225. doi:10.1111/acel.12160

Van Der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605. doi:10.1007/s10479-011-0841-3

Van Essen DC, Ugurbil K (2012) The future of the human connectome. Neuroimage 62:1299–1310

Varadan V, Miller DM, Anastassiou D (2006) Computational inference of the molecular logic for synaptic connectivity in C. elegans. Bioinformatics 22:e497–e506. doi:10.1093/bioinformatics/btl224

Vermunt MW, Reinink P, Korving J et al (2014) Large-scale identification of coregulated enhancer networks in the adult human brain. Cell Rep 9:767–779. doi:10.1016/j.celrep.2014.09.023

Veronese M, Zanotti-Fregonara P, Rizzo G et al (2016) Measuring specific receptor binding of a PET radioligand in human brain without pharmacological blockade: the genomic plot. Neuroimage 130:1–12. doi:10.1016/j.neuroimage.2016.01.058

Vértes PE, Rittman T, Whitaker KJ et al (2016) Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks. Philos Trans R Soc Lond B Biol Sci 371:735–769. doi:10.1098/rstb.2015.0362

Visel A, Thaller C, Eichele G (2004) GenePaint.org: an atlas of gene expression patterns in the mouse embryo. Nucleic Acids Res 32:D552–D556. doi:10.1093/nar/gkh029

Voineagu I, Wang X, Johnston P et al (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474:380–384. doi:10.1038/nature10110

Wagner F, French L, Veh RW (2016) Transcriptomic-anatomic analysis of the mouse habenula uncovers a high molecular heterogeneity among neurons in the lateral complex, while gene expression in the medial complex largely obeys subnuclear boundaries. Brain Struct Funct 221:39–58. doi:10.1007/s00429-014-0891-9

Wang H, Nie F, Huang H et al (2012) Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. Bioinformatics 28:i127–i136. doi:10.1093/bioinformatics/bts228

Wang YXR, Waterman MS, Huang H (2014) Gene coexpression measures in large heterogeneous samples using count statistics. Proc Natl Acad Sci 111:16371–16376. doi:10.1073/pnas.1417128111

Wang GZ, Belgard TG, Mao D et al (2015) Correspondence between resting-state activity and brain gene expression. Neuron 88:659–666. doi:10.1016/j.neuron.2015.10.022

Weinstein JN, Collisson EA, Mills GB et al (2013) The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45:1113–1120. doi:10.1038/ng.2764

Welter D, MacArthur J, Morales J et al (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001–D1006. doi:10.1093/nar/gkt1229

Whitaker KJ, Vértes PE, Romero-Garcia R et al (2016) Adolescence is associated with transcriptionally patterned consolidation of the hubs of the human brain connectome. Proc Natl Acad Sci. doi:10.1073/PNAS.1601745113

Willsey AJ, Sanders SJ, Li M et al (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155:997–1007. doi:10.1016/j.cell.2013.10.020

Wolf L, Goldberg C, Manor N et al (2011) Gene expression in the rodent brain is associated with its regional connectivity. PLoS Comput Biol 7:e1002040. doi:10.1371/journal.pcbi.1002040

Xiong HY, Alipanahi B, Lee LJ et al (2014) The human splicing code reveals new insights into the genetic determinants of disease. Science 347:1254806. doi:10.1126/science.1254806

Xu X, Coats JK, Yang CF et al (2012) Modular genetic control of sexually dimorphic behaviors. Cell 148:596–607. doi:10.1016/j.cell.2011.12.018

Yan J, Du L, Kim S et al (2014) Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. Bioinformatics 30:i564–i571. doi:10.1093/bioinformatics/btu465

Yang Y, Han L, Yuan Y et al (2014) Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat Commun 5:1–9. doi:10.1038/ncomms4231

Yarkoni T, Poldrack RA, Nichols TE et al (2011) Large-scale automated synthesis of human functional neuroimaging data. Nat Methods 8:665–670. doi:10.1038/nmeth.1635

Zalachoras I, Houtman R, Meijer OC (2013) Understanding stress-effects in the brain via transcriptional signal transduction pathways. Neuroscience 242:97–109. doi:10.1016/j.neuroscience.2013.03.038

Zaldivar A, Krichmar JL (2013) Interactions between the neuromodulatory systems and the amygdala: exploratory survey using the Allen Mouse Brain Atlas. Brain Struct Funct 218:1513–1530. doi:10.1007/s00429-012-0473-7

Zeisel A, Muñoz-Manchado AB, Codeluppi S et al (2015) Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 347:1138–1142. doi:10.1126/science.aaa1934

Zeng T, Li R, Mukkamala R et al (2015) Deep convolutional neural networks for annotating gene expression patterns in the mouse brain. BMC Bioinform 16:1–10. doi:10.1186/s12859-015-0553-9

Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4:Article17. doi:10.2202/1544-6115.1128

Zhang B, Gaiteri C, Bodea LG et al (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 153:707–720. doi:10.1016/j.cell.2013.03.030

Zhang Y, Chen K, Sloan SA et al (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci 34:1–19. doi:10.1523/JNEUROSCI.1860-14.2014

Ziats MN, Rennert OM (2013) Identification of differentially expressed microRNAs across the developing human brain. Mol Psychiatry. doi:10.1038/mp.2013.93