# Analyzing spatial transcriptomics and neuroimaging data in neurodegenerative diseases

Keo, D.

# Chapter 7
# Discussion

Despite many efforts to identify genes associated with neurodegenerative diseases, the disease etiology leading to the formation of protein inclusions and neuronal cell death largely remains unclear. This is mainly due to the complexity of both polygenic and environmental factors that contribute to the progression of neurodegenerative diseases. Expression profiles of samples derived from patients are affected by ongoing inflammation, oxidative stress, and other immune mechanisms and it is not known whether such molecular mechanisms are a cause or consequence of the disease. By using different approaches to combine neuroimaging data with a spatial gene expression atlas of the human brain, we revealed healthy state transcriptomic signatures occurring in brain regions that are selectively vulnerable in neurodegenerative diseases. We showed that structural brain networks that are associated with gray matter loss in PD patients are enriched for the expression of cholinergic genes. In structural networks associated with gray matter loss in HD, we found strong co-expression between polyglutamine (polyQ) genes *HTT*, *ATXN2*, and *ATN1*. Similar relationships were found in anatomical brain structures that are known to be affected in HD. In the stress response network of individuals at risk of schizophrenia, we found that upregulated genes were associated with psychiatric disorders. The integrated analysis of both gene expression data and neuroimaging data revealed that the expression of genes involved in cellular maintenance mechanisms are correlated with cortical thickness in PD patients. Finally, we identified a module of dopaminergic genes for which its expression is correlated with the PD Braak staging of brain regions. Overall, we showed how the AHBA can be combined with brain phenotypes observed in neurodegenerative diseases such as gray matter loss, neuronal loss, accumulation of disease-specific protein aggregates, and changes in functional activity or cortical thickness. Our findings point towards local molecular events and enable a better understanding of the spatial organization of brain functions that are impaired in neurodegenerative diseases.

### 7.1 Sampling resolution of brain transcriptomic data

Spatial transcriptomics can reveal gene expression patterns that are indicative of local gene functions. Although case-control studies are important to understand gene expression changes in health and disease, more profound differences in gene expression are found between brain regions than between disease conditions [1]. This supports the idea that genes fulfill specific functions in different brain regions. The AHBA allows analyzing gene expression at an unprecedented spatial resolution. Yet, due to this high resolution, results cannot be directly compared to other datasets with a case-control setting, since these datasets usually lack this high sampling resolution. This emphasizes the need for high-resolution sampling datasets in both control and patient data. The unavailability of high-resolution gene expression data from PD patients makes it difficult to compare expression patterns from the AHBA to data from PD patients. Nevertheless, in Chapter 6, we analyzed two datasets of healthy individuals and two datasets of PD patients that had samples from several brain regions associated with Braak stages. Here, we also found more differentially expressed genes between brain regions than between disease conditions in two PD datasets

with samples from several brain regions that are involved in the Braak staging scheme. Interestingly, this is still true when correcting the analyses for cell-type composition.

While neurodegenerative diseases are age-related diseases, we did not take the age difference between elderly with PD and healthy adult donors from the AHBA into account. Moreover, the individual variability not only depends on age, but also gender, genetic background and other factors that were also not taken into account in our analyses, all because of the limited number of brain donors in the AHBA. Finding donors representing a healthy brain is challenging and requires thorough screening and quality control tests to make sure they are eligible for inclusion. Besides that, some brain regions are also vulnerable to neurodegeneration during healthy aging. Hence, to better understand the molecular processes underlying neurodegenerative diseases, it is important to understand healthy aging in relation to a disease. This will help to even better understand early changes in disease and enable diagnosis before disease onset.

### 7.2 Imaging cohorts of PD are heterogeneous

Approaches to combine neuroimaging with brain-wide transcriptomic data allow analyzing the functional organization of gene expression across the brain, which is important to better understand patterns of neurodegeneration. Although many brain regions have been associated with the pathology of PD, it is not well understood what determines the typical patterns of atrophy in all PD subtypes or differences in atypical PD. Imaging cohorts of HD patients are better characterized genetically, compared to cohorts of PD patients, as HD diagnosis is confirmed based on genetic tests to determine whether the CAG-repeat length in the *HTT* gene is expanded. Since the CAG-repeat length determines the age of onset and therefore also the severity and progression of the disease, this determines the rate of brain atrophy and causes differences in MRI scans of HD patients [2,3]. As such, imaging studies should take into account the CAG repeat length. PD cohorts likely have more heterogeneous groups of patients, as PD diagnosis cannot be confirmed with genetic tests yet and is nowadays based on the observation of clinical symptoms during life. Moreover, PD is a complex disorder in which not all patients have the same symptoms and the severity of symptoms varies substantially. True PD diagnosis can only be confirmed after death upon pathological examination. Interestingly, it may turn out that a patient actually suffered a similar but different disease, such incidental Lewy body disease, or dementia with Lewy bodies. This impacts the group coherence of a PD cohort and lowers the significance of results in an MRI study. Thus, it is important to acknowledge that PD is a complex disorder with a wide spectrum of symptoms and that this heterogeneity may influence findings.

### 7.3 Study design and interpretation

To analyze the transcriptome of a group of samples, gene expression levels need to be compared to a control group. While this is straightforward in case-control studies, in our studies of the healthy human brain, we assessed differential expression between the region of interest (associated with the disease) and a control region (assumed to not be associated with disease). It is not always clear whether a region is really unaffected in a disease. Therefore, choosing the control region for comparison depends on the research question

that has to be answered and will influence findings. For example, studies should take into account that expression patterns within regions may be dominated by differently sized or distinct anatomical structures, such as the cortex or cerebellum being quite different from the other brain regions. Hence, it might not be clear whether the comparison with a control region might result in biases in the analysis or that the results indeed are biologically associated with the disease. One way to solve this is to repeat the analysis with and without such regions and find the overlap in results as we did when analyzing the stress network of individuals at risk of schizophrenia in Chapter 4. In this study, we also excluded the cerebellum from this analysis because it has as transcriptomic signature that is quite heterogeneous and very distinct from the rest of the brain. When analyzing polyQ co-expression patterns in the cerebellum in Chapter 3, we excluded the cerebellar nuclei from the cerebellar cortex as these samples showed strong expression differences. Depending on the study design, brain regions may be excluded if their distinct expression can cause biases in the analysis.

For functional interpretation of results, we relied on databases with functional annotation of genes describing molecular processes, biological components, or pathways. In general, we found that genes enriched in brain regions associated with HD and PD were related to lysosomal, mitochondrial and DNA repair pathways, ubiquitin, and the cell cycle, which have been described before in HD and PD studies. Genes that have been well-studied are highly annotated, while other genes received less attention. This annotation inequality may lead researchers to focus mostly on richly annotated genes, while other genes with statistical significance or large effect sizes may be neglected [4]. In this case, data-driven studies fail to identify unknown mechanisms involved in disease. As such, it is good practice that studies report all significant findings so these can always be searched for in the future by scientists that are interested in unraveling the role of particular genes.

### 7.4 Expression of genetic risk factors

Rare genetic variants that are highly penetrant have been identified in large families with PD and more common variants with smaller effect sizes were discovered in genome wide association studies (GWAS) of sporadic PD. How these mutation variants lead to the molecular consequences observed in PD generally remains unclear. One of the many hypotheses is that these mutations result in misfolded toxic proteins that affect other proteins and gradually spread through the brain [5]. Analyzing the spatial expression of genetic risk factors allows to better understand the consequence of genetic risk variants and how they influence the progression to neurodegenerative disease. In our PD studies results, we analyzed the whole genome and looked up whether our findings included PD-related genes. For some PD-related genes we could indeed find an association with regional vulnerability, but for many PD-related genes we could not find interesting expression patterns across the brain. One reason for this could be that we looked at gene expression in healthy adults and the effects of variants on gene expression may only be apparent in PD patients or elderly. On the other hand, it is not known whether the donors from the AHBA that were considered neurologically healthy could have developed PD at a later stage in life.

Moreover, it is possible that the identified variants in GWAS of PD affect other genes than those that were linked in those studies. To pinpoint variants and genes that may be associated with a trait, GWAS rely on linkage disequilibrium patterns and information on the functional consequences of the variant. During this analysis step, GWAS hits can be erroneously mapped to genes.

Furthermore, GWAS have been mostly facilitated by inexpensive SNP arrays that are designed to target common variants across the whole genome [6]. Arrays like the NeuroX-chip have been designed to also include more rare risk variants that may be related to neurodegenerative diseases [7], but these arrays still require predesigned DNA probes to target the mutations. Future studies will likely make more use of next-generation sequencing techniques, such as whole genome sequencing (WGS) which can cover more variation in the genome and also the spectrum of minor allele frequencies of variants. GWAS using WGS increases the power and precision of analyses which is likely due to more accurate determination of genotypes [8].

For complex disorders like PD, many loci contribute to the genetic variation observed in PD and the proportion of variance explained by individual variants is small. The polygenic risk score is a popular method to assess the status of multiple disease-related variants and the aggregated effect size to highlight an individual's risk to develop a disease [9]. Moreover, polygenic risk scores (PRS) can help to better understand the shared genetic architecture of neurodegenerative diseases by determining whether variants with pleiotropic effects identified in one disease can lead to an increased risk for another disease [10]. It would be interesting to see how polygenic effects of common and rare variants can change gene expression levels of disease-related tissues. A promising direction is to use RNA-sequencing to perform expression quantitative trait loci (eQTL) mapping and find regulatory variants that can explain variation in gene expression levels. Future studies may consider using a PRS calculated based on effect sizes obtained with eQTL instead of GWAS and assess the polygenic effect across populations, multiple tissues, cell-types, or even single cells. In addition, an omnigenic model integrating the effect of rare and common variants along with gene co-expression networks can aid in better understanding the genetic architecture of complex disorders like PD [11].

### 7.5 Gene network analysis

Co-expression is the correlation between the expression patterns of two genes and can be used to construct weighted gene networks that reflect functional associations between genes. One of the goals in co-expression analysis is the detection of gene modules that represent tightly connected subnetworks of co-expressed genes and allows inferring gene function with the guilt-by-association principle. While many studies analyzed gene co-expression across samples from individuals, here, we exploit the AHBA to analyze spatial co-expression across samples from different brain regions. Since gene expression can be highly tissue-specific we assessed spatial co-expression patterns that could be indicative of a relationship between two genes with different levels of interactions across the brain.

In our studies, we used Pearson's correlation as a measure for co-expression which can capture co-expression patterns that are informative, but can only capture monotonic linear relationships. Other popular measures of co-expression are Spearman's rank correlation, mutual information and biweight midcorrelation [12,13], but the efficiency of different methods depends on the data properties and varies with biological processes [14]. More recently, it has been demonstrated that with increasing sample sizes Pearson's correlation coefficient with highest reciprocal ranking is well-suited to create robust gene co-expression networks [15] and the quality of co-expression networks can be further improved by down-sampling the expression dataset and integrating smaller networks into stronger networks [16].

There are several ways to determine the threshold used for constructing gene co-expression networks. These include setting a hard threshold at a co-expression cut-off value or determining the statistical significance of the correlation and set a *P*-value cut-off. In both cases the cut-off is arbitrary chosen and not necessarily biologically relevant. WGCNA proposed a method where the cut-off threshold is selected by choosing a soft thresholding power such that the gene network approximates a scale-free topology [17]. It is often claimed that real-world networks are scale-free, meaning that a fraction of nodes with degree $k$ follows a power-law $k^{-\alpha}$. However, there is evidence across social, biological, and technological domains that scale-free networks are empirically rare and that the power law is not a good fit for network degree distribution [18]. Furthermore, the power-transformation may put more emphasis on stronger associations and mitigate weaker associations by raising the co-expression similarity to a power, but we argue that this transformation is equivalent to changing the hierarchical tree cutting threshold to obtain larger clusters. A higher power applied to the similarity matrix will result in a smaller number of clusters, which are essentially superclusters of the clusters obtained with a power-transformation. Therefore, we believe that soft-thresholding has no additional value to our analyses.

### 7.6 Correcting for cell-type composition in bulk tissue

The expression of cell-type markers can be analyzed to understand the distribution of different cell-types across the brain. Here, we have assessed the presence of cell-type markers among our results to find out which cell-types are related to the brain region of interest. While the AHBA has a high sampling resolution, samples were collected from bulk tissues. This means that one sample is characterized by the composition of present cell-types which likely affects gene expression measurements. Brain regions of patients with neurodegenerative diseases show changes in cell-type composition compared to age-matched controls. During disease progression neurons are lost but other cell-types become more abundant such as astrocytes and microglia. Therefore, it is important to find differentially expressed genes in case-control studies that do not result from differences in cell-type composition between patients and controls.

Multiple cell-type deconvolution methods have been proposed to analyze differential expression while correcting for cell-type composition between samples from patients and controls. We have explored one of these cell-type deconvolution methods to correct for cell-type composition between different brain regions as the cell-type composition likely differs between brain regions according to their structural and anatomical functions. Since the cell-type composition is unknown for samples from bulk tissue, the correction methods require cell-type markers to estimate the cell-type composition of a sample. The existing methods use linear models to correct fold-changes in cell-type specific expression between two groups of samples. The proportions of each cell-type can be estimated by taking the mean expression of markers for a specific cell-type. Some methods use PCA or SVD to calculate the eigengene, but the eigengene expression level of a cell-type cannot be relatively compared to the eigengene expression of other cell-types, and therefore the gene expression of all cell-types should not be summed in a linear equation. Additionally, cell-type proportions should be non-negative and sum up to one. Methods to deal with these constrains include non-negative matrix factorization and non-negative least squares [19].

Another problem that deconvolution methods deal with is the choice of reference markers. There are only few robust cell-type markers for which their exclusive expression in a specific cell-type is certain. Finding an optimal set of cell-type markers is not straightforward because the definition of a cell-type remains an open question in biological research. Cells of the central nervous system are divided into neurons and glia each consisting of many subtypes with their own molecular properties. The main cell-types that are recognized have been defined based on morphological and physiological features, but more recent studies have identified many more subclasses based on the gene expression of groups of cells to understand their molecular properties. In our studies with the AHBA, we relied on cell-type markers that have been defined in other studies with mouse brains. Cells were disassociated and sorted to find sets of cells with a common transcriptomic signature that is different from other cells. While most genes may serve as orthologs for other species, some genes may be absent or perform different functions in the human brain [20]. Better markers are needed that can robustly identify the presence of specific cell-types. Advances in single cell technologies enable the transcriptomic analysis of single cells and the growing interest in single cell analysis is rapidly leading to the discovery of more cell-types and a better understanding the cellular heterogeneity within the brain. Several methods have been proposed to infer cell-type composition of RNA-seq bulk tissue using single cell RNA-seq (scRNA-seq) data as a reference [21–25]. A benchmarking study of deconvolution methods explained relevant factors that should be taken into account when using cell-type deconvolution methods, mainly the input data should be kept in a linear scale, and missing cell-types in the reference lead to erroneous estimations of cell-type proportions [26].

### 7.7 Single cell transcriptomics

Single cell transcriptomics can exploit the cellular diversity and unravel the cell-type composition of selected brain areas. To identify cell-types and their associated transcriptome, most studies look for cell-types that have already been defined in literature.

For example, a study that identified cell-types from single cell human data still relied on prior knowledge of mouse cell-types [27]. To identify groups of cells with similar transcriptomes, clustering methods are used to construct a cell-type taxonomy tree and can therefore identify known and new cell-types [28]. Once the cell-types have been established, cell-type classifiers can be applied to unseen single cell data to label cells based on their transcriptomic signature. Since the advance of scRNA-seq, several classification methods for single cell data have been developed that seem to perform well on multiple single cell datasets [29]. With these recent advances in single cell analysis, many subclasses of known cell-types have been discovered expanding our knowledge on cellular diversity. However, subclasses at the bottom of the cell hierarchy are based on the analysis of only few cells, and these detailed findings may not be reproducible. In addition, newly identified cell-types may be falsely discovered when there are unforeseen biological factors that influence the transcriptome of as cell, such as the transitional state during cell cycle processes. For example, the transcriptome of cells changes to perform functions related to cell division or cell development. The Allen institute for Brain Science (AIBS) provides a cell-type database that contains electrophysiological, morphological, and transcriptomic data from single cells in the human and mouse brain. This database allows analyzing multiple properties of the cell to help researchers gain more insights into cell-types characteristics.

Single cell studies of PD are mostly done with mouse models of PD or induced pluripotent stem cells (iPSC) from PD patients that are differentiated into dopaminergic cells to study them *in vitro*. To date, scRNA-seq data for human striatum or substantia nigra is very limited, but more data of disease-related tissues is expected in the near future. Analyzing differences in cell-type composition of specific brain regions in PD and control can aid in better understanding how different cell-types play a role in neurodegeneration. The analysis of transcriptomic data from brain single cells and bulk tissue of case-control studies has led to the identification of cell-types that are susceptible in neurological disorders, such as epilepsy, schizophrenia and AD [30]. The numerous possibilities of single cell analysis holds promise for the future to better understand cellular changes between health and disease and we expect that more studies of neurodegenerative diseases will make use of scRNA-seq either from patient or healthy controls. Currently, single cell samples are collected from selected brain areas to understand the regional diversity of cells. Whole brain single cell analysis is not yet feasible as an adult human brain consists of 100 billion cells which will generate an enormous amount of data. Most of the data may not even contain any useful information for which complex methods are needed to provide useful insights. Although, single cell analyses suffers from sparsity and low-throughput, increasing interest and research efforts may help to improve the sequencing throughput of single cell analysis in the near future.

## 7.8 Spatial transcriptomics

In our studies, we analyzed spatial transcriptomics across the whole brain, however small-scale spatial expression patterns can provide more detailed analysis of smaller tissues. Emerging technologies in the field of spatial transcriptomics extend on microscopy methods to study single cells *in situ* while retaining the spatial context [31–33]. Using these

visual methods enables profiling RNA while it is in the tissue to capture spatial heterogeneity of smaller pieces of tissue, for example the transcriptomic differences between cortical layers. For this purpose, a tissue is cut into sections and slices, stained for a specific transcript using fluorescent hybridization probes, and visualized under a microscope to localize and quantify gene expression. Combinatorial approaches enable the analysis of multiple gene transcripts on one piece of tissue, but hybridization approaches are still limited in the number of genes; now up to 10,000 genes can be measured on a single tissue [34].Although scRNA-seq can measure gene expression over the whole genome, the spatial information is lost when single cells are dissected. To overcome the limitations of both scRNA-seq and hybridization technologies in spatial transcriptomics, new technologies have been developed to transfer RNA from tissue sections onto a surface with DNA barcoded beads for spatial indexing and genome-wide analysis using scRNA-seq [35,36]. Moreover, computational methods have been developed to infer the spatial locations of dissociated scRNA-seq samples by using information from complementary *in situ* hybridization data [37–39]. Spatial transcriptomics is a fast moving field and continuous efforts provide hope for the future to have a spatial gene expression atlas of the human brain at the single cell level. Having such information available can help to elucidate brain-wide as well as local molecular mechanisms and is a promising direction to study cellular differences in neurodegenerative diseases and health.

### 7.9 Common mechanisms in neurodegenerative diseases and healthy aging

Each neurodegenerative disease is characterized by different associated symptoms and pathology. But next to the degeneration of neuronal cells, there are also many common mechanisms between neurodegenerative diseases. Multiple studies, including our own, pointed towards similar functions that are disrupted in neurodegenerative diseases, such as ubiquitination, oxidative stress, and mitochondrial dysfunction. Similar deficits in brain functions are thought to underlie impaired movement and cognition in HD and PD. One thing that is common between our studies in HD and PD and other studies of neurodegenerative diseases is that resulting genes are generally associated with DNA repair mechanisms and protein degradation pathways. There is also an overlap in symptoms and the pathology between neurodegenerative diseases, e.g., Lewy bodies have also been found in patients with Alzheimer's disease, and tau and β-amyloid inclusion can also appear in PD. In addition, there are co-occurring diseases and deficits such as dementia, mental disorders, and cognitive impairments. The term parkinsonism or parkinsonian syndrome describes the combination of symptoms of PD that may also occur in other diseases. The fact that there is no clear boundary between neurodegenerative diseases, also explains why there are different forms of PD with different symptoms. Neurodegenerative diseases may have different causes, but the functional organization of the brain may eventually be disrupted in similar ways. Future studies should focus on both common mechanisms and differences in neurodegenerative and neurological disorders. To do so, studies may rely on meta-analysis approaches to combine the analysis of multiple cohorts. Ideally, samples should also be collected from both hemispheres, as asymptotic symptoms seem to be apparent in all neurodegenerative diseases.

Neurodegeneration also occurs during healthy aging and the differences between normal aging and neurodegenerative diseases are not fully understood. To further unravel the cause of neurodegenerative diseases, future studies may focus on analyzing spatial gene expression in brains considered in extremely opposite conditions. More research is being conducted into differences between extreme healthy elderly and patients with neurodegenerative disease which could better expose genes with a higher significance of differential expression. Samples from patients are often derived in the late stage of the disease, while PD and HD can also occur at early ages. This is because early symptoms are harder to distinguish from normal aging as the decline in motor functions is still relatively small. Case-control studies only allow for binary outcomes, but multi-cohort studies may as well focus on analyzing the gene expression differences in ordinal groups of individuals: early onset, late onset, healthy elderly, and extreme healthy elderly. Ordinal analysis approaches may be used for this purpose. While groups should be clearly defined, patients with multiple conditions that are considered in between disease boundaries should also be included as long as samples are well annotated. This will help to better understand the wide spectrum of PD cases.

## 7.10 Conclusion and future outlook

We showed that the AHBA is useful for analyzing the transcriptome of vulnerable brain regions in neurodegenerative diseases. How these vulnerable brain regions are defined depends on what is currently known about the disease pathology and remains a debatable line of topic. Diagnostic gold standards for PD remain an issue and there is a need to extend our knowledge about the pathology to be able to intervene early in the course of the disease. Future studies will make more use of multi-omics data to reveal better insights into the different molecular mechanism, such as the integration of transcriptomics with proteomics and epigenomics. This holds promise for the future to have such technologies and methods available. Large projects like these require collaborations with experts from different fields of neurobiology and computational biology. To make this possible scientists have to work in close collaboration and have to be multidisciplinary to allow for effective communication between different fields. Furthermore, emerging technologies such as scRNA-seq and machine learning will take its place in PD research. By analyzing the spatial transcriptomics of the healthy brain we revealed known and new genes that may be involved in neurodegenerative diseases, but to confirm a relationship to PD or HD, results need to be validated in wet-lab experiments with samples that represent health and disease conditions. While post-mortem human tissues are scarce and animal models and cell lines of PD do not well translate to human PD, studies will make more use of brain organoids that are three-dimensional structures generated from iPSC. Finally, by combining spatial transcriptomics of the healthy brain with neuroimaging data, we revealed that molecular mechanisms such as mitochondrial function and cellular stress response may be involved in neurodegenerative diseases, but may also be essential for maintaining health and increased longevity.

## References


[1] M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, et al., "The human transcriptome across tissues and individuals," *Science* 348, 660–665 (2015).

[2] S. M. D. Henley, E. J. Wild, N. Z. Hobbs, R. I. Scahill, G. R. Ridgway, D. G. MacManus, R. A. Barker, N. C. Fox, and S. J. Tabrizi, "Relationship between CAG repeat length and brain volume in premanifest and early Huntington's disease," *J. Neurol.* 256, 203–212 (2009).

[3] H. H. Ruocco, L. Bonilha, L. M. Li, I. Lopes-Cendes, and F. Cendes, "Longitudinal analysis of regional grey matter loss in Huntington disease: Effects of the length of the expanded CAG repeat," *J. Neurol. Neurosurg. Psychiatry* 79, 130–135 (2008).

[4] W. A. Haynes, A. Tomczak, and P. Khatri, "Gene annotation bias impedes biomedical research," *Sci. Rep.* 8, 1–7 (2018).

[5] C. G. Chung, H. Lee, and S. B. Lee, "Mechanisms of protein toxicity in neurodegenerative diseases," *Cell. Mol. Life Sci.* 75, 3159–3180 (2018).

[6] P. M. Visscher, N. R. Wray, Q. Zhang, P. Sklar, M. I. Mccarthy, M. A. Brown, and J. Yang, "10 Years of GWAS discovery: Biology, function, and translation," *Am. J. Hum. Genet.* 101, 5–22 (2017).

[7] M. A. Nalls, J. Bras, D. G. Hernandez, M. F. Keller, E. Majounie, A. E. Renton, M. Saad, I. Jansen, R. Guerreiro, et al., "NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases," *NBA* 36, 1605.e7-1605.e12 (2015).

[8] J. Höglund, N. Rafati, M. Rask-Andersen, S. Enroth, T. Karlsson, W. E. Ek, and Å. Johansson, "Improved power and precision with whole genome sequencing data in genome-wide association studies of inflammatory biomarkers," *Sci. Rep.* 9, 1–14 (2019).

[9] C. Blauwendraat, M. A. Nalls, and A. B. Singleton, "The genetic architecture of Parkinson's disease," *Lancet Neurol.* 19, 170–178 (2020).

[10] L. Ibanez, F. H. G. Farias, U. Dube, K. A. Mihindukulasuriya, and O. Harari, "Polygenic risk scores in neurodegenerative diseases: a review," *Curr. Genet. Med. Rep.* 7, 22–29 (2019).

[11] E. A. Boyle, Y. I. Li, and J. K. Pritchard, "An expanded view of complex traits: From polygenic to omnigenic," *Cell* 169, 1177–1186 (2017).

[12] L. Song, P. Langfelder, and S. Horvath, "Comparison of co-expression measures: Mutual information, correlation, and model based indices," *BMC Bioinformatics* 13 (2012).

[13] P. E. Meyer, F. Lafitte, and G. Bontempi, "Minet: A R/bioconductor package for inferring large transcriptional networks using mutual information," *BMC Bioinformatics* 9, 461 (2008).

[14] S. Kumari, J. Nie, H. S. Chen, H. Ma, R. Stewart, X. Li, M. Z. Lu, W. M. Taylor, and H. Wei, "Evaluation of gene association methods for coexpression network construction and biological knowledge discovery," *PLoS One* 7, e50411 (2012).

[15] F. Liesecke, D. Daudu, R. D. De Bernonville, S. Besseau, M. Clastre, V. Courdavault, J. O. De Craene, J. Crèche, N. Giglioli-Guivarc'h, et al., "Ranking genome-wide correlation measurements improves microarray and RNA-seq based global and targeted co-expression networks," *Sci. Rep.* 8, 1–16 (2018).

[16] F. Liesecke, J. O. De Craene, S. Besseau, V. Courdavault, M. Clastre, V. Vergès, N. Papon, N. Giglioli-Guivarc'h, G. Glévarec, et al., "Improved gene co-expression network quality through expression dataset down-sampling and network aggregation," *Sci. Rep.* 9, 1–16 (2019).

[17] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics* 9, 1–13 (2008).

[18] A. D. Broido and A. Clauset, "Scale-free networks are rare," *Nat. Commun.* 10, 1–10 (2019).

[19] F. Avila Cobos, J. Vandesompele, P. Mestdagh, and K. De Preter, "Computational deconvolution of transcriptomics data from mixed cell populations," *Bioinformatics* 34, 1969–1979 (2018).

[20] R. D. Hodge, T. E. Bakken, J. A. Miller, K. A. Smith, E. R. Barkan, L. T. Graybuck, J. L. Close, B. Long, N. Johansen, et al., "Conserved cell types with divergent features in human versus mouse cortex," *Nature* 573, 61–68 (2019).

[21] Y. Cao, Y. Lin, J. T. Ormerod, P. Yang, J. Y. H. Yang, and K. K. Lo, "scDC: single cell differential composition analysis," *BMC Bioinformatics* 20, 721 (2019).

[22] X. Wang, J. Park, K. Susztak, N. R. Zhang, and M. Li, "Bulk tissue cell type deconvolution with multi-subject single-cell expression reference," *Nat. Commun.* 10, 1–9 (2019).

[23] D. Tsoucas, R. Dong, H. Chen, Q. Zhu, G. Guo, and G. Yuan, "Accurate estimation of cell-type composition from gene expression data," *Nat. Commun.* 10, 1–9 (2019).

[24] R. Du, V. Carey, and S. T. Weiss, "deconvSeq: deconvolution of cell mixture distribution in sequencing data," *Bioinformatics*, 35, 5095–5102 (2019).

[25] B. Jew, M. Alvarez, E. Rahmani, Z. Miao, A. Ko, J. H. Sul, K. H. Pietiläinen, P. Pajukanta, E. Halperin, et al., "Accurate estimation of cell composition in bulk expression through robust integration of single-cell information," *Nat. Commun.* 11, 1–11 (2020).

[26] F. A. Cobos, J. Alquicira-hernandez, and J. Powell, "Comprehensive benchmarking of computational deconvolution of transcriptomics data," *bioRxiv* (2020).

[27] S. Darmanis, S. A. Sloan, Y. Zhang, M. Enge, C. Caneda, L. M. Shuer, M. G. Hayden Gephart, B. A. Barres, S. R. Quake, et al., "A survey of human brain transcriptome diversity at the single cell level," *Proc. Natl. Acad. Sci.* 112, 7285–7290 (2015).

[28] B. Tasic, V. Menon, T. N. Nguyen, T. K. Kim, T. Jarsky, Z. Yao, B. Levi, L. T. Gray, S. A. Sorensen, et al., "Adult mouse cortical cell taxonomy revealed by single cell transcriptomics," *Nat. Neurosci.* 19, 335–346 (2016).

[29] T. Abdelaal, L. Michielsen, D. Cats, D. Hoogduin, H. Mei, M. J. T. Reinders, and A. Mahfouz, "A comparison of automatic cell identification methods for single-cell RNA-sequencing data," *bioRxiv*, 644435 (2019).

[30] N. G. Skene and S. G. N. Grant, "Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment," *Front. Neurosci.* 10, 1–11 (2016).

[31] S. Shah, E. Lubeck, W. Zhou, and L. Cai, "In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus Sheel," *Neuron* 92, 342–357 (2016).

[32] J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, T. C. Ferrante, R. Terry, B. M. Turczyk, J. L. Yang, H. S. Lee, et al., "Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues," *Nat. Protoc.* 10, 442–458 (2015).

[33] K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang, "Spatially resolved, highly multiplexed RNA profiling in single cells," *Science* 348, 1360–1363 (2015).

[34] C. H. L. Eng, M. Lawson, Q. Zhu, R. Dries, N. Koulena, Y. Takei, J. Yun, C. Cronin, C. Karp, et al., "Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+," *Nature* 568, 235–239 (2019).

[35] S. G. Rodriques, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, et al., "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution," *Science* 363, 1463–1467 (2019).

[36] S. Vickovic, G. Eraslan, F. Salmén, J. Klughammer, L. Stenbeck, D. Schapiro, T. Äijö, R. Bonneau, L. Bergenstråhle, et al., "High-definition spatial transcriptomics for in situ tissue profiling," *Nat. Methods* 16, 987–990 (2019).

[37] R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev, "Spatial reconstruction of single-cell gene expression data," *Nat. Biotechnol.* 33, 495–502 (2015).

[38] M. Nitzan, N. Karaiskos, N. Friedman, and N. Rajewsky, "Charting a tissue from single-cell transcriptomes," *bioRxiv* (2018).

[39] M. Nitzan, N. Karaiskos, N. Friedman, and N. Rajewsky, "Gene expression cartography," *Nature* 576, 132–137 (2019).