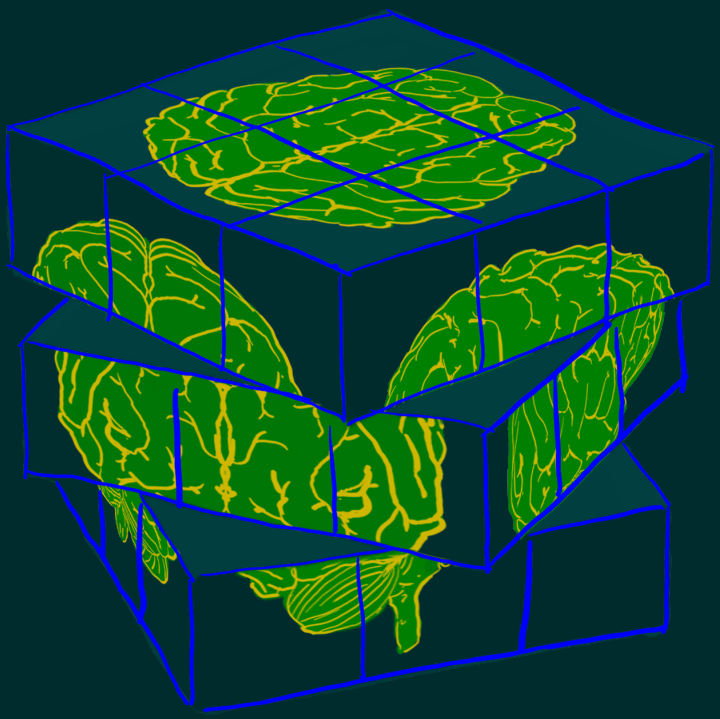
Analyzing spatial transcriptomics and neuroimaging data in neurodegenerative diseases

Keo, D.

ANALYZING SPATIAL TRANSCRIPTOMICS AND NEUROIMAGING DATA IN NEURODEGENERATIVE DISEASES



ARLIN KEO

Analyzing spatial transcriptomics and neuroimaging data in neurodegenerative diseases

Proefschrift

Ter verkrijging van de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 3 december 2020
klokke 11:15 uur

door

Dar-Lin Keo-den Boer

geboren te Oss
in 1988

Promotor:	Prof. dr. ir. B. P. F. Lelieveldt
Co-promotor:	Prof. dr. ir. M. J. T. Reinders Dr. ir. A. M. E. T. A. Mahfouz
Leden promotiecommissie:	Prof. dr. ir. O. C. Meijer Prof. dr. ir. M. P. van den Heuvel *Amsterdam University Medical Center* Prof. dr. ir. E. M. Hol *University Medical Center Utrecht*

Analyzing spatial transcriptomics and neuroimaging data in neurodegenerative diseases

Dar-Lin Keo-den Boer

ISBN: 978-94-6423-044-4

This work was performed at the Leiden Computational Biology Center, Leiden University Medical Center, the Netherlands.

Printing of this thesis was kindly supported by Bontius Stichting and Parkinson Vereniging.

Contents

Chapter 1
Introduction

1.1 MOLECULAR MECHANISMS OF NEURODEGENERATIVE DISEASES

1.1.1 NEURODEGENERATIVE DISEASES

The life expectancy of the global population is expected to keep increasing, especially in the western world where people live longer due to better health care and improved living conditions [1]. The downside is that increased age comes with age-related neurodegenerative diseases [2] that are characterized by the progressive loss of neurons in the central nervous system, eventually leading to deficits in brain functions affecting movement and cognition [3]. Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease (AD) and affects 1% of the population at the age of 65 years [4]. Recently, it has been reported that PD is the fastest growing neurological disorder in the world and the number of cases is expected to keep increasing [5]. PD is characterized by the loss of dopaminergic neurons in the substantia nigra leading to motor dysfunction e.g., slowness, and rigidity. Currently available treatments help to control motor symptoms, but cannot slow down disease progression [6]. It is well-recognized that alterations within the basal ganglia neuronal circuit play a role in the etiology of PD, but the basal ganglia is also involved in the etiology of Huntington's disease (HD) [7,8]. Due to cell loss in the basal ganglia, both neurodegenerative diseases may be accompanied by neuropsychiatric symptoms including depression and schizophrenia, which drastically impacts the quality of life [9,10]. Computational tools enable the analysis of large biological datasets to better understand the molecular and clinical traits of neurodegenerative diseases. In this thesis, we mainly focused our research on PD, but we also studied HD and the risk to develop schizophrenia. Our research findings can aid to provide new insights into the etiology of PD and other neurodegenerative diseases with an outlook to improved treatment and preferably the early prevention of PD.

1.1.2 AGING VERSUS NEURODEGENERATION

Healthy aging and neurodegenerative diseases are generally associated with a decline in resistance to cellular stress that may lead to cell death affecting brain function [11]. Initially, PD patients show similar rates of dopaminergic cell loss as to healthy individuals, but as time progresses neurodegeneration is accelerated and the first signs of motor symptoms start to appear (Figure 1.1) [12]. The pre-motor stage is associated with mild motor symptoms, but non-motor features may also occur early in the course of the disease, e.g., symptoms of pain, depression, cognitive dysfunction, dementia, sleep disturbance, and constipation [13,14]. In the early stage of PD, symptoms are often confused with signs of aging and may therefore go unnoticed. As the disease progresses and symptoms start to impair daily activities, PD patients are diagnosed based on clinical assessment scores for motor and non-motor symptoms [15,16]. Generally, when a patient is first diagnosed, a substantial proportion of dopaminergic neurons in the substantia nigra has already been lost, and neurodegeneration has also spread to other regions of the central nervous system [6]. Findings over the past years have revealed that PD is a complex disease with some clinical challenges, for example the inability to make a definitive diagnosis at the early stages of PD and difficulties in managing the symptoms at later stages [17]. In general, the

pathogenesis of PD is multifactorial and results from a combination of aging, and genetic and environmental factors [18].

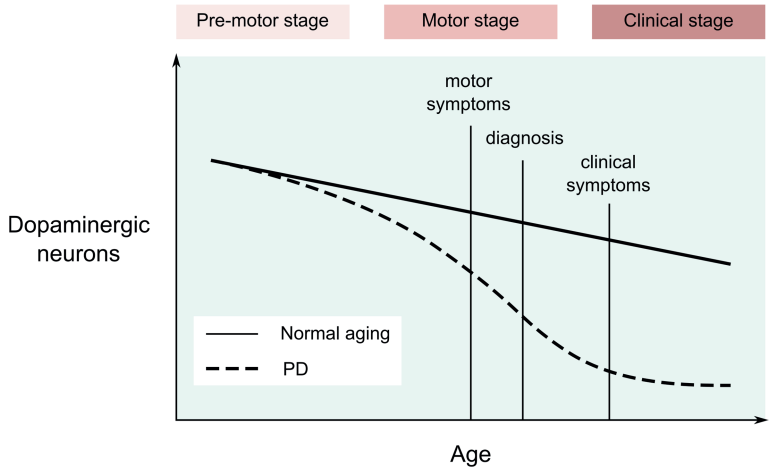


Figure 1.1 Schematic representation of the progression of the level of dopaminergic neurons during normal aging versus aging in Parkinson's disease (PD).

1.1.3 NEUROPATHOLOGY AND CAUSE OF PARKINSON'S DISEASE

PD diagnosis is essentially clinical but has to be confirmed upon post-mortem neuropathological examination by confirming the loss of dopaminergic neurons in the substantia nigra and the presence of protein inclusions called Lewy bodies (LBs) [19]. The Braak staging scheme has been proposed to assess the staging of the brain pathology in PD for neuropathological diagnosis [20]. According to Braak's theory, LBs first appear in the brainstem and then spread to selective brain regions during the course of the disease, manifesting in motor and non-motor symptoms (Figure 1.2). PD comprises a spectrum of disorders with different subtypes, for which most also share the LB pathology [21]. LBs are composed mainly of α-synuclein proteins encoded by the gene *SNCA* that plays a regulatory role in dopamine homeostasis [22]. There are also PD cases where LBs are not found in the brain. These patients were usually early-onset (before the age of 40 years), slow progressing, and neuronal loss was mostly restricted to the substantia nigra [23]. Aside from PD, there are a number of other atypical parkinsonian syndromes (e.g., dementia with Lewy bodies, multiple system atrophy, progressive supranuclear palsy, and corticobasal degeneration) that show abnormal protein depositions of α-synuclein or tau together with parkinsonism; a combination of PD-related symptoms, namely bradykinesia with rigidity, tremor, or postural instability [24]. Point mutations in the *SNCA* gene as well as duplication and triplication variants have been associated with familial PD, but how these genetic variants results in LB pathology remains unclear [25]. It has been suggested that the spreading of LB pathology seems to occur across anatomical brain networks and is mostly explained by α-synuclein expression [26]. Mutations in *SNCA* are involved in both familial and sporadic PD, but it is unclear whether *SNCA* has a causative role or it is just a bystander of disease

mechanisms [21]. Other genetic mutations in more than 20 genes have been identified in familial and population studies of PD patients, including rare monogenic causes [27–30], however only 10% of PD cases is ascribed to monogenic mutations [31]. The heritability of risk to develop PD is 60% which can be partly explained by pathogenic mutations and the major risk genes, but the remaining 40% is due to yet unidentified genes [32].

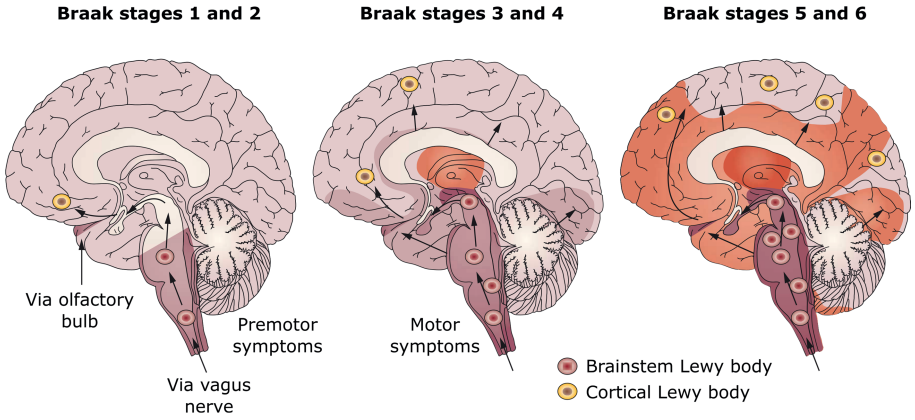


Figure 1.2 Braak staging scheme of Parkinson's disease. The six Braak stages describe the spread of Lewy bodies from the brainstem to limbic and cortical regions and are associated with the progression of clinical symptoms. Image adapted from Doty (2012) [33].

1.1.4 GENETIC CAUSE OF HUNTINGTON'S DISEASE

HD may serve as a model to study neurodegenerative diseases as it is monogenic, dominantly inherited, and also a disorder of protein misfolding [34]. The striatum is most severely affected in HD, but other brain areas also show neuronal loss [35]. Interestingly, HD and PD are characterized by opposite movement disorders: slowness and rigidity are symptoms observed in PD, while excessive and uncontrolled movements are typical in HD [36]. The nigrostriatal circuit, connecting the striatum and substantia nigra, plays a role in the initiation or termination of voluntary movements and is impaired in disorders like PD and HD (Figure 1.3) [37]. The cause of HD is a CAG-trinucleotide repeat expansion in the *HTT* gene coding for polyglutamine (polyQ) amino acids in the huntingtin protein and leads to improper folding of mutant huntingtin proteins (Figure 1.4) [38]. The CAG-trinucleotide normally occurs with up to 35 repeats in tandem, while an increased number of repeats of 40 or more causes HD [39]. Interestingly, the size of the repeat expansion has a strong negative correlation with the age at onset of the disease [40] and the expanded CAG-repeat length in other genes also determines the age at onset of a wide spectrum of other neurodegenerative diseases including multiple spinocerebellar ataxias (SCAs), dentatorubralpallidoluysian atrophy (DRPLA), and spinobulbar muscular atrophy (SBMA) which are also characterized by progressive motor and cognitive deficits [41]. Furthermore, it has been shown that the age of onset in HD is also influenced by the CAG-repeat variations in other genes associated with polyQ diseases suggesting that there are interactions between polyQ genes [42]. Defects in DNA repair mechanisms are thought to underlie the age at onset of HD and SCAs

[43]. In addition, movement disorders that appear in PD and HD, including parkinsonism and chorea, are also frequent in SCAs [44]. This suggests that PD and polyQ diseases share common disease mechanisms and that genetic factors can influence disease onset.

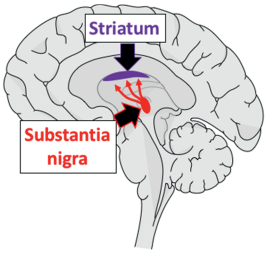


Figure 1.3 The nigrostriatal pathway is one of the four dopamine pathways and connects the striatum to the substantia nigra. Image credit: Pharmwiki [45].

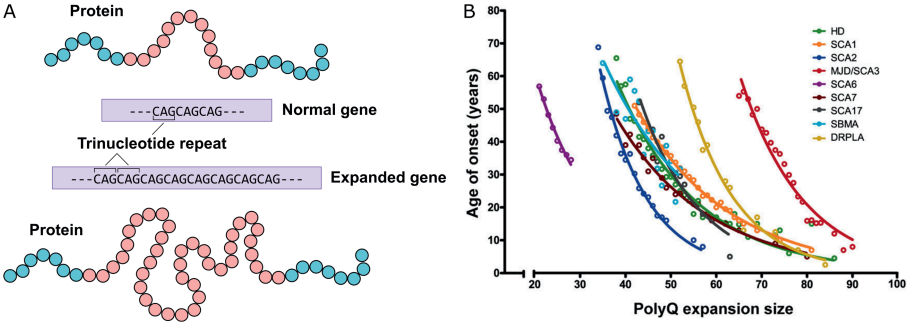


Figure 1.4 CAG-trinucleotide repeat expansions in polyglutamine (polyQ) diseases. (A) CAG-repeats are translated to polyglutamine amino acids, but an expansion in the number of repeats causes the misfolding of proteins. (B) Age at onset is strongly influenced by the CAG-repeat size in the causative gene of nine polyQ diseases. Circles depict the mean age at onset for a given expansion size based on multiple cohorts of patients. Image adapted from Kuiper et al. (2017) [46].

1.2 SPATIAL TRANSCRIPTOMICS TO STUDY NEURODEGENERATIVE DISEASES

1.2.1 TRANSCRIPTOMICS OF PARKINSON'S DISEASE

Although several genetic and environmental factors have been associated with the development of PD, the underlying pathobiology remains poorly understood. Transcriptomics can yield new insights into the disease mechanism by measuring genome-wide expression profiles of samples derived from brain tissue. The whole human genome is estimated to have ~20,000 protein encoding genes for which the amount of RNA transcripts per gene can be measured with microarray chips or RNA-sequencing techniques. The goal of most transcriptomic studies in PD is to find genes that are differentially expressed between PD patients and age-matched controls and assess whether they are enriched for genes associated with known biological pathways. The majority of these studies analyzed samples from the substantia nigra; a structure of the basal ganglia that shows loss of

dopaminergic neurons in PD. Transcriptomic studies of PD have been limited by the small number of samples (less than 40 PD cases) due to difficulties in sample selection, namely the limited availability of well-characterized patients, and cases and controls that should be well matched based on age, ethnicity, and gender [47]. To overcome this limitation, animal models have been developed to mimic the clinical symptoms of PD, however it remains unclear to what extent the underlying gene expression patterns reflect those in human PD or how well these animal models mimic phenotypes observed in PD [48,49]. Gene expression profiles have also been retrieved from blood samples of patients, however these samples do not reflect the transcriptomic signatures of the brain [47,50–52]. Furthermore, researchers have used induced pluripotent stem cells (iPSC) from patients to generate specific neuronal cell-types, such as dopaminergic cells, but iPSC technology is still challenging and poorly understood by scientists [53,54]. Another limitation of transcriptomic studies of PD is that brain samples are derived from patients in the late stage of the disease and therefore they cannot investigate the early changes that lead to the development of PD [55]. Improving our understanding of PD in its prodromal phase during midlife is essential to enable early stage PD diagnosis [14]. Finally, transcriptomic data from PD brains that are currently available do not cover many brain regions, and there are only few PD studies that analyzed multiple brain regions from the same donors [56]. Taken together, more extensive brain-wide gene expression analyses can help to understand the spatial context of gene expression and their functions that may be dysregulated in neurodegenerative diseases.

1.2.2 Transcriptome map of the healthy brain

Spatial transcriptomics is a method where gene expression is analyzed in a spatial context to understand how gene transcription activity is organized across a tissue or organ, such as the brain. Generally, it seems that gene expression varies more between brain regions than between individuals [57]. The Allen Human Brain Atlas (AHBA) is a high resolution gene expression atlas covering the whole brain [58]. It was established by collaborative efforts of multi-disciplinary scientists and made publicly available for neurobiological research. Almost 4,000 samples were collected from post-mortem brains of six healthy adult donors without a neurological or neuropsychiatric history (Figure 1.5). To map brain-wide transcriptomic profiles, brains were carefully dissected while keeping track of anatomical annotations and spatial coordinates based on magnetic resonance imaging (MRI) scans. This brain-wide gene expression dataset helps to unravel functions of different brain structures and their spatial organization. Since its publication the AHBA has been used extensively to yield new insights into diseases of the central nervous system such as migraine [59], schizophrenia [60,61], and AD [62–65]. By analyzing the transcriptome of brain regions that are vulnerable in disease, studies were able to establish the expression patterns of disease-implicated genes and provide new insights into disease mechanisms. In studies of neurodegenerative diseases, brain regions of interest were defined based on the vulnerability to neurodegeneration and/or protein aggregation from imaging studies or pathological findings. Altogether, the AHBA has offered new opportunities to link molecular function to brain organization and defects in neurological diseases and can also be of great value to unravel the molecular mechanisms underlying vulnerable brain regions in PD or HD.

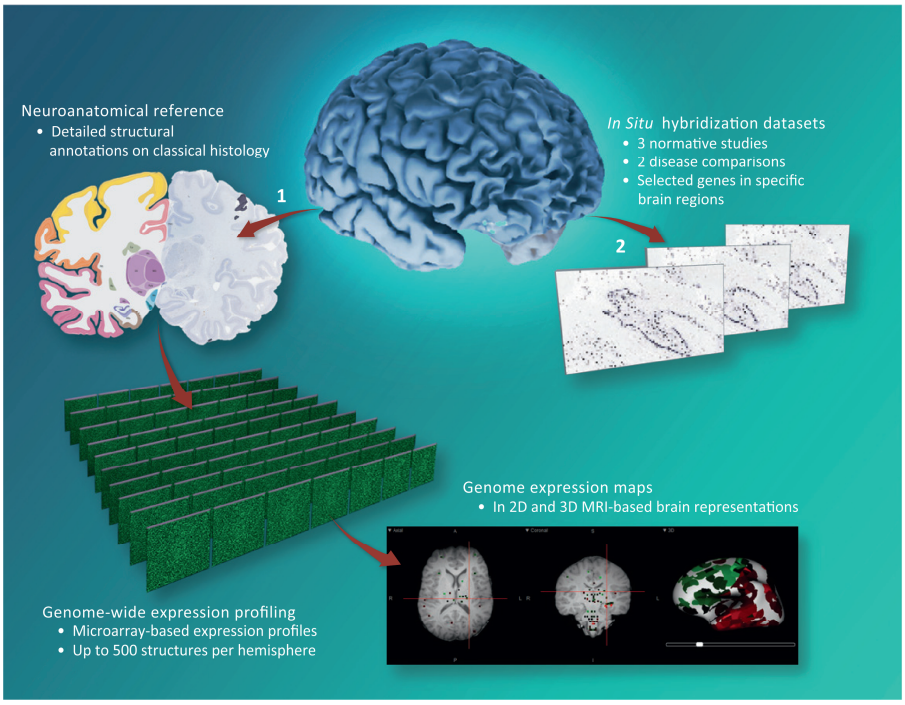


Figure 1.5 Allen Human Brain Atlas. Gene expression map of the healthy human brain. Image credit: Shen et al. (2012) [66].

1.3 Computational approaches in integrating genomics with neuroimaging

1.3.1 Phenotypic information from neuroimaging

Neuroimaging is a method to study functional neuroanatomy, brain-behavior relationships and the pathophysiology of brain disorders [67], and thus provides useful phenotypic features of the brain. It allows detecting differences in imaging features between a group of patients and controls in vivo to assess, for example, the loss of gray matter in PD brains. Differences between disease cases and controls are statistically assessed to highlight brain areas with significant structural or functional changes associated with the disease. The two main imaging techniques are structural MRI, that measures morphological properties of the brain (e.g., cortical thickness and volume), and functional MRI (fMRI), that measures blood oxygen flow to identify functionally connected brain regions based on their activity during task-based performance or in resting state. In PD imaging studies, the region of interest often includes anatomical structures of the basal ganglia, nigrostriatal projections, the cortex and other regions involved in the motor circuit [68,69]. Some MRI studies did not focus only on specific anatomical structures, but analyzed the entire brain to unbiasedly identify new brain regions vulnerable to atrophy in PD [70]. Structural covariance is an MRI method to identify anatomical networks of the brain based on gray matter variation across a population of individuals [71]. The co-varying regions in structural covariance networks

(SCNs) are thought to arise from functionally connected brain regions. Both functional and structural networks define a set of connected brain regions that can be compared between PD and controls to find associations with gray matter loss that is indicative of atrophy within such networks. Interestingly, several PD imaging studies highlighted relationships between structural or functional MRI changes in PD and the severity of clinical symptoms [72–78]. In addition, neuroimaging genetics studies analyzed MRI differences between groups of patients with different mutations associated with subtypes of PD [79–81] or AD [82]. It would be interesting to combine spatial transcriptomics with neuroimaging of PD brains to reveal deeper insights into the molecular mechanisms; however transcriptomic studies of PD have been limited to few selected brain regions.

1.3.2 NEUROIMAGING AND TRANSCRIPTOMICS TO STUDY NEURODEGENERATIVE DISEASES

Analyzing the transcriptional activity of structural or functional brain networks could reveal new insights into the molecular organization of the brain. Single photon emission computed tomography (SPECT) and positron emission tomography (PET) allow whole brain imaging of molecular changes in patients in vivo, but are limited to the analysis of a single receptor, transporter, or enzyme [83]. The AHBA provides a whole-genome expression map of the healthy brain that can be combined with neuroimaging data to correlate transcriptomic signatures with phenotypic features of the brain. Each sample in the AHBA is annotated with spatial MNI coordinates that can be used to map the gene expression samples to brain regions defined by MRI, for example to determine which samples fall within or outside a region of interest. In recent years, studies have used varying approaches to combine neuroimaging data with the AHBA, but the processing of gene expression data has been highly variable [84]. In general, the AHBA is used to examine gene expression profiles across brain regions or networks that were identified in neuroimaging studies to find correlations with imaging features, such as network function or gray matter loss. In studies of neurological disorders, genes may behave differently across disease-related regions compared to control regions that are considered unaffected. The control region can be the whole brain, or a region known to not be involved in the disease. The differential activity of these genes that are correlated with imaging features inform about the cellular mechanisms that may be involved in disease. Gene co-expression analysis is a powerful robust method to understand transcriptomic organization [85]. Co-expression of samples across genes has revealed a network of samples that mimicked a network detected with fMRI [86]. Alternatively, the correlation between genes across a set of samples measures whether two genes behave similarly. This information can be used to create a network of putative gene interactions. Commonly, gene co-expression patterns have been analyzed in samples from specific tissues across individuals, but can also be examined spatially across samples from different brain regions. Analyzing spatial gene co-expression patterns in selected brain regions informs about the spatial organization of molecular pathways and on how gene interactions are regulated differently in different brain regions. Based on their co-expression similarity, genes may be clustered into groups referred to as co-expression modules. Whole brain analysis of both imaging features and gene expression patterns allows finding correlations between two data modalities [87–89]. Previous studies that combined neuroimaging data with spatial transcriptomics have shown that there are

multiple ways to analyze gene expression patterns in brain regions defined by imaging technologies (Figure 1.6). In other words, analyzing the healthy transcriptome of brain regions associated with neurodegenerative diseases have shown to improve our understanding of genes and their role in disease pathology.

1.3.3 Functional interpretation of findings

Genes resulting from computational analyses need to be investigated for the presence of interesting genes for which a functional role has been earlier described. Functional enrichment analysis assesses whether there is an overrepresentation of groups of genes associated with specific functions, pathways, diseases, or cell-types. There exists multiple curated databases with gene annotations that can be used for gene set enrichment analysis (the Gene Ontology resource [90,91], Reactome pathway database [92], KEGG [93], DisGENet [94], etc.) to aid in the functional interpretation of transcriptomic findings. Gene markers that are known to be uniquely expressed in specific cell-types can be used to identify the presence of cell-types. Immunohistochemical staining of cell-type markers is extensively used to reveal the presence and abundance of specific cell-types in brain slices, such as staining for *TH* (tyrosine hydroxylase) to reveal the loss of dopaminergic cells in PD [95]. In neurodegenerative diseases like PD and HD there is remarkable loss of neuronal cells compared to healthy individuals [96]. Interestingly, cell-types associated with immune responses (e.g., astrocytes, microglia) are more abundant in the neurodegenerative brain. Therefore, studies that perform transcriptomic analysis of bulk tissue samples between patients and controls have to be aware of the different cell populations within samples. Also, when comparing tissues from different brain regions, an overrepresentation of cell-type markers may reflect the anatomical organization of brain regions, for example gene markers for neuronal cells are highly expressed in the synapse-dense cortex. The cell-type composition has, with no doubt, an influence on the measured gene expression profile, but normally the true cell-type compositions of bulk tissue samples are unknown. Several methods have been proposed that use cell-type markers to estimate cell-type composition and correct differential expression analysis results based on these estimations [97–100].

While there are many genetic risk factors associated with PD, their functional role in the development of PD remains unclear. Several studies that combined neuroimaging data with the AHBA focused on the brain-wide expression of genetic risk factors and revealed regional gene activity that was correlated with imaging features in PD [101,102]. The expression pattern of genetic risk factors can aid to better understand how genetic mutations lead to molecular deficits in affected brain regions [103]. While genetic risk factors provide a direct link to the disease, whole transcriptomic analysis reveals more interesting genes that can explain the molecular mechanisms underlying neurodegenerative diseases. Previous studies have shown the various possibilities to analyze neuroimaging data with brain-wide transcriptomic data, and provide examples of how to use similar computational tools to unravel molecular mechanisms underlying brain regions and networks that are vulnerable to PD and HD.

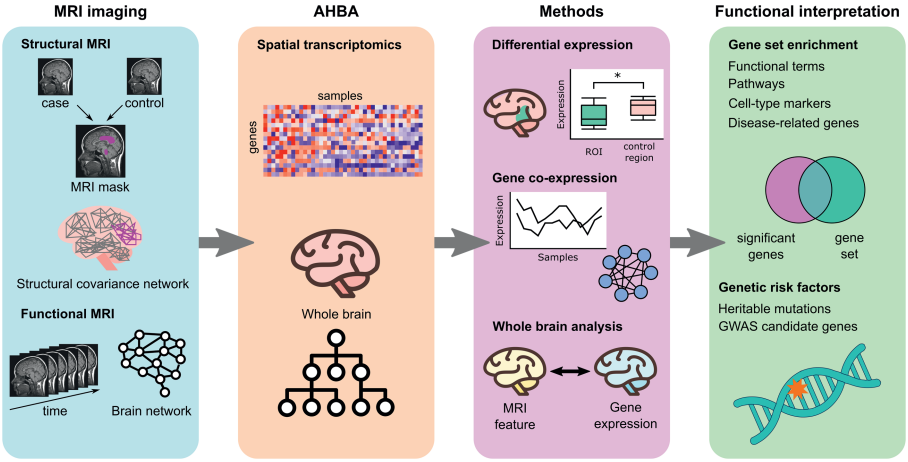


Figure 1.6 An overview of computational approaches to combine neuroimaging data with spatial transcriptomics from the AHBA to study neurodegenerative diseases. Neuroimaging provides phenotypic information and can be used to create structural or functional networks and find changes between disease cases and controls. Combining neuroimaging data with spatial transcriptomics enables the analysis of gene expression in brain regions that are vulnerable in disease. Brain samples from the AHBA were collected according to a hierarchical ontology. The whole brain is on top of the hierarchy and splits into more detailed brain structures. Common methods to analyze neuroimaging with spatial transcriptomics are differential gene expression in regions of interest (MRI mask), gene co-expression across selected samples, and whole brain analysis of both data modalities using an integrated approach. For functional interpretation, findings are assessed for the overrepresentation of known gene sets and/or genetic risk factors associated with the disease.

1.4 CONTRIBUTION OF THESIS

This thesis describes our research on exploring the relationship between the gene expression profile of different brain regions and age-related neurodegenerative disorders. We mainly focused our research on PD, but also used similar computational tools to understand molecular mechanism underlying HD and polyQ genes, as well as schizophrenia. Our main hypothesis is that certain genes are important to maintain normal molecular functions in the healthy state, while dysfunctioning of these genes may underlie the vulnerability to neurodegeneration in specific brain regions. We sought to answer multiple research questions to detect such genes and consequently improve our understanding of PD, and other neurodegenerative diseases:

- Which genes are differentially expressed in brain regions associated with disease?
- Which biological functions or pathways are associated with these genes?
- How is the spatial expression of genetic risk factors involved in disease?
- Which molecular interactions are associated with specific brain regions?

We used and developed various bioinformatics tools to combine neuroimaging data with spatial transcriptomic data to reveal genes of interest and its associated functions that may

be involved in the pathology of neurodegenerative diseases. In each chapter, we exploited the AHBA to understand gene expression patterns across the healthy brain as a guide to better understand the molecular mechanisms underlying neurodegenerative diseases. With this approach, we rely on previous studies that have shown that spatial gene expression patterns can reveal new insights into the functional organization of the brain, and how dysfunctional regulation may lead to the development of neurodegenerative diseases.

In Chapter 2, we used spatial transcriptomics to study molecular functions of structural brain networks that may explain patterns of neurodegeneration in PD. Comparative analysis between the transcriptomics signatures of multiple SCNs obtained from MRI revealed the presence of cholinergic genes that may be associated with patterns of atrophy in PD.

In Chapter 3, we proposed that polyQ genes likely interact because CAG-repeat lengths within these genes strongly influences the age at onset of HD. We revealed co-expression between *HTT* and polyQ genes *ATN1* and *ATXN2* in anatomical structures involved in HD and a region associated with neurodegeneration in HD based on MRI images of patients. The co-expression patterns suggest potential interactions between the three polyQ genes in brain regions that are vulnerable in HD and the shared co-expressed genes among the three polyQ genes suggests the involvement of DNA repair pathways.

In Chapter 4, we used similar computational tools to investigate the molecular mechanisms underlying a stress network defined by task-based fMRI in individuals at risk of schizophrenia. Genes differentially expressed within the network showed associations with stress-related psychiatric disorders, neuronal cell populations, and neurotransmitter receptors. Many of the differentially expressed genes have been described to interact with the HPA-axis, a neuro-endocrine system that controls hormonal stress.

In Chapter 5, we propose that brain-wide gene expression patterns can predict cortical atrophy in PD brains and possibly the relationship between cortical atrophy and the severity of clinical symptoms. In this study, we used a multimodal approach to understand the behavior of groups of genes instead of individual genes and their correlation with patterns of neurodegeneration. Regional cortical thickness changes in PD were correlated with pathways associated with the maintenance of cellular health. Without any assumptions on the involvement of genes and vulnerable brain regions in PD, our findings highlight the activity of biological pathways across the cortex that are correlated with cellular maintenance mechanisms that previously have been associated with PD.

In Chapter 6, we used the Braak staging scheme for PD to define a set of brain regions that are known to be progressively affected during the different clinical stages of PD. The gradual spreading of Lewy bodies might be explained by the underlying molecular mechanisms of the involved brain regions. We revealed genes, including *SNCA* and other genetic risk factors, for which their healthy expression pattern correlated with the Braak stages of brain regions. These correlation patterns were not preserved in Braak stage-

involved regions of PD patients. Additionally, we highlighted two co-expression modules involved in dopamine biosynthesis and blood-oxygen control whose expression signature was shown to be correlated with Braak stages. Understanding the transcriptomic differences between brain regions with different vulnerabilities to PD provided new insights into the progression of PD.

Each chapter helps to understand the spatial organization of gene transcription activity and how dysregulation of molecular processes may lead to symptoms apparent in neurodegenerative diseases. Finally, we discuss the contribution of our work to HD and PD research and our perspective on future research to improve our understanding of neurodegenerative diseases.

REFERENCES

[1] United Nations, Department of Economic and Social Affairs, and Population Division, "World Population Ageing 2015 (ST/ESA/SER.A/390)," 2015.

[2] A. Reeve, E. Simcox, and D. Turnbull, "Ageing and Parkinson's disease: Why is advancing age the biggest risk factor?," *Ageing Res. Rev.* 14, 19–30 (2014).

[3] H. M. Gao and J. S. Hong, "Why neurodegenerative diseases are progressive: uncontrolled inflammation drives disease progression," *Trends Immunol.* 29, 357–365 (2008).

[4] L. M. L. de Lau and M. M. B. Breteler, "Epidemiology of Parkinson's disease," *Lancet Neurol.* 5, 525–535 (2006).

[5] E. R. Dorsey, T. Sherer, M. S. Okun, and B. R. Bloem, "The emerging evidence of the Parkinson pandemic," *J. Parkinsons. Dis.* 8, S3–S8 (2018).

[6] R. Balestrino and A. H. V. Schapira, "Parkinson disease," *Eur. J. Neurol.* 27, 27–42 (2020).

[7] T. Macpherson and T. Hikida, "Role of basal ganglia neurocircuitry in the pathology of psychiatric disorders," *Psychiatry Clin. Neurosci.* 73, 289–301 (2019).

[8] J. A. Obeso, M. C. Rodriguez-Oroz, M. Stamelou, K. P. Bhatia, and D. J. Burn, "The expanding universe of disorders of the basal ganglia," *Lancet* 384, 523–531 (2014).

[9] C. M. Eddy, E. G. Parkinson, and H. E. Rickards, "Changes in mental state and behaviour in Huntington's disease," *The Lancet Psychiatry* 3, 1079–1086 (2016).

[10] C. Winograd-Gurvich, P. B. Fitzgerald, N. Georgiou-Karistianis, J. L. Bradshaw, and O. B. White, "Negative symptoms: A review of schizophrenia, melancholic depression and Parkinson's disease," *Brain Res. Bull.* 70, 312–321 (2006).

[11] T. Wyss-Coray, "Ageing, neurodegeneration and brain rejuvenation," *Nature* 539, 180–186 (2016).

[12] C. A. Haaxma, "New perspectives on preclinical and early stage Parkinson's disease" (2011).

[13] A. H. V. Schapira, K. R. Chaudhuri, and P. Jenner, "Non-motor features of Parkinson disease," *Nat. Rev. Neurosci.* 18, 435–450 (2017).

[14] A. Kilzheimer, T. Hentrich, S. Burkhardt, and J. M. Schulze-Hentrich, "The Challenge and Opportunity to Diagnose Parkinson's Disease in Midlife," *Front. Neurol.* 10 (2019).

[15] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, et al., "Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Mov. Disord.* 23, 2129–2170 (2008).

[16] J. F. van der Heeden, J. Marinus, P. Martinez-Martin, and J. J. van Hilten, "Evaluation of severity of predominantly non-dopaminergic symptoms in Parkinson's disease: The SENS-PD scale," *Park. Relat. Disord.* 25, 39–44 (2016).

[17] L. V. Kalia and A. E. Lang, "Parkinson's disease," *Lancet* 386, 896–912 (2015).

[18] S. Y. Pang, P. W. Ho, H. Liu, C. Leung, L. Li, E. Eun, and S. Chang, "The interplay of aging , genetics and environmental factors in the pathogenesis of Parkinson ' s disease," 1–11 (2019).

[19] O. B. Tysnes and A. Storstein, "Epidemiology of Parkinson's disease," *J. Neural Transm.* 124, 901–905 (2017).

[20] H. Braak, K. Del Tredici, U. Rüb, R. A. I. De Vos, E. N. H. Jansen Steur, and E. Braak, "Staging of brain pathology related to sporadic Parkinson's disease," *Neurobiol. Aging* 24, 197–211 (2003).

[21] P. Riederer, D. Berg, N. Casadei, F. Cheng, J. Classen, C. Dresel, W. Jost, R. Krüger, T. Müller, et al., "α-Synuclein in Parkinson's disease: causal or bystander?," *J. Neural Transm.* 126, 815–840 (2019).

[22] L. L. Venda, S. J. Cragg, V. L. Buchman, and R. Wade-Martins, "α-Synuclein and dopamine at the crossroads of Parkinson's disease," *Trends Neurosci.* 33, 559–568 (2010).

[23] T. Kunath, A. Natalwala, C. Chan, Y. Chen, B. Stecher, M. Taylor, S. Khan, and M. M. K. Muqit, "Are *PARKIN* patients ideal candidates for dopaminergic cell replacement therapies?," *Eur. J. Neurosci.* 49, 453–462 (2019).

[24] J. Levin, A. Kurz, T. Arzberger, A. Giese, and G. U. Höglinger, "The Differential Diagnosis and Treatment of Atypical Parkinsonism," *Dtsch. Arztebl. Int.* 113, 61–70 (2016).

[25] D. J. Surmeier, J. A. Obeso, and G. M. Halliday, "Selective neuronal vulnerability in Parkinson disease," *Nat. Rev. Neurosci.* 3, 973–982 (2016).

[26] M. X. Henderson, E. J. Cornblath, A. Darwich, B. Zhang, H. Brown, R. J. Gathagan, R. M. Sandler, D. S. Bassett, J. Q. Trojanowski, et al., "Spread of α-synuclein pathology through the brain connectome is modulated by selective vulnerability and predicted by network analysis," *Nat. Neurosci.* 22, 1248–1257 (2019).

[27] D. Chang, M. A. Nalls, I. B. Hallgrímsdóttir, J. Hunkapiller, M. van der Brug, F. Cai, G. A. Kerchner, G. Ayalon, B. Bingol, et al., "A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci," *Nat. Genet.* 49, 1511–1516 (2017).

[28] M. A. Nalls, N. Pankratz, C. M. Lill, C. B. Do, D. G. Hernandez, M. Saad, A. L. Destefano, E. Kara, J. Bras, et al., "Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease," *Nat. Publ. Gr.* 46, 989–993 (2014).

[29] S. Olgiati, M. Quadri, and V. Bonifati, "Genetics of Movement Disorders in the Next-Generation Sequencing Era," *Mov. Disord.* 31, 458–470 (2016).

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

[31] H. Deng, P. Wang, and J. Jankovic, "The genetics of Parkinson disease," *Ageing Res. Rev.* 42, 72–85 (2018).

[32] T. H. Hamza and H. Payami, "The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors," *J. Hum. Genet.* 55, 241–243 (2010).

[33] R. L. Doty, "Olfactory dysfunction in Parkinson disease," *Nat. Rev. Neurol.* 8, 329–339 (2012).

[34] C. A. Ross and S. J. Tabrizi, "Huntington's disease: From molecular pathogenesis to clinical treatment," *Lancet Neurol.* 10, 83–98 (2011).

[35] M. Arrasate and S. Finkbeiner, "Protein aggregates in Huntington's disease," *Exp. Neurol.* 238, 1–11 (2012).

[36] H. H. P. Nguyen and M. A. Cenci, *Behavioral Neurobiology of Huntington's Disease and Parkinson's Disease*, H. H. P. Nguyen and M. A. Cenci, Eds. (2015).

[37] X. Jin and R. M. Costa, "Start/stop signals emerge in nigrostriatal circuits during sequence learning," *Nature* 466, 457–462 (2010).

[38] C. a Ross, E. H. Aylward, E. J. Wild, D. R. Langbehn, J. D. Long, J. H. Warner, R. I. Scahill, B. R. Leavitt, J. C. Stout, et al., "Huntington disease: natural history, biomarkers and prospects for therapeutics.," *Nat. Rev. Neurol.* 10, 204–216 (2014).

[39] T. Alexi, C. V. Borlongan, R. L. M. Faull, C. E. Williams, R. G. Clark, P. D. Gluckman, and P. E. Hughes, "Neuroprotective strategies for basal ganglia degeneration: Parkinson's and Huntington's diseases," *Prog. Neurobiol.* 60, 409–470 (2000).

[40] J. M. Lee, E. M. Ramos, J. H. Lee, T. Gillis, J. S. Mysore, M. R. Hayden, S. C. Warby, P. Morrison, M. Nance, et al., "CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion," *Neurology* 78, 690–695 (2012).

[41] M.-U. Manto, "The wide spectrum of spinocerebellar ataxias (SCAs).," *Cerebellum* 4, 2–6 (2005).

[42] G. Stuitje, M. J. van Belzen, S. L. Gardiner, W. M. C. van Roon-Mom, M. W. Boogaard, S. J. Tabrizi, R. A. C. Roos, N. A. Aziz, and N. A. Aziz, "Age of onset in Huntington's disease is influenced by CAG repeat variations in other polyglutamine disease-associated genes," *Brain*, 10–12 (2017).

[43] L. Jones, H. Houlden, and S. J. Tabrizi, "DNA repair in the trinucleotide repeat disorders," *Lancet Neurol.* 16, 1234 (2017).

[44] J. van Gaalen, P. Giunti, and B. P. van de Warrenburg, "Movement disorders in spinocerebellar ataxias," *Mov. Disord.* 26, 792–800 (2011).

[45] Pharmwiki, "Parkinson Disease: Neurologic Pathways & Drug Targets," <http://tmedweb.tulane.edu/pharmwiki/doku.php/treating_parkinson_s_disease>.

[46] E. F. E. Kuiper, E. P. de Mattos, L. B. Jardim, H. H. Kampinga, and S. Bergink, "Chaperones in polyglutamine aggregation: Beyond the Q-stretch," *Front. Neurosci.* 11, 1–11 (2017).

[47] G. Borrageiro, W. Haylett, S. Seedat, H. Kuivaniemi, and S. Bardien, "A review of genome-wide transcriptomics studies in Parkinson's disease," *Eur. J. Neurosci.* 47, 1–16 (2018).

[48] E. Oerton and A. Bender, "Concordance analysis of microarray studies identifies representative gene expression changes in Parkinson's disease: a comparison of 33 human and animal studies," *BMC Neurol.* 17, 1–14 (2017).

[49] T. C. Burns, M. D. Li, S. Mehta, A. J. Awad, and A. A. Morgan, "Mouse models rarely mimic the transcriptome of human neurodegenerative diseases : A systematic bioinformatics-based critique of preclinical models," *Eur. J. Pharmacol.*, 1–16 (2015).

[50] J. A. Santiago and J. A. Potashkin, "Blood transcriptomic meta-analysis identifies dysregulation of hemoglobin and iron metabolism in Parkinson' Disease," *Front. Aging Neurosci.* 9, 1–8 (2017).

[51] C. R. Scherzer, A. C. Eklund, L. J. Morse, Z. Liao, J. J. Locascio, D. Fefer, M. A. Schwarzschild, M. G. Schlossmacher, M. A. Hauser, et al., "Molecular markers of early Parkinson's disease based on gene expression in blood," *Proc. Natl. Acad. Sci. U. S. A.* 104, 955–960 (2007).

[52] E. Mutez, L. Larvor, F. Leprêtre, V. Mouroux, D. Hamalek, J.-P. Kerckaert, J. Pérez-Tur, N. Waucquier, C. Vanbesien-Mailliot, et al., "Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation," *NBA* 32, 1839–1848 (2011).

[53] S. K. Negi and C. Guda, "Global gene expression profiling of healthy human brain and its application in studying neurological disorders," *Sci. Rep.* 7, 1–12 (2017).

[54] A. E. Omole, A. Omotuyi, and J. Fakoya, "Ten years of progress and promise of induced pluripotent stem cells: historical origins, characteristics, mechanisms, limitations, and potential applications," *PeerJ* 6, e4370 (2018).

[55] J. Kelly, R. Moyeed, C. Carroll, D. Albani, and X. Li, "Gene expression meta-analysis of Parkinson's disease and its relationship with Alzheimer's disease," *Mol. Brain* 12, 1–10 (2019).

[56] B. E. Riley, S. J. Gardai, D. Emig-Agius, M. Bessarabova, A. E. Ivliev, B. Schüle, J. Alexander, W. Wallace, G. M. Halliday, et al., "Systems-based analyses of brain regions functionally impacted in Parkinson's disease reveals underlying causal mechanisms," *PLoS One* 9, 1–14 (2014).

[57] 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).

[58] M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C.-K. K. Lee, et al., "Canonical genetic signatures of the adult human brain," *Nat. Neurosci.* 18, 1832–1844 (2015).

[59] E. Eising, S. M. H. Huisman, A. Mahfouz, L. S. Vijfhuizen, V. Anttila, B. S. Winsvold, and T. Kurth, "Gene co-expression analysis identifies brain regions and cell types involved in migraine pathophysiology: a GWAS-based study using the Allen Human Brain Atlas," *Hum. Genet.* 135, 425–439 (2016).

[60] F. Liu, H. Tian, J. Li, S. Li, and C. Zhuo, "Altered voxel-wise gray matter structural brain networks in schizophrenia: Association with brain genetic expression pattern," *Brain Imaging Behav.* 13, 493–502 (2019).

[61] I. A. C. Romme, M. A. De Reus, R. A. Ophoff, R. S. Kahn, M. P. van den Heuvel, M. A. de Reus, R. A. Ophoff, R. S. Kahn, and M. P. van den Heuvel, "Connectome disconnectivity and cortical gene expression in patients with schizophrenia," *Biol. Psychiatry* 81, 495–502 (2017).

[62] S. M. H. Huisman, A. Mahfouz, N. K. Batmanghelich, B. P. F. Lelieveldt, and M. J. T. Reinders, "A structural equation model for imaging genetics using spatial transcriptomics," *Brain Informatics* 5 (2018).

[63] G. V Roshchupkin, H. H. Adams, S. J. van der Lee, M. W. Vernooij, C. M. van Duijn, A. G. Uitterlinden, A. van der Lugt, A. Hofman, W. J. Niessen, et al., "Fine-mapping the effects of Alzheimer's disease risk loci on brain morphology," *Neurobiol. Aging* 48, 204–211 (2016).

[64] R. J. Mullins, M. Mustapic, E. J. Goetzl, and D. Kapogiannis, "Exosomal biomarkers of brain insulin resistance associated with regional atrophy in Alzheimer's disease," *Hum. Brain Mapp.* 1940, 1933–1940 (2017).

[65] R. Freer, P. Sormanni, G. Vecchi, P. Ciryam, C. M. Dobson, and M. Vendruscolo, "A protein homeostasis signature in healthy brains recapitulates tissue vulnerability to Alzheimer's disease," *Sci. Adv.* 2, 1–8 (2016).

[66] E. H. Shen, C. C. Overly, and A. R. Jones, "The Allen Human Brain Atlas Comprehensive gene expression mapping of the human brain," *Trends Neurosci.* 35, 711–714 (2012).

[67] V. I. Müller, E. C. Cieslik, A. R. Laird, P. T. Fox, J. Radua, D. Mataix-Cols, C. R. Tench, T. Yarkoni, T. E. Nichols, et al., "Ten simple rules for neuroimaging meta-analysis," *Neurosci. Biobehav. Rev.* 84, 151–161 (2018).

[68] N. W. Sterling, M. M. Lewis, G. Du, X. Huang, P. State, and U. S. H. Medical, "Structural imaging and Parkinson's disease : moving toward quantitative markers of disease progression," *J. Park. Dis.* 6, 557–567 (2016).

[69] C. P. Weingarten, M. H. Sundman, P. Hickey, and N. kuei Chen, "Neuroimaging of Parkinson's disease: Expanding views," *Neurosci. Biobehav. Rev.* 59, 16–52 (2015).

[70] P. L. Pan, W. Song, and H. F. Shang, "Voxel-wise meta-analysis of gray matter abnormalities in idiopathic Parkinson's disease," *Eur. J. Neurol.* 19, 199–206 (2012).

[71] A. Alexander-Bloch, J. N. Giedd, and E. Bullmore, "Imaging structural co-variance between human brain regions," *Nat. Rev. Neurosci.* 14, 322–336 (2013).

[72] L. J. de Schipper, J. van der Grond, J. Marinus, J. M. L. Henselmans, and J. J. van Hilten, "Loss of integrity and atrophy in cingulate structural covariance networks in Parkinson's disease," *NeuroImage Clin.* 15, 587–593 (2017).

[73] K. Rosenberg-Katz, T. Herman, Y. Jacob, E. Kliper, N. Giladi, and J. M. Hausdorff, "Subcortical volumes differ in Parkinson's disease motor subtypes: New insights into the pathophysiology of disparate symptoms," *Front. Hum. Neurosci.* 10, 1–9 (2016).

[74] M. Wang, S. Jiang, Y. Yuan, L. Zhang, J. Ding, J. Wang, J. Zhang, K. Zhang, and J. Wang,

"Alterations of functional and structural connectivity of freezing of gait in Parkinson's disease," *J. Neurol.* 263, 1583–1592 (2016).

[75] D. Aarsland, B. Creese, M. Politis, K. R. Chaudhuri, D. H. Ffytche, D. Weintraub, and C. Ballard, "Cognitive decline in Parkinson disease," *Nat. Rev. Neurol.* 13, 217–231 (2017).

[76] D. Zheng, C. Chen, W. C. Song, Z. Q. Yi, P. W. Zhao, J. G. Zhong, Z. Y. Dai, H. C. Shi, and P. L. Pan, "Regional gray matter reductions associated with mild cognitive impairment in Parkinson's disease: A meta-analysis of voxel-based morphometry studies," *Behav. Brain Res.* 371, 111973 (2019).

[77] O. Lucas-Jiménez, N. Ojeda, J. Peña, M. Díez-Cirarda, A. Cabrera-Zubizarreta, J. C. Gómez-Esteban, M. Á. Gómez-Beldarrain, and N. Ibarretxe-Bilbao, "Altered functional connectivity in the default mode network is associated with cognitive impairment and brain anatomical changes in Parkinson's disease," *Park. Relat. Disord.* 33, 58–64 (2016).

[78] H. Wilson, F. Niccolini, C. Pellicano, and M. Politis, "Cortical thinning across Parkinson's disease stages and clinical correlates," *J. Neurol. Sci.* 398, 31–38 (2019).

[79] J. H. Won, M. Kim, B. Y. Park, J. Youn, and H. Park, "Effectiveness of imaging genetics analysis to explain degree of depression in Parkinson's disease," *PLoS One* 14, 1–18 (2019).

[80] J. P. M. van der Vegt, B. F. L. van Nuenen, B. R. Bloem, C. Klein, and H. R. Siebner, "Imaging the impact of genes on Parkinson's disease," *Neuroscience* 164, 191–204 (2009).

[81] S. E. Winder-Rhodes, A. Hampshire, J. B. Rowe, J. E. Peelle, T. W. Robbins, A. M. Owen, and R. A. Barker, "Association between *MAPT* haplotype and memory function in patients with Parkinson's disease and healthy aging individuals," *Neurobiol. Aging* 36, 1519–1528 (2015).

[82] M. Lorenzi, A. Altmann, B. Gutman, S. Wray, C. Arber, D. P. Hibar, N. Jahanshad, J. M. Schott, D. C. Alexander, et al., "Susceptibility of brain atrophy to *TRIB3* in Alzheimer's disease, evidence from functional prioritization in imaging genetics," *Proc. Natl. Acad. Sci.* 115, 3162–3167 (2018).

[83] G. Pagano, F. Niccolini, and M. P. C, "Imaging in Parkinson's disease," *Clin. Med.* 16, 371–375 (2016).

[84] A. Arnatkevičiūtė, B. D. Fulcher, and A. Fornito, "A practical guide to linking brain-wide gene expression and neuroimaging data," *Neuroimage* 189, 353–367 (2019).

[85] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Stat. Appl. Genet. Mol. Biol.* 4 (2005).

[86] J. Richiardi, A. Altmann, A.-C. Milazzo, C. Chang, T. Chakravarty, M. Mallar Banaschewski, G. J. Barker, A. L. W. Bokde, U. Bromberg, C. Büchel, et al., "Correlated gene expression supports synchronous activity in brain networks," *Science* 348, 11–14 (2015).

[87] R. Romero-Garcia, K. J. Whitaker, F. Váša, J. Seidlitz, M. Shinn, P. Fonagy, R. J. Dolan, P. B. Jones, I. M. Goodyer, et al., "Structural covariance networks are coupled to expression of genes enriched in supragranular layers of the human cortex," *Neuroimage* 171, 256–267 (2018).

[88] P. E. Vértes, T. Rittman, K. J. Whitaker, R. Romero-Garcia, F. Váša, M. G. Kitzbichler, K. Wagstyl, P. Fonagy, R. J. Dolan, et al., "Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks," *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150362 (2016).

[89] K. J. Whitaker, P. E. Vértes, R. Romero-Garcia, F. Váša, M. Moutoussis, G. Prabhu, N. Weiskopf, M. F. Callaghan, K. Wagstyl, et al., "Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome," *Proc. Natl. Acad. Sci.* 113, 9105–9110 (2016).

[90] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, et al., "Gene Ontology: tool for the unification of biology," *Nat. Genet.* 25, 25–29 (2011).

[91] The Gene Ontology Consortium, "The Gene Ontology Resource: 20 years and still GOing

strong," *Nucleic Acids Res.* 47, D330–D338 (2019).

[92] B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, et al., "The reactome pathway knowledgebase," 498–503 (2020).

[93] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Res.* 28, 27–30 (2000).

[94] J. Piñero, À. Bravo, N. Queralt-Rosinach, A. Gutiérrez-Sacristán, J. Deu-Pons, E. Centeno, J. García-García, F. Sanz, and L. I. Furlong, "DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Res.* 45, D833–D839 (2017).

[95] V. M. Pickel, T. H. Joh, P. M. Field, C. G. Becker, and D. J. Reis, "Cellular localizatin of tyrosine hydrozylase by immunohistochemistry," *J. Histochem. Cytochem.* 23, 1–12 (1975).

[96] A. Capurro, L.-G. Bodea, P. Chaefer, R. Luthi-Carter, and V. M. Perreau, "Computational deconvolution of genome wide expression data from Parkinson's and Huntington's disease brain tissues using population-specific expression analysis," *Front. Neurosci.* 8, 1–12 (2015).

[97] A. Kuhn, D. Thu, H. J. Waldvogel, R. L. M. Faull, and R. Luthi-Carter, "Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain," *Nat. Methods* 8, 945–947 (2011).

[98] M. Chikina, E. Zaslavsky, and S. C. Sealfon, "CellCODE: A robust latent variable approach to differential expression analysis for heterogeneous cell populations," *Bioinformatics* 31, 1584–1591 (2015).

[99] 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).

[100] A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, and A. A. Alizadeh, "Robust enumeration of cell subsets from tissue expression profiles," *Nat. Methods* 12, 453–457 (2015).

[101] T. Rittman, M. Rubinov, P. E. Vértes, A. X. Patel, C. E. Ginestet, B. C. P. Ghosh, R. A. Barker, M. G. Spillantini, E. T. Bullmore, et al., "Regional expression of the *MAPT* gene is associated with loss of hubs in brain networks and cognitive impairment in Parkinson disease and progressive supranuclear palsy," *Neurobiol. Aging* 48, 153–160 (2016).

[102] B. S. Freeze, D. Acosta, S. Pandya, Y. Zhao, and A. Raj, "Regional expression of genes mediating trans-synaptic alpha-synuclein transfer predicts regional atrophy in Parkinson disease," *NeuroImage Clin.* 18, 456–466 (2018).

[103] A. Ramasamy, D. Trabzuni, S. Guelfi, V. Varghese, C. Smith, R. Walker, T. De, U. K. Brain, E. Consortium, et al., "Genetic variability in the regulation of gene expression in ten regions of the human brain," *Nat. Publ. Gr.* 17, 1418–1428 (2014).

Chapter 2

Cholinergic genes in the healthy brain are differentially expressed in regions that exhibit gray matter loss in Parkinson's disease

Arlin Keo

Oleh Dzyubachyk

Jeroen van der Grond

Anne Hafkemeijer

Wilma D.J. van de Berg

Jacobus J. van Hilten

Marcel J. T. Reinders

Ahmed Mahfouz

This chapter was published online in: *bioRxiv* (2019), doi: 10.1101/2019.12.17.875880
(in submission)
Supplementary material is available online at:
https://www.biorxiv.org/content/10.1101/2019.12.17.875880v1.supplementary-material

Abstract

Structural covariance networks are able to identify functionally organized brain regions by gray matter volume covariance across a population. We examined the transcriptomic signatures of such anatomical networks in the healthy brain using post-mortem microarray data from the Allen Human Brain Atlas. A previous study revealed that a posterior cingulate network and anterior cingulate network showed decreased gray matter in brains of Parkinson's disease patients. Therefore, we examined these two anatomical networks to understand the underlying molecular processes that may be involved in Parkinson's disease. Whole brain transcriptomics from the healthy brain revealed upregulation of genes associated with serotonin, GPCR, GABA, glutamate, and RAS signaling pathways. Our results also suggest involvement of the cholinergic circuit, in which genes *NPPA*, *SOSTDC1*, and *TYRP1* may play a functional role. Finally, both networks were enriched for genes associated with neuropsychiatric disorders that overlap with Parkinson's disease symptoms. The identified genes and pathways contribute to healthy functions of the posterior and anterior cingulate networks and disruptions to these functions may in turn contribute to the pathological and clinical events observed in Parkinson's disease.

2.1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder characterized by the impairment of diverse motor and non-motor symptoms that get progressively worse over time [1]. The decline in clinical performance has been associated with changes in morphological properties of structural and functional neuroimaging networks [2–4]. In turn, studies have investigated the relationship between imaging networks and genetic risk factors associated with PD to provide new insights into the pathogenesis of PD [5–8]. However, less is known about the functions that underlie the spatial organization of brain regions contributing to PD. To identify the molecular mechanisms underlying changes in structural and functional networks in PD, imaging data has been integrated with brain-wide healthy gene expression from the Allen Human Brain Atlas (AHBA) [9,10]. Regional brain atrophy in PD patients was correlated with the expression of genes implicated in trans-synaptic alpha-synuclein transfer [11] and a loss of regional connectivity in PD patients was correlated with the regional expression of *MAPT* in the healthy brain [12]. These studies showed that combining imaging data in PD and gene expression from the healthy brain can shed light on the molecular mechanisms underlying the morphological differences between PD and controls.

Structural covariance networks (SCNs) identify brain regions that co-vary in gray matter volume across a population and can reveal functional network organizations [13]. SCNs have been shown to be dysregulated in different neurological disorders [14–18], and gray matter variations in SCNs can be explained by transcriptomic similarity and structural connectivity [19,20]. Hafkemeijer *et al.* [21] identified nine SCNs based on gray matter variation among healthy middle-aged to older adults. Gray matter volume in four of these nine networks was negatively associated with age: a subcortical network, sensorimotor network, posterior cingulate networks, and anterior cingulate network. Two of these networks were found to show loss of gray matter volume in PD patients beyond the effects of aging: the posterior cingulate network and anterior cingulate network [2]. Atrophy within these two networks was also associated with cognitive impairment and daytime sleepiness, respectively. Together these studies revealed how brain networks change in aging and PD, but the molecular mechanisms contributing to the relevant SCNs remain unclear.

Here, we investigated the transcriptomic signatures of the anterior and posterior cingulate networks within the healthy brain. By integrating the nine SCNs with spatial gene expression data from the Allen Human Brain Atlas, we showed that genes highly expressed in the posterior and anterior cingulate networks were associated with multiple neurotransmitter signaling pathways as well as with memory-related, pain-related, and neuropsychiatric disorders. In addition, both networks showed high expression of cholinergic marker genes that are known to act as regulators of extracellular signaling. Our results provide new insights into the molecular processes underlying anatomical network function and aids in better understanding the selective progression of PD.

2.2 Materials and Methods

2.2.1 Transcriptomic data preprocessing

To understand transcriptomic signatures of nine anatomical networks of the healthy brain, we analyzed gene expression data from the AHBA, a post-mortem microarray data set of 3,702 anatomical brain regions from six non-neurological individuals (5 males and 1 female, mean age 42, range 24–57 years) [10]. Normalized gene expression from the AHBA was downloaded from http://human.brain-map.org/. To filter and map probes to genes, the data was concatenated across the six donors. We removed 10,521 probes with missing Entrez IDs, and 6,068 probes with low presence as they were expressed above background in <1% of samples (PA-call containing presence/absence flag) [10]. The remaining 44,072 probes were mapped to 20,017 genes with unique Entrez IDs using the *collapseRows*-function in R-package WGCNA v1.64.1 [22] as follows: (i) if there is one probe, that one probe is chosen, (ii) if there are two probes, the one with maximum variance across all samples is chosen (method="maxRowVariance"), (iii) if there are more than two probes, the probe with the highest connectivity (summed adjacency) is chosen (connectivityBasedCollapsing = TRUE).

For visualization of gene expression in heatmaps, data was Z-score normalized across all samples for each brain donor separately. Heatmaps were plotted using R-package ComplexHeatmap v2.0.0 [23]. Genes were clustered using complete linkage with Euclidean distances. The same color scale for gene expression was used for all heatmaps.

2.2.2 Mapping AHBA samples to SCNs of the healthy brain

We focused on anatomical networks that were previously defined based on whole brain gray matter volume covariation in 370 middle-aged to older adults between 45 and 85 years; for more detailed information on the networks see Hafkemeijer et al. 2014. Nine networks were defined and named according to the presence of the main structures: thalamus (network A), lateral occipital cortex (Network B), posterior cingulate cortex (Network C), anterior cingulate cortex (network D), temporal pole (network E), putamen (network F), and cerebellum (networks G, H, and I). The same networks were previously investigated for loss of integrity in 159 PD patients from the same age range; for demographic and clinical information see de Schipper et al. 2017. All samples from each one of the six donors in AHBA were mapped to regions defined by the nine SCNs in MNI coordinate space.

2.2.3 Differential expression analysis

For differential expression analysis we focused on the posterior cingulate network (Network C) and anterior cingulate network (Network D) that were previously associated with gray matter loss in PD [2]. Gene expression in each of the two networks C and D was compared to the other 7 networks together (A, B, E, F, G, H, and I). A two-tailed *t*-test was used for each gene and the analysis was done separately for each donor from AHBA. Since the microarray data was $\log_2$-transformed, the mean expression difference is interpreted as the $\log_2$-transformed fold-change (FC). The effect sizes for each one of the six donors were combined by meta-analysis (metafor R-package 2.0). For the meta-analysis, a random effects model

was applied which assumes that each brain is considered to be from a larger population of brains and therefore takes the within-brain and between-brain variance into account. The between-brain variance (tau^2) was estimated with the Dersimonian-Delaird model. Variances and confidence intervals were obtained using the escalc-function. The significance of summary effect sizes was assessed through a two-sided *t*-test (H_o: FC=0; unequal variances). *P*-values of the effect sizes were Benjamini-Hochberg (BH) corrected for all 20,017 genes. Genes were differentially expressed within the posterior cingulate network or the anterior cingulate network compared to the other networks combined when the absolute fold-change (FC) > 1 and the BH-corrected *P*-value < 0.05.

2.2.4 PATHWAY ANALYSIS

Pathway analysis was done with the ReactomePA R-package version 1.28 using the function *enrichPathway* searching for human pathways. All 20,017 genes in the AHBA dataset were set as background genes. Pathways with a minimum size of 10 genes were significant when the BH-corrected $P < 0.05$.

2.2.5 CELL-TYPE MARKER ENRICHMENT

Gene markers for 28 cell-types were downloaded from the NeuroExpresso database (http://neuroexpresso.org/) using markers from all brain regions. These have been identified in a cross-laboratory dataset of cell-type specific transcriptomes from the mouse brain [24]. To assess their expression, Entrez IDs of the mouse cell-type specific markers were converted to human homologs (homologene R-package version 1.4) and filtered for genes present in the AHBA dataset (Supplementary Table 1). Two markers with different mouse gene IDs (14972, *H2-K1*, microglial, and 15006, *H2-Q1* serotonergic), were converted to the same human gene ID (3105, *HLA-A*), and therefore removed before analysis. For cell-type enrichment, we assessed which cell-type markers were overrepresented among the differentially expressed genes. For 17 cell-types that had at least six markers (astrocyte, Bergmann, cerebellar granule, dentate granule, ependymal, GabaReln, hypocretinergic, microglia, activated microglia, deactivated microglia, noradrenergic, oligo, purkinje, serotonergic, spinal cord cholinergic, spiny, and thalamus cholinergic), we assessed the significance with the hypergeometric test and *P*-values were corrected for all 17 cell-types (BH-corrected $P < 0.05$).

2.2.6 ENRICHMENT OF DISEASE-ASSOCIATED GENES

Differentially expressed genes were also assessed for the overrepresentation of disease-associated genes from DisGeNET [25]. A table of 628,685 gene-disease associations were obtained from DisGeNET version 6.0 (July, 2019) from http://www.disgenet.org/. A hypergeometric test was used to assess the significance of overlapping genes ($P < 0.05$), and *P*-values were BH-corrected for 24,166 diseases. The odds ratio (OR) for cell-type and disease enrichment was calculated using the DescTools R-package.

2.2.7 CODE AVAILABILITY

Scripts to run all analyses can be found online at https://github.com/arlinkeo/pd_scn.

2.3 RESULTS

2.3.1 TRANSCRIPTOMICS OF THE POSTERIOR AND ANTERIOR CINGULATE NETWORKS

We analyzed the transcriptomes of healthy subjects across nine anatomical networks defined by structural covariance of gray matter volume among healthy middle-aged to older adults [21]. For this we used the AHBA microarray dataset of spatial gene expression in post-mortem brains of six non-neurological donors and samples were mapped to each one of the nine networks A-I (Table 2.1) based on their spatial location. We focused on the posterior cingulate network (Network C) and the anterior cingulate network (Network D) that showed loss of gray matter in PD patients (Figure 2.1) [2] and characterized their transcriptional signatures by comparing them to the remaining seven networks together.

Table 2.1 Number of samples from the AHBA that fall within networks A-I.

Donors	Network								
	A	B	C	D	E	F	G	H	I
Donor 9861	72	67	157	47	74	90	26	39	83
Donor 10021	79	46	121	65	49	84	25	55	91
Donor 12876	37	24	57	28	42	45	6	17	25
Donor 14380	38	33	52	30	45	61	7	27	53
Donor 15496	34	24	41	21	39	55	13	24	69
Donor 15697	49	20	38	33	47	64	29	37	49
Total	309	214	466	224	296	399	106	199	370

A: Thalamus; B: Lateral occipital cortex, C: Posterior cingulate cortex, D: Anterior cingulate cortex, E: Temporal pole; F: Putamen; G, H, I: Cerebellum.

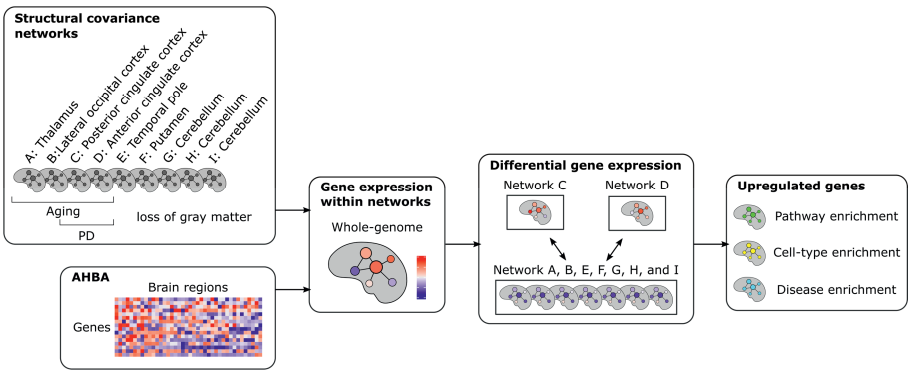


Figure 2.1 Study overview. Transcriptomic data from AHBA were mapped to nine anatomical networks that have been defined based on healthy subjects. The posterior cingulate network (Network C) and anterior cingulate network (Network D) have been associated with gray matter loss in PD, while the seven remaining networks were not related to PD. We compared gene expression in network C and D to gene expression in networks A, B, E, F, G, H, and I together. Upregulated genes were assessed for the overrepresentation of pathway-specific genes, cell-type marker genes, and disease-associated genes.

Whole genome differential expression analysis showed a large overlap of genes that were differentially expressed in the same direction in the two networks. We found that 73 genes in the posterior cingulate network and 39 genes in anterior cingulate network were downregulated, of which 25 genes overlapped between both networks (Figure 2.2AB, Supplementary Table 2 and 3). Furthermore, 200 genes in the posterior cingulate network and 269 genes in anterior cingulate network were upregulated, for which 144 genes overlapped (Supplementary Table 4 and 5). Among the differentially expressed genes in the posterior and anterior cingulate networks, no PD-implicated genes were found that arouse from familial and genome-wide association studies [26–28].

For functional interpretation of the differentially upregulated genes we further assessed their associated pathways (see Methods, Supplementary Table 6). As both networks C and D shared many differentially expressed genes, they also shared similar pathways: transcriptional regulation by *MECP2*, GPCR signaling, voltage gated potassium channels, and neurotransmitter receptor and postsynaptic signal transmission (Figure 2.2C). These pathways are also hierarchically related to each other based on the ontology of the Reactome Pathway Database. The posterior cingulate network was additionally related to more specific pathways such as lysosphingolipid and LPA receptors, GABA receptor activation, RAS-signaling mediated by NMDA receptors, glutamate binding, activation of AMPA receptors and synaptic plasticity, and long-term potentiation. The anterior cingulate cortex was additionally associated with serotonin receptors.

2.3.2 CELL-TYPE ENRICHMENT IN ANATOMICAL NETWORKS

The composition of specific cell-types can shape the transcriptomic features of anatomical networks. Therefore, we analyzed whether genes differentially expressed in the posterior and anterior cingulate networks were enriched for cell-type specific marker genes from the NeuroExpresso database [24]. To assess the expression of each cell-type, we averaged the expression of the marker genes associated with that cell-type. Both the posterior and anterior cingulate networks showed high expression of marker genes for brainstem cholinergic cells, GabaSSTReln, GabaVIPReln, glutamatergic, and pyramidal cells (Figure 2.3 and Supplementary Figure 1).

Among the differentially upregulated genes in the posterior and anterior cingulate networks, we found 10 marker genes representing six cell-types: astrocyte, Bergmann, GabaVIPReln, hypocretinergic, pyramidal, and thalamus cholinergic (Table 2.2). Markers that were significantly upregulated in the posterior cingulate network were also significantly upregulated in the anterior cingulate network. In both networks, the 10 markers were highly expressed in cortical regions, including the cingulate gyrus, and lowly expressed in limbic regions (Figure 2.4 and Supplementary Figure 2).

Only genes upregulated in the anterior cingulate gyrus were significantly enriched for a cell-type, namely thalamus cholinergic cells (OR = 17.12 and P = 2.01e-02). The responsible markers *NPPA*, *SOSTDC1*, and *TYRP1* showed high expression within the anterior cingulate gyrus network, as well as in most parts of the posterior cingulate gyrus network (Figure 2.4).

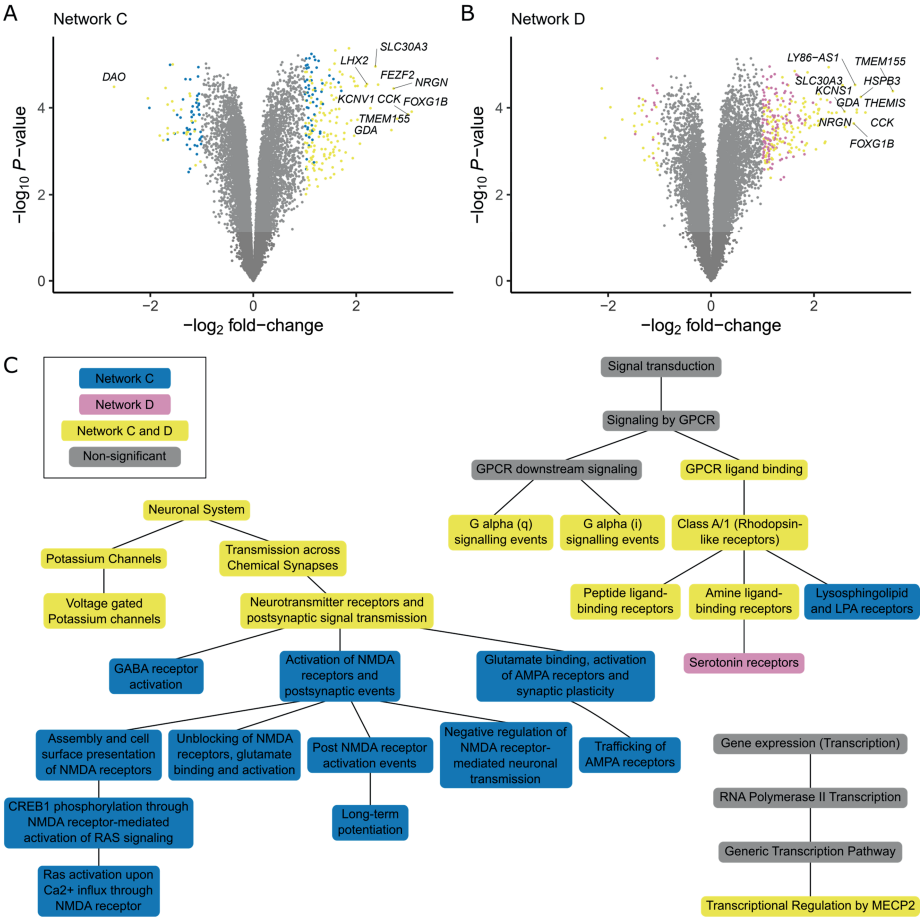


Figure 2.2 Differential expressed genes and associated pathways. Genes were analyzed for differential expression in (A) the posterior cingulate network (Network C) and (B) the anterior cingulate network (Network D). Effect sizes were summarized across the six healthy donors of AHBA with meta-analysis. For all genes (points) the $\log_2$ fold-change (FC; x-axis) and -$\log_{10}$ of nominal *P*-values (y-axis) are shown. Significant differentially expressed genes (*t*-test, BH-corrected $P < 0.05$, and $|FC| > 1$) are unique for each network (blue and purple points) or significant in both networks (yellow points). The top 10 genes with the highest absolute FC are labeled for each network and highly overlap between both networks. (C) Pathway analysis of differentially upregulated genes in the posterior cingulate network and anterior cingulate network. Network C and Network D shared similar pathways (yellow) that are hierarchically organized in the Reactome database. The posterior cingulate network showed more specific associations with pathways involved in neurotransmitter receptors and postsynaptic signal transmission (blue). The anterior cingulate network was more specifically associated with serotonin receptors (purple). See Supplementary Table 6 for gene counts and BH-corrected *P*-values.

Interestingly, while other thalamus cholinergic marker genes showed high expression in limbic samples and low expression in cortical samples within both networks, *NPPA*, *SOSTDC1*, and *TYRP1* showed opposite expression patterns with low expression in limbic samples, including the thalamus, and high expression in cortical samples (Supplementary Figure 3).

Table 2.2 Differentially expressed cell-type marker genes in the posterior cingulate network (Network C) and anterior cingulate network (Network D). Fold-change (FC) and Benjamini-Hochberg (BH) corrected *P*-value are shown for cell-type marker genes that were differentially expressed in the two networks compared to the remaining networks. FC > 1 and BH < 0.05 are highlighted in red text.

Gene	Cell-type	Network C		Network D	
		FC	BH	Estimate	BH
LHX2	Astrocyte	2.21	3.92E-03	2.00	6.46E-03
IGFBP2	Astrocyte	0.69	5.80E-02	1.18	1.78E-02
RORB	Astrocyte	0.82	3.09E-02	1.19	1.39E-02
WIF1	Bergmann	1.02	8.74E-03	1.03	7.95E-03
VIP	GabaVIPReln	1.67	4.23E-03	1.85	6.89E-03
PCSK1	Hypocretinergic	1.15	1.25E-02	1.57	1.06E-02
NEUROD6	Pyramidal	1.90	4.78E-03	1.92	6.76E-03
NPPA	ThalamusCholin	1.64	6.98E-03	2.09	6.39E-03
TYRP1	ThalamusCholin	0.81	2.41E-02	1.43	9.82E-03
SOSTDC1	ThalamusCholin	0.83	1.21E-02	1.14	6.39E-03

Fold-change (FC) and Benjamini-Hochberg (BH) corrected *P*-value for cell-type markers genes that were differentially expressed in network C and D compared to the remaining networks. FC > 1 and BH < 0.05 are highlighted in red.

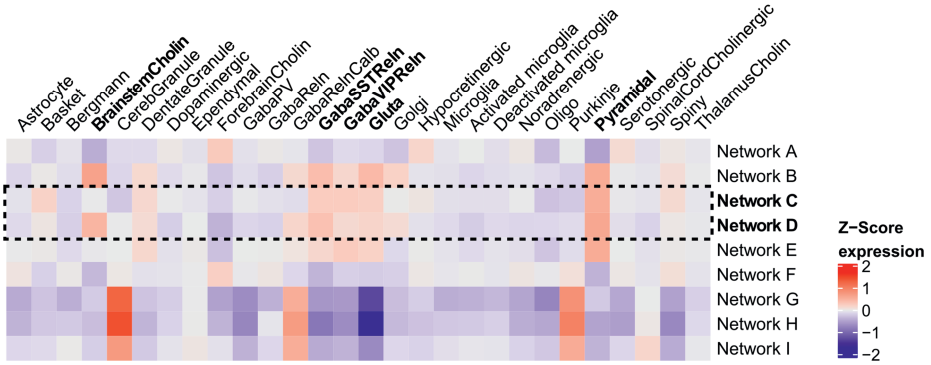


Figure 2.3 Expression of cell-types in anatomical networks. Gene expression was Z-scored and averaged across cell-type specific markers, across samples within anatomical networks, and across the six donors in the AHBA. Networks G, H, and I are cerebellar networks and thus showed distinct expression patterns. The posterior and anterior cingulate networks (Network C and Network D) showed high expression of marker genes for brainstem cholinergic cells, GabaSSTReln, GabaVIPReln, glutamatergic cells, and pyramidal cells. Gene expression heatmaps for each donor are shown in Supplementary Figure 1.

2.3.3 Cingulate networks are enriched for genes associated with cognitive disorders

Dysregulation of functional networks may result in a broader spectrum of disorders than PD. Therefore, we assessed which disease-associated genes from DisGeNET were overrepresented among the differentially upregulated genes in the posterior cingulate network as well as the anterior cingulate network. Since both networks shared many upregulated genes, similar disease-associations were also found. We found that genes upregulated in both networks were significantly associated with epileptic and non-epileptic seizures, many mental disorders (bipolar, panic, autistic, cocaine-related, (age-related) memory, mood, major depressive, and anxiety disorder), pain and schizophrenia (Figure 2.5). The posterior cingulate network was more related to memory and pain-related disorders, while the anterior cingulate network was more related to mental and neuropsychiatric disorders.

2.4 Discussion

The posterior and anterior cingulate networks have been previously associated with decreased gray matter in PD patients. We examined transcriptomic signatures of both networks in the healthy brain to identify molecular mechanisms underlying these two regions. Pathway analysis revealed genes related to gPCR signaling, transcriptional regulation by *MECP2*, and neurotransmitter receptors and postsynaptic signal transmission. We only found significant enrichment of cell-types for genes upregulated in the anterior cingulate gyrus, which were the thalamus cholinergic marker genes. Upon further examination the specific genes were also highly expressed in the posterior cingulate cortex, although not significantly. Moreover, our results showed that both SCNs are associated with multiple neurotransmitter signaling pathways, e.g., serotonin, GPCR, GABA, glutamate, and RAS.

2.4.1 Cholinergic function in PD

Genes that were highly expressed in the anterior cingulate network were significantly enriched for thalamus cholinergic markers, specifically: *NPPA*, *SOSTDC1*, and *TYRP1*. These marker genes, together with other markers of this cell-type, were previously defined based on their expression in cholinergic cells from the mouse thalamus, more specifically the habenula [24]. According to the AHBA ontology the habenula is not part of the thalamus. In this study, most thalamus cholinergic marker genes indeed showed high expression in human thalamic regions. However, *NPPA*, *SOSTDC1*, and *TYRP1* unexpectedly showed opposite expression patterns with mainly high expression in cortical regions and low expression in limbic regions, including the thalamus. Cholinergic circuits are key in cognitive functions and cholinergic denervation of the cortex and thalamus in PD patients may contribute to the transition from PD to PD with dementia [29]. We found that glutamatergic and GABAergic marker genes were also highly expressed within the posterior and anterior cingulate networks, although statistical significance could not be assessed due to the small number of marker genes for these cell-types. Interestingly, acetylcholine release by cholinergic neurons affects glutamatergic and GABAergic signaling by altering the synaptic excitability [30,31]. Moreover, it is thought that dysfunction of cholinergic circuits contributes to cognitive decline associated with neurodegenerative diseases [29].

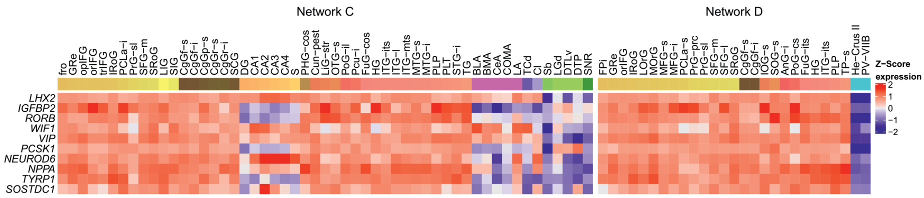


Figure 2.4 Expression of differentially upregulated cell-type marker genes in the posterior cingulate network (Network C) and anterior cingulate network (Network D). Heatmaps of differentially expressed marker genes (rows) are shown for one of the six donors in the Allen Human Brain Atlas (donor 10021). Samples from different anatomical substructures within the networks are color annotated (columns). Expression was averaged across samples from an anatomical substructure with the same acronym ignoring left and right hemisphere annotations. See Supplementary Figure 2 for heatmaps for all six donors from the AHBA and Supplementary Table 7 for full names of the region-specific acronyms.

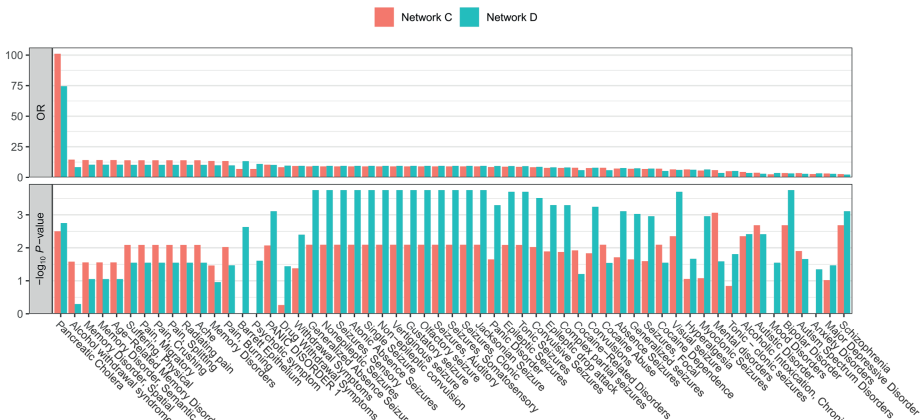


Figure 2.5 Disease associations of the posterior cingulate network (Network C) and anterior cingulate network (Network D). Differentially upregulated genes in each network were assessed for the enrichment of disease-associated genes from DisGeNET (hypergeometric test, BH-corrected $P < 0.05$). Top plot shows odds ratios (ORs) for the number of overlapping genes, and bottom plot shows the significance of overlap indicated with $-\log_{10}$ P-values (y-axis). Disorders (columns) are sorted based on highest ORs in either one of the networks.

2.4.2 *NPPA*, *SOSTDC1*, and *TYRP1*

Cholinergic marker genes *NPPA*, *SOSTDC1*, *TYRP1* were highly expressed in the posterior cingulate network and anterior cingulate network of the healthy brain compared to the other seven SCNs. While the functions of these genes likely involve cholinergic signaling, several studies suggest they also function as extracellular regulators of multiple other signaling pathways, including cAMP, Wnt, and β-catenin signaling [32–37].

NPPA (natriuretic peptide precursor A) and other natriuretic peptides are thought to be involved in a wide range of functions, including neurovascular functions, blood-brain barrier, brain homeostasis, neuroprotection, and synaptic transmission by regulating the release and re-uptake of neurotransmitters such as noradrenalin, dopamine and glycine [38]. Impaired function of natriuretic peptides in brains of AD patients could accelerate neurodegeneration and may impair structural integrity of the brain leading to a higher risk of cognitive decline [39]. Our results suggest that *NPPA* might similarly be involved in PD pathogenesis given its high expression within the anterior and posterior cingulate networks.

SOSTDC1 (sclerostin domain-containing 1) is known as a negative regulator of bone morphogenetic protein (BMP) and Wnt-signaling, but recent studies also show that *SOSTDC1* regulates natural killer cell maturation and cytotoxicity [35]. An increased number of natural killer cells have been found in PD, but the actual relevance with PD risk is still unclear [40]. The BMP signaling pathway promotes the development of midbrain dopaminergic neurons [41], in which *SOSTDC1* may play a role. Furthermore, *SOSTDC1* was upregulated in the striatum of Parkinsonian rats that were treated by subthalamic nucleus high frequency stimulation, and is therefore suggested to have neuroprotective effects [42].

TYRP1 (tyrosinase-related protein 1) produces melanocytes-specific proteins involved in the biosynthesis of melanin in brain, skin and eyes [43,44]. Melanoma and PD share genes involved in the synthesis of melanin and dopamine, including *SNCA* which encodes the α-synuclein protein found in Lewy bodies [45]. Furthermore, neuromelanin is produced almost exclusively in human catecholaminergic neurons and is responsible for the pigmentation of dopaminergic neurons of the substantia nigra, and noradrenergic neurons of the locus cereleus [46]. It is considered to be protective due to its ability to chelate metals, especially iron which increases with age [46].

2.4.3 Disease-associations

The posterior and anterior cingulate networks shared similar highly expressed genes and were likewise associated with similar diseases. Both SCNs represent anatomical networks that function normally in healthy brains, but their activity is reduced in aging and PD [2,21]. As part of the default mode network, both the posterior and anterior cingulate cortex have been shown to be dysregulated in neuropsychiatric disorders [47,48]. Based on our analysis of transcriptomic signatures in the healthy brain, we found that the posterior cingulate network showed stronger associations with memory and pain-related disorders compared to the anterior cingulate networks which showed stronger associations with mental and

neuropsychiatric disorders. Our findings suggest that genes involved in multiple signaling pathways, such as serotonin, GPCR, GABA, glutamate, and RAS, contribute to healthy functions of the posterior and anterior cingulate networks.

2.5 Acknowledgments

We thank L. E. Jonkman for her critical insight on the manuscript. This research received funding from The Netherlands Technology Foundation (STW), as part of the STW project 12721 (Genes in Space). O. Dzyubachyk received funding from The Dutch Research Council (NWO) project 17126 (3DOmics). W. D. J. van de Berg received funding from Alzheimer Netherlands (ISAO #14536) and LECMA (#14797) to study transcriptome datasets in the context of Parkinson's and Alzheimer's disease and was financially supported by grants from Amsterdam Neuroscience; Dutch Research council (ZonMW); Stichting Parkinson Fonds; Alzheimer association; the MJ Fox foundation and Rotary Aalsmeer-Uithoorn. W. D. J. van de Berg performed contract research and consultancy for Hoffmann-La Roche; Lysosomal Therapeutics; CHDR; Cross beta Sciences and received research consumables from Hoffmann-La Roche and Prothena. Prof. J.J. van Hilten received grants from Alkemade-Keuls Foundation; Stichting Parkinson Fonds (Optimist Study); The Netherlands Organisation for Health Research and Development (#40-46000-98-101); NWO (#628.004.001); Hersenstichting; AbbVie; Hoffmann-La-Roche; Lundbeck; and Centre of Human Drug Research outside the submitted work.

References

[1] M. Goedert, M. G. Spillantini, K. Del Tredici, and H. Braak, "100 years of Lewy pathology," *Nat. Rev. Neurol.* 9, 13–24 (2012).

[2] L. J. de Schipper, J. van der Grond, J. Marinus, J. M. L. Henselmans, and J. J. van Hilten, "Loss of integrity and atrophy in cingulate structural covariance networks in Parkinson's disease," *NeuroImage Clin.* 15, 587–593 (2017).

[3] O. Lucas-Jiménez, N. Ojeda, J. Peña, M. Díez-Cirarda, A. Cabrera-Zubizarreta, J. C. Gómez-Esteban, M. Á. Gómez-Beldarrain, and N. Ibarretxe-Bilbao, "Altered functional connectivity in the default mode network is associated with cognitive impairment and brain anatomical changes in Parkinson's disease," *Park. Relat. Disord.* 33, 58–64 (2016).

[4] M. Wang, S. Jiang, Y. Yuan, L. Zhang, J. Ding, J. Wang, J. Zhang, K. Zhang, and J. Wang, "Alterations of functional and structural connectivity of freezing of gait in Parkinson's disease," *J. Neurol.* 263, 1583–1592 (2016).

[5] J. P. M. van der Vegt, B. F. L. van Nuenen, B. R. Bloem, C. Klein, and H. R. Siebner, "Imaging the impact of genes on Parkinson's disease," *Neuroscience* 164, 191–204 (2009).

[6] D. Aarsland, B. Creese, M. Politis, K. R. Chaudhuri, D. H. Ffytche, D. Weintraub, and C. Ballard, "Cognitive decline in Parkinson disease," *Nat. Rev. Neurol.* 13, 217–231 (2017).

[7] F. Sampedro, J. Marín-Lahoz, S. Martínez-Horta, J. Pagonabarraga, and J. Kulisevsky, "Reduced gray matter volume in cognitively preserved COMT ^{158}Val/Val Parkinson's disease patients and its association with cognitive decline," *Brain Imaging Behav.* (2019).

[8] S. E. Winder-Rhodes, A. Hampshire, J. B. Rowe, J. E. Peelle, T. W. Robbins, A. M. Owen, and R. A. Barker, "Association between *MAPT* haplotype and memory function in patients with Parkinson's disease and healthy aging individuals," *Neurobiol. Aging* 36, 1519–1528 (2015).

[9] A. Arnatkevičiūtė, B. D. Fulcher, and A. Fornito, "A practical guide to linking brain-wide gene expression and neuroimaging data," *Neuroimage* 189, 353–367 (2019).

[10] M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C.-K. K. Lee, et al., "Canonical genetic signatures of the adult human brain," *Nat. Neurosci.* 18, 1832–1844 (2015).

[11] B. S. Freeze, D. Acosta, S. Pandya, Y. Zhao, and A. Raj, "Regional expression of genes mediating trans-synaptic alpha-synuclein transfer predicts regional atrophy in Parkinson disease," *NeuroImage Clin.* 18, 456–466 (2018).

[12] T. Rittman, M. Rubinov, P. E. Vértes, A. X. Patel, C. E. Ginestet, B. C. P. Ghosh, R. A. Barker, M. G. Spillantini, E. T. Bullmore, et al., "Regional expression of the *MAPT* gene is associated with loss of hubs in brain networks and cognitive impairment in Parkinson disease and progressive supranuclear palsy," *Neurobiol. Aging* 48, 153–160 (2016).

[13] A. Alexander-Bloch, J. N. Giedd, and E. Bullmore, "Imaging structural co-variance between human brain regions," *Nat. Rev. Neurosci.* 14, 322–336 (2013).

[14] E. M. Coppen, J. van der Grond, A. Hafkemeijer, S. A. R. B. Rombouts, and R. A. C. Roos, "Early grey matter changes in structural covariance networks in Huntington's disease," *NeuroImage Clin.* 12, 806–814 (2016).

[15] C. W. Huang, S. W. Hsu, S. J. Tsai, N. C. Chen, M. E. Liu, C. C. Lee, S. H. Huang, W. N. Chang, Y. T. Chang, et al., "Genetic effect of interleukin-1 beta (C-511T) polymorphism on the structural covariance network and white matter integrity in Alzheimer's disease," *J. Neuroinflammation* 14, 1–13 (2017).

[16] R. N. Spreng and G. R. Turner, "Structural Covariance of the Default Network in Healthy and Pathological Aging," *J. Neurosci.* 33, 15226–15234 (2013).

[17] A. Alexander-Bloch, A. Raznahan, E. Bullmore, and J. Giedd, "The convergence of maturational change and structural covariance in human cortical networks," *J. Neurosci.* 33, 2889–2899 (2013).

[18] F. Liu, H. Tian, J. Li, S. Li, and C. Zhuo, "Altered voxel-wise gray matter structural brain networks in schizophrenia: Association with brain genetic expression pattern," *Brain Imaging Behav.* 13, 493–502 (2019).

[19] Y. Yee, D. J. Fernandes, L. French, J. Ellegood, L. S. Cahill, D. A. Vousden, L. Spencer Noakes, J. Scholz, M. C. van Eede, et al., "Structural covariance of brain region volumes is associated with both structural connectivity and transcriptomic similarity," *Neuroimage* 179, 357–372 (2018).

[20] R. Romero-Garcia, K. J. Whitaker, F. Váša, J. Seidlitz, M. Shinn, P. Fonagy, R. J. Dolan, P. B. Jones, I. M. Goodyer, et al., "Structural covariance networks are coupled to expression of genes enriched in supragranular layers of the human cortex," *Neuroimage* 171, 256–267 (2018).

[21] A. Hafkemeijer, I. Altmann-schneider, A. J. M. De Craen, P. E. Slagboom, J. van der Grond, and S. A. R. B. Rombouts, "Associations between age and gray matter volume in anatomical brain networks in middle-aged to older adults," *Aging Cell* 13, 1068–1074 (2014).

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

[23] Z. Gu, R. Eils, and M. Schlesner, "Complex heatmaps reveal patterns and correlations in multidimensional genomic data," 2847–2849 (2016).

[24] B. O. Mancarci, L. Toker, S. J. Tripathy, B. Li, B. Rocco, E. Sibille, and P. Pavlidis, "Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data," *eNeuro* 4, 1–20 (2017).

[25] J. Piñero, À. Bravo, N. Queralt-Rosinach, A. Gutiérrez-Sacristán, J. Deu-Pons, E. Centeno, J. García-García, F. Sanz, and L. I. Furlong, "DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Res.* 45, D833–D839 (2017).

[26] V. Bonifati, "Genetics of Parkinson's disease – state of the art, 2013," *Parkinsonism Relat.*

Disord. 20, S23–S28 (2014).

[27] M. A. Nalls, N. Pankratz, C. M. Lill, C. B. Do, D. G. Hernandez, M. Saad, A. L. Destefano, E. Kara, J. Bras, et al., "Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease," *Nat. Publ. Gr.* 46, 989–993 (2014).

[28] D. Chang, M. A. Nalls, I. B. Hallgrímsdóttir, J. Hunkapiller, M. van der Brug, F. Cai, G. A. Kerchner, G. Ayalon, B. Bingol, et al., "A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci," *Nat. Genet.* 49, 1511–1516 (2017).

[29] E. C. Ballinger, M. Ananth, D. A. Talmage, and L. W. Role, "Basal forebrain cholinergic circuits and signaling in cognition and cognitive Decline," *Neuron* 91, 1199–1218 (2016).

[30] A. J. Granger, N. Mulder, A. Saunders, and B. L. Sabatini, "Cotransmission of acetylcholine and GABA," *Neuropharmacology* 100, 40–46 (2015).

[31] J. J. D. Buendia, L. Tiroshi, W. Chiu, and J. A. Goldberg, "Selective remodeling of glutamatergic transmission to striatal cholinergic interneurons after dopamine depletion," *Eur. J. Neurosci.* 49, 824–833 (2019).

[32] M. K. Kutchko and J. Siltberg-Liberles, "Metazoan innovation : from aromatic amino acids to extracellular signaling," *Amino Acids* 45, 359–367 (2013).

[33] Y. Bansho, J. Lee, E. Nishida, and M. Nakajima-Koyama, "Identification and characterization of secreted factors that are upregulated during somatic cell reprogramming," *FEBS Lett.* 591, 1584–1600 (2017).

[34] B. M. Brenner, B. J. Ballermann, M. E. Gunning, and M. L. Zeidel, "Diverse biological actions of atrial natriuretic peptide," *Physiol. Rev.* 70, 665–699 (1990).

[35] A. J. Millan, S. R. Elizaldi, E. M. Lee, J. O. Aceves, D. Murugesh, G. G. Loots, and J. O. Manilay, "Sostdc1 regulates NK cell maturation and cytotoxicity," *J. Immunol.* 202, 2296–2306 (2019).

[36] P. De Vito, "Atrial natriuretic peptide: An old hormone or a new cytokine?," *Peptides* 58, 108–116 (2014).

[37] T. Hirobe, "How are proliferation and differentiation of melanocytes regulated?," *Pigment Cell Melanoma Res.* 24, 462–478 (2011).

[38] S. Mahinrad, A. J. M. de Craen, S. Yasar, D. van Heemst, and B. Sabayan, "Natriuretic peptides in the central nervous system: Novel targets for cognitive impairment," *Neurosci. Biobehav. Rev.* 68, 148–156 (2016).

[39] S. Mahinrad, M. Bulk, I. van der Velpen, A. Mahfouz, W. van Roon-Mom, N. Fedarko, S. Yasar, B. Sabayan, D. van Heemst, et al., "Natriuretic peptides in post-mortem brain tissue and cerebrospinal fluid of non-demented humans and Alzheimer's disease patients," *Front. Neurosci.* 12, 1–12 (2018).

[40] S. Jiang, H. Gao, Q. Luo, P. Wang, and X. Yang, "The correlation of lymphocyte subsets, natural killer cell, and Parkinson's disease: a meta-analysis," *Neurol. Sci.* 38, 1373–1380 (2017).

[41] V. M. Jovanovic, A. Salti, H. Tilleman, K. Zega, M. M. Jukic, H. Zou, R. H. Friedel, N. Prakash, S. Blaess, et al., "BMP/SMAD pathway promotes neurogenesis of midbrain dopaminergic neurons in vivo and in human induced pluripotent and neural stem cells," *J. Neurosci.* 38, 1662–1676 (2018).

[42] S. Lortet, E. Lacombe, N. Boulanger, P. Rihet, C. Nguyen, L. K. Le Goff, and P. Salin, "Striatal Molecular Signature of Subchronic Subthalamic Nucleus High Frequency Stimulation in Parkinsonian Rat," *PLoS One* 8, e60447 (2013).

[43] N. Wang and D. N. Hebert, "Tyrosinase maturation through the mammalian secretory pathway: Bringing color to life," *Pigment Cell Res.* 19, 3–18 (2006).

[44] H. Lu, L. Li, E. R. Watson, R. W. Williams, E. E. Geisert, M. M. Jablonski, and L. Lu, "Complex interactions of Tyrp1 in the eye.," *Mol. Vis.* 17, 2455–2468 (2011).

[45] T. Pan, J. Zhu, W. J. Hwu, and J. Jankovic, "The role of alpha-synuclein in melanin synthesis

in melanoma and dopaminergic neuronal cells," *PLoS One* 7, 3–10 (2012).

[46] B. Pavan and A. Dalpiaz, "Odorants could elicit repair processes in melanized neuronal and skin cells," *Neural Regen. Res.* 12, 1401–1404 (2017).

[47] D. Öngür, M. Lundy, I. Greenhouse, A. K. Shinn, V. Menon, B. M. Cohen, and P. F. Renshaw, "Default mode network abnormalities in bipolar disorder and schizophrenia," *Psychiatry Res. - Neuroimaging* 183, 59–68 (2010).

[48] S. J. Broyd, C. Demanuele, S. Debener, S. K. Helps, C. J. James, and E. J. S. Sonuga-Barke, "Default-mode brain dysfunction in mental disorders: A systematic review," *Neurosci. Biobehav. Rev.* 33, 279–296 (2009).

Chapter 3
Co-expression patterns between *ATN1* and *ATXN2* coincide with brain regions affected in Huntington's disease

Arlin Keo

N. Ahmad Aziz

Oleh Dzyubachyk

Jeroen van der Grond

Willeke M.C. Roon-Mom

Boudewijn P.F. Lelieveldt

Marcel J.T. Reinders

Ahmed Mahfouz

This chapter was published in: *Front. Mol. Neurosci.* 10, 1–13 (2017), doi: 10.3389/fnmol.2017.00399.
Supplementary material is available online at: https://www.frontiersin.org/articles/10.3389/fnmol.2017.00399/full#supplementary-material

Abstract

Cytosine-adenine-guanine (CAG) repeat expansions in the coding regions of nine polyglutamine (polyQ) genes (*HTT*, *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *ATN1*, *AR*, and *TBP*) are the cause of several neurodegenerative diseases including Huntington's disease (HD), six different spinocerebellar ataxias (SCAs), dentatorubral-pallidoluysian atrophy, and spinobulbar muscular atrophy. The expanded CAG repeat length in the causative gene is negatively related to the age-at-onset (AAO) of clinical symptoms. In addition to the expanded CAG repeat length in the causative gene, the normal CAG repeats in the other polyQ genes can affect the AAO, suggesting functional interactions between the polyQ genes. However, there is no detailed assessment of the relationships among polyQ genes in pathologically relevant brain regions.

We used gene co-expression analysis to study the functional relationships among polyQ genes in different brain regions using the Allen Human Brain Atlas (AHBA), a spatial map of gene expression in the healthy brain. We constructed co-expression networks for seven anatomical brain structures, as well as a region showing a specific pattern of atrophy in HD patients detected by magnetic resonance imaging (MRI) of the brain. In this HD-associated region, we found that *ATN1* and *ATXN2* were co-expressed and shared co-expression partners which were enriched for DNA repair genes. We observed a similar co-expression pattern in the frontal lobe, parietal lobe, and striatum in which this relation was most pronounced. Given that the co-expression patterns for these anatomical structures were similar to those for the HD-associated region, our results suggest that their disruption is likely involved in HD pathology. Moreover, *ATN1* and *ATXN2* also shared many co-expressed genes with *HTT*, the causative gene of HD, across the brain. Although this triangular relationship among these three polyQ genes may also be dysregulated in other polyQ diseases, stronger co-expression patterns between *ATN1* and *ATXN2* observed in the HD-associated region, especially in the striatum, may be more specific to HD.

3.1 INTRODUCTION

Polyglutamine (polyQ) diseases are a family of nine neurodegenerative disorders caused by a cytosine-adenine-guanine (CAG) trinucleotide repeat expansion in the coding region of one of the polyQ disease-associated genes. PolyQ diseases include Huntington's disease (HD), six spinocerebellar ataxias (SCAs), dentatorubral-pallidoluysian atrophy (DRPLA), and spinobulbar muscular atrophy (SBMA), each with its own causative gene: *HTT*, *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *ATN1*, and *AR*, respectively [1]. With the exception of SBMA, all polyQ diseases are inherited in an autosomal dominant manner [1]. The CAG repeat region is translated into a stretch of glutamine amino acids, also referred to as the polyQ tract [2,3]. In HD patients, the polyQ expansion in the huntingtin protein causes neurodegeneration that affects the striatum most severely and results in cognitive, psychiatric as well as motor disturbances and gait abnormalities [4,5]. It is thought that the expansion of the polyQ tract causes the protein to misfold and aggregate, and therefore loses its normal function. Hence, these mutant proteins also become toxic components as they trigger the misfolding of other proteins [6].

In HD and SCAs, longer CAG repeat lengths in the causative polyQ genes are associated with an earlier age-at-onset (AAO) [7,8]. In HD, up to 75% of the variability in AAO can be explained by the *HTT* CAG repeat length [7], while in SCA1, SCA2, SCA3, SCA6, and SCA7, the CAG repeat in the causative gene explains between 32 and 80% of the AAO variability [8]. There is also evidence suggesting that, in addition to the expanded CAG repeat length in the causative gene, the normal CAG repeat lengths in other non-causative polyQ genes affect the AAO in HD and SCAs [7–12]. For example, clinical observations show that the AAO in HD is not only affected by the expanded CAG repeat length in *HTT*, but also by the normal CAG repeat length in *ATN1* and *ATXN1* negatively affecting the AAO [7]. The observation that the number of CAG repeats in multiple polyQ genes can affect the AAO in HD or SCAs suggests that polyQ gene products are functionally interacting.

In addition to the suggested genetic associations based on shared effects on AAO, similar mechanisms likely contribute to polyQ disease pathogenesis. These mechanisms include misfolding of the disease protein, deleterious protein interactions, transcriptional dysregulation, mitochondrial dysfunction, aberrant neuronal signaling, cellular protein homeostasis impairment, and RNA toxicity [1]. Despite the similarities, these mechanisms seem to affect specific brain regions depending on the particular polyQ disease. To get a better understanding of how the CAG expansion in *HTT* affects the brain, multiple studies used genome-wide expression analysis of post-mortem samples collected from different brain regions. By comparing the brain region-specific expression profiles of HD patients and healthy controls, the highest number of differentially expressed genes was found in the caudate nucleus and to a lesser extent in the cerebellum [13]. This demonstrates that differential gene expression in the HD brain has a regional pattern corresponding to the known pattern of neuropathology. A more recent study used longitudinal RNA-sequencing expression data of HD knock-in mice with increasing CAG repeat lengths, and using weighted gene co-expression network analysis (WGCNA), they studied length-dependence

of molecular networks in HD-relevant brain regions [14]. They found one striatal module which showed strong patterns of downregulation that were both CAG length- and age-dependent.

In addition to patient postmortem samples and mouse models, characterizing expression patterns of wild-type genes in the healthy brain can provide useful insights into the molecular and pathological changes regarding polyQ diseases. Using gene expression data from ten different brain tissues and 101 healthy individuals from the UK Brain Expression Consortium [15], Bettencourt et al. analyzed the co-expression relationships of SCA genes in the healthy brain using WGCNA [16]. Two cerebellar modules were enriched in SCA transcripts of which one module seemed preserved across all brain regions and the other module seemed unique to the cerebellum. The study suggests that genes co-expressed with SCA genes across all brain regions give rise to more complex phenotypes (other ataxia syndromes and neurodegenerative disorders), while cerebellum-specific co-expressing genes results in pure ataxia phenotypes. While previous studies have focused on region-specific gene expression differences between patients and healthy controls [13,17], and co-expression relationships between SCA genes in the healthy brain [16], there is no detailed assessment of the relationships among polyQ genes in pathologically relevant regions of the brain.

We aim to find common mechanisms through which the nine polyQ genes could interact with each other in order to understand how these interactions could affect neuropathology in polyQ disorders. Our approach consisted of interrogating healthy brain gene expression data from the Allen Human Brain Atlas (AHBA) [18] as its high spatial resolution allows localizing interactions to specific brain regions. PolyQ genes with similar expression patterns in a specific brain area suggest their co-involvement in functions specific to that brain area. Using the AHBA to relate biological functions to disease and pinpoint pathways to specific regions of the brain has been reported previously, for example through evaluation of normal activity in the brain of genes associated to migraine [19], and autism [20–22]. Here, we follow a similar approach to detect functional relationships between the nine polyQ genes. We focus on a brain region consisting of several brain areas which were recently shown to be most severely affected in HD through an MRI-guided unbiased approach [23]. In addition, the analysis was repeated for seven anatomical brain structures to assess whether the relationships between polyQ genes within the HD-associated region are reflected in other anatomical brain structures (Figure 3.1). We analyzed gene co-expression networks within the different brain regions to identify region-specific relations among polyQ genes and their underlying functional mechanisms.

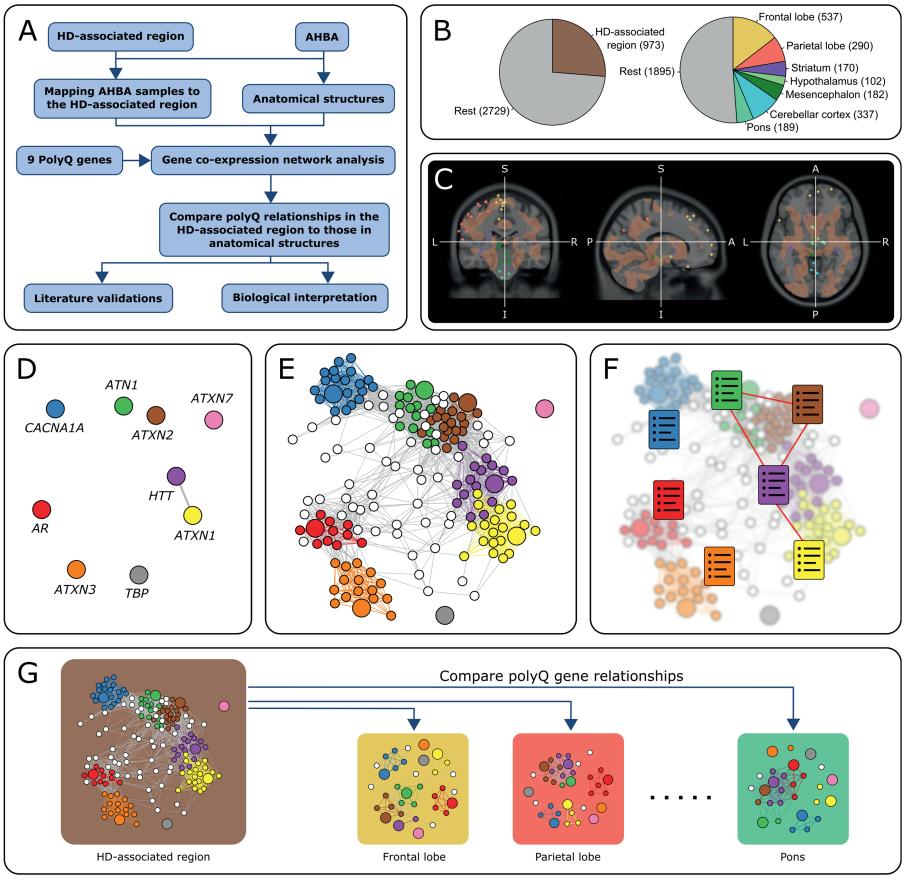


Figure 3.1 Overview of the approach. (A) Flowchart of the general approach in this study using co-expression analysis. (B) Number of samples in all six donors. The left chart shows the number of samples in the HD-associated region and the rest of the brain. The right chart shows the number of samples in seven anatomical brain structures and the rest of the brain. There are 3,702 samples in total. (C) Mapping AHBA samples to the HD-associated region. AHBA samples were spatially mapped to the HD-associated region (brown) which was based on an MRI study that looked at gray matter changes between HD patients and controls [23]. In this MRI image, we discarded all the samples that did not belong to one of the predefined regions. The colored points correspond to samples collected from anatomical structures that fall either inside or outside the HD-associated region. (D-F) Co-expression networks for the HD-associated region and each anatomical structure were constructed by retaining links between strongly correlated gene pairs (Pearson's $r > 0.5$). PolyQ genes are colored and have a larger node size. The relationships between polyQ genes were analyzed based on three methods: (D) Direct co-expression between polyQ genes. (E) Subnetworks of genes co-expressed with polyQ genes. Nodes are colored according to the co-expressed polyQ gene. (F) Functional enrichment of the set of genes co-expressed with each of the polyQ genes (colored charts). Functional links between polyQ genes were assessed based on the overlap of enriched functional terms. (G) PolyQ gene relationships in the HD-associated region were then compared to those in networks constructed for different anatomical structures.

3.2 Material and methods

3.2.1 Spatial gene expression in the healthy human brain

We used the publicly available gene expression data set from AHBA [18] to exploit possibilities to study neurodegenerative diseases in the healthy brain. The spatially high-resolution data set contains genome-wide microarray expression of 3,702 samples collected from 6 healthy adult donors (5 males and 1 female, mean age 42, range 24-57 years). For each brain, 363-946 samples are available. In two out of the six brains, samples were collected from both hemispheres, while in the remaining four brains, samples were collected from the left hemisphere only.

The expression of several genes in the AHBA was assessed using multiple probes. We selected one probe per gene as follows: (1) if there was one probe for a gene, this probe was selected; (2) if there were two probes for a gene, we selected the probe with the highest variance (measured per brain and then averaged across the six brains); (3) if there were more than two probes for a gene, we chose the probe with the highest connectivity (measured as the sum of the Pearson correlations per brain and then averaged across the six brains). These steps resulted in 19,992 genes selected for further analysis.

3.2.2 Mapping AHBA samples to the HD-associated region

In our analysis, we included a brain region affected in HD patients compared to healthy controls defined according to gray matter changes in MRI scans [23]. Several structural co-variance networks (SCNs), corresponding to discontinuous brain areas, were associated with HD. The SCN analysis was applied on whole brain images rather than predefined regions [24] to reveal a map of regions affected in HD. Two disjoint brain areas, referred to as the caudate nucleus and hippocampal network, showed strong significant association with pre-manifest HD and manifest HD patients. The caudate nucleus network includes the nucleus accumbens, pallidum, putamen, and precuneus. The hippocampal network includes the parahippocampal gyrus, cerebellum, pallidum, and planum polare. We combined these two regions into one, hereafter named "HD-associated region". AHBA samples were mapped to the HD-associated region using the MNI coordinate space. Based on this mapping, we identified AHBA samples located inside the HD-associated region that exhibits significant gray matter volume changes in HD compared to healthy controls.

3.2.3 Differential gene expression in the HD-associated region

We examined gene expression differences between samples located inside and outside the HD-associated region. Differential expression was assessed for each brain separately using two one-tailed Mann-Whitney *U*-tests to identify up- and downregulated genes. Genes with a *P*-value lower than 0.025 in either the upper tail or lower tail in five out of six brains were considered differentially expressed.

3.2.4 Region-specific co-expression analysis

We used Pearson's correlation (*r*) as a measure of co-expression between two genes. Gene pairs showing a co-expression greater than 0.5 were considered to be related as they have similar expressions across tissue-specific samples. The co-expression threshold was based

on our observation of the polyQ co-expression distributions in different regions (Supplementary Figure 1). For each polyQ gene, a correlation threshold of 0.5 on average coincides with selecting the top 5% most co-expressed genes. Similar patterns of the number of co-expressed genes were observed when the co-expression threshold was set at 0.4 and 0.6 (Supplementary Figure 2). Co-expression was calculated for all possible gene pairs across samples inside the HD-associated region as well as samples representing one of the following brain structures: frontal lobe, parietal lobe, striatum, hypothalamus, mesencephalon, cerebellar cortex, and pons. Co-expression was first calculated per donor and then averaged across all donors to one consensus matrix containing all pairwise gene co-expressions.

3.2.5 FUNCTIONAL ENRICHMENT OF CO-EXPRESSED GENES

We assessed enrichment in functional terms for sets of genes using DAVID Bioinformatics resources [25,26]. Enrichment analysis was done for genes differentially expressed in the HD-associated region (up- and downregulated) and sets of genes co-expressed with one of the nine polyQ genes. We selected the following annotation categories: GOTERM_BP_ALL, GOTERM_MF_ALL, GOTERM_CC_ALL. Enrichment analysis was done using Entrez IDs as gene identifiers and all 19,992 genes in the AHBA were added as a background list. Functional terms were selected when the Benjamini-Hochberg (BH) corrected *P*-value was lower than 0.05 and when at least two genes were present in the gene ontology category. To summarize the functional terms enriched in genes differentially expressed between the HD-region and the rest of the brain, we used the clustering function from DAVID and retained only terms with a cluster enrichment score > 2.

3.2.6 POLYQ INTERACTIONS BASED ON CO-EXPRESSION

We used three different representations of the relationships between polyQ genes, each highlighting a different way of how polyQ genes could be related. (1) Direct co-expression: co-expression between two polyQ genes indicates a strong interaction based on their spatial expression patterns. (2) Shared co-expression: a significant overlap of the two co-expressed gene sets of two polyQ genes suggests that these two polyQ genes are indirectly related to each other through other genes with which they interact. (3) Functional overlap: if two polyQ genes indirectly interact, the overlap between the functional terms enriched in their corresponding sets of co-expressed genes point out whether the two polyQ genes are also functionally related. We assessed the significance of shared co-expressed genes between two polyQ genes using one-tailed Fisher's exact test. We considered a functional overlap between two polyQ genes when their respective gene sets share at least 10 enriched functional terms.

3.2.7 NETWORK ANALYSIS

The polyQ gene network for the HD-associated region was visualized using Gephi [27] and the rgexf R-package. In the network, nodes represent polyQ genes (each with a unique color) as well as genes co-expressed with polyQ genes (each assigned the color of the polyQ gene with which it is co-expressed). The node color is mixed when genes are co-expressed with

multiple polyQ genes. Edges are colored by the polyQ gene that they connect to. We used 'Fruchterman-Reingold' and 'label adjust' options for the layout. The circular co-expression plots were visualized using Cytoscape [28] and the RCy3 R-package [29]. Each node represents one of the nine polyQ genes and node colors indicate their mean expression across samples specific to the indicated regions per donor and then averaged across the six brains. Expression levels increase from blue to white to red. Edge thickness indicates the co-expression strength between two genes, the number of shared co-expressed genes, or number of shared functional terms between two polyQ gene sets.

3.2.8 Enrichment of ubiquitin ligases and DNA binding genes

Genes encoding ubiquitin conjugating enzymes (UBE2) were downloaded from the HUGO Nomenclature Committee [30] and includes 41 genes. DNA repair and ubiquitination gene sets, consisting of 125 and 40 genes, respectively, were obtained from the Molecular Signature Database (MSigDB) [31].

3.3 Results

3.3.1 *ATXN2* is differentially expressed in the HD-associated region

To assess normal gene expression changes in regions affected by HD, we examined the differential expression of genes in AHBA samples split according to being inside or outside the HD-associated region. Among the polyQ genes under study, only *ATXN2* was significantly downregulated in five out of six brains (Supplementary Figure 3). All other polyQ genes did not show consistent expression changes across the six brains. The analysis yielded 2,812 (14.1% of all 19,992 genes) significant genes (P-value < 0.025; one-sided Mann-Whitney U-test; Bonferroni-corrected). Out of the differentially expressed genes, 711 genes showed a significant higher expression in the HD-associated region (upregulated) and 2,101 genes showed a lower expression (downregulated). The set of downregulated genes was enriched in terms related to cytoplasm, mitochondrial processes, cellular component organization, DNA damage recognition, synapses, autophagy, and metabolic processes (BH-corrected P < 0.05; term cluster enrichment score > 2; Supplementary Table 1). These biological functions have been implicated in HD before [5,13,32–34]. Interestingly, there were no functional terms enriched in the set of upregulated genes. From Supplementary Figure 3 it can also be inferred that *ATXN2* had medium expression both inside and outside the HD-associated region in all six brains. Compared to all polyQ genes, *ATN1* had the highest expression across the brain, while *ATXN3* had the lowest expression.

3.3.2 *ATN1*, *ATXN2* and *HTT* have the highest connectivity among polyQ genes

We examined the functional relations among the polyQ genes by analyzing their spatial co-expression using the high-resolution AHBA. Co-expression of the polyQ genes was analyzed in the healthy brain using the AHBA samples within regions associated with gray matter changes in HD patients (HD-associated region). In addition, we analyzed seven anatomical brain structures to assess whether polyQ gene co-expression patterns in the HD-associated

region are reflected within these structures. All anatomical structures had a sufficient number of samples to perform co-expression analysis (Table 3.1).

Table 3.1 Number of samples for the HD-associated region and seven anatomical brain structures across donors.

Donor	HD-associated region	Frontal lobe	Parietal Lobe	Striatum	Hypo-thalamus	Mesen-cephalon	Cerebellar cortex	Pons
9861	280	161	81	48	9	48	41	52
10021	183	135	57	46	22	61	76	50
12876	99	47	37	16	17	9	40	6
14380	139	70	42	24	20	22	44	26
15496	125	64	32	18	16	24	60	29
15697	147	60	41	18	18	18	76	26
Total samples	973	537	290	170	102	182	337	189

The HD-associated region overlaps with parts of different anatomical structures. None of the examined anatomical structures were overrepresented in the HD-associated region (Supplementary Table 2). We examined the cerebellar cortex instead of the cerebellum because of strong expression differences between the cerebellar cortex and cerebellar nuclei samples [35]. The cerebellar nuclei were left out of the co-expression analysis as they had too few samples. Since the HD-associated region showed little overlap of samples with anatomical structures, co-expression patterns in this region are not dominated by samples from any specific anatomical structures.

The number of co-expressed genes varied per polyQ gene and brain region (Figure 3.2). *ATN1* had the highest number of co-expressed genes in all examined regions except the mesencephalon and pons where only *HTT* had more co-expressed genes. The largest number of genes co-expressed with *ATN1* was observed in the striatum (3,428 genes), followed by the parietal lobe (3,332), and the frontal lobe (1,971). *ATXN2* had the second largest number of co-expressed genes in the parietal lobe, followed by the striatum and the frontal lobe. *TBP* had only one gene co-expressing in the cerebellar cortex. For all sets of genes co-expressed with polyQ genes we performed functional enrichment analysis to obtain sets of functional GO terms (BH-corrected $P < 0.05$). This was repeated for all inspected brain regions (Supplementary Figure 4). Similar patterns were observed for the number of functional terms and their respective gene set sizes, with *ATN1*, *ATXN2*, and *HTT* showing the highest number of GO terms.

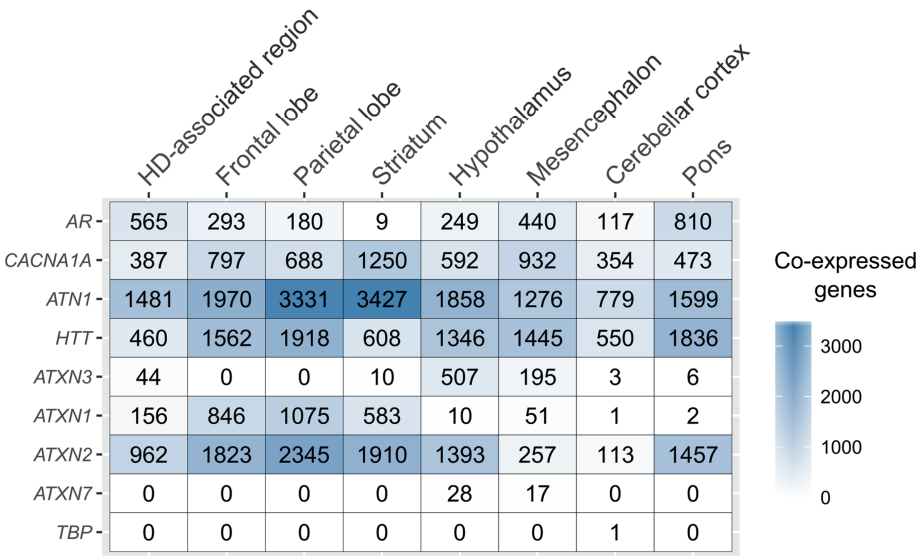


Figure 3.2 PolyQ co-expressed gene sets across the brain. The number of genes co-expressed with each of the polyQ genes (rows) for the different, anatomical brain structures (columns).

3.3.3 STRONG CONNECTIVITY BETWEEN *ATN1* AND *ATXN2* IN THE HD-ASSOCIATED REGION

The co-expression network of polyQ genes within the HD-associated region includes all genes that are co-expressed with one of the nine polyQ genes (Figure 3.3). The numbers of shared co-expressed genes for different brain regions are shown in Figure 3.4A, and their respective co-expressed genes in the HD-associated region are listed in Supplementary Table 3. The functional overlap between two polyQ genes (Figure 3.4B) was measured as the number of shared enriched functional terms between their corresponding sets of co-expressed genes (Supplementary Figure 4).

In the HD-associated region, *ATN1-ATXN2* was the only polyQ gene pair that was directly co-expressed ($r = 0.53$, Figure 3.5). The same pair also shared most (490) co-expressed genes among all polyQ gene pairs. *TBP* and *ATXN7* were not co-expressed with other genes at all. While both genes show low mean expression levels in the examined regions, the variance is similar to other polyQ genes. Their low expression levels indicate that *TBP* and *ATXN7* are less active in these brain regions, suggesting that they are not functionally related to other polyQ genes, at least in the examined regions. The *AR* and *ATXN3* gene sets showed more indirect and distant co-expressions with other polyQ genes (Supplementary Figure 5). *HTT* shared many co-expressed genes with *ATN1*, *ATXN2*, and *ATXN1* (Figure 3.4A).

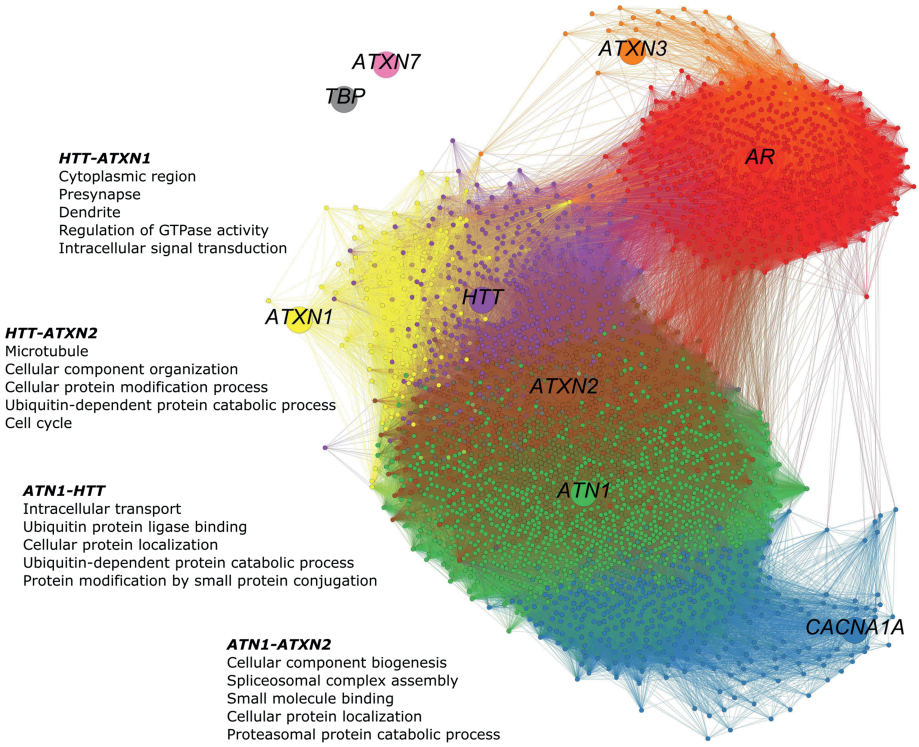


Figure 3.3 PolyQ gene co-expression network in the HD-associated region. The network consists of 3,368 nodes representing genes. The poly genes are shown as larger nodes, each with a unique color. Genes were considered co-expressed when the calculated Pearson's correlations exceed 0.5. Smaller nodes are genes that are co-expressed with at least one of the nine polyQ genes. Their colors indicate the polyQ gene with which they are co-expressed. Genes co-expressed with multiple polyQ genes have a mixed color. Edges are colored according to the colors of the nodes they connect. In the HD-associated region, *ATN1-ATXN2* is the only gene pair that is directly co-expressed, while other polyQ genes are more distantly related through indirect relationships. In this network, 41 genes co-expressed with *HTT*, *ATN1*, and *ATXN2* and 488 co-expressed with *ATN1* and *ATXN2*.

3.3.4 PolyQ co-expression relationships overlap between the HD-associated region and anatomical structures involved in HD-pathology

Co-expression relationships of polyQ genes in the HD-associated region were also observed in other anatomical structures. The direct co-expression between *ATN1* and *ATXN2* in the HD-associated region was also observed in the frontal lobe ($r = 0.53$), parietal lobe ($r = 0.52$), as well as the striatum ($r = 0.56$) that showed the highest correlation (Figure 3.5A). In the striatum, *ATN1-ATXN2* was also the only co-expressed polyQ gene pair similar to the HD-associated region. *ATN1* and *ATXN2* had the highest correlations in the striatum and they also had the highest number of overlapping co-expressed genes and functional terms in the striatum. This indicates that the co-expression patterns between *ATN1* and *ATXN2* found in the HD-associated region are particularly pronounced in the striatum.

There is a significant overlap between sets of co-expressed genes of *HTT*, *ATN1*, and *ATXN2* across all studied regions (Figure 3.4A and Figure 3.5B). When considering the overlap between sets of enriched functional GO terms (Figure 3.4B and Figure 3.5C), the triangular relationship between these three genes is only observed in the frontal and parietal lobes of the cerebral cortex. In the striatum, we observe a very strong functional overlap between *ATN1* and *ATXN2*. Together, these observations suggest a strong regional and functional relatedness between the three genes, rather than just the measured co-expression.

In addition, *ATXN1* shows significant overlap of functional terms with *HTT* in the HD-associated region, frontal lobe, parietal lobe, and striatum. *ATXN1* was also found to share functional terms with *ATN1*, *ATXN2*, and *HTT*, forming a clique of four polyQ genes in the parietal lobe. Finally, the co-expression patterns in the HD-associated region seemed to be dominated by the co-expression patterns within the frontal lobe, parietal lobe, and striatum, which were also the regions described as part of the HD-associated region in the imaging study (Supplementary Figure 6) [23].

3.3.5 *HTT*, *ATN1*, AND *ATXN2* ARE ASSOCIATED THROUGH BINDING, LOCALIZATION, AND REGULATION OF CELLULAR COMPONENTS

To gain insight into the functional relationships between *HTT*, *ATN1*, and *ATXN2*, we examined the overlap of functional terms between all three genes (Supplementary Table 4). These three genes share 15 functional terms, in the frontal and the parietal lobe, showing that they are involved in binding of cellular components. In the striatum, they are involved in the positive regulation of catalytic activity and molecular function, while in the hypothalamus, they are involved in DNA repair and cell cycle process and regulation. In the pons, they are involved in the cytoskeleton and regulation of cellular component organization. In the HD-associated region, they are involved in ubiquitin protein ligase binding, microtubule cytoskeleton, and cellular protein and macromolecule localization. In the mesencephalon and cerebellar cortex *HTT*, *ATN1*, and *ATXN2* do not share functional terms together. Together these overlapping functions between *HTT*, *ATN1*, and *ATXN2* suggest that their connectivity in the aforementioned brain regions is involved in the binding, localization, and regulation of cellular components.

We focused on genes involved in ubiquitination or DNA repair pathways, because these pathways have been previously associated with polyQ diseases [16,36,37]. In the HD-associated region, eight *UBE2* genes that are part of the ubiquitin conjugating enzymes gene family, *UBE2G2*, *UBE2I*, *UBE2C*, *CDC34*, *UBE2W*, *UBE2Z*, *UBE2E3*, and *UBE2D4*, co-expressed with polyQ genes. Furthermore, several polyQ genes co-expressed with DNA repair genes, especially *HTT*, *ATN1*, and *ATXN2*. Most DNA repair and ubiquitination genes were co-expressed with *ATN1* in the striatum (34 and 11 genes, respectively; Supplementary Figure 7). In summary, ubiquitin-related genes and DNA repair genes co-expressed with highly connected polyQ genes *HTT*, *ATN1*, and *ATXN2* in regions associated with HD.

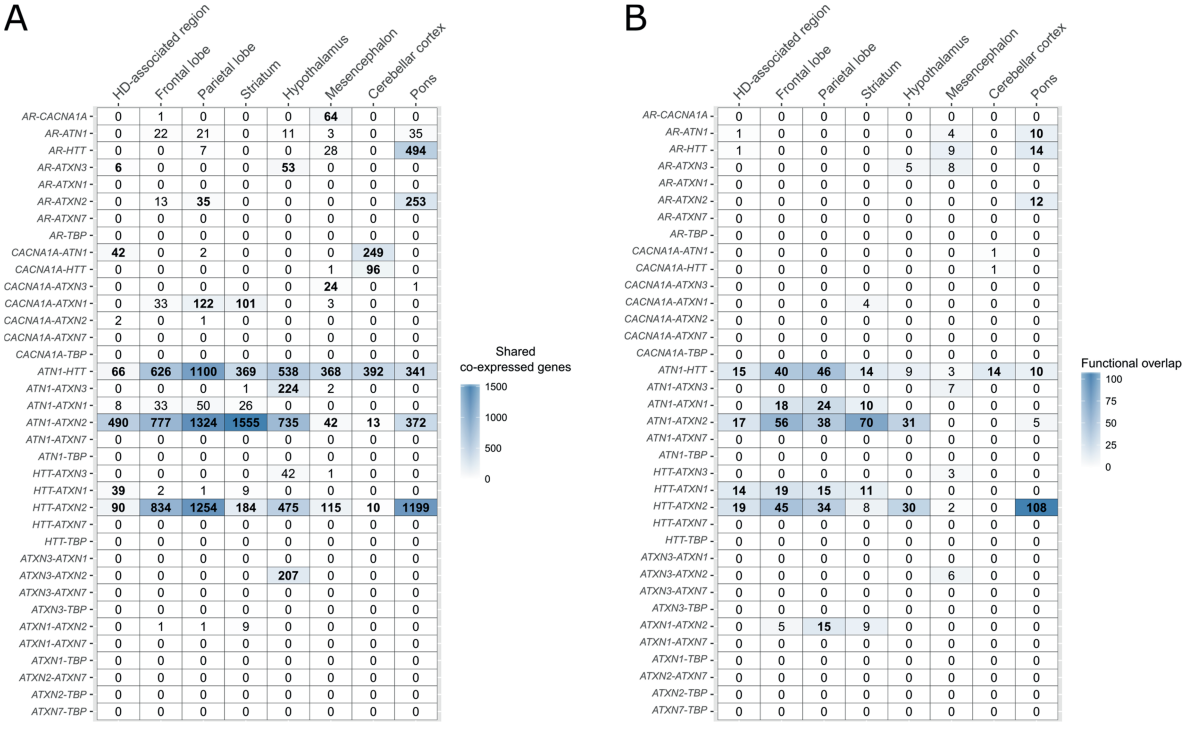


Figure 3.4 Overlap of co-expressed gene sets and functional terms between polyQ genes across the brain. (A) Number of shared co-expressed genes between each polyQ gene pair in different brain regions. Significant overlap of co-expressed genes is indicated in bold numbers (Fisher's exact test; $P < 0.05$). The respective genes co-expressed with two polyQ genes in the HD-associated region are given in Supplementary Table 3. (B) Functional overlap between each polyQ gene pair in different brain regions measured as the number of shared enriched functional terms between their corresponding sets of co-expressed genes. Functional overlaps of at least 10 functional terms are indicated in bold.

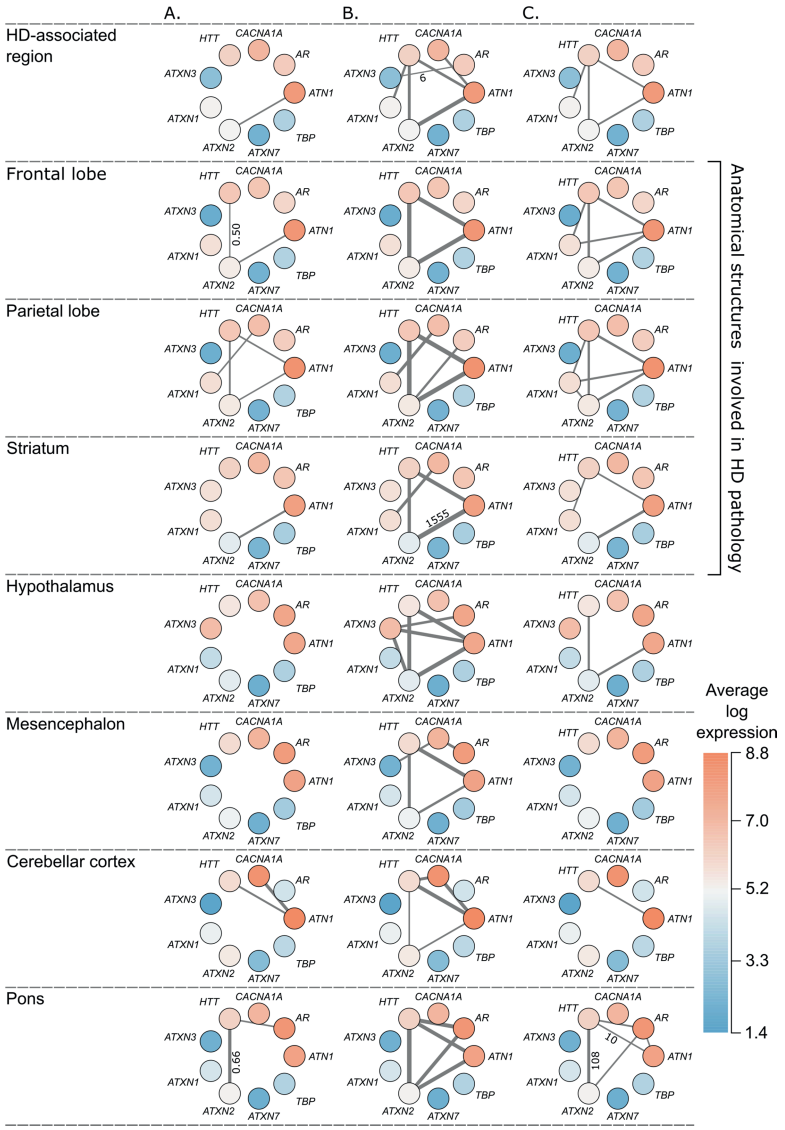


Figure 3.5 Overview of polyQ relationships in the HD-associated region and seven anatomical structures. Each node represents a polyQ gene and the color represents the expression of that polyQ gene averaged across region-specific samples and the six donors (red indicates high expression and blue low expression). Edge width indicates the strength of the detected relationship between polyQ genes. Each column represents a different type of relationship: (A) direct co-expression between polyQ genes, edge width corresponds to the co-expression value; (B) shared co-expression, edge width indicates the overlap of co-expressed genes, only for significant overlaps ($P < 0.05$); (C) functional overlap, edge width indicates the overlap size between two polyQ gene sets shown only for overlaps of at least 10 enriched terms. For each type of relationship the lowest and highest value are given along the thinnest and thickest edge.

3.4 Discussion

Polyglutamine diseases share a similar genetic basis and their phenotypes, such as AAO, are affected by CAG repeat variations in the normal range in polyQ genes other than the causative gene [7–12]. The causative polyQ genes are thought to modulate protein interactions through the role of CAG tracts in stabilizing protein interactions [38]. We analyzed the relationships between nine polyQ genes based on their co-expression within different regions of the healthy human brain.

We performed co-expression analysis in a brain region previously associated with HD specific damage in the cortical and subcortical gray matter. By combining imaging data from HD mutation carriers with spatial gene expression patterns from healthy brains, we were able to study the role of interactions between polyQ genes in the neuropathology of HD. Seven anatomical brain structures were also included in the analysis to examine regional differences compared to the HD-associated region. The anatomical structures also included the striatum, the main affected structure in HD. However, co-expression patterns in the striatum may be more specific to the structure itself rather than to HD pathology.

Among the nine polyQ genes, the relationship between *ATN1* and *ATXN2* was most pronounced based on our three representations of gene associations: it was the only gene pair that directly co-expressed, shared the most co-expressed genes, and also showed functional overlap in the HD-associated region. The same polyQ gene pair also shared co-expressed genes, and a functional overlap with *HTT*, suggesting functional connectivity with the causative gene of HD. Furthermore, out of the nine polyQ genes, only *ATXN2* had a significant lower expression in the HD-associated region compared to the rest of the brain. An expanded polyQ tract in the huntingtin protein may dysregulate pathways in which all three polyQ genes are involved. The function of *ATXN1* may also be affected in HD as this gene also shared co-expressed genes and showed functional overlap with *HTT* in the HD-associated region.

The co-expression patterns among polyQ genes within the HD-associated region overlapped with those observed within anatomical structures involved in HD neuropathology. In the striatum, the main structure affected in HD, *ATN1* and *ATXN2* were directly co-expressed and shared co-expressed genes. This suggests that this gene relationship has a more important role in HD than in other polyQ diseases. For the frontal and parietal lobes, *ATN1* and *ATXN2* were directly co-expressed and shared functional terms together with *HTT*, similar as for the patterns observed for the HD-associated region. The frontal lobe includes the primary motor cortex BA4, and is located next to the parietal lobe, which includes other motor areas. The motor cortex, involved in planning, control, and execution of voluntary movement, has been previously implicated in HD [39–42]. The motor cortex is also known to be directly connected to the caudate nucleus of the striatum [43]. In a previous study, HD gene expression profiles in the caudate nucleus and motor cortex were strikingly similar [13], suggesting that similar molecular mechanisms in different brain regions are involved in neurodegeneration. These mechanisms may include polyQ genes as we find similar co-

expression patterns in the frontal lobe and striatum between *HTT*, *ATN1*, *ATXN2*, and *ATXN1* (Figure 3.5).

There was a significant overlap in co-expressed genes between *HTT*, *ATN1*, and *ATXN2* in all examined brain regions. It would be interesting to evaluate whether this triangular relation between *HTT*, *ATN1*, and *ATXN2* is dysregulated in all polyQ diseases, especially HD, DRPLA, and SCA2. In addition, we observed strong connectivity of *ATXN2* and *AR* with *HTT* in the pons that has been described to undergo atrophy in SCA2 [44]. This relationship may possibly be disrupted in the pons of SCA2 patients. Altogether, these region-specific interactions demonstrate that co-expression analysis can reveal many interesting relations between genes based on their spatial information.

Accumulating evidence suggests that normal CAG repeat size variations in polyQ genes could act as genetic modifiers of AAO in different polyQ diseases. For example, the normal CAG repeat length in *ATXN3*, known to have a deubiquitinating function [45], has been found to have a positive effect on HD progression as it was associated with a later AAO [12]. There is also evidence for a polyQ-length dependent interference of both the mutant and normal proteins with a range of other protein binding partners [46–52]. Mutant proteins can interact either more strongly or weakly with other polyQ containing proteins and thereby inhibit there physiological function [49,52]. A case in this regard is a recent study which showed that the polyQ domain of wild-type ataxin 3 enables it to interact with beclin 1, a key initiator of autophagy, allowing the deubiquitinase activity of ataxin 3 to protect beclin 1 from proteasome-mediated degradation and thereby enabling autophagy [49]. They demonstrated that mutant huntingtin polyQ fragments competed with ataxin 3 for interaction with beclin 1 in a polyQ-length dependent manner, thereby inhibiting autophagy and contributing to neuronal dysfunction. This finding may explain the protective role of larger *ATXN3* CAG repeat sizes in HD [12]. In our analysis we did not find any relation in co-expression between *HTT* and *ATXN3* in the examined regions. This might be due to the polyQ length-dependent competing between huntingtin and ataxin 3 being relevant only with pathologically expanded polyQ tracts, which were not present in AHBA. Furthermore, atrophin 1 and the androgen receptor have been found to bind to beclin 1 in a polyQ length-dependent way. Among the polyQ genes and examined brain regions in our study, we found that *BECN1* is only co-expressed with *ATN1* in the pons. Several putative interactions with *HTT* reported in previous studies were also supported by the co-expression analysis in our study. In the HD-associated region, *ATN1* and *ATXN2* shared co-expression partners with *HTT* and both genes were also functionally related to *HTT*. For each of the genes it was shown before that variations in the normal CAG repeat length, together with the expanded CAG repeat in *HTT*, affect the AAO in HD [7]. The co-expression patterns between *HTT*, *ATN1*, and *ATXN1* suggest a functional relationship in brain regions that are involved in HD pathology.

Ubiquitin may be involved in HD pathogenesis through the relationships between *HTT*, *ATN1*, and *ATXN2*. The ubiquitin-proteasome system (UPS) has been linked extensively to

the pathogenesis of neurodegenerative diseases, including HD and other SCAs [1,6,36,53–55]. Aggregates in polyQ diseases show parts of ubiquitin and several important homeostatic proteins [56]. We found that E3 ubiquitin ligase genes *UBR4* and *CHIP/STUB1*, and ubiquitin-activating *UBA1* gene to be co-expressed with *HTT*, *ATN1*, and *ATXN2*. This suggests that the three polyQ genes contribute to the UPS in the HD-associated region of the healthy brain. Biochemical properties of *CHIP* are well-studied and this gene is known to interact with mutant polyQ proteins (including *HTT*, *ATXN1*, *ATXN3*, and *AR*) suppressing polyQ aggregation by promoting proteasomal degradation [1,2,57–59]. *CHIP* has also been implicated in several SCAs indicating it is directly involved in disease pathology [60]. The gene *UB4R* has a fundamental role in calcium signaling and neuronal survival and may contribute to neurodegenerative conditions [61]. This gene has been identified as a candidate locus in early-onset episodic ataxia [62]. It is not known whether recognition of aberrant proteins by ubiquitin ligases is beneficial or disadvantageous [63]. *UBA1* is an important regulator of cellular protein homeostasis and contributes to the pathogenesis in SBMA and HD [32,64]. Altogether, we found genes involved in the UPS that co-expressed with multiple polyQ genes which have been associated with similar pathologies in neurodegenerative disorders. The UPS is involved in protein homeostasis and may be affected early in HD through strong regional interactions. If there are changes in huntingtin protein function due to polyQ expansion, then the interaction between *ATN1* and *ATXN2* is likely to be affected. The low expression of these genes in certain brain regions could indicate their increased vulnerability to dysregulations of the UPS. We suspect co-expressed polyQ genes to have a role in AAO, as they and their shared co-expression partners seem to be involved in functions that have been associated to polyQ diseases.

There are several issues that limit a one-to-one comparison between the co-expression relationships reported in the present study and phenotype-genotype associations described previously [7–12]. First, we do not expect all phenotype-genotype associations to be explained by changes in co-expression patterns. Second, there is a wide degree of heterogeneity in both the data and methods used in the genetic association studies [7–12], limiting the possibility to combine their findings. Third, thus far, not all possible pairs of polyQ genes have been tested in the genetic association studies, highlighting the need for more comprehensive genetic association studies in larger cohorts of patients with polyQ disorders.

To validate our findings in other brain gene expression data we need a dataset with a spatial sampling resolution that allows co-expression analysis within substructures of the brain. Although the UK Brain Expression Consortium [15] sampled multiple (10) brain regions per healthy individual, the expression within a substructure, e.g., the cerebellum, is still represented by a single sample. Gene expression data of HD patients may be used to observe whether co-expression patterns are altered in disease state. However, existing datasets are particularly rich in the number of individuals they sampled and not brain regions [13]. This captures variation across individuals, while in our gene co-expression networks we capture spatial variation within a brain region of interest (e.g., striatum).

3.5 Conclusion

We showed that polyQ genes are co-expressed in the healthy brain and that their relationships are also specific to certain brain regions including a region associated with HD. Our aim was to find co-expression patterns between polyQ genes in different brain regions.The co-expression networks are likely altered in polyQ diseases due to interaction changes, especially in brain regions associated with the disease. The fact that these findings could not be validated on other expression datasets implies the importance of follow-up studies to understand more about the mechanisms behind polyQ diseases. We show that gene expression in the healthy brain may render specific regions vulnerable to expression changes based on gene co-expression networks.

3.6 Acknowledgements

We would like to thank S. M. H. Huisman for helpful discussions of the differential expression analysis. This research has received partial funding from The Netherlands Technology Foundation (STW), as part of the STW Project 12721 (Genes in Space) under the Imaging Genetics (IMAGENE) Perspective programme, and from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement 604102. NAA is supported by a VENI-grant (#91615080) from the Netherlands Organization of Scientific Research and a Marie Sklodowska-Curie Individual Fellowship grant from the European Union (Horizon 2020, #701130).

References

[1] A. J. Williams and H. L. Paulson, "Polyglutamine neurodegeneration: protein misfolding revisited," *Trends Neurosci.* 31, 521–528 (2008).

[2] N. R. Jana, P. Dikshit, A. Goswami, S. Kotliarova, S. Murata, K. Tanaka, and N. Nukina, "Co-chaperone CHIP associates with expanded polyglutamine protein and promotes their degradation by proteasomes," *J. Biol. Chem.* 280, 11635–11640 (2005).

[3] B. P. C. van de Warrenburg, R. J. Sinke, C. C. V. Bemelmans, and H. Scheffer, "Spinocerebellar ataxias in the Netherlands," *Neurology* 58, 702–708 (2002).

[4] J. S. Paulsen, J. D. Long, C. A. Ross, D. L. Harrington, C. J. Erwin, J. K. Williams, H. J. Westervelt, H. J. Johnson, E. H. Aylward, et al., "Prediction of manifest Huntington's disease with clinical and imaging measures: A prospective observational study," *Lancet Neurol.* 13, 1193–1201 (2014).

[5] D. C. Rubinsztein, "Lessons from animal models of Huntington's disease," *Trends Genet.* 18, 202–209 (2002).

[6] M. Arrasate and S. Finkbeiner, "Protein aggregates in Huntington's disease," *Exp. Neurol.* 238, 1–11 (2012).

[7] D. Hmida-Ben Brahim, M. Chourabi, S. Ben Amor, I. Harrabi, S. Trabelsi, M. Haddaji-Mastouri, M. Gribaa, S. Sassi, F. E. Gahbiche, et al., "Modulation at age of onset in tunisian Huntington disease patients: implication of new modifier genes," *Genet. Res. Int.* 2014, 210418 (2014).

[8] S. Tezenas, A. Durr, P. Bauer, K. P. Figueroa, C. Mariotti, B. P. C. van de Warrenburg, L. Orsi, and P. Giunti, "Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes," *Brain* 137, 2444–2455 (2014).

[9] Z. Chen, C. Zheng, Z. Long, L. Cao, X. Li, H. Shang, X. Yin, B. Zhang, J. Liu, et al., "(CAG)n loci as genetic modifiers of age-at-onset in patients with Machado-Joseph disease from mainland China," *Brain*, 1–5 (2016).

[10] R. M. De Castilhos, G. V. Furtado, T. C. Gheno, P. Schaeffer, A. Russo, O. Barsottini, J. L. Pedroso, D. Z. Salarini, F. R. Vargas, et al., "Spinocerebellar ataxias in Brazil - Frequencies and modulating effects of related genes," *Cerebellum* 13, 17–28 (2014).

[11] S. M. Pulst, N. Santos, D. Wang, H. Yang, D. Huynh, L. Velazquez, and K. P. Figueroa, "Spinocerebellar ataxia type 2: PolyQ repeat variation in the CACNAIA calcium channel modifies age of onset," *Brain* 128, 2297–2303 (2005).

[12] G. Stuitje, M. J. van Belzen, S. L. Gardiner, W. M. C. van Roon-Mom, M. W. Boogaard, S. J. Tabrizi, R. A. C. Roos, N. A. Aziz, and N. A. Aziz, "Age of onset in Huntington's disease is influenced by CAG repeat variations in other polyglutamine disease-associated genes," *Brain*, 10–12 (2017).

[13] A. Hodges, A. D. Strand, A. K. Aragaki, A. Kuhn, T. Sengstag, G. Hughes, L. A. Elliston, C. Hartog, D. R. Goldstein, et al., "Regional and cellular gene expression changes in human Huntington's disease brain," *Hum. Mol. Genet.* 15, 965–977 (2006).

[14] P. Langfelder, J. P. Cantle, D. Chatzopoulou, N. Wang, F. Gao, I. Al-Ramahi, X. Lu, E. M. Ramos, K. El-zein, et al., "Integrated genomics and proteomics define huntingtin CAG length – dependent networks in mice," *Nat. Neurosci.*, 1–11 (2016).

[15] D. Trabzuni, M. Ryten, R. Walker, C. Smith, S. Imran, A. Ramasamy, M. E. Weale, and J. Hardy, "Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies," *J. Neurochem.* 119, 275–282 (2011).

[16] C. Bettencourt, M. Ryten, P. Forabosco, S. Schorge, J. Hersheson, J. Hardy, and H. Houlden, "Insights from cerebellar transcriptomic analysis into the pathogenesis of ataxia.," *JAMA Neurol.* 71, 831–839 (2014).

[17] A. Neueder and G. P. Bates, "A common gene expression signature in Huntington's disease patient brain regions," *BMC Med. Genom.* 7, 1–23 (2014).

[18] M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C.-K. K. Lee, et al., "Canonical genetic signatures of the adult human brain," *Nat. Neurosci.* 18, 1832–1844 (2015).

[19] E. Eising, S. M. H. Huisman, A. Mahfouz, L. S. Vijfhuizen, V. Anttila, B. S. Winsvold, and T. Kurth, "Gene co-expression analysis identifies brain regions and cell types involved in migraine pathophysiology: a GWAS-based study using the Allen Human Brain Atlas," *Hum. Genet.* 135, 425–439 (2016).

[20] A. Mahfouz, M. N. Ziats, O. M. Rennert, B. P. F. Lelieveldt, and M. J. T. Reinders, "Shared Pathways Among Autism Candidate Genes Determined by Co-expression Network Analysis of the Developing Human Brain Transcriptome," *J. Mol. Neurosci.* 57, 580–594 (2015).

[21] A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, et al., "Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism," *Cell* 155, 997–1007 (2013).

[22] N. N. Parikshak, R. Luo, A. Zhang, H. Won, J. K. Lowe, V. Chandran, S. Horvath, and D. H. Geschwind, "Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism," *Cell* 155, 1008–1021 (2013).

[23] E. M. Coppen, J. van der Grond, A. Hafkemeijer, S. A. R. B. Rombouts, and R. A. C. Roos, "Early grey matter changes in structural covariance networks in Huntington's disease," *NeuroImage Clin.* 12, 806–814 (2016).

[24] L. Minkova, S. B. Eickhoff, A. Abdulkadir, C. P. Kaller, J. Peter, E. Scheller, J. Lahr, R. A. Roos, A. Durr, et al., "Large-scale brain network abnormalities in Huntington's disease revealed by structural covariance," *Hum. Brain Mapp.* 37, 67–80 (2016).

[25] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nat. Protoc.* 4, 44–57 (2008).

[26] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Res.* 37, 1–13 (2009).

[27] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," *Third Int. AAAI Conf. Weblogs Soc. Media*, 361–362 (2009).

[28] P. Shannon, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Res.* 13, 2498–2504 (2003).

[29] P. T. Shannon, M. Grimes, B. Kutlu, J. J. Bot, and D. J. Galas, "RCytoscape: tools for exploratory network analysis," *BMC Bioinformatics* 14, 217 (2013).

[30] K. A. Gray, B. Yates, R. L. Seal, M. W. Wright, and E. A. Bruford, "Genenames.org: the HGNC resources in 2015," *Nucleic Acids Res.* 43, 1079–1085 (2015).

[31] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, and M. A. Gillette, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide," *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550 (2005).

[32] E. J. N. Groen and T. H. Gillingwater, "UBA1: At the crossroads of ubiquitin homeostasis and neurodegeneration," *Trends Mol. Med.* 21, 622–632 (2015).

[33] L. Jones, H. Houlden, and S. J. Tabrizi, "DNA repair in the trinucleotide repeat disorders," *Lancet Neurol.* 16, 1234 (2017).

[34] T. Seredenina and R. Luthi-Carter, "What have we learned from gene expression profiles in Huntington's disease?," *Neurobiol. Dis.* 45, 83–98 (2012).

[35] S. M. H. Huisman, B. van Lew, A. Mahfouz, N. Pezzotti, L. Michielsen, A. Vilanova, M. J. T. Reinders, H. Thomas, and P. F. Lelieveldt, "BrainScope : interactive visual exploration of the spatial and temporal human brain transcriptome," *Nucleic Acids Res.* 45, e83 (2017).

[36] C. Bettencourt, D. Hensman-Moss, M. Flower, S. Wiethoff, A. Brice, C. Goizet, G. Stevanin, and G. Koutsis, "DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases," *Ann. Neurol.* 79, 983–990 (2016).

[37] Z. Chen, C. Wang, C. Zheng, Z. Long, L. Cao, X. Li, H. Shang, X. Yin, B. Zhang, et al., "Ubiquitin-related network underlain by (CAG)n loci modulate age at onset in Machado-Joseph disease," *Brain* 140, 1–5 (2017).

[38] M. H. Schaefer, E. E. Wanker, and M. A. Andrade-Navarro, "Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks," *Nucleic Acids Res.* 40, 4273–4287 (2012).

[39] L. Fogassi and G. Luppino, "Motor functions of the parietal lobe," *Curr. Opin. Neurobiol.* 15, 626–631 (2005).

[40] G. Pfurtscheller and A. Berghold, "Patterns of cortical activation during planning of voluntary movement," *Electroencephalogr. Clin. Neurophysiol.* 72, 250–258 (1989).

[41] C. M. Stinear, J. P. Coxon, and W. D. Byblow, "Primary motor cortex and movement prevention: Where Stop meets Go," *Neurosci. Biobehav. Rev.* 33, 662–673 (2009).

[42] P. Nachev, C. Kennard, and M. Husain, "Functional role of the supplementary and pre-supplementary motor areas.," *Nat. Rev. Neurosci.* 9, 856–869 (2008).

[43] R. Muhammad, J. D. Wallis, and E. K. Miller, "A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum.," *J. Cogn. Neurosci.* 18, 974–989 (2006).

[44] S. H. Ying, S. I. Choi, S. L. Perlman, R. W. Baloh, D. S. Zee, and A. W. Toga, "Pontine and cerebellar atrophy correlate with clinical disability in SCA2," *Neurology* 66, 424–426 (2006).

[45] C. A. Matos, S. de Macedo-Ribeiro, and A. L. Carvalho, "Polyglutamine diseases: The special case of ataxin-3 and Machado-Joseph disease," *Prog. Neurobiol.* 95, 26–48 (2011).

[46] C. L. Benn, T. Sun, G. Sadri-Vakili, K. N. McFarland, D. P. DiRocco, G. J. Yohrling, T. W. Clark, B. Bouzou, and J.-H. J. Cha, "Huntingtin modulates transcription, occupies gene promoters in vivo, and binds directly to DNA in a polyglutamine-dependent manner.," *J. Neurosci.* 28, 10720–10733 (2008).

[47] L. S. Kaltenbach, E. Romero, R. R. Becklin, R. Chettier, R. Bell, A. Phansalkar, A. Strand, C. Torcassi, J. Savage, et al., "Huntingtin interacting proteins are genetic modifiers of neurodegeneration," *PLoS Genet.* 3, 689–708 (2007).

[48] X.-J. Li, M. Friedman, and S. Li, "Interacting proteins as genetic modifiers of Huntington disease.," *Trends Genet.* 23, 531–533 (2007).

[49] A. Ashkenazi, C. F. Bento, T. Ricketts, M. Vicinanza, F. Siddiqi, M. Pavel, F. Squitieri, M. C. Hardenberg, S. Imarisio, et al., "Polyglutamine tracts regulate beclin 1-dependent autophagy," *Nature* 545, 108–111 (2017).

[50] M. Yokoshi, Q. Li, M. Yamamoto, H. Okada, Y. Suzuki, and Y. Kawahara, "Direct binding of ataxin-2 to distinct elements in 3' UTRs promotes mRNA stability and protein expression," *Mol. Cell* 55, 186–198 (2014).

[51] J. Lim, J. Crespo-Barreto, P. Jafar-Nejad, A. B. Bowman, R. Richman, D. E. Hill, H. T. Orr, and H. Y. Zoghbi, "Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1," *Nature* 452, 713–718 (2008).

[52] L. A. Becker, B. Huang, G. Bieri, R. Ma, D. A. Knowles, P. Jafar-Nejad, J. Messing, H. J. Kim, A. Soriano, et al., "Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice," *Nature* 544, 367–371 (2017).

[53] N. P. Dantuma and L. C. Bott, "The ubiquitin-proteasome system in neurodegenerative diseases: precipitating factor, yet part of the solution.," *Front. Mol. Neurosci.* 7, 70 (2014).

[54] Z. Ortega and J. J. Lucas, "Ubiquitin-proteasome system involvement in Huntington's disease," *Front. Mol. Neurosci.* 7, 1–11 (2014).

[55] G. Atkin and H. Paulson, "Ubiquitin pathways in neurodegenerative disease.," *Front. Mol. Neurosci.* 7, 63 (2014).

[56] A. B. Bowman, S. Y. Yoo, N. P. Dantuma, and H. Y. Zoghbi, "Neuronal dysfunction in a polyglutamine disease model occurs in the absence of ubiquitin-proteasome system impairment and inversely correlates with the degree of nuclear inclusion formation," *Hum. Mol. Genet.* 14, 679–691 (2005).

[57] E. F. E. Kuiper, E. P. de Mattos, L. B. Jardim, H. H. Kampinga, and S. Bergink, "Chaperones in polyglutamine aggregation: Beyond the Q-stretch," *Front. Neurosci.* 11, 1–11 (2017).

[58] V. M. Miller, R. F. Nelson, C. M. Gouvion, A. Williams, E. Rodriguez-lebron, S. Q. Harper, B. L. Davidson, M. R. Rebagliati, and H. L. Paulson, "CHIP suppresses polyglutamine aggregation and toxicity in vitro and in vivo," *J. Neurosci.* 25, 9152–9161 (2005).

[59] H. Adachi, M. Waza, K. Tokui, M. Katsuno, M. Minamiyama, F. Tanaka, M. Doyu, and G. Sobue, "CHIP overexpression reduces mutant androgen receptor protein and ameliorates phenotypes of the spinal and bulbar muscular atrophy transgenic mouse model," *J. Neurosci.* 27, 5115–5126 (2007).

[60] S. M. Ronnebaum, C. Patterson, and J. C. Schisler, "Emerging evidence of coding mutations in the ubiquitin-proteasome system associated with cerebellar ataxias.," *Hum. Genome Var.* 1, 14018 (2014).

[61] K. Parsons, Y. Nakatani, and M. D. Nguyen, "P600/UBR4 in the central nervous system," *Cell. Mol. Life Sci.* 72, 1149–1160 (2015).

[62] J. Conroy, P. McGettigan, R. Murphy, D. Webb, S. M. Murphy, B. McCoy, C. Albertyn, D. McCreary, C. McDonagh, et al., "A novel locus for episodic ataxia: UBR4 the likely candidate," *Eur. J. Hum. Genet.* 22, 505–510 (2014).

[63] D. Chhangani, N. R. Jana, and A. Mishra, "Misfolded proteins recognition strategies of e3 ubiquitin ligases and neurodegenerative diseases," *Mol. Neurobiol.* 47, 302–312 (2013).

[64] H. Y. Liu and C. M. Pfleger, "Mutation in E1, the ubiquitin activating enzyme, reduces drosophila lifespan and results in motor impairment," *PLoS One* 8, 1–10 (2013).

Chapter 4
Molecular characterization of the stress network in the human brain

Mandy Meijer

Arlin Keo

Judith M.C. van Leeuwen

Oleh Dzyubachyk

Onno C. Meijer

Christiaan H. Vinkers

Ahmed Mahfouz

This chapter was published online in: *bioRxiv* (2020), doi: 10.1101/661587 (in submission).
Supplementary material is available online at:
https://www.biorxiv.org/content/10.1101/661587v2.supplementary-material

Abstract

The biological mechanisms underlying inter-individual differences in human stress reactivity remain poorly understood. We aimed to identify the molecular underpinning of neural stress sensitivity. Linking mRNA expression data from the Allen Human Brain Atlas to task-based fMRI revealed 201 differentially expressed genes in cortex-specific brain regions differentially activated by stress in individuals with low or high stress sensitivity. These genes are associated with stress-related psychiatric disorders (e.g., schizophrenia and anxiety) and include markers for specific neuronal populations (e.g., *ADCYAP1*, *GABRB1*, *SSTR1*, and *TNFRSF12A*), neurotransmitter receptors (e.g., *GRIN3A*, *SSTR1*, *GABRB1*, and *HTR1E*), and signaling factors that interact with the corticosteroid receptor and hypothalamic-pituitary-adrenal axis (e.g., *ADCYAP1*, *IGSF11, and PKIA*). Overall, the identified genes potentially underlie altered stress reactivity in individuals at risk for psychiatric disorders and play a role in mounting an adaptive stress response, making them potentially druggable targets for stress-related diseases.

4.1 INTRODUCTION

Stress is a major risk factor for the development of a wide range of psychiatric disorders, including schizophrenia and depression [1]. Inter-individual differences in how the brain responds to stress depend on intrinsic (e.g., genetic and developmental) as well as on extrinsic (e.g., hormonal) factors [2]. The neural correlates underlying stress reactivity are currently a growing topic of investigation [3–5]. In healthy individuals, acute stress causes a shift in neural networks by suppressing the executive control network and activating the salience network and default mode network (DMN) [6]. One hypothesis is that stress vulnerability is the result of maladaptive changes in the dynamic response of these neural networks, either during the acute phase, during the recovery period in the aftermath of stress, or both[2]. Moreover, acute social stress deactivates the DMN in the aftermath of stress during emotion processing in healthy controls but not in siblings of schizophrenia patients who are at-risk for several psychiatric disorders [7][8]. Yet, the molecular mechanisms underlying differences in brain reactivity to stress in humans remain unknown as access to the tissue of interest in humans is limited.

Nevertheless, stress-related brain regions and networks as identified by fMRI can be further characterized based on transcriptomic signatures. Mapping gene expression atlases of the healthy brain to imaging data allows the identification of the molecular mechanisms underlying imaging phenotypes. Previous studies have identified gene expression patterns associated with structural brain changes in autism spectrum disorders, Huntington's disease and the onset of schizophrenia [9–12]. Similarly, mapping resting-state fMRI and connectivity data onto gene expression atlases has led to identification of molecular profiles underlying these fMRI networks [13–15].

In this study, we examined the putative molecular signatures of brain regions linked to stress reactivity. We linked gene expression data from the Allen Human Brain Atlas (AHBA) to an fMRI-stress network (Figure 4.1). In short, we found that the stress network was enriched for genes associated to specific subtypes of neurons (*i.e.* components of the cortical circuitry) with genetic relevance for psychiatric disorders, and for signaling factors and proteins that interact with the activation of the Hypothalamic-Pituitary-Adrenal axis (HPA-axis) and response to glucocorticoids. These all constitute potential targets for directed pharmacotherapy in stress related disorders.

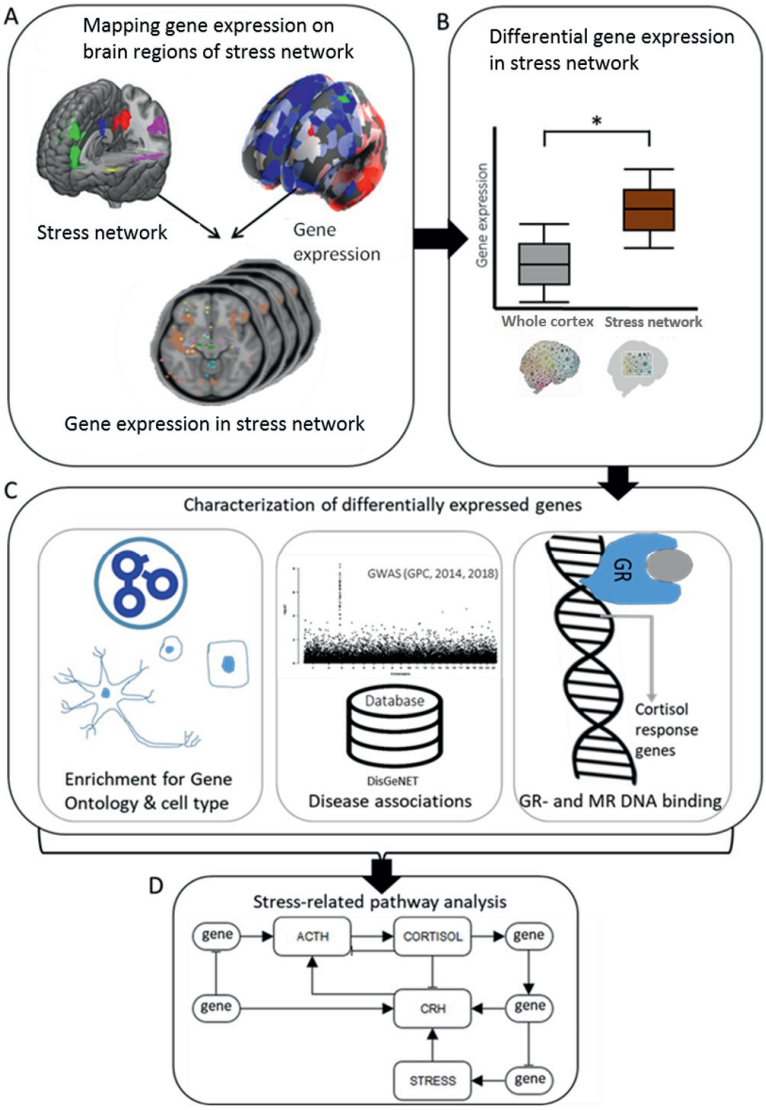


Figure 4.1 Study overview. (A) Cortical brain regions vulnerable to stress (= stress network) during an emotion processing task were assessed in an fMRI study. All brain regions showed higher stress-induced brain activity following an acute social stressor in at risk individuals (healthy siblings of schizophrenia patients). The fMRI data was mapped to the AHBA resulting in an overlay of the fMRI and gene expression data. (B) With this overlay, differential gene expression between the brain regions vulnerable to stress and the rest of the cortex were assessed. (C) Differentially expressed genes were consequently characterized by identifying enrichment for gene ontology and cell type markers, associations with stress-related diseases and enrichment for cortisol responsive genes. (D) Information provided by the previous analyses was used to build a model of a molecular pathway underlying human stress reactivity.

4.2 METHODS

4.2.1 DEFINING THE STRESS NETWORK

Based on a previous study, we selected brain regions that were differentially affected by stress in individuals with high and low stress sensitivity [6]. In this study, there were four experimental groups: control-no-stress (n=19), control-stress (n=20), sibling-no-stress (n=20) and sibling-stress (n=19) (Table 4.1). Before scanning, participants in the stress groups underwent a Trier Social Stress Test [16] and 30 minutes after the onset of the test, participants performed an emotion-processing task in the magnetic resonance imaging (MRI) scanner based on the International Affective Picture System [17] during which pictures were presented that had to be rated as either neutral, positive or negative. All participants in this experiment gave written informed consent and the experiment was approved by the Utrecht Medical Center ethical review board and performed according to the guidelines for Good Clinical Practice and the declaration of Helsinki. Based on a 2×2 ANOVA (control/sibling × stress/no-stress) voxel-wise analysis, several brain regions that responded differently to all pictures after acute social stress in siblings compared to healthy individuals were identified. These regions include key nodes of the DMN (posterior cingulated cortex/precuneus and medial prefrontal cortex) and salience network (anterior insula), as well as the superior temporal gyrus, middle temporal gyrus, middle cingulate gyrus, ventrolateral prefrontal cortex, precentral gyrus and cerebellar vermis (Figure 4.2A). We selected and present in the figures the cortex-specific brain regions for the initial analyses to prevent that our results are being driven by differences between the cortex and subcortex. Analyses on all brain regions in the stress network can be found in the supplementary text.

Table 4.1 Group characteristics of fMRI study

	Con-no-stress	Con-stress	Sib-no-stress	Sib-stress	*P*-value
N	19	20	20	19	
Age (years)	32.6 (8.5)	34.8 (9.1)	33.8 (10.8)	32.5 (7.4)	0.836[a]
Handedness (% right)	89.5	95	70	89.5	0.194[b]
Educational level	7.6 (2.7)	7.1 (1.9)	7.0 (1.6)	7.4 (1.5)	0.688[a]
Body Mass Index	24.1 (2.7)	24.2 (2.1)	24.0 (3.0)	24.9 (3.9)	0.774[a]
Ethnicity (% Caucasian)	84.2	90	90	84.2	0.900[b]
Smoker (% yes)	5.3	35	30	31.6	0.132[b]

Con = control; sib = sibling of schizophrenia patient.

Mean values (SD) are denoted for age, education, and body mass index. All other values are reported in frequency.

a = one-way-ANOVA

b = chi square test

4.2.2 ALLEN HUMAN BRAIN ATLAS (AHBA)

Gene expression data from six healthy brains were acquired from the AHBA [18]. In this microarray dataset, probes were mapped to genes as previously described [19]. Z-scores for normalized gene expression levels from the AHBA were calculated separately for each of the six individual brains. Gene expression data were linked to an fMRI-based stress network according to the MNI coordinate system, such that samples of the AHBA exactly overlap with the corresponding fMRI voxels. For all samples in the AHBA, we determined whether they were located in the cortical stress network for all six donors separately. The gene expression levels of the AHBA samples were extracted and resulted in expression data of 19,992 genes in 111 and 1839 brain samples in- and outside the cortical stress network, respectively.

4.2.3 DIFFERENTIAL GENE EXPRESSION IN THE CORTICAL STRESS NETWORK

To identify genes differentially expressed between the cortical stress network and the rest of the cortex, we analyzed each of the six brain donors separately. Differential expression was determined for the cortical stress network altogether as one mask. For each gene, we combined effect sizes (difference in mean expression between the brain stress network and the rest of the brain) across donors using a meta-analysis approach from the 'metafor' 2.0-0 R-package. In brief, a random effects model was used, taking into account the within-brain and between-brain variance, which was estimated with the Dersimonian-Laird model. Variances and confidence intervals needed for the meta-analysis were calculated using the escalc-function. Genes were considered to be differentially expressed at a Benjamini-Hochberg (BH) adjusted *P*-value < 0.05.

We also performed analysis on the whole brain (differentially expressed genes BH-adjusted *P*-value < 0.05 and $\log_2$ fold-change (FC) > |1|). Given the large difference in the transcriptional profile of the cerebellum compared to the rest of the brain [20], we excluded the cerebellum from the whole brain analysis. In addition, we performed the differential expression analysis between samples inside and outside the stress network for each of the following brain regions separately: cerebral cortex (Cx), frontal gyrus (FG), cingulate gyrus (CgG), cerebellum (Cb), and the hippocampal formation (HiF). Other anatomical regions contained too few samples (< 2 in the mask) to perform the analysis on these particular structures separately.

We used a bootstrapping approach to assess the robustness of our results with respect to the imbalance between the number of AHBA samples inside and outside the cortical stress network (111 inside and 1839 outside). We randomly selected 111 samples from the whole cortex, regardless of their location inside or outside the stress network and compared gene expression profiles of these brain samples with the original set of 111 samples inside the cortical stress network. We repeated this process 1000 times to assess the reproducibility of the differentially expressed genes.

4.2.4 GENE ONTOLOGY (GO) ENRICHMENT ANALYSIS

To characterize the functionality of the differentially expressed genes, a GO enrichment analysis was performed. The list of unranked differentially expressed genes was uploaded to GOrilla (Gene Ontology Enrichment Analysis and Visualization Tool) [21]. As a background list, the top 20% of genes with the highest expression level in the cortex was used, to correct for non-selective ontologies. GO terms were considered significant when the *P*-values < 0.001 (Fisher's exact test) after BH-correction.

4.2.5 CELL TYPE ENRICHMENT ANALYSIS

We assessed whether the differentially expressed genes were enriched for cell type markers [22]. Genes with a 20-fold higher expression in neurons (628 marker genes), oligodendrocytes (186 marker genes), astrocytes (332 marker genes), microglia (520 marker genes) and endothelial cells (456 marker genes) were considered to be markers for that cell type. Since most of our AHBA samples were located inside the cortex, we used a set of brain-region-specific markers and focused on 18 cortical cell types [23]. Details on markers can be found on https://pavlab.msl.ubc.ca/data-and-supplementary-information/supplement-to-mancarci-et-al-neuroexpresso. Finally, to assess which neuronal cell types might be involved in stress sensitivity, single cell RNA sequencing data of the middle temporal gyrus of the human neocortex from the Allen Brain Institute [24] (http://celltypes.brain-map.org/rnaseq/human) were used. The sum of the $\log_{10}$ values of the counts per differentially expressed gene were calculated for each cell cluster separately.

4.2.6 ENRICHMENT ANALYSIS OF DISEASE-ASSOCIATED GENES

To assess whether the differentially expressed genes are associated to stress-related psychiatric disorders, a disease-associated gene enrichment analysis was performed based on existing Genome-Wide Association Studies (GWAS) including schizophrenia [25,26], Bipolar Disorder [27], and Major Depressive Disorder [28], and stress-related diseases such as Post-Traumatic Stress Disorder, as well as non-stress-related diseases (e.g., Huntington and osteoporosis) based on disease gene sets from DisGeNET [29]. As non-disease control conditions, genes associated to height and waste-hip ratio were included in the analysis [30,31]. The schizophrenia, MDD and BP GWAS loci were considered to be associated if they reached genome-wide significance of $P < 5^{*}10^{-8}$. Intersections of loci based on GENCODE with UCSC hg19/NCBI build 37 position were used to map loci to risk genes by the authors of the GWAS [25,27,28]. These annotations were used for the enrichment analyses. All genes assessed in the AHBA that were not associated to a disease or trait were used as background test in the Fisher-test. FDR-correction was applied over the amount of enrichment tests.

To assess the enrichment of disease-related gene sets in intercellular signaling genes, neuropeptides and receptor genes were selected from the differentially expressed genes. Odds ratios (ORs) were calculated for the set of neuropeptides and receptors for each disease as a measurement of effect size, (i.e. the increased chance of a peptide or receptor being present in the set of differentially expressed genes). For this, the number of receptors found within the trait was compared to all the receptors that were measured in the AHBA (1203 receptors), based on the gene annotation of the AHBA. Gene names that included the

word 'receptor' were selected and this list was manually verified whether the gene was a receptor or a modulator. The ORs for the neuropeptides were calculated in the same way, based on a list of neuropeptides available from NeuroPep [32].

4.2.7 Mineralocorticoid and glucocorticoid DNA binding loci

MR and GR DNA binding loci under stress in the rat hippocampus were previously assessed [33]. We identified sets of genes with GR-specific, MR-specific and GR-MR-overlapping DNA binding loci, i.e. potential target genes. To predict glucocorticoid sensitivity of our differentially expressed genes, we assessed whether these sets of targets were enriched among the differentially expressed genes.

4.2.8 Enrichment statistics for GO, cell type, disease-associated genes and receptor binding

Enrichments were assessed based on Fisher's Exact Tests and odds ratios (ORs) were calculated as a measurement of effect size for the enrichments. An OR of 1 indicates no effects, whereas an OR > 1 and 0 < OR < 1 reflects enrichment and depletion, respectively. All *P*-values were corrected for multiple testing using the Benjamini-Hochberg (BH) method and a BH corrected *P*-value < 0.05 was considered to be significant, unless stated otherwise.

4.3 Results

4.3.1 Differentially expressed genes in the stress network

We identified the gene expression signatures of the cortical stress network with altered stress-induced activity by determining which genes are differentially expressed in the stress network compared to the rest of the cortex. Using a meta-analysis approach to combine results across all donors of the AHBA (n = 6), we identified 201 differentially expressed genes (BH-adjusted $P < 0.05$, Figure 4.2B; Supplementary Table 1). Among those genes, 177 were higher expressed, while the other 24 genes were lower expressed in the cortical stress network compared to the rest of the cortex (Figure 4.2C). Using a bootstrapping approach (see 2.3), we found the identified set of genes to be highly robust to the imbalance between the number of AHBA samples inside and outside the stress network (in 83% of our 1000 iterations, only genes out of our initial 201 differentially expressed gene list were found).

We also identified the gene expression signatures of the stress network with altered stress-induced activity by determining which genes are differentially expressed in the stress network compared to the rest of the brain minus the cerebellum. Using the same meta-analysis approach, we identified 261 differentially expressed genes (BH-adjusted $P < 0.05$ and $\log_2 FC > |1|$). A full description of the results, including tables and figures, can be found in the supplementary text. However, due to the higher representation of cerebral cortex samples in the brain regions vulnerable to stress (109 out of 127; 91%) compared to the rest of the brain minus the cerebellum (1,950 out of 3,225; 60%), differentially expressed genes in the whole brain stress network were also differentially expressed between cortical and non-cortical samples (222 out of 261 genes were also differentially expressed in the top 10% difference between cortex and non-cortex, $P < 0.00001$). Therefore, we chose to focus on the cortex-specific stress network.

The differentially expressed genes in the cortical stress network generally showed high expression values in the cortex but not the hippocampus, and mostly low expression levels in non-cortical areas (Figure 4.2D). The two most differentially expressed genes in the stress-specific cortical regions are Tumor necrosis factor receptor superfamily member 12A (*TNFRSF12A)* (BH-adjusted *P*-value = 0.006, $\log_2$(FC) = -0.24) and Exosome Component 6 Pseudogene (*LOC392145*) (BH-adjusted *P*-value = 0.009, $\log_2$(FC) = 0.38). *TNFRSF12A* may module cellular adhesion to matrix proteins, whereas *LOC392145* is an mRNA transport regulator.

4.3.2 FUNCTIONALITY AND CELL-TYPE SPECIFICITY OF DIFFERENTIALLY EXPRESSED GENES IN THE STRESS NETWORK

A GO term enrichment analysis was performed to assess whether the differentially expressed genes in the stress network are enriched for specific functions. The differentially expressed genes were enriched for GO terms involved in neuronal development and neurogenesis, synaptic signal transmission, and glutamate receptor signaling (Figure 4.3A and Supplementary Table 1). Genes involved in most processes based on GO terms (at least assigned to five out of ten GO terms) include *SHANK, GRIN3A, CNTN4* and *ADCYAP1*. Enrichment analysis for cellular components indicated that the proteins coded by the differentially expressed genes were mainly found at the synapse, reflecting both the high expression of the genes in the synapse-dense cerebral cortex, and a potential role for synaptic proteins as determinants for the differential activation.

Next, we identified the specific cell types underlying the differential gene expression levels in the cortical stress network using enrichment analysis of cortical cell-type markers [22]. Enrichment was found for neuronal cell markers (BH-adjusted $P = 5^{*}10^{-5}$), including *ADCYAP1, DPYSL3, INSM1, PKIA, SSTR1, NOL4, BAIAP3, KCNB2, FAM65B, ABLIM3, TEKT2, SHANK1, DACT1, PCBP3, SCN3B, LMO3, CA10, LRRTM4, SYT16, GPRIN1, TMEM200A, LRRC3B, GRIN3A,* and *PNCK*. However, no specific subtype of neuron was in particular enriched. Enrichment was also found for astrocytes (*ATP2B4, PTCH1, FABP7, IGSF11, KCNN3, GRM3, GABRB1, PTX3,* BH-adjusted *P* = 0.024). The list of differentially expressed genes included a few microglia (*TMEM52, F13A1, MSH5, ARHGAP4, NOD2,* and *TNFRSF12A*), endothelial cell (*LAMA1, LAMB1, C10rf115, DOK4*, and *MICB*), and oligodendrocyte markers (*EFNB3* and *TYRO3*), although not significantly enriched (BH-adjusted *P* = 0.924). Moreover, we found that neuronal markers showed a partially overlapping distribution in a t-Distributed Stochastic Neighbor Embedding (t-SNE) map of all genes across the whole brain as the differentially expressed genes in, indicating that neuronal markers and the differentially expressed genes show the same expression patterns across cortical areas (Figure 4.3B) and thus differential activity may depend on neuronal gene expression.

Using a human-specific single cell RNA-sequencing data of the medio-temporal gyrus [24], we found the differentially expressed genes to be mainly enriched in glutamatergic excitatory neurons compared to GABAergic and non-neuronal cells, using a Wilcoxon rank test (*P*-value = $2.2^{*}10^{-}16$, Figure 4.3C).

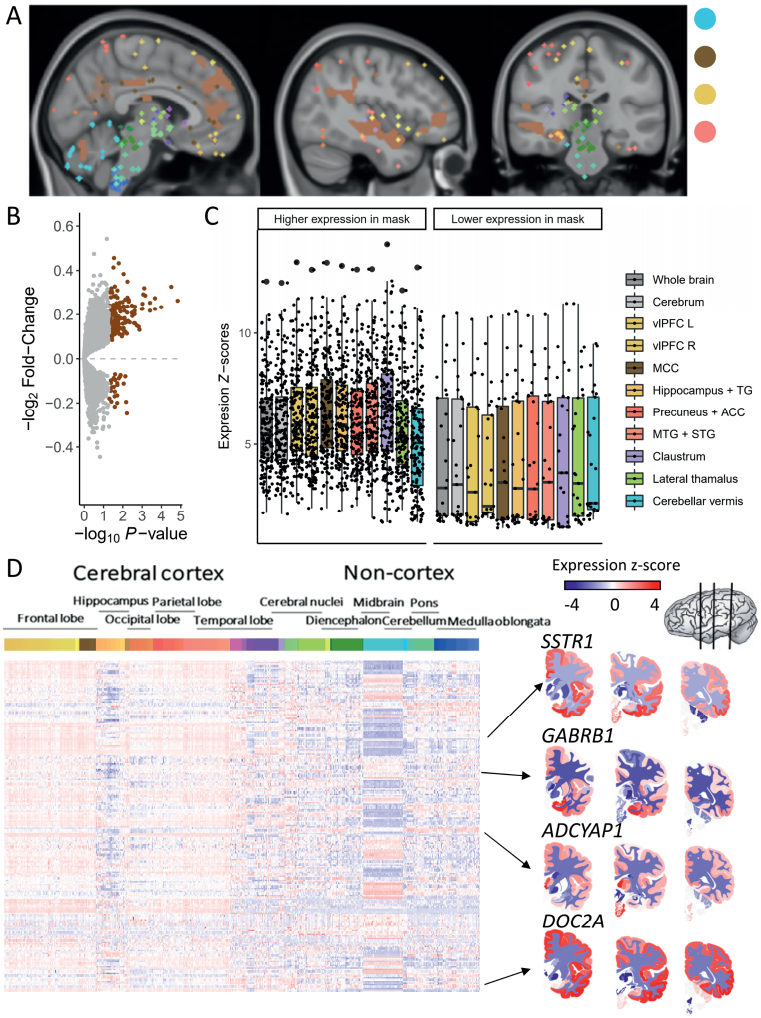


Figure 4.2 Differentially expressed genes in brain regions vulnerable to stress can be identified using gene expression atlases. (A) Brain regions in the stress network are present throughout the brain (including cerebellum, cingulate gyrus, frontal gyrus, temporal gyrus, and hippocampal formation). For the analysis, all regions in the cortex were combined. (B) Differential gene expression was determined for the cortical stress network compared to the rest of the cortex. Significant genes (BH-adjusted *P*-value < 0.05 have higher expression in the stress network. Grey dots represent non-significant genes and brown dots represent significant genes based on meta-analysis across all six AHBA donors. (C) The box plots show the expression of the higher (left) and lower (right) expressed genes compared to the rest of the cortex in the brain regions of interest from the stress network. (D) In the whole brain, differentially expressed genes show mostly high expression levels in the cortex and low expression levels in non-cortical brain regions. In the heatmap, each row represents a gene and each column represents a sample and all samples of the AHBA are illustrated here. On the right, coronal brain sections for the genes *SSTR1*, *GABRB1*, *ADCYAP1*, and *DOC2A* are presented. Colors indicates high (red) and low (blue) expression levels.

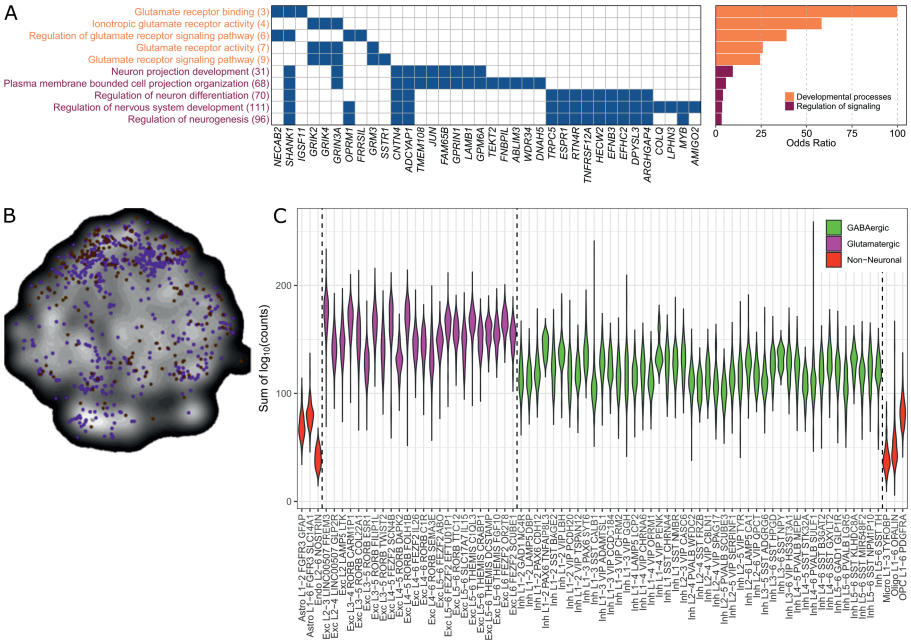


Figure 4.3 Functionality of differentially expressed genes in the stress network. (A) Differentially expressed genes annotated to one of the GO terms were assigned to multiple GO terms and thus involved in multiple processes. Between parentheses, the total number of genes assigned to the GO term is depicted. On the right side of the graph, ORs are displayed per GO term. (B) Differentially expressed genes (brown), neuronal marker genes (purple) and overlapping genes (yellow) are plotted in a t-SNE plot generated using BrainScope.nl [20], where points close together represent genes with similar gene expression profiles. The differentially expressed genes show a similar profile in the t-SNE plot as neuronal cell markers (purple). (C) The sum of the $\log_{10}$ values of the counts per gene is plotted for each cell cluster. Green clusters belong to GABAergic cells, purple clusters to glutamatergic cells and red clusters to non-neuronal cells.

4.3.3 DIFFERENTIALLY EXPRESSED GENES IN STRESS NETWORK ARE ASSOCIATED TO STRESS-RELATED DISEASES

We hypothesized that the differentially expressed genes in the stress network would be associated to the genetic background of psychiatric disorders, particularly for stress-related brain disorders, as stress plays a major role in the development of these disorders. Using genetic variants from GWAS of the Genomics of Psychiatry Consortium [25,26], we assessed whether schizophrenia-associated risk loci are enriched in the set of differentially expressed genes. Indeed, schizophrenia risk genes were enriched in the differentially expressed genes in the stress network (Fisher Exact test, BH-adjusted *P*-value = 0.015). The schizophrenia risk genes *CNTN4*, *GRM3*, *FUT9*, *SATB2*, *GPM6A*, *COQ10B*, *DOC2A*, and *NISCH* were present in our differentially expressed genes, and all except one (*COQ10B*) were higher expressed in the cortical brain regions vulnerable to stress. Based on a recent GWAS across multiple

psychiatric disorder, multiple pleiotropic risk genes were identified [34]. Furthermore, gene-disease associations from DisGeNet, a manually curated database, were used to assess risk gene enrichment for psychiatric, brain and non-brain diseases and non-disease traits. Enrichment was found for neuropsychiatric disorders (schizophrenia, bipolar disorder, and autism spectrum disorder) and other brain diseases (Parkinson's disease). However, no gene enrichment was found for non-brain diseases (e.g., osteoporosis) and non-disease traits (e.g., height and waste-hip-ratio; Figure 4.4). Thus, differentially expressed genes in the stress network are predominantly involved in genes relevant for stress-related diseases but not in non-brain-related disorders and traits.

Interestingly, the set of 201 differentially expressed genes in the stress network included a considerable number of receptors. Apart from their use as markers for specific cell types (e.g., *ADCYAP1*, *GABRB1*, *SSTR1*, and *TNFRSF12A*), these are important for signaling in the brain (e.g., *ADCYAP1* modulates glutamatergic signaling [35] and the HPA-axis response [36]) and some of them are known to be involved in the regulation of stress [36,37]. Therefore, we assessed whether there were more receptors in our set of genes associated to psychiatric disorders than you would expect by chance. We found higher odds ratios for brain and psychiatric disorders, with the biggest effect sizes in psychiatric disorders (Figure 4.4). Effect sizes for receptor enrichment in Epilepsy (*P*-value < $1^{*}10^{-5}$, OR = 9.07), Huntington (*P*-value = 0.007, OR = 5.18), Obsessive Compulsive Disorder (*P*-value = 0.004, OR = 7.18), Major Depressive Disorder (*P*-value = $1^{*}10^{-5}$, OR = 13.5), Bipolar Disorder (*P*-value = $1^{*}10^{-5}$, OR = 26.8), and Schizophrenia (*P*-value = 0.03, OR = 3.82) were significant.

4.3.4 CORTISOL SENSITIVITY OF THE STRESS NETWORK

The enrichment of the neuronal GO terms in our set of genes and the association with stress-related diseases indicates that the differentially expressed genes in the stress network are relevant for stress and may be responsive to the pivotal stress hormone cortisol. To investigate glucocorticoid sensitivity, we compared our list of differentially expressed genes with genes that show a DNA binding site for the glucocorticoid- and/or mineralocorticoid receptors (GR and MR) in the rat hippocampus by Chromatin Immunoprecipitation sequencing after stimulation with the endogenous steroid corticosterone [33]. Differentially expressed genes that showed DNA binding loci for both the GR and MR are: *SLC26A4*, *IGSF11*, *GRIK4*, *SCN3B*, *GABRB1*, *CCDC85A*, *KIRREL3*, *HECW2* and *PKIA* (9/459 genes with binding sites, BH-adjusted *P*-value for enrichment = 0.038). There was no significant enrichment for either GR binding (*OPRM1*, *PTER*, *FUT9*, *EPB41L4B*, *PKLR*, *GPM6A*, and *GSC;* 7/704 genes with binding sites, BH-adjusted p- value = 0.98) or MR binding exclusively (*ADCYAP1*, *LAMA1*, *CNTN4*, *XPO1*, *RGS12*, *MGRN1*, *CHST15*, and *ANKLE2*; 8/1247 genes with binding sites, BH-adjusted *P*-value = 0.18). These results indicate that differentially expressed genes in brain the stress network are enriched for DNA-binding loci of that (in the rat) can be bound by both the GR and the MR.

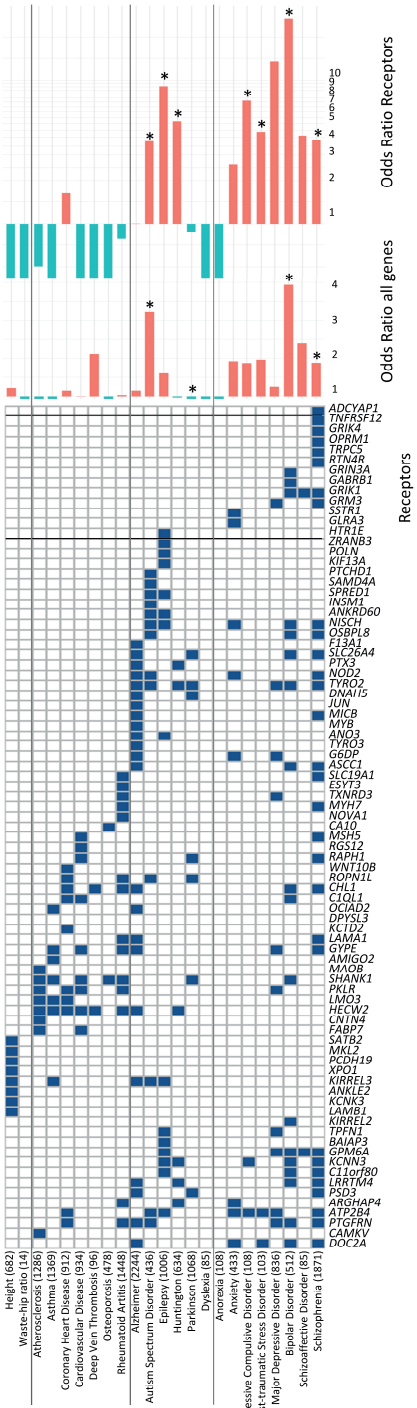


Figure 4.4 Differentially expressed genes in the stress network and stress-related psychiatric disorders. Disease risk gene enrichment was performed for the differentially expressed genes. The diseases are clustered as non-brain related disease, brain disease and psychiatric disease. As a none-disease-associated set of genes, waste-hip ratio and height were used. Numbers between the parentheses indicate the number of genes known to be associated with the disease, based on DisGeNet. The effect size of the gene enrichment is presented at the middle part of the figure and considered significant when the BH-adjusted *P*-value < 0.05 (*). Blue bars mean depletion of genes, whereas red bars indicate enrichment of genes in a trait. ORs for the amount of receptors in the set of differentially expressed genes are depicted for every trait (shown on the right side of the graph).

4.4 DISCUSSION

In this study, we identified genes and pathways in the cortical stress network based on an fMRI-based study involving acute stress exposure. By combining fMRI data to gene expression data, we found 201 differentially expressed genes involved in neuronal processes and enriched in stress-related psychiatric disorders. Moreover, the enriched genes included several neuropeptides and neurotransmitter receptors with regulation by both the GR and MR and substantial links to HPA-axis activity. This gene set uncovered by combining human gene expression and neuroimaging results give important new insights into the putative neural populations and mechanisms underlying stress vulnerability in humans.

Our results point to the involvement of (cortical) cell type markers in differential stress reactivity. For example, we found enrichment for some astrocyte markers, which among others modulate glutamate metabolism and transmission [38]., and there is evidence from both human and rodent models that they may play a role in stress-related disorders [39]. Moreover, the differentially expressed genes are in general highly expressed in excitatory glutamatergic compared to inhibitory GABAergic neurons [24]. Thus, glutamate signaling seems to be involved in a more global level, whereas GABA-related mechanisms that may underlie differential reactivity to stress are limited to a specific subset of GABA-ergic neurons. Specific targeting of these GABAergic populations, based on their receptor repertoire, may help to separate primary from secondary changes in the cortical circuitry.

For genes that do not represent specific neuronal subtypes, changed expression levels may reflect differential responsiveness based on more generic signaling pathways. This may, in particular, be the case for the identified stress-related genes with a genetic association to schizophrenia. *OPRM1* encodes for a mu-opioid receptor, which has been shown to interact with glutamate to adapt to chronic drug abuse, a stress-related disorder [40]. Moreover, mu-opioid receptors are known to modulate the HPA-axis [41].

Genes with high expression levels in the regions vulnerable to stress include neuropeptides and neurotransmitter receptors, which may be directly targeted to modify the activity of these brain regions. *SST1* codes for the somatostatin receptors, a neuropeptide produced in the hypothalamus. This neuropeptide is known to attenuate the stress response, by counteracting CRH signaling via the SST1 receptor [42]. Also a number of serotonergic, GABAergic, and glutamatergic receptors are differentially represented in the stress network. All these factors may well have a role in regulating neuronal network activity during maladaptive stress responses [43–45]. Of note, the excitatory 5-HT1E receptors are overrepresented in brain regions that failed to shut off after stress in at-risk subjects. Antagonism of 5-HT2A is common between several antipsychotic and antidepressant drugs, and normalizing the activity of these brain regions after stressor exposure may be part of their therapeutic mechanism. However, the exact function of the 5-HT1E receptors are unknown, but *HTR1E* is a candidate gene for several stress-related disorders [46–48].

The enrichment analysis of gene ontology terms suggests that the list of differentially expressed genes play a role in stress vulnerability and risk for psychiatric disorders. For

example, prenatal chronic stress has consequences on nervous system development as shown in mice [49–51]. Moreover, disruption of neuronal plasticity [52]. is induced by a prolonged stressor and is a common symptoms of stress-related psychiatric disorders [53].

Furthermore, we found that differentially expressed genes in the stress network are enriched for DNA-binding loci of both the GR and the MR based on experimental data in rats. GR is thought to facilitate recovery and adaptation in the aftermath of stress [54] and polymorphisms as well as post-translational modifications alter susceptibility for stress-related psychiatric disorders [55,56]. The MR has been shown to facilitate stress reactivity [57]. The link with GR and MR suggests that it related to factors related to systemic adaptations, even though we cannot know to what extent these loci actually reflect target genes.

We found a significant overlap between the genes found to be differentially expressed in the whole stress network and those found to be differentially expressed in the region-specific stress-network analysis of which some are known to be involved in stress-related phenotypes [58,59]. The differences between the results obtained in these experiments can be partially explained by the fact that the AHBA samples were collected using bulk sequencing which does not allow the detection of differences across individual cell populations [60]. With the increasingly availability of single cell data we will have enough resolution to detect more subtle differences within the cortex, but for now, human brain single cell data is very limited [24,61–63]. Moreover, previous studies has shown that structures within the cortex are relatively similar in terms of gene expression [14]. Therefore, the finding of 201 differentially expressed genes, point out to a true difference in the cortical stress-network and all other cortical brain regions. The non-overlapping genes from the combined analysis of cortical and non-cortical samples might be driven by anatomical differences, although it is complex to entangle the true biological signals from anatomy-driven signals.

We do not know whether the differentially expressed genes are subject to genetic regulation and whether they show differential translational responses. Furthermore, we could not infer causality, but rather association of genes with stress-sensitivity. In this regard, it will be of considerable interest to further study the genes that have been linked to psychiatric disorders, as genetic variation may, in fact, lead to abnormal expression of the genes we identified. It will also be of interest to study epigenetic regulation of the genes of interest and gene-environment interactions [64–67].

Given that we assessed gene expression levels in the healthy brain, it is challenging to interpret the differences in high and low expression levels and the meaning in diseased brains. High expression levels of the genes in the stress network do not necessarily mean that stress sensitivity is a result of the high gene expression *per se*. It might be the ability to regulate neurobiological processes via direct neurotransmitter and receptor signaling or the ability to indirectly regulate changes in gene expression [68]. Moreover, we have to take into

consideration that we identified genes that already show low baseline expression levels in the brain.

Another limitation of this study is that the low number of samples in some brain regions did not allow the analysis of differential expression within these regions. For example, the precuneus and the angular gyrus were underrepresented in the AHBA (n = 7 for both regions), but harbored great changes according to the fMRI signal. However, there were sufficient brain samples available from the AHBA to analyze brain regions vulnerable to stress altogether. Moreover, the stress-network we defined was based on 78 males. Given the relatively small sample size, replication in a bigger independent cohort should be awaited. Furthermore, the six donors were five males and one female. It is important to take donor's sex effect into account, since there is a sex difference in the development and symptoms of stress-related diseases [69,70]. Therefore, we checked whether gene expression levels were different for the female donor compared to the male donors. We did not find gender effects of gene expression levels of the differentially expressed genes. To maximize the number of samples, we decided to include the female donor in our analyses. It has to be taken into account, however, that the outcome of the performed task might be different across the genders [71]. This implies that our results cannot be generalized over the whole population, but are rather reflective for males, since the stress-network we identified could also be male-specific. We also acknowledge that we do not know how the stress network would look like in individuals at risk for other psychiatric disorder. Moreover, brain regions differentially activated by acute stress are specific for the emotion processing task. Therefore, we might have missed some relevant brain structures, and thus genes, that might have become active during another task under stressful conditions. Lastly, the stress network that was used in this paper was based on data from siblings of patients with schizophrenia. Even though stress is a transdiagnostic factor and relevant for all psychiatric disorders [72], we cannot directly extrapolate the stress network to other psychiatric disorders such as depression and bipolar disorder. There is increasing research into risk groups for these disorders, but to our knowledge, direct comparisons on brain-related stress sensitivity between (risk groups of) across psychiatric disorders are lacking.

To our knowledge, this is the first study to map gene expression atlases to task-based fMRI data in order to identify the molecular mechanisms underlying human stress reactivity in relation to the risk to develop psychiatric disorders. Here, we show that this method can aid in disentangling the molecular underpinnings of specific tasks and traits. We showed that genes possibly underlying stress reactivity are also associated with neuronal cell type markers (e.g., glutaminergic excitatory neurons), stress-related disease, GR and MR responsiveness and HPA-axis activity. We identified several neuropeptides and receptors as important players. These identified systems are not only important to understand the underlying mechanisms of stress vulnerability, but can also be used to develop new drug targets. Therefore, identification of novel drug targets involved in stress vulnerability would be of great interest for the development of new therapies in stress-related psychopathology.

4.5 Acknowledgements

The authors would like to thank S. M. H. Huisman for his advice and help on some of the statistical aspects of the study and M. P. van den Heuvel for his advice on the manuscript.

References

[1] H. Tost, F. A. Champagne, and A. Meyer-Lindenberg, "Environmental influence in the brain, human welfare and mental health," *Nat. Neurosci.* 18, 1421–1431 (2015).

[2] R. H. DeRijk, "Single nucleotide polymorphisms related to HPA axis reactivity," *Neuroimmunomodulation* 16, 340–352 (2009).

[3] A. Luque-García, V. Teruel-Martí, S. Martínez-Bellver, A. Adell, A. Cervera-Ferri, and J. Martínez-Ricós, "Neural oscillations in the infralimbic cortex after electrical stimulation of the amygdala. Relevance to acute stress processing," *J. Comp. Neurol.* 526, 1403–1416 (2018).

[4] H. Zhu, Y. Li, M. Yuan, Z. Ren, C. Yuan, Y. Meng, J. Wang, W. Deng, C. Qiu, et al., "Increased functional segregation of brain network associated with symptomatology and sustained attention in chronic post-traumatic stress disorder," *J. Affect. Disord.* 247, 183–191 (2019).

[5] L. D. Godoy, M. T. Rossignoli, P. Delfino-Pereira, N. Garcia-Cairasco, and E. H. de Lima Umeoka, "A comprehensive overview on stress neurobiology: basic concepts and clinical implications," *Front. Behav. Neurosci.* 12, 127 (2018).

[6] J. van Oort, I. Tendolkar, E. J. Hermans, P. C. Mulders, C. F. Beckmann, A. H. Schene, G. Fernández, and P. F. van Eijndhoven, "How the brain connects in response to acute stress: A review at the human brain systems level," *Neurosci. Biobehav. Rev.* 83, 281–297 (2017).

[7] J. M. C. van Leeuwen, M. Vink, G. Fernández, E. J. Hermans, M. Joëls, R. S. Kahn, and C. H. Vinkers, "At-risk individuals display altered brain activity following stress," *Neuropsychopharmacology* 43, 1954–1960 (2018).

[8] S. V Argyropoulos, S. Landau, S. Kalidindi, T. Toulopoulou, D. J. Castle, R. M. Murray, and M. M. Picchioni, "Twins discordant for schizophrenia: psychopathology of the non-schizophrenic co-twins," *Acta Psychiatr. Scand.* 118, 214–219 (2008).

[9] R. Romero-Garcia, V. Warrier, E. T. Bullmore, S. Baron-Cohen, and R. A. I. Bethlehem, "Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism," *Mol. Psychiatry* 24, 1053–1064 (2019).

[10] K. J. Whitaker, P. E. Vértes, R. Romero-Garcia, F. Váša, M. Moutoussis, G. Prabhu, N. Weiskopf, M. F. Callaghan, K. Wagstyl, et al., "Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome," *Proc. Natl. Acad. Sci.* 113, 9105–9110 (2016).

[11] J. Ritchie, S. P. Pantazatos, and L. French, "Transcriptomic characterization of MRI contrast with focus on the T1-w/T2-w ratio in the cerebral cortex," *Neuroimage* 174, 504–517 (2018).

[12] P. McColgan, S. Gregory, K. K. Seunarine, A. Razi, M. Papoutsi, E. Johnson, A. Durr, R. A. C. C. Roos, B. R. Leavitt, et al., "Brain Regions Showing White Matter Loss in Huntington's Disease Are Enriched for Synaptic and Metabolic Genes," *Biol. Psychiatry* 83, 1–10 (2017).

[13] P. E. Vértes, T. Rittman, K. J. Whitaker, R. Romero-Garcia, F. Váša, M. G. Kitzbichler, K. Wagstyl, P. Fonagy, R. J. Dolan, et al., "Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks," *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150362 (2016).

[14] M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C.-K. K. Lee, et al., "Canonical genetic signatures of the adult human brain," *Nat. Neurosci.* 18, 1832–1844 (2015).

[15] I. A. C. Romme, M. A. De Reus, R. A. Ophoff, R. S. Kahn, M. P. van den Heuvel, M. A. de Reus, R. A. Ophoff, R. S. Kahn, and M. P. van den Heuvel, "Connectome disconnectivity and cortical gene expression in patients with schizophrenia," *Biol. Psychiatry* 81, 495–502 (2017).

[16] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology* 28, 76–81 (1993).

[17] P. J. Lang, M. M. Bradley, B. N. Cuthbert, and others, "International affective picture system (IAPS): Technical manual and affective ratings," *NIMH Cent. Study Emot. Atten.* 1, 39–58 (1997).

[18] M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. van de Lagemaat, K. A. Smith, A. Ebbert, et al., "An anatomically comprehensive atlas of the adult human brain transcriptome," *Nature* 489, 391–399 (2012).

[19] A. Keo, N. A. Aziz, O. Dzyubachyk, J. van der Grond, W. M. C. van Roon-Mom, B. P. F. Lelieveldt, M. J. T. Reinders, and A. Mahfouz, "Co-expression Patterns between *ATN1* and *ATXN2* Coincide with Brain Regions Affected in Huntington's Disease," *Front. Mol. Neurosci.* 10, 1–13 (2017).

[20] S. M. H. Huisman, B. van Lew, A. Mahfouz, N. Pezzotti, L. Michielsen, A. Vilanova, M. J. T. Reinders, H. Thomas, and P. F. Lelieveldt, "BrainScope : interactive visual exploration of the spatial and temporal human brain transcriptome," *Nucleic Acids Res.* 45, e83 (2017).

[21] E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists," *BMC Bioinformatics* 10, 1–7 (2009).

[22] Y. Zhang, K. Chen, S. A. Sloan, M. L. Bennett, A. R. Scholze, S. O'Keeffe, H. P. Phatnani, P. Guarnieri, C. Caneda, et al., "An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex," *J. Neurosci.* 34, 11929–11947 (2014).

[23] B. O. Mancarci, L. Toker, S. J. Tripathy, B. Li, B. Rocco, E. Sibille, and P. Pavlidis, "Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data," *eNeuro* 4, 1–20 (2017).

[24] 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).

[25] S. W. G. of the Psychiatric Genomics Consortium and others, "Biological insights from 108 schizophrenia-associated genetic loci.," *Nature* 511, 421–427 (2014).

[26] A. F. Pardiñas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D. Cameron, et al., "Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection," *Nat. Genet.* 50, 381–389 (2018).

[27] E. A. Stahl, G. Breen, A. J. Forstner, A. McQuillin, S. Ripke, V. Trubetskoy, M. Mattheisen, Y. Wang, J. R. I. Coleman, et al., "Genome-wide association study identifies 30 loci associated with bipolar disorder," *Nat. Genet.* 51, 793–803 (2019).

[28] N. R. Wray, S. Ripke, M. Mattheisen, M. Trzaskowski, E. M. Byrne, A. Abdellaoui, M. J. Adams, E. Agerbo, T. M. Air, et al., "Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression," *Nat. Genet.* 50, 668–681 (2018).

[29] J. Piñero, À. Bravo, N. Queralt-Rosinach, A. Gutiérrez-Sacristán, J. Deu-Pons, E. Centeno, J. García-García, F. Sanz, and L. I. Furlong, "DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Res.* 45, D833–D839 (2017).

[30] Y.-J. Lin, W.-L. Liao, C.-H. Wang, L.-P. Tsai, C.-H. Tang, C.-H. Chen, J.-Y. Wu, W.-M. Liang, A.-R. Hsieh, et al., "Association of human height-related genetic variants with familial short stature in Han Chinese in Taiwan," *Sci. Rep.* 7, 1–7 (2017).

[31] I. M. Heid, A. U. Jackson, J. C. Randall, T. W. Winkler, L. Qi, V. Steinthorsdottir, G. Thorleifsson, M. C. Zillikens, E. K. Speliotes, et al., "Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution," *Nat. Genet.* 42, 949–960 (2010).

[32] Y. Wang, M. Wang, S. Yin, R. Jang, J. Wang, Z. Xue, and T. Xu, "NeuroPep: a comprehensive resource of neuropeptides," *Database* 2015, bav038 (2015).

[33] L. T. C. M. van Weert, J. C. Buurstede, A. Mahfouz, P. S. M. Braakhuis, J. A. E. Polman, H. C. M. Sips, B. Roozendaal, J. Balog, E. R. de Kloet, et al., "NeuroD factors discriminate mineralocorticoid from glucocorticoid receptor DNA binding in the male rat brain," *Endocrinology* 158, 1511–1522 (2017).

[34] P. H. Lee, V. Anttila, H. Won, Y. C. A. Feng, J. Rosenthal, Z. Zhu, E. M. Tucker-Drob, M. G. Nivard, A. D. Grotzinger, et al., "Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders," *Cell* 179, 1469-1482.e11 (2019).

[35] P. J. Magistretti, J.-R. Cardinaux, and J.-L. Martin, "VIP and PACAP in the CNS: Regulators of Glial Energy Metabolism and Modulators of Glutamatergic Signaling," *Ann. N. Y. Acad. Sci.* 865, 213–225 (1998).

[36] N. Stroth, Y. Holighaus, D. Ait-Ali, and L. E. Eiden, "PACAP: a master regulator of neuroendocrine stress circuits and the cellular stress response," *Ann. N. Y. Acad. Sci.* 1220, 49 (2011).

[37] R. Giordano, A. Picu, L. Bonelli, M. Balbo, R. Berardelli, E. Marinazzo, G. Corneli, E. Ghigo, and E. Arvat, "Hypothalamus-pituitary-adrenal axis evaluation in patients with hypothalamo-pituitary disorders: comparison of different provocative tests," *Clin. Endocrinol. (Oxf).* 68, 935–941 (2008).

[38] Y.-Y. Mei, D. C. Wu, and N. Zhou, "Astrocytic regulation of glutamate transmission in schizophrenia," *Front. psychiatry* 9, 544 (2018).

[39] C. Murphy-Royal, G. R. Gordon, and J. S. Bains, "Stress-induced structural and functional modifications of astrocytes—Further implicating glia in the central response to stress," *Glia* 67, 1806–1820 (2019).

[40] J. L. Scavone, E. Asan, and E. J. Van Bockstaele, "Unraveling glutamate-opioid receptor interactions using high-resolution electron microscopy: Implications for addiction-related processes," *Exp. Neurol.* 229, 207–213 (2011).

[41] E. C. Degli Uberti, F. Petraglia, M. Bondanelli, A. L. Guo, A. Valentini, S. Salvadori, M. Criscuolo, R. E. Nappi, and A. R. Genazzani, "Involvement of μ-opioid receptors in the modulation of pituitary-adrenal axis in normal and stressed rats," *J. Endocrinol. Invest.* 18, 1–7 (1995).

[42] A. Stengel and Y. F. Taché, "Activation of brain somatostatin signaling suppresses CRF receptor-mediated stress response," *Front. Neurosci.* 11, 231 (2017).

[43] V. Krishnan, "Defeating the fear: New insights into the neurobiology of stress susceptibility," *Exp. Neurol.* 261, 412–416 (2014).

[44] M. R. Picciotto, M. J. Higley, and Y. S. Mineur, "Acetylcholine as a neuromodulator: cholinergic signaling shapes nervous system function and behavior," *Neuron* 76, 116–129 (2012).

[45] Y. S. Mineur, A. Obayemi, M. B. Wigestrand, G. M. Fote, C. A. Calarco, A. M. Li, and M. R. Picciotto, "Cholinergic signaling in the hippocampus regulates social stress resilience and anxiety-and depression-like behavior," *Proc. Natl. Acad. Sci.* 110, 3573–3578 (2013).

[46] E. F. Sener, S. Taheri, M. C. Sahin, K. K. Bayramov, M. K. Marasli, G. Zararsiz, E. Mehmetbeyoglu, D. B. Oztop, M. Canpolat, et al., "Altered global mRNA expressions of pain and aggression related genes in the blood of children with autism spectrum disorders," *J. Mol. Neurosci.* 67, 89–96 (2019).

[47] M. Nishioka, M. Bundo, S. Koike, R. Takizawa, C. Kakiuchi, T. Araki, K. Kasai, and K. Iwamoto, "Comprehensive DNA methylation analysis of peripheral blood cells derived from patients with first-episode schizophrenia," *J. Hum. Genet.* 58, 91–97 (2013).

[48] E. Baca-Garcia, C. Vaquero-Lorenzo, M. M. Perez-Rodriguez, M. Gratacòs, M. Bayés, R. Santiago-Mozos, J. M. Leiva-Murillo, M. de Prado-Cumplido, A. Artes-Rodriguez, et al., "Nucleotide variation in central nervous system genes among male suicide attempters," *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* 153, 208–213 (2010).

[49] L. N. Giordana, A. A. Bozzo, D. S. Cots, I. Monedero Cobeta, A. Rolando, D. Borghi, T. Diaz, H. F. Gauna, and M. C. Romanini, "The effect of chronic stress on prenatal development of the central nervous system," *Biotech. Histochem.* 90, 146–151 (2015).

[50] A. Miranda and N. Sousa, "Maternal hormonal milieu influence on fetal brain development," *Brain Behav.* 8, e00920 (2018).

[51] S. Entringer, C. Buss, and P. D. Wadhwa, "Prenatal stress, development, health and disease risk: A psychobiological perspective—2015 Curt Richter Award Paper," *Psychoneuroendocrinology* 62, 366–375 (2015).

[52] D. J. Christoffel, S. A. Golden, and S. J. Russo, "Structural and synaptic plasticity in stress-related disorders," *Rev. Neurosci.* 22, 535–549 (2011).

[53] A. Etkin, A. Gyurak, and R. O'Hara, "A neurobiological approach to the cognitive deficits of psychiatric disorders," *Dialogues Clin. Neurosci.* 15, 419 (2013).

[54] M. Joëls, R. A. Sarabdjitsingh, and H. Karst, "Unraveling the time domains of corticosteroid hormone influences on brain activity: rapid, slow, and chronic modes," *Pharmacol. Rev.* 64, 901–938 (2012).

[55] A. T. Spijker and E. F. C. van Rossum, "Glucocorticoid receptor polymorphisms in major depression: Focus on glucocorticoid sensitivity and neurocognitive functioning," *Ann. N. Y. Acad. Sci.* 1179, 199–215 (2009).

[56] C. A. McClung, "Glucocorticoid receptor function and resilience: a tale of mice and men," *Biol. Psychiatry* 77, 310–311 (2015).

[57] F. ter Heegde, R. H. De Rijk, and C. H. Vinkers, "The brain mineralocorticoid receptor and stress resilience," *Psychoneuroendocrinology* 52, 92–110 (2015).

[58] S. M. Wojcik, M. Tantra, B. Stepniak, M. Kwun-nok, K. Müller-Ribbe, M. Begemann, A. Ju, S. Papiol, A. Ronnenberg, et al., "Genetic markers of a Munc13 protein family member, *BAIAP3*, are gender specifically associated with anxiety and benzodiazepine abuse in mice and humans," *Mol. Med.* 19, 135–148 (2013).

[59] A. Savarese and A. W. Lasek, "Regulation of anxiety-like behavior and *Crhr1* expression in the basolateral amygdala by LMO3," *Psychoneuroendocrinology* 92, 13–20 (2018).

[60] B. Tasic, Z. Yao, L. T. Graybuck, K. A. Smith, T. N. Nguyen, D. Bertagnolli, J. Goldy, E. Garren, M. N. Economo, et al., "Shared and distinct transcriptomic cell types across neocortical areas," *Nature* 563, 72–78 (2018).

[61] B. B. Lake, S. Chen, B. C. Sos, J. Fan, G. E. Kaeser, Y. C. Yung, T. E. Duong, D. Gao, J. Chun, et al., "Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain," *Nat. Biotechnol.* 36, 70–80 (2018).

[62] N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, et al., "Massively parallel single-nucleus RNA-seq with DroNc-seq," *Nat. Methods* 14, 955–958 (2017).

[63] 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).

[64] C. H. Vinkers, A. L. Kalafateli, B. P. F. Rutten, M. J. Kas, Z. Kaminsky, J. D. Turner, and M. P. M. Boks, "Traumatic stress and human DNA methylation: a critical review," *Epigenomics* 7, 593–608 (2015).

[65] I. Zovkic, J. P. Meadows, G. A. Kaas, and J. D. Sweatt, "Interindividual variability in stress susceptibility: a role for epigenetic mechanisms in PTSD," *Front. Psychiatry* 4, 60 (2013).

[66] A. M. Stankiewicz, A. H. Swiergiel, and P. Lisowski, "Epigenetics of stress adaptations in the brain," *Brain Res. Bull.* 98, 76–92 (2013).

[67] T. Klengel and E. B. Binder, "Epigenetics of stress-related psychiatric disorders and gene × environment interactions," *Neuron* 86, 1343–1357 (2015).

[68] V. Krishnan, M.-H. Han, D. L. Graham, O. Berton, W. Renthal, S. J. Russo, Q. LaPlant, A. Graham, M. Lutter, et al., "Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions," *Cell* 131, 391–404 (2007).

[69] A. Riecher-Rössler, "Sex and gender differences in mental disorders," *The Lancet Psychiatry* 4, 8–9 (2017).

[70] D. A. Bangasser, S. R. Eck, A. M. Telenson, and M. Salvatore, "Sex differences in stress regulation of arousal and cognition," *Physiol. Behav.* 187, 42–50 (2018).

[71] M. E. Kret and B. De Gelder, "A review on sex differences in processing emotional signals," *Neuropsychologia* 50, 1211–1221 (2012).

[72] J. V Zorn, R. R. Schür, M. P. Boks, R. S. Kahn, M. Joëls, and C. H. Vinkers, "Cortisol stress reactivity across psychiatric disorders: a systematic review and meta-analysis," *Psychoneuroendocrinology* 77, 25–36 (2017).

Chapter 5
Transcriptomic signatures associated with regional cortical thickness changes in Parkinson's disease

Arlin Keo
Oleh Dzyubachyk
Jeroen van der Grond
Jacobus J. van Hilten
Marcel J. T. Reinders
Ahmed Mahfouz

This chapter was published online in: *bioRxiv* (2020), doi: 10.1101/2020.06.19.158808 (in submission).
Supplementary material is available online at:
https://www.biorxiv.org/content/10.1101/2020.06.19.158808v1.supplementary-material

ABSTRACT

Cortical atrophy is a common manifestation in Parkinson's disease (PD), particularly in later disease stages. Here, we investigated patterns of cortical thickness using T1-weighted anatomical MRI data of 149 PD patients and 369 controls. To elucidate the molecular underpinnings of cortical thickness changes in PD, we performed an integrated analysis of brain-wide healthy transcriptomic data from the Allen Human Brain Atlas and neuroimaging features. For this purpose, we used partial least squares regression to identify gene expression patterns correlated with cortical thickness changes. In addition, we identified gene expression patterns underlying the relationship between cortical thickness and clinical domains of PD. Our results show that genes whose expression in the healthy brain is associated with cortical thickness changes in PD are enriched in biological pathways related to sumoylation, regulation of mitotic cell cycle, mitochondrial translation, DNA damage responses, and ER-Golgi traffic. The associated pathways were highly related to each other and all belong to cellular maintenance mechanisms. The expression of genes within most pathways was negatively correlated with cortical thickness changes, showing higher expression in regions associated with decreased cortical thickness (atrophy). On the other hand, sumoylation pathways were positively correlated with cortical thickness changes, showing higher expression in regions with increased cortical thickness (hypertrophy). Our findings suggest that alterations in the balanced interplay of these mechanisms play a role in changes of cortical thickness in PD and possibly influence motor and cognitive functions.

5.1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder characterized by a progressive loss of dopaminergic and non-dopaminergic neurons in the brain and peripheral and autonomic nervous system [1]. Cortical atrophy occurs during the later disease stages and has been associated with cognitive decline, including executive, attentional, memory, and visuospatial deficits [2,3]. Although MRI studies of patient brains have tried to link regional cortical atrophy to clinical features of the disease [4–8], little is known about the pathobiology that underlies the selective cortical vulnerability in PD.

Analyzing the transcriptome in vulnerable cortical regions may aid in better understanding the underlying molecular mechanisms of atrophy in PD. Although gene expression data of human post-mortem PD brains is available, most findings relate to studies that focused only on one or few coarse brain regions [9]. To perform whole brain analysis of both gene expression and imaging data, studies turn to the Allen Human Brain Atlas (AHBA), a high resolution gene expression atlas covering the entire brain of six adult donors without any history of neurological disorders [10,11]. The AHBA has been combined with functional MRI data of PD patients and revealed that the regional expression of *MAPT*, but not *SNCA*, correlated with the loss of regional connectivity [12]. Using a similar approach, correlations were identified between a cortical atrophy pattern and the regional expression of 17 genes implicated in PD [13]. Although both studies used spatial transcriptomics to explore gene expression across the whole brain, they only analyzed the expression of a limited set of genes that are of interest to PD, e.g., those that are known as genetic risk factors.

To investigate the relationship between high dimensional genome-wide expression patterns and imaging data, multivariate analysis methods are required. Partial least squares (PLS) regression has been used to perform simultaneous analysis of brain-wide gene expression from the AHBA and neuroimaging data of adolescents, healthy adults, and Huntington's disease [14–16]. The PLS approach allows the linking of multiple predictor variables (genes) and multiple response variables (imaging features) and deals with multicollinearity by projecting variables to a smaller set of components that are maximally correlated between both datasets. Thus, PLS is an attractive model to identify gene expression patterns associated with imaging features.

Here, we exploited PLS regression to find transcriptomic signatures that are related to changes in cortical thickness (CT) in PD. MRI data was obtained from patients and age-matched controls to find CT changes across all cortical regions. Gene expression samples from healthy donors in the AHBA were anatomically mapped to the cortical regions to find brain-wide gene expression patterns predictive of the CT changes observed in PD patients. In addition, we assessed the relationships between CT and clinical scores in PD patients and used a second PLS model to find expression patterns associated with these relationships across all cortical regions. With these models we address three research questions: (1) Which cortical regions show CT changes in PD, (2) Which genes and biological pathways show expression patterns associated with these regional changes, and (3) Which molecular mechanisms underlie the relationships between CT and clinical scores in PD. To answer

these questions, we explored the whole transcriptome in cortical regions of the healthy brain to find expression signatures predictive of imaging features in PD.

5.2 Materials and methods

5.2.1 MRI data acquisition

MRI images of 149 PD patients (mean age = 64.8 years; 65.7% male) were obtained from a cross-sectional cohort study and is part of the 'PROfiling PARKinson's disease' (PROPARK) study [17]. PD patients were recruited from the outpatient clinic for Movement Disorders of the Department of Neurology of the Leiden University Medical Center and nearby university and regional hospitals. All participants fulfilled the United Kingdom Parkinson's Disease Society Brain Bank criteria for idiopathic PD [18]; written consent was obtained from all participants. The Medical Ethics Committee of the LUMC approved the study. Three-dimensional T1-weighted anatomical images were acquired on a 3 Tesla MRI scanner (Philips Achieva, Best, the Netherlands) using a standard 32-channel whole-head coil. Acquisition parameters were: repetition time = 9.8 ms, echo time = 4.6 ms, flip angle = 8°, field of view 220 × 174 × 156 mm, 130 slices with a slice thickness of 1.2 mm with no gap between slices, resulting in a voxel size of 1.15 mm × 1.15 mm × 1.20 mm.

Three-dimensional T1-weighted images from 369 controls (mean age = 65.7 years; 48.1% male) were acquired in a different cohort [19], where all imaging was performed on a whole body 3 Tesla MRI scanner (Philips Medical Systems, Best, the Netherlands), using the following imaging parameters: TR = 9.7 ms, TE = 4.6 ms, FA = 8°, FOV = 224 × 177 × 168 mm. The anatomical images covered the entire brain with no gap between slices resulting in a nominal voxel size of 1.17 × 1.17 × 1.4 mm. Acquisition time was approximately 5 min.

5.2.2 Cortical thickness changes in segmented cortical regions

CT in cortical regions of PD patients and controls was determined using cortical parcellation implemented in FreeSurfer version 5.3.0 [20]. The FreeSurfer algorithm automatically parcellates the cortex and assigns a neuroanatomical label to each location on a cortical surface model based on probabilistic information. The parcellation scheme of the Desikan–Killiany atlas was used to divide the cortex into 34 regions per hemisphere [21].

To assess CT changes between patients (149) and controls (369), a two-tailed *t*-test assuming unequal variances was applied in SPSS Statistics version 23. *P*-values were corrected for multiple testing across 68 cortical regions using the Benjamini-Hochberg (BH) method. A two-tailed *t*-test was also used to assess CT differences between the left and right hemisphere for each one of the 34 cortical regions, with *P*-values being BH-corrected across the 34 cortical regions.

5.2.3 Clinical scores

All patients underwent standardized assessments, and an evaluation of demographic and clinical characteristics [17]. MDS-UPDRS is a clinical rating scale consisting of four parts: (I) Non-motor Experiences of Daily Living; (II) Motor Experiences of Daily Living; (III) Motor Examination; (IV) Motor Complications [22]. UPDRSTOTSCR is the total score of all four

parts. The SENS-PD scale is a composite score of non-dopaminergic symptoms [23], LED is the levodopa equivalent dose [24], and MMSE is the mini-mental state examination [25].

5.2.4 Relationship between CT and clinical scores

We used CT data and clinical scores from 149 PD patients to determine the relationships between CT and clinical domains. We selected 9 clinical features with numeric (non-nominal) values for which scores were available for 82-123 patients: AGEONSET, SENSPDSC, MDS_UPDRS_3, MMSE, LED, MDS_UPDRS_1, MDS_UPDRS_2, MDS_UPDRS_4, and UPDRSTOTSCR (Supplementary Figure 1).

The correlation between CT and the scores of each clinical feature individually was determined across patients by applying linear regression. To obtain maximum correlation, separate linear regression models were used for each combination of a region and clinical feature:

$$CT_i = \alpha + \beta_1 K_j + \beta_2 Age + \varepsilon \quad (1)$$

where CT_i is the CT of one region i across patients, K_j is the score of one clinical feature j across patients. Age is taken into account to correct for the age of patients. α is the background term, β_1 is the regression coefficient for K_j, β_2 is the regression coefficients for Age, and ε is the residual. The regression coefficient β_1 was used to determine the relationship between CT and clinical domain scores, and assessed for statistical significance where *P*-values were BH-corrected for 34 regions and 9 clinical features (*t*-test, H_0: $\beta_1 = 0$, $P < 0.05$).

5.2.5 Mapping transcriptomic data to cortical regions

We downloaded normalized gene expression data from the Allen Human Brain Atlas (AHBA; http://human.brain-map.org/), a microarray data set of 3,702 anatomical brain regions from six non-neurological individuals (5 males and 1 female, mean age 42, range 24–57 years [10]). To analyze the transcriptome in the cortical regions, we used the mapping of AHBA samples to cortical regions in neuroimaging data proposed in [11], where they applied Freesurfer on T1 MRIs of the six donors in the AHBA to segment the cortical regions according to the Desikan-Killiany atlas. AHBA samples were mapped to 34 cortical regions from the left hemisphere, since for only two out of six brains samples were collected from both hemispheres and for four brains they only sampled from the left hemisphere. By only analyzing the left hemisphere, we assumed that there are small to no differences in gene expression between the left and right hemisphere [10]. Samples were assigned to a segmented cortical region when their MNI coordinates corresponds to a voxel within a parcel, including samples that are up to 2 mm away from any voxel in the parcel. In total 1,284 samples from the AHBA were assigned to the 34 cortical regions.

5.2.6 Partial least squares (PLS) model-1 and model-2

We used PLS regression (R-package *pls* 2.7) to find gene expression patterns across the 34 cortical regions that are predictive of gray matter atrophy and possibly their relationship to scores of nine clinical domains (Supplementary Methods). PLS regression and principal component analysis regression are both methods where the original measurements are

projected to latent variables to study the data in reduced dimensions (Figure 5.1A). PLS however, projects variables from each dataset to latent variables such that they are maximally correlated between two datasets X and Y (Figure 5.1B). In this study, the predictor X is a gene expression matrix of 34 regions (n) in the left hemisphere and all 20,017 genes (m) and is used to predict imaging variables (p) in the same set of 34 cortical regions. For each cortical region and each gene, expression levels were averaged across samples that fall within that cortical region and then averaged across the six donors from the AHBA, such that the input matrix of predictor variables contains one expression value for every gene per cortical region. We implemented two PLS models (Figure 5.1C): one single-response PLS model, *model-1*, to predict CT changes, measured as the t-statistics of ΔCT between PD patients and controls, and one multi-response PLS model, *model-2*, to predict the correlation between CT and clinical scores in PD patients, measured as the t-statistics of the coefficients β_1 in Equation 1.

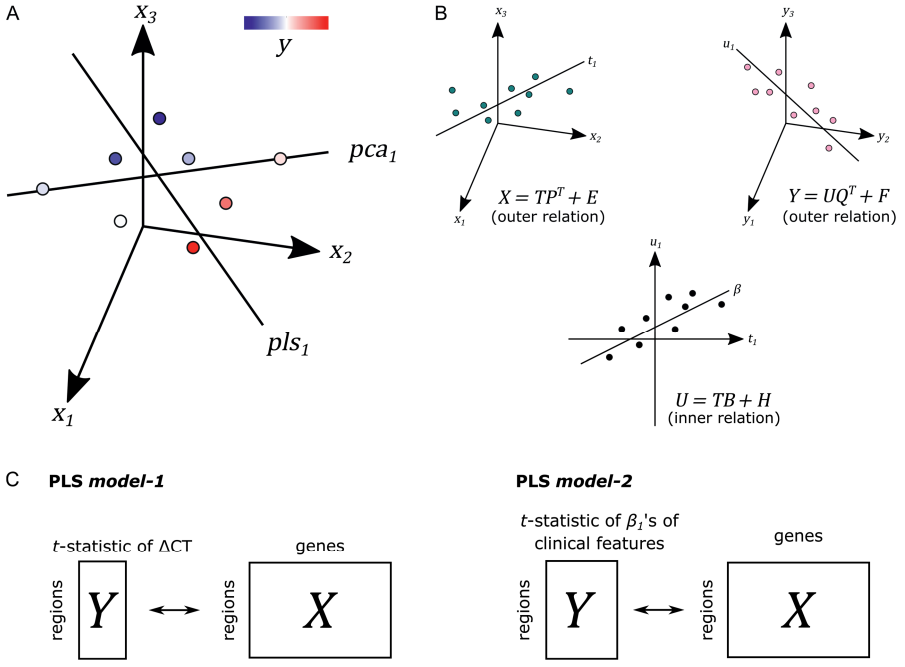


Figure 5.1 Principal of partial least squares regression (PLS). (A) Principal component analysis (PCA) and PLS project measurements to a new latent space. Unlike PCA, PLS tries to find a latent space that is maximally correlated with another measurement $\boldsymbol{y}$ from dataset $\boldsymbol{Y}$ on the same samples. (B) The first latent component $\boldsymbol{t_1}$ of dataset $\boldsymbol{X}$ is maximally correlated with the first latent factor $\boldsymbol{u_1}$ of dataset $\boldsymbol{Y}$. $\boldsymbol{T}$ and $\boldsymbol{U}$ scores determine the outer relations of individual datasets in the model. The coefficient $\boldsymbol{\beta}$ determines the inner relation between both datasets X and Y in the model (more details in Supplementary Methods). (C) In PLS model-1, we used regional gene expression as input to predict the regional t-statistics of ΔCT. Given the PLS model, $\boldsymbol{R}$ in Equation 5 and 6 in Supplementary Methods is used as gene weights. In PLS model-2, we used the same input to predict the t-statistics of correlation coefficients $\boldsymbol{\beta_1}$ of clinical features from Equation 1.

5.2.7 Pathway enrichment

Pathway enrichment analysis was done using gene set enrichment analysis (GSEA) and 2,225 pathways from the Reactome database in ReactomePA R-package version 1.28. Genes were ranked based on their weights to each PLS component; R in Equation 5 and 6 in Supplementary Methods. Pathways were significant when the FDR-adjusted $P < 0.05$.

5.2.8 Data and code availability

Transcriptomic data from the AHBA is available at http://human.brain-map.org/. All scripts were run in R version 1.6 and can be found online at https://github.com/arlinkeo/pd_pls.

5.3 Results

5.3.1 CT changes between PD patients and controls

We analyzed CT changes between PD patients and healthy controls (ΔCT) as a measure for gray matter loss (Figure 5.1C). Each of the 68 cortical regions from both hemispheres was assessed, for which ΔCT was statistically significant in 10 cortical regions (t-test, BH-corrected $P < 0.05$; Figure 5.2A and Supplementary Table 1). The lateral occipital cortex showed decreased CT in patients compared to controls in both the left hemisphere and right hemisphere. The left caudal anterior cingulate, right isthmus cingulate, and right pericalcarine also showed decreased CT in patients. Cortical regions with increased CT in patients included the pars opercularis from both the left hemisphere and right hemisphere, the right rostral middle frontal cortex, right temporal pole, and right superior temporal cortex. In general, we observed more decreased CT (atrophy) in caudal regions of the cortex compared to rostral regions that showed increased CT (hypertrophy).

5.3.2 CT changes between hemispheres in PD

Clinical symptoms appear asymmetrical at disease onset with the left hemisphere being more susceptible to degeneration than the right [26]. To assess whether this asymmetry is reflected also in the observed atrophy patterns, we compared the CT between the left and right hemisphere for each of the 34 cortical regions in PD patients. We found six cortical regions that showed significant hemispheric differences (BH-corrected $P < 0.05$; Figure 5.2B and Supplementary Table 2). For five out of six significant regions, CT was indeed smaller in the left hemisphere compared to the right: banks of superior temporal sulcus, entorhinal cortex, temporal pole, medial orbitofrontal cortex, and lateral occipital cortex. For the lateral orbitofrontal cortex, the CT was larger in the left hemisphere compared to the right.

5.3.3 Gene expression patterns predictive of CT changes in PD patients

To identify the molecular mechanisms underlying CT changes in PD, we integrated the imaging features with brain-wide gene expression profiles from the AHBA (Figure 5.1C). Using PLS *model-1* (see Methods), the expression of all 20,017 genes in 34 brain regions from the left hemisphere was used as predictor variables and we used the t-statistics of ΔCT between PD patients and controls in the 34 regions (Supplementary Table 1) as a single response variable. The number of AHBA samples varied between 0 and 92 for each one of the six brain donors and 34 cortical regions (Supplementary Table 3).

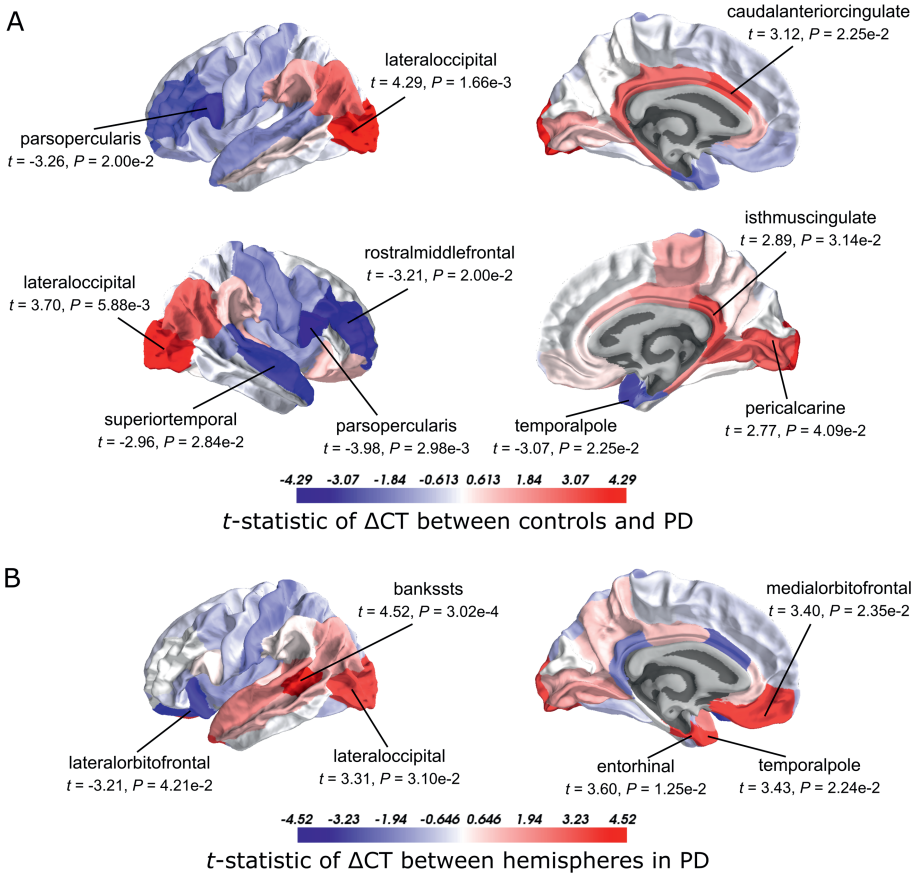


Figure 5.2 t-Statistics of cortical thickness changes (ΔCT) across cortical regions. (A) CT was assessed between PD patients and healthy controls. Higher t-statistics (red) indicate a larger CT in controls compared to the CT in patients and thus corresponds to cortical atrophy. (B) CT in the left hemisphere compared to the right hemisphere in PD patients. Higher t-statistics (red) indicate a larger CT in the right hemisphere compared to the left hemisphere and thus corresponds to cortical atrophy in the left hemisphere. *P*-values are BH-corrected and significant regions ($P < 0.05$) are labeled.

The PLS components that explain maximum covariance between the input space and the response variable are derived from successively deflated predictor and response matrices. Hence, the first component of the predictor matrix, *component-1*, has maximum covariance with the first component of the response matrix, and the second component of the predictor matrix, *component-2*, has maximum covariance with the second component of the response matrix, etc. Since PLS *model-1* has a single response variable, *component-1* of the response

matrix is equal to a scaled version of the single response variable. As such, we only examined PLS *component-1* of the predictor matrix (additional checking with leave-one-out cross-validation showed that the optimal number of components is indeed one, Supplementary Figure 2).

The scores of PLS *component-1* of the predictor variables (genes) showed a caudal-to-rostral expression pattern (Figure 5.3A) that was correlated with CT changes in PD brains (Figure 5.3B), i.e. gene expression of PLS *component-1* was high in caudal regions associated with atrophy and low in regions associated with hypertrophy. The Pearson correlation between the PLS *component-1* scores of the predictor variables (gene expression) and the response variable (*t*-statistics of ΔCT) was 0.58, and explained 20.5% of the variance in gene expression and 34.2% of the variance in CT changes. Cortical atrophy was highest in the lateral occipital cortex and related to high PLS *component-1* scores. The pericalcarine showed the highest PLS *component-1* score. These results showed that the expression profiles of a weighted combination of genes can be predictive of CT changes in PD.

5.3.4 Functionality of genes predictive of CT changes

A PLS component of the predictor variables is a linear combination of weighted gene expression. We used the gene weights of PLS *component*-1 to perform GSEA analysis and revealed significant enrichment of 90 pathways, which were among others involved in DNA damage checkpoints, stabilization of p53, regulation of apoptosis, mitochondrial translation, and SUMOylation of chromatin organization proteins (Supplementary Table 4). High overlap of genes between the enriched pathways suggested that these functional processes are highly related to each other (Supplementary Figure 3).

Significant pathways are either positively or negatively correlated with CT changes based on the median weight of genes within pathways. Out of the 90 pathways that were significantly enriched, three pathways were positively correlated with the *t*-statistic of ΔCT. These included SUMOylation of chromatin organization proteins, signaling by cytosolic FGFR1 fusion mutants, and class C/3 (Metabotropic glutamate/pheromone receptors). Higher mean expression of genes within these three pathways is related to cortical atrophy (higher *t*-statistics of ΔCT); as apparent in the lateral occipital cortex (Figure 5.3C and Supplementary Figure 4). The positive correlation also indicates that a lower expression of these pathways is related to cortical hypertrophy (lower *t*-statistics of ΔCT). We found 87 negatively correlated pathways (median gene weight < 0). These pathways seem to play a role in the mitochondrial regulation of mitosis as we found pathways for mitochondrial translation, the regulation of mitotic cell cycle, p53-(in)dependent DNA damage checkpoints, and the degradation of mitotic proteins, such as cyclins A, and D. In general, the mean expression of genes in the negatively correlated pathways was high in cortical regions that showed hypertrophy, such as the pars opercularis or the entorhinal cortex.

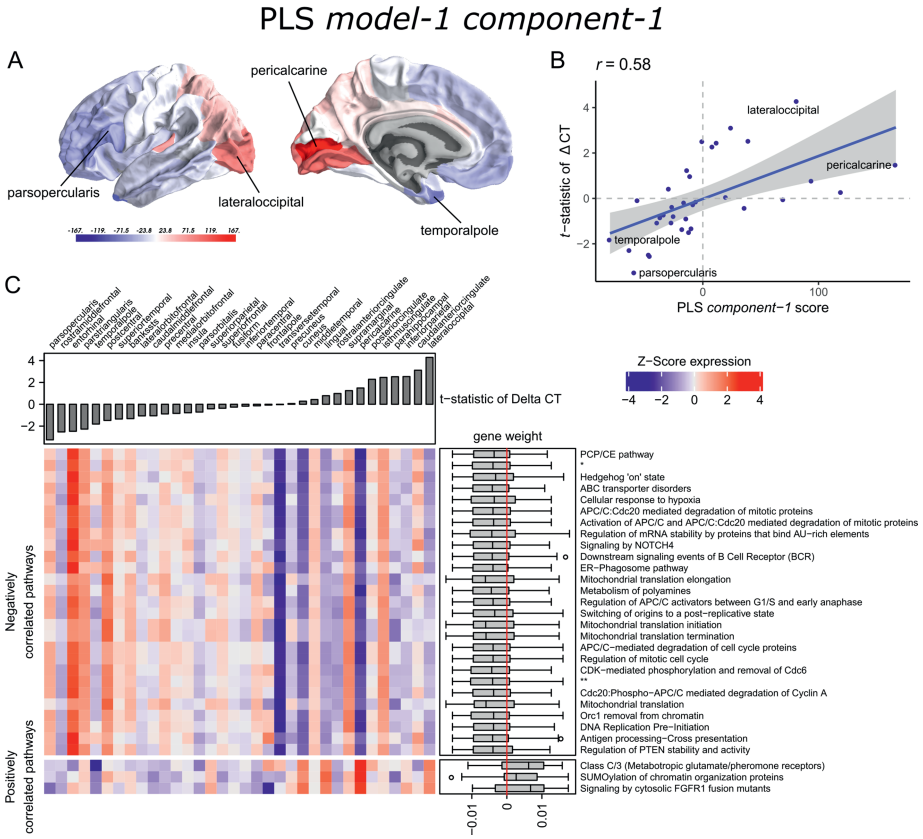


Figure 5.3 Transcriptional signatures to predict t-statistics of ΔCT between PD patients and controls in PLS model-1. (A) PLS component-1 scores of predictor variables (gene expression) visualized in cortical regions (lateral and medial view of the left hemisphere). (B) Regression fit of the latent predictor variable, PLS component-1 scores, with the single response variable, CT changes in PD measured as the t-statistics of ΔCT between PD patients (149) and controls (369) across the 34 cortical regions. (C) Mean expression of genes in the top 30 significant pathways (rows) across cortical regions (columns). A complete heatmap with all significant pathways is given in Supplementary Figure 4. The correlation between transcriptomic signatures and CT changes in PD across cortical regions is predicted by the gene weights for PLS component-1 shown in boxplots for each pathway where the median weight is either negative or positive. Negatively correlated pathways show high gene expression in regions with low t-statistics of ΔCT and gene expression decreases in regions with higher t-statistics of ΔCT. In our analysis, negative t-statistics correspond to increased CT (cortical hypertrophy) and positive t-statistics of ΔCT correspond to decreased CT (cortical atrophy). Positively correlated pathways show low expression in regions with low t-statistics of ΔCT and expression increases in regions with higher t-statistics of ΔCT. * = APC:Cdc20 mediated degradation of cell cycle proteins prior to satisfation of the cell cycle checkpoint. ** = APC/C:Cdh1 mediated degradation of Cdc20 and other APC/C:Cdh1 targeted proteins in late mitosis/early G1.

5.3.5 RELATIONSHIPS BETWEEN CLINICAL SCORES AND CORTICAL THICKNESS

Next, we set to understand the relationship between CT in 34 cortical regions and clinical scores of PD patients. Linear regression was used to predict clinical scores from CT across patients and obtain regression coefficients, β_1, for each cortical region and clinical domain (Equation 1). We assessed the *t*-statistics of the regression coefficients instead of the coefficients β_1 themselves (H_0: $\beta_1 = 0$) (Figure 5.4). Negative *t*-statistics showed that most combinations of cortical regions and clinical features are negatively correlated. For all clinical features, higher scores also indicate more severe symptoms, except for MMSE scores where lower scores indicate more severe symptoms, and thus showed positive relationships with CT. In most regions, age at onset (AGEONSET) also showed positive relationships with CT, indicating that age at onset has an effect on the loss of CT. While these general interpretations apply to most cortical regions, some regions showed different relationships with CT. For example, CT in the rostral anterior cingulate is negatively related to age at onset, and positively related to MDS-UPDRS 4 scores.

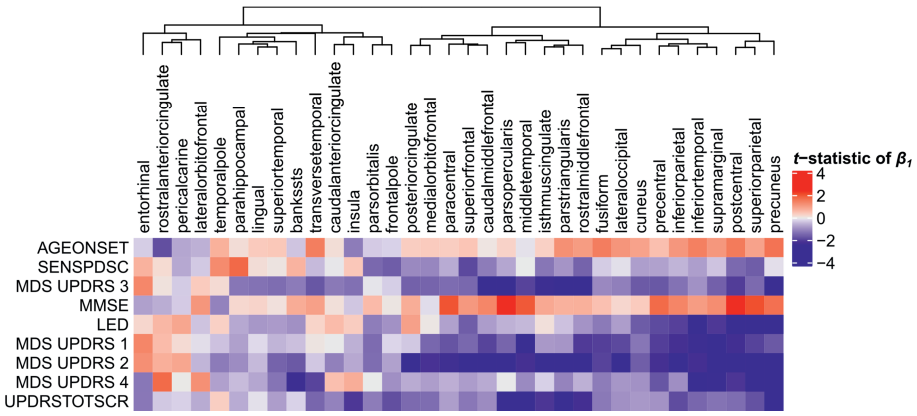


Figure 5.4 Relationship between clinical scores and CT across PD patients. Linear regression was used to predict clinical scores from CT across at most 123 PD patients (Supplementary Figure 1). Separate models were used for each clinical feature (row) and cortical region (column) to obtain regression coefficients, see Equation 1. The heatmap shows the two-sided t-statistics of the regression coefficient when tested for H0: $\boldsymbol{\beta_1} = 0$. Regions (columns) are clustered based on complete linkage of the Euclidean distance of the t-statistics of $\boldsymbol{\beta_1}$.

5.3.6 GENES PREDICTIVE OF RELATIONSHIPS BETWEEN CLINICAL SCORES AND CORTICAL THICKNESS

With PLS *model-2*, we examined gene expression patterns that are predictive of the relationship between CT and clinical scores measured as *t*-statistics of the correlation coefficients $\boldsymbol{\beta_1}$ in Equation 1 (Figure 5.1C). We selected the first two PLS components for further analysis, which explained 36% of the variance of the predictor variables and 37% of the variance of the response variables (Supplementary Figure 5). PLS *component-1* scores of the predictor variables showed a ventral-to-dorsal gene expression pattern (Figure 5.5A) that is correlated with the PLS *component-1* scores of the response variables (Pearson $r = 0.76$,

Figure 5.5B). The dorsal regions include the postcentral gyrus which is part of the primary somatosensory cortex. PLS *component-2* scores of the predictor variables showed a caudal-to-rostral gene expression pattern (Pearson $r = 0.56$, Figure 5.6A) that is correlated with the PLS *component-2* scores of the response variables (Figure 5.6B). Moreover, we assessed PLS *component-3* (Pearson $r = 0.76$ between the predictors and response variables), which additionally explained 9% variance of the predictor variables and 11% variance of the response variables. However, further analysis revealed there were no enriched pathways for *component-3* limiting the functional interpretation of this component.

PLS *component-1* and *component-2* of the predictor variables showed 144 and 230 significantly enriched pathways, respectively, with 54 overlapping pathways between the two components (Supplementary Table 5 and 6). Both components showed a cluster of related pathways involved in anterograde and retrograde transport between Golgi and endoplasmic reticulum (ER), and asparagine N-linked glycosylation (Supplementary Figure 6 and 7). Other pathways that overlapped between the two components included macroautophagy, mitochondrial translation, mitochondrial biogenesis, mitochondrial protein import, DNA damage/telomere stress induced senescence, oxidative stress induced senescence, and protein localization.

Furthermore, PLS *component-1* showed enrichment of pathways involved in tRNA and rRNA processing in the nucleus and mitochondrion, voltage-gated potassium channels, uptake and actions of bacterial toxins, and interleukin signaling. PLS *component-2* showed strong enrichment of neutrophil degranulation, DNA replication, p53-(in)dependent DNA damage response, and chaperonin-mediated protein folding, and tubulin folding. Notably, the gene expression pattern of PLS *component-2* was also associated with several sumoylation pathways and pathways involved in mitotic cell cycles and the degradation of mitotic proteins (Supplementary Table 6).

The enriched pathways for PLS *component-1 and component-2* either showed negative or positive median gene weights that inform about the sign of the correlation between genes within a pathway and the PLS component score of the response variables (Figure 5.5C, Figure 5.6C, Supplementary Figure 8 and 9). For example, the expression of genes within pathways relating to mitochondrial processes increases for higher PLS *component-1* scores of the response variables. We further assessed PLS *component-1 and component-2* scores of the predictor variables and their correlation with each individual response variable, which are the clinical features and their relationship with CT in PD patients (Figure 5.7). The rostral-to-dorsal expression pattern of PLS *component-1* is highly predictive of the relationship between CT and MMSE score in patients (Pearson's $r = 0.71$). Thus, pathways associated with PLS *component-1* may play an important role in cognitive circuits, which seems to be apparent based on their expression in the postcentral gyrus, but also the entorhinal cortex. PLS *component-2* scores showed low correlations with the clinical features and their relation with CT across cortical regions, and suggests weak associations between the expression patterns of PLS *component-2* and the response variables.

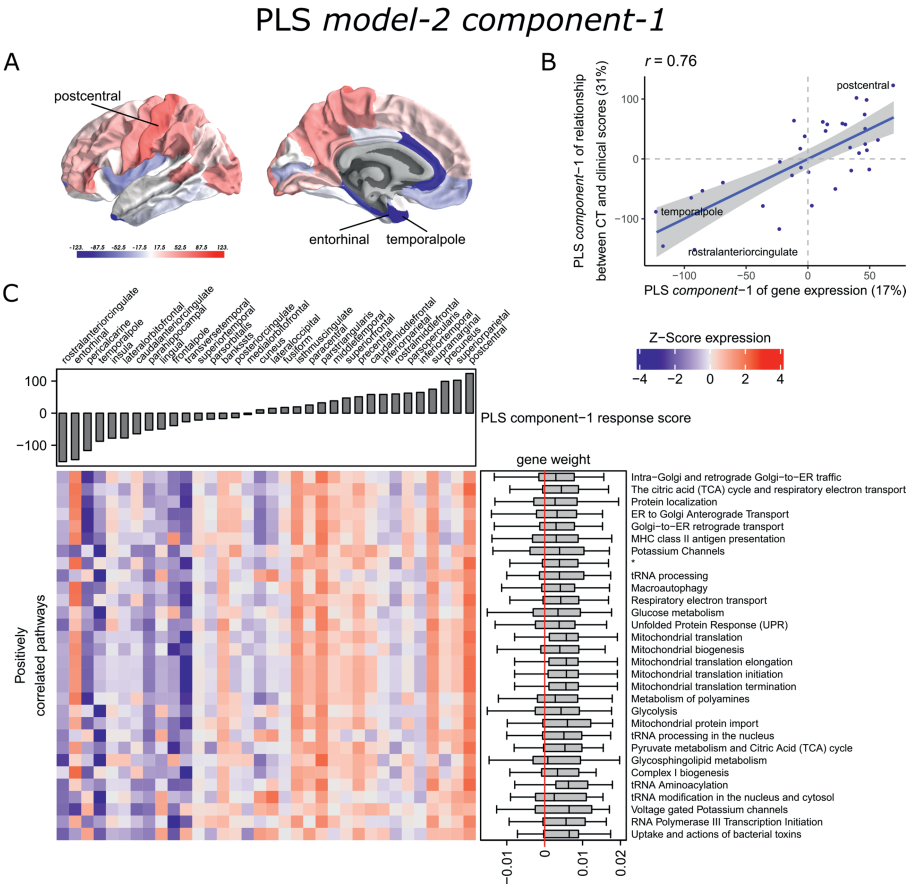


Figure 5.5 Transcriptional signatures of PLS *component-1* in PLS *model-2* predictive of the relationship between cortical thickness (CT) and clinical scores. (A) PLS scores for PLS *component-1* of the predictor variables (gene expression) and (B) its correlation with PLS *component-1* of the response variables (*t*-statistic of $\boldsymbol{\beta_1}$ in Equation 1). Axes show the percentage of explained variance for each component; *r* indicates the Pearson correlation. (C) Mean expression across cortical regions (columns) of genes in the top 30 significant pathways (rows). A complete heatmap with all significant pathways is given in Supplementary Figure 8. * = Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.

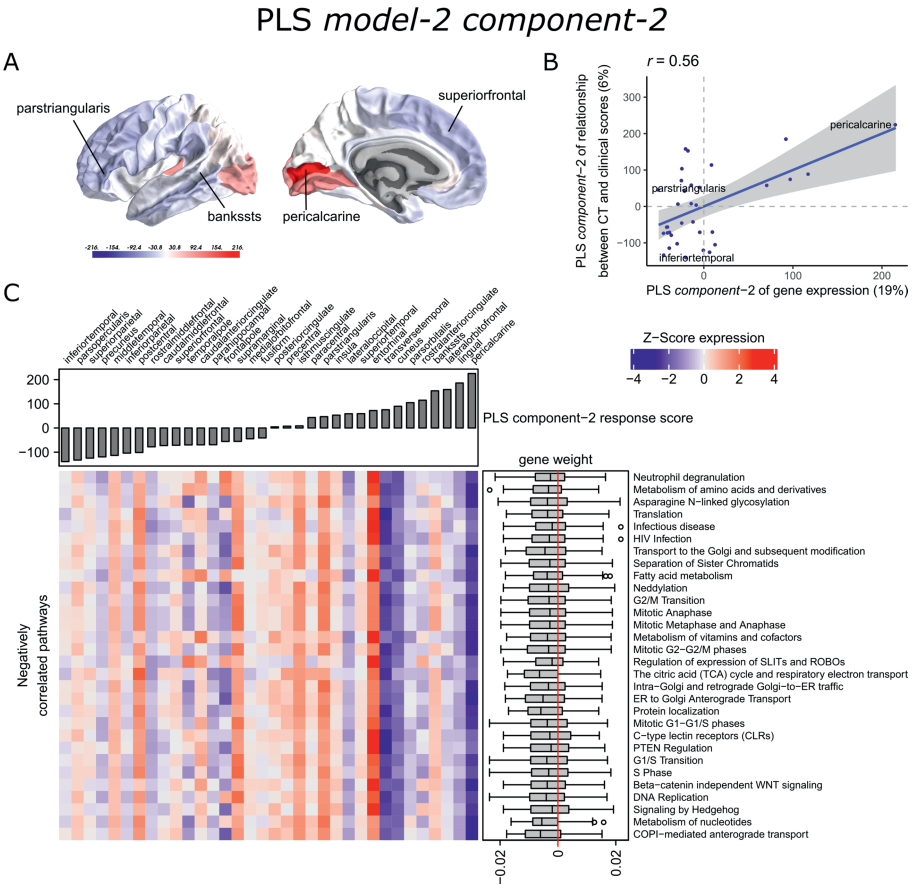


Figure 5.6 Transcriptional signatures of PLS *component-2* in PLS *model-2* predictive of the relationship between cortical thickness (CT) and clinical scores. (A) PLS scores for PLS *component-2* of the predictor variables (gene expression) and (B) its correlation with PLS *component-1* of the response variables (*t*-statistic of $\boldsymbol{\beta_1}$ in Equation 1). Axes show the percentage of explained variance for each component; *r* indicates the Pearson correlation. (C) Mean expression across cortical regions (columns) of genes in the top 30 significant pathways (rows). A complete heatmap with all significant pathways is given in Supplementary Figure 9.

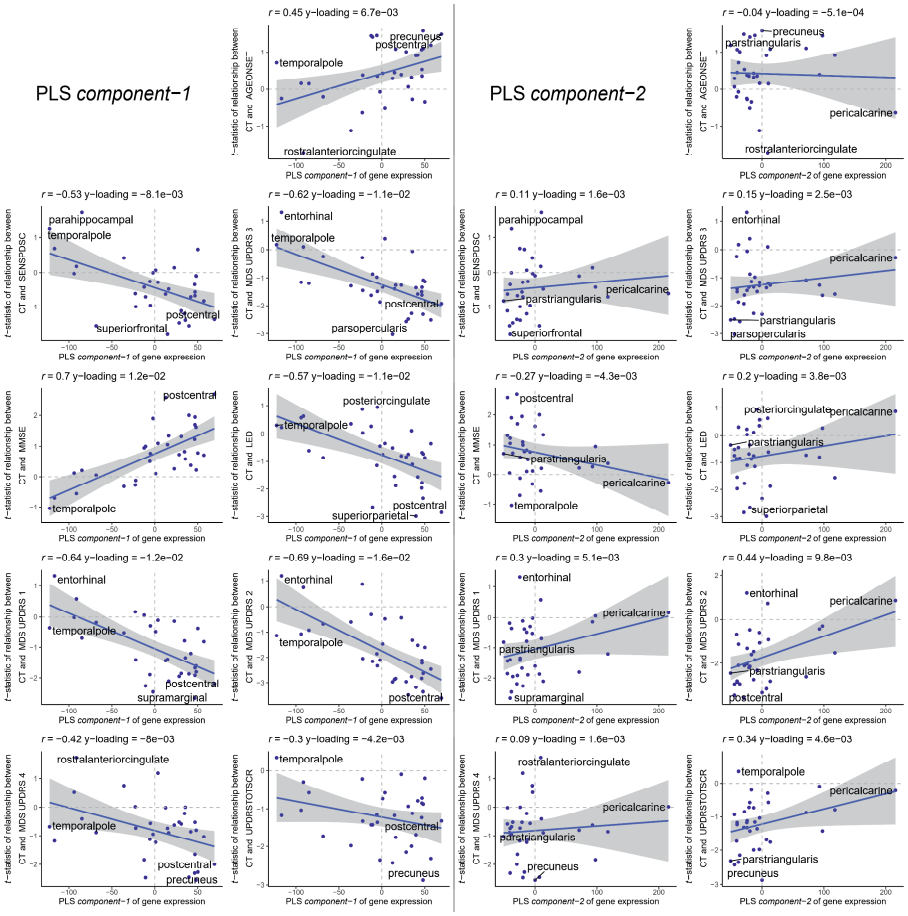


Figure 5.7 Correlations between PLS *model-2 component-1 and component-2* scores of the predictor variables and individual response variables. Each plot shows the correlation between the predictor variables of gene expression (x-axis) and the response variables which are the relationships between CT and scores of a clinical feature across cortical regions (y-axis). On top of each plot, the Pearson correlation and the Y-loadings (Q in Eq. 4 and 8) are shown; both values tell something about the sign (-/+) and magnitude (high/low) of the correlation. Each point or sample is one of the 34 cortical regions. Regions are labeled for those with minimum or maximum value along one of the axes.

5.4 Discussion

We found a caudal-to-rostral gene expression pattern that was correlated with CT changes in PD (PLS *model-1*); cortical atrophy was found in caudal regions while rostral regions showed cortical hypertrophy. This transcriptional signature was highly enriched for genes in biological pathways associated with mitochondrial translation and mitotic cell cycle regulation. We also found a ventral-to-dorsal and caudal-to-rostral gene expression pattern that was correlated with the relationship between CT and clinical domains of PD (PLS *model-*

2). Both transcriptional signatures were associated with similar pathways, including macroautophagy and Golgi-ER trafficking, and may be involved in the effect of CT on clinical scores, namely MMSE scores for cognitive assessment.

The CT analyses between disease conditions and hemispheres in patients revealed cortical regions that are susceptible to atrophy. Cortical atrophy in PD commonly occurs asymmetrical, with a preference for the left hemisphere, particularly in the early disease stages [26–29]. Here, we showed that five out of six regions with significant CT changes between hemispheres, indeed revealed more atrophy in the left hemisphere. Two cortical regions that showed significant changes between patients and controls, also showed changes between the left and right hemisphere. Our findings are in line with those of a previous study showing that cortical atrophy in PD most prominently affects the lateral occipital cortex, particularly in the left hemisphere [13]. The temporal pole showed hypertrophy in patients compared to controls, which was only significant in the right hemisphere. However, our analysis between hemispheres of PD brains suggests that the left temporal pole is more susceptible to CT loss than the right hemisphere. The remaining regions that were susceptible to CT changes showed atrophy in either the left or right hemisphere; however differences between hemispheres in patients could not be confirmed. All 10 regions that were different between patients and controls, except the pericalcarine, were earlier identified as part of two structural covariance networks that were related to gray matter atrophy in the same PD dataset as in this study [17]. Overall, we observed atrophy in caudal regions, which earlier has been associated with late stage PD [26].

With our findings of the PLS models we interpret gene expression patterns of the healthy brain in relation to imaging features observed in PD. The six adult donors of the AHBA had no known neuropsychiatric or neuropathological history [30], however it is unknown whether these individuals could have developed neurodegenerative diseases later in life. The observed spatial gene expression patterns reflect the physiological conditions in the adult healthy brain and are informative of important molecular mechanisms that are vulnerable in PD. The biological pathways found for PLS *model-1* were closely related as they shared many similar genes. These interrelated pathways suggest a strong functional relationship between molecular processes involving mitotic cell cycle, mitochondrial translation, transport between ER and Golgi, DNA damage checkpoints, and sumoylation. We found that differential regulation of these molecular processes across the brain was associated with CT changes observed in PD. Similar pathways were found in PLS *model-2* with multiple response variables corresponding to the relationships between CT and nine clinical domain scores in PD.

There is evidence that impaired cell cycle control plays a role in the pathogenesis of neurodegenerative diseases. In healthy conditions, differentiated neuronal cells become quiescent cells that cannot re-enter the cell cycle, however in neurodegenerative diseases they are reactivated which is associated with increased cell death [31]. Cell cycle checkpoints are controlled by cyclins that guide the cell from one phase to the next phase and its expression can induce cell cycle re-initiation [32]. Here, we found that regional expression

of pathways associated with the degradation of cell cycle proteins in healthy conditions were negatively correlated with CT changes in PD, i.e. higher expression was associated with cortical hypertrophy in rostral regions such as the pars opercularis and temporal pole. Reversely, we observed low expression of protein degradation pathways in caudal regions that were associated with atrophy, and therefore suggests that regions with low expression are more vulnerable to improper degradation of cell cycle proteins leading to cell cycle initiation. This indicates that regions with low expression of such essential pathways are predisposed to neurodegeneration.

We found that the expression of several pathways associated with DNA replication and p53-(in)dependent DNA damage responses and checkpoints were correlated with CT changes. DNA replication during the S-phase may control the survival of post-mitotic cells by DNA repair mechanisms or apoptosis followed by DNA damage, which seems to be the case in neurodegenerative diseases [33]. Furthermore, DNA damage response signaling can be modulated by tumor suppressor p53 and may also contribute to apoptosis in aging and age-related neurodegenerative disorders [34]. These pathways showed similar expression patterns as those associated with the mitotic cell cycle, and therefore a lower expression of these DNA damage response pathways in caudal regions is related to cortical atrophy in PD.

Similar caudal-to-rostral expression patterns were found for pathways associated with mitochondrial translation. Increased risk for PD has been associated with mutations in *SNCA*, *PARK2* (parkin), *PINK1*, *DJ-1*, and *LRRK2* which have been linked to mitochondrial function and oxidative stress [35]. *PINK1* and parkin mediates clearance of damaged mitochondria by mitophagy and may therefore influence mitotic cell cycle progression [36]. *PINK1* also regulates both retrograde and anterograde axonal transport of mitochondria via axonal microtubules [37] The interaction between *PINK1* and parkin is likely involved in mitochondrial quality control mechanisms, where anterograde transport of damaged mitochondria is reduced and retrograde transport is enhanced for elimination by mitophagy in the neuronal cell body [38].

A cluster of pathways involved in ER-Golgi traffic were found enriched for PLS *model-2 component-1 and component-2*, and involved both ER-to-Golgi anterograde and Golgi-to-ER retrograde transport. *Component-1* showed a ventral-to dorsal gene expression pattern that was associated with higher correlations between CT and clinical scores, namely, the mental state of PD patients and the performance of motor functions. The pathways involved in ER-Golgi traffic were notably high expressed in the postcentral gyrus which contains the somatosensory cortex that is known for its role in processing sensory information and the regulation of emotion [39]. Our results suggest that genes in ER-Golgi traffic pathways are important for cognitive functions controlled by the postcentral gyrus. Genes involved in ER-Golgi vesicle trafficking have the ability to modify α-synuclein toxicity in yeast [40]. Moreover, fragmentation to the Golgi apparatus has been associated with the accumulation of aberrant proteins in neurodegenerative diseases, including α-synuclein [41]. A study in yeast models has showed that α-synuclein expression modulates ER stress signaling response and inhibits viral infections and viral replication [42]. We found several pathways

associated to HIV and influenza infections that were correlated to the relationship between CT and clinical scores. Another pathway that shared overlapping genes with those involved in ER-Golgi traffic was asparagine N-linked glycosylation, which is a biochemical linkage important for the structure and function of proteins. The *N*-glycosylated proteins are synthesized essentially in the ER and Golgi through sequential reactions and aberrant glycolysation of proteins may lead to inflammation and mitochondrial dysfunction in PD and consequently to a cellular overload of dysfunctional proteins [43].

We found that the expression of genes involved in sumoylation of chromatin organization proteins was correlated with CT changes, i.e. higher expression within caudal brain regions, such as the pericalcarine and the lateral occipital cortex, was associated with greater atrophy in PD. Therefore, higher activity of sumoylation events may play a role in the regional vulnerability to neurodegeneration observed in PD. On the other hand, lower expression of these pathways, such as in the pars opercularis, was associated with hypertrophy in rostral regions, suggesting that lower expression of sumoylation pathways has a protective effect. Additionally, the higher expression of sumoylation pathways was associated with higher correlations between CT and clinical scores as projected by PLS *component-2* in *model-2*. Sumoylation involves small ubiquitin-like modifier (SUMO) proteins that increase in response to cellular stress, such as DNA damage and oxidative stress, and can promote α-synuclein aggregation and Lewy body formation [44–46]. Several proteins associated with inherited forms of PD are targets modified by SUMO regulating mitochondrial processes, these include α-synuclein, DJ-1, and parkin [47]. Sumoylation has been associated with several diseases, including cancers, cardiac diseases, and neurodegenerative diseases [48]. In cancer, sumoylation mediates cell cycle progression and plays an essential role during mitosis [49]. SUMO seems to promote cell death mediated by the p53 tumor suppressor protein, which may be responsible for the cell death of dopaminergic neurons in PD [44]. Our findings are in support of these hypotheses, and further suggest that sumoylation is important in specific cortical regions that are atrophic in PD, such as the lateral occipital cortex.

Spatial gene expression data from PD brains are limited in the number of brain donors and brain regions, which is mainly due to the limited availability of well-defined post-mortem PD patients. Therefore, we used healthy gene expression from the AHBA to perform unbiased whole brain and whole transcriptome analysis. Gene expression for all the six healthy adult donors in AHBA was only available for the left hemisphere. Therefore, this study was restricted to the analysis of the left hemisphere when combining gene expression with MRI data. Furthermore, it is generally assumed that gene expression changes with age, however due to the limited number of brain donors in the AHBA, age-related differences in gene expression were not taken into account. In addition, MRI data from the patient and control groups were collected in different cohorts where different MRI scanners were used. However, both datasets were processed separately to obtain CT measurements per region. Thus, these morphological features could be directly compared between the two groups. Finally, to determine whether genes and pathways truly have predictive power of imaging features, both PLS models need to be validated with independent imaging cohorts of PD.

5.5 CONCLUSION

We set out to find biological explanations for the selective regional vulnerability in PD by correlating healthy gene expression in cortical regions with CT changes in PD observed as atrophy and hypertrophy in neuroimaging data. We found genes that point towards pathways involved in cellular maintenance mechanisms that are well known in PD and other neurodegenerative diseases, but were shown to be differently regulated across the brain. Sumoylation pathways showed opposite expression patterns across the brain compared to pathways associated with the regulation of mitotic cell cycle, p53-(in)dependent DNA damage response, mitochondrial translation, and ER-Golgi trafficking (Figure 5.8). Nevertheless, all the enriched pathways were highly interconnected as shown by the number of shared genes and suggest a balanced interplay between sumoylation events and the other molecular mechanisms that seem to be important in controlling CT in different cortical regions. Moreover, we propose that dysfunctions of these pathways may impair motor and cognitive functions in PD as a consequence of cortical atrophy.

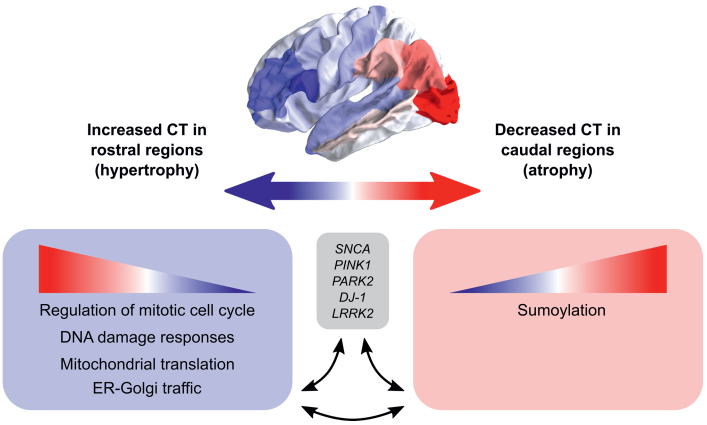


Figure 5.8 Schematic overview of the balance between biological pathways and their influence on CT across cortical brain regions. The big arrow indicates the caudal-to-rostral (red-to-blue) or rostral-to-caudal (blue-to-red) change in CT across cortical brain regions of PD patients with red indicating decreased CT (atrophy) in caudal regions and blue indicating increased CT (hypertrophy) in rostral regions. Genes within pathways associated with sumoylation showed that the expression of these genes within the pathways increases from rostral to caudal regions. Other biological pathways that were correlated with CT changes in PD included regulation of mitotic cell cycle, mitochondrial translation, DNA damage responses, and ER-Golgi traffic, and the involved genes showed decreasing expression patterns from rostral to caudal regions (or increasing from caudal to rostral regions). All enriched pathways shared many common genes and were generally associated with cellular maintenance mechanisms. Literature studies suggest that these biological pathways may be involved in the pathobiology of PD through their interaction with genetic risk variants.

5.6 ACKNOWLEDGEMENTS

We would like to thank Laura J. de Schipper for discussions on atrophy patterns. This research received funding from The Netherlands Technology Foundation (STW), as part of

the STW Project 12721 (Genes in Space). Dr. Oleh Dzyubachyk received funding from The Dutch Research Council (NWO) project 17126 (3DOmics). Prof. J.J. van Hilten received grants from Alkemade-Keuls Foundation; Stichting Parkinson Fonds (Optimist Study); The Netherlands Organisation for Health Research and Development (#40-46000-98-101); The Netherlands Organisation for Scientific Research (#628.004.001); Hersenstichting; AbbVie; Hoffmann-La-Roche; Lundbeck; and Centre of Human Drug Research outside the submitted work.

References

[1] E. C. Hirsch, S. Vyas, and S. Hunot, "Neuroinflammation in Parkinson's disease," *Park. Relat. Disord.* 18, S210–S212 (2012).

[2] D. Aarsland, B. Creese, M. Politis, K. R. Chaudhuri, D. H. Ffytche, D. Weintraub, and C. Ballard, "Cognitive decline in Parkinson disease," *Nat. Rev. Neurol.* 13, 217–231 (2017).

[3] H. Wilson, F. Niccolini, C. Pellicano, and M. Politis, "Cortical thinning across Parkinson's disease stages and clinical correlates," *J. Neurol. Sci.* 398, 31–38 (2019).

[4] X. Li, Y. Xing, A. Martin-Bastida, P. Piccini, and D. P. Auer, "Patterns of grey matter loss associated with motor subscores in early Parkinson's disease," *NeuroImage Clin.* 17, 498–504 (2018).

[5] M. Wang, S. Jiang, Y. Yuan, L. Zhang, J. Ding, J. Wang, J. Zhang, K. Zhang, and J. Wang, "Alterations of functional and structural connectivity of freezing of gait in Parkinson's disease," *J. Neurol.* 263, 1583–1592 (2016).

[6] D. Zheng, C. Chen, W. C. Song, Z. Q. Yi, P. W. Zhao, J. G. Zhong, Z. Y. Dai, H. C. Shi, and P. L. Pan, "Regional gray matter reductions associated with mild cognitive impairment in Parkinson's disease: A meta-analysis of voxel-based morphometry studies," *Behav. Brain Res.* 371, 111973 (2019).

[7] B. Chen, S. Wang, W. Sun, X. Shang, H. Liu, G. Liu, J. Gao, and G. Fan, "Functional and structural changes in gray matter of parkinson's disease patients with mild cognitive impairment," *Eur. J. Radiol.* 93, 16–23 (2017).

[8] K. Rosenberg-Katz, T. Herman, Y. Jacob, E. Kliper, N. Giladi, and J. M. Hausdorff, "Subcortical volumes differ in Parkinson's disease motor subtypes: New insights into the pathophysiology of disparate symptoms," *Front. Hum. Neurosci.* 10, 1–9 (2016).

[9] E. Oerton and A. Bender, "Concordance analysis of microarray studies identifies representative gene expression changes in Parkinson's disease: a comparison of 33 human and animal studies," *BMC Neurol.* 17, 1–14 (2017).

[10] M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C.-K. K. Lee, et al., "Canonical genetic signatures of the adult human brain," *Nat. Neurosci.* 18, 1832–1844 (2015).

[11] A. Arnatkevičiūtė, B. D. Fulcher, and A. Fornito, "A practical guide to linking brain-wide gene expression and neuroimaging data," *Neuroimage* 189, 353–367 (2019).

[12] T. Rittman, M. Rubinov, P. E. Vértes, A. X. Patel, C. E. Ginestet, B. C. P. Ghosh, R. A. Barker, M. G. Spillantini, E. T. Bullmore, et al., "Regional expression of the *MAPT* gene is associated with loss of hubs in brain networks and cognitive impairment in Parkinson disease and progressive supranuclear palsy," *Neurobiol. Aging* 48, 153–160 (2016).

[13] B. S. Freeze, D. Acosta, S. Pandya, Y. Zhao, and A. Raj, "Regional expression of genes mediating trans-synaptic alpha-synuclein transfer predicts regional atrophy in Parkinson disease," *NeuroImage Clin.* 18, 456–466 (2018).

[14] K. J. Whitaker, P. E. Vértes, R. Romero-Garcia, F. Váša, M. Moutoussis, G. Prabhu, N. Weiskopf, M. F. Callaghan, K. Wagstyl, et al., "Adolescence is associated with genomically

patterned consolidation of the hubs of the human brain connectome," *Proc. Natl. Acad. Sci.* 113, 9105–9110 (2016).

[15] P. E. Vértes, T. Rittman, K. J. Whitaker, R. Romero-Garcia, F. Váša, M. G. Kitzbichler, K. Wagstyl, P. Fonagy, R. J. Dolan, et al., "Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks," *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150362 (2016).

[16] P. McColgan, S. Gregory, K. K. Seunarine, A. Razi, M. Papoutsi, E. Johnson, A. Durr, R. A. C. C. Roos, B. R. Leavitt, et al., "Brain regions showing white matter loss in Huntington's disease are enriched for synaptic and metabolic Genes," *Biol. Psychiatry* 83, 1–10 (2017).

[17] L. J. de Schipper, J. van der Grond, J. Marinus, J. M. L. Henselmans, and J. J. van Hilten, "Loss of integrity and atrophy in cingulate structural covariance networks in Parkinson's disease," *NeuroImage Clin.* 15, 587–593 (2017).

[18] W. R. Gibb and A. J. Lees, "The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease," *J. Neurol. Neurosurg. Psychiatry* 51, 745–752 (1988).

[19] I. Altmann-schneider, A. J. M. De Craen, P. E. Slagboom, R. G. J. Westendorp, A. van Buchem, A. B. Maier, and J. van der Grond, "Brain tissue volumes in familial longevity: the Leiden Longevity Study," *Aging Cell* 11, 933–939 (2012).

[20] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *PNAS* 97, 11050–11055 (2000).

[21] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage* 31, 968–980 (2006).

[22] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, et al., "Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Mov. Disord.* 23, 2129–2170 (2008).

[23] J. F. van der Heeden, J. Marinus, P. Martinez-Martin, and J. J. van Hilten, "Evaluation of severity of predominantly non-dopaminergic symptoms in Parkinson's disease: The SENS-PD scale," *Park. Relat. Disord.* 25, 39–44 (2016).

[24] C. L. Tomlinson, R. Stowe, S. Patel, C. Rick, R. Gray, and C. E. Clarke, "Systematic review of levodopa dose equivalency reporting in Parkinson's disease," *Mov. Disord.* 25, 2649–2653 (2010).

[25] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "'Mini-mental state' A practical method for grading the cognitive stte of patients for the clinician," *J. Psychiatr. Res.* 12, 189–198 (1975).

[26] D. O. Claassen, K. E. McDonell, M. Donahue, S. Rawal, S. A. Wylie, J. S. Neimat, H. Kang, P. Hedera, D. Zald, et al., "Cortical asymmetry in Parkinson's disease: early susceptibility of the left hemisphere," *Brain Behav.* 6, 1–10 (2016).

[27] E. Mak, J. Zhou, L. C. S. Tan, W. L. Au, Y. Y. Sitoh, and N. Kandiah, "Cognitive deficits in mild Parkinson's disease are associated with distinct areas of grey matter atrophy," *J. Neurol. Neurosurg. Psychiatry* 85, 576–580 (2014).

[28] A. Brück, T. Kurki, V. Kaasinen, T. Vahlberg, and J. O. Rinne, "Hippocampal and prefrontal atrophy in patients with early non-demented Parkinson's disease is related to cognitive impairment," *J. Neurol. Neurosurg. Psychiatry* 75, 1467–1469 (2004).

[29] J. B. Pereira, P. Svenningsson, D. Weintraub, K. Brønnick, A. Lebedev, E. Westman, and D. Aarsland, "Initial cognitive decline is associated with cortical thinning in early Parkinson disease," *Neurology* 82, 2017–2025 (2014).

[30] M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. van de Lagemaat, K. A. Smith, A. Ebbert, et al., "An anatomically comprehensive atlas of the adult human brain transcriptome," *Nature* 489, 391–399 (2012).

[31] D. J. Bonda, G. Casadesus, X. Zhu, and M. A. Smith, "Review: Cell cycle aberrations and neurodegeneration," *Neuropathol Appl Neurobiol*. 36, 157–163 (2018).

[32] C. C. Walton, W. Zhang, I. Patiño-parrado, E. Barrio-alonso, J. Garrido, and J. M. Frade, "Primary neurons can enter M-phase," 1–15 (2019).

[33] P. Tokarz, K. Kaarniranta, and J. Blasiak, "Role of the cell cycle re-initiation in DNA damage response of post-mitotic cells and its implication in the pathogenesis of neurodegenerative diseases," *Aging Dis*. 19, 131–140 (2016).

[34] A. Mohammadzadeh, M. Mirza-aghazadeh-attari, S. Hallaj, and M. Majidinia, "Crosstalk between P53 and DNA damage response in ageing," *DNA Repair (Amst)*. 80, 8–15 (2019).

[35] M. H. Yan, X. Wang, and X. Zhu, "Mitochondrial defects and oxidative stress in Alzheimer disease and Parkinson disease," *Free Radic. Biol. Med*. 62, 90–101 (2013).

[36] S. A. Sarraf, D. P. Sideris, N. Giagtzoglou, L. Ni, M. W. Kankel, A. Sen, L. E. Bochicchio, C.-H. Huang, S. C. Nussenzweig, et al., "PINK1/Parkin influences cell cycle by sequestering TBK1 at damaged mitochondria, inhibiting mitosis," *Cell Rep*. 29, 225–235 (2019).

[37] S. Liu, T. Sawada, S. Lee, W. Yu, G. Silverio, P. Alapatt, I. Millan, A. Shen, W. Saxton, et al., "Parkinson's disease-associated kinase PINK1 regulates miro protein level and axonal transport of mitochondria," *PLOS Genet* 8, 15–17 (2012).

[38] E. Lionaki, M. Markaki, K. Palikaras, and N. Tavernarakis, "Mitochondria, autophagy and age-associated neurodegenerative diseases: New insights into a complex interplay," *Biochim. Biophys. Acta* 1847, 1412–1423 (2015).

[39] E. Kropf, S. K. Syan, L. Minuzzi, and B. N. Frey, "From anatomy to function: the role of the somatosensory cortex in emotional regulation," *Brazilian J. Psychiatry* 41, 261–269 (2019).

[40] A. A. Cooper, A. D. Gitler, A. Cashikar, C. M. Haynes, K. J. Hill, B. Bhullar, K. Liu, K. Xu, K. E. Strathearn, et al., "a-Synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's Models," *Science* 313, 324–328 (2006).

[41] J. Fan, Z. Hu, L. Zeng, W. Lu, X. Tang, J. Zhang, and T. Li, "Golgi apparatus and neurodegenerative diseases," *Int. J. Dev. Neurosci*. 26, 523–534 (2008).

[42] E. L. Beatman, A. Massey, K. D. Shives, K. S. Burrack, M. Chamanian, and T. E. Morrison, "Alpha-synuclein expression restricts RNA viral infections in the brain," *J. Virol*. 90, 2767–2782 (2016).

[43] P. A. Q. Videira and M. Castro-Caldas, "Linking glycation and glycosylation with inflammation and mitochondrial dysfunction in Parkinson's disease," *Front. Neurosci*. 12, 1–20 (2018).

[44] K. Eckermann, "SUMO and Parkinson's Disease," *Neuromol Med* 15, 737–759 (2013).

[45] S. Bologna and S. Ferrari, "It takes two to tango: Ubiquitin and SUMO in the DNA damage response," *Front. Genet*. 4, 1–18 (2013).

[46] R. Rott, R. Szargel, V. Shani, H. Hamza, M. Savyon, F. Abd, and R. Bandopadhyay, "SUMOylation and ubiquitination reciprocally regulate α-synuclein degradation and pathological aggregation," *PNAS* 114, 1–6 (2017).

[47] A. C. Guerra de Souza, R. D. Prediger, and H. Cimarosti, "SUMO-regulated mitochondrial function in Parkinson's disease," *J. Neurochem*. | 137, 673–686 (2016).

[48] Y. Yang, Y. He, X. Wang, G. He, P. Zhang, H. Zhu, N. Xu, and S. Liang, "Protein SUMOylation modification and its associations with disease" (2017).

[49] K. Eifler and A. C. O. Vertegaal, "SUMOylation-mediated regulation of cell cycle progression and cancer SUMO: a ubiquitin-like modifier that regulates nuclear processes," *Trends Biochem. Sci*. 40, 779–793 (2015).

Chapter 6
Transcriptomic signatures of brain regional vulnerability to Parkinson's disease

Arlin Keo
Ahmed Mahfouz
Angela M.T. Ingrassia
Jean-Pascal Meneboo
Celine Villenet
Eugénie Mutez
Thomas Comptdaer
Boudewijn P.F. Lelieveldt
Martin Figeac
Marie-Christine Chartier-Harlin
Wilma D.J. van de Berg
Jacobus J. van Hilten
Marcel J.T. Reinders

This chapter was published in: *Commun. Biol.* 3, 1–12 (2020), doi: 10.1038/s42003-020-0804-9
Supplementary material is available online at: https://www.nature.com/articles/s42003-020-0804-9#Sec22

Abstract

The molecular mechanisms underlying caudal-to-rostral progression of Lewy body pathology in Parkinson's disease remain poorly understood. Here, we identified transcriptomic signatures across brain regions involved in Braak Lewy body stages in non-neurological adults from the Allen Human Brain Atlas. Among the genes that are indicative of regional vulnerability, we found known genetic risk factors for Parkinson's disease: *SCARB2*, *ELOVL7*, *SH3GL2*, *SNCA*, *BAP1*, and *ZNF184*. Results were confirmed in two datasets of non-neurological subjects, while in two datasets of Parkinson's disease patients we found altered expression patterns. Co-expression analysis across vulnerable regions identified a module enriched for genes associated with dopamine synthesis and microglia, and another module related to the immune system, blood-oxygen transport, and endothelial cells. Both were highly expressed in regions involved in the preclinical stages of the disease. Finally, alterations in genes underlying these region-specific functions may contribute to the selective regional vulnerability in Parkinson's disease brains.

6.1 Introduction

Parkinson's disease (PD) is characterized by a temporal caudal-rostral progression of Lewy body (LB) pathology across a selected set of nuclei in the brain [1]. The distribution pattern of LB pathology is divided into six Braak stages based on accumulation of the protein α-synuclein – the main component of LBs and Lewy neurites – in the brainstem, limbic, and neocortical regions [1]. These six Braak stages indicate affected regions throughout the progression of PD with the region involved in Braak stage 1 being first affected and the region involved in Braak stage 6 being last affected. Thus, the Braak staging scheme points out vulnerable brain regions involved in disease progression and the sequential order of their vulnerability. Different hypotheses have been brought forward to explain the evolving LB pathology across the brain, including: retrograde transport of pathological α-synuclein via neuroanatomical networks, α-synuclein's prion-like behavior, and cell- or region-autonomous factors [2,3]. Yet, the mechanisms underlying the selective vulnerability of brain regions to LB pathology remain poorly understood, limiting the ability to diagnose and treat PD.

Multiplications of the *SNCA* gene encoding α-synuclein are relatively common in autosomal dominant PD and *SNCA* dosage has been linked to the severity of PD [4,5]. For other PD-associated variants, e.g., *GBA* and *LRRK2*, their role in progressive α-synuclein accumulation is less clear, although they have been associated with mitochondrial (dys)function and/or protein degradation pathways [6–8]. On the other hand, transcriptomic changes between PD patients and non-neurological controls of selected brain regions, e.g., the substantia nigra, have identified several molecular mechanisms underlying PD pathology, including synaptic vesicle endocytosis [9–11]. However, post-mortem human brain tissue of well-characterized PD patients and controls is scarce, usually focuses on a select number of brain regions, and have a limited coverage of patients with different Braak LB stages, resulting in low concordance of findings across different studies [12].

Spatial gene expression patterns in the human brain have been studied to unravel the pathogenic mechanisms underlying amyloid-β and tau pathology progression in Alzheimer's disease, revealing proteins that co-aggregate with amyloid-β and tau, and protein homeostasis components [13,14]. This highlights the value of analyzing spatial transcriptomics to study the pathobiology in neurodegenerative diseases. Interestingly, by integrating Allen Human Brain Atlas (AHBA) gene expression data [15] with magnetic resonance imaging of PD patients, the regional expression pattern of *MAPT* and *SNCA* was associated with loss of functional connectivity in PD [16], and regional expression of synaptic transfer genes was related to regional gray matter atrophy in PD [17]. This combined gene-MRI analysis illustrates the importance of local gene expression changes on functional brain networks. More detailed knowledge about the spatial organization of transcriptomic changes in physiological and pathological conditions may aid in understanding these changes on a functional level during disease progression in PD.

In the present study, we analyzed the transcriptome of brain regions involved in Braak LB stages [18] of non-neurological adult donors from the AHBA to reveal molecular factors underlying selective vulnerability to LB pathology during PD progression. We validated our findings in two independent non-neurological datasets (the Genotype-Tissue Expression project (GTEx) [19] and UK Brain Expression Consortium (UKBEC) [20]). Further, we showed that Braak stage-related genes (BRGs) are indeed progressively disrupted in patients with incidental Lewy body disease (iLBD; assumed to represent the pre-clinical stage of PD [11,21]) and PD. The observed transcriptomic signatures of vulnerable brain regions pointed towards the dopamine biosynthetic process and oxygen transport that were highly expressed in brain regions related to the preclinical stages of PD. Together, our analyses provide important insights that enable a better understanding of the biological mechanisms underlying disease progression.

6.2 Results

6.2.1 Study overview

The PD Braak staging scheme defines a temporal order of brain regions affected during the progression of the disease [18]. Based on the sequence of events as postulated by Braak et al. [1], we hypothesized that genes whose expression patterns increase or decrease across regions involved in the Braak staging scheme might contribute to higher vulnerability to LBs in PD brains (Figure 6.1). Based on this assumption, we aimed to find (1) which genes are involved, (2) which modules of interacting genes are involved, and (3) which biological processes contribute to this vulnerability. We analyzed the regions of interest using a microarray dataset of anatomical brain regions from six individuals without any known neuropsychiatric or neurological background from the AHBA [15]. Therefore, we first assigned 2,334 out of 3,702 brain samples to Braak stage-related regions R1-R6 [18]: myelencephalon (medulla, R1), pontine tegmentum including locus coeruleus (R2), substantia nigra, basal nucleus of Meynert, CA2 of hippocampus (R3), amygdala, occipito-temporal gyrus (R4), cingulate gyrus, temporal lobe (R5), frontal lobe including the olfactory area, and parietal lobe (R6) (Supplementary Table 1, and Supplementary Figure 1).

6.2.2 PD Braak stage-related genes

To identify genes with expression patterns that are associated with selective vulnerability to PD, i.e. BRGs, we correlated gene expression with the label of these vulnerable regions as defined by Braak stage. To ensure that genes have large expression differences across regions, we assessed differential expression between all pairs of Braak stage-related regions R1-R6, and found most significant changes between regions related to the most distant stages: R1 versus R5 and R1 versus R6 ($|\text{fold-change (FC)}| > 1$, Benjamini-Hochberg (BH)-corrected $P < 0.05$, *t*-test; Supplementary Figure 2). Thus, in the selection of BRGs, we also focused on the FC between the disease-related end points R1 and R6.

BRGs were selected based on (i) the highest absolute Braak label correlation ($|r|$), (ii) highest absolute FC between R1 and R6 ($|FC_{R1\text{-}R6}|$), and (iii) smallest BH-corrected *P*-values of the FC (P_{FC}). The top 10% (2,001) ranked genes for each one of the three criteria resulted in genes

with $|r| > 0.66$, $|FC_{R1\text{-}R6}| > 1.33$, and $P_{FC} < 0.00304$ (Figure 6.2A). The overlap of the three sets of top 10% ranked genes resulted in 960 BRGs, with 348 negatively and 612 positively correlated genes showing a decreasing ($r < 0$) or increasing ($r > 0$) expression pattern across regions R1-R6, respectively (Figure 6.2B and C, Supplementary Figure 3, and Supplementary Data 1). Negatively correlated BRGs were significantly enriched for gene ontology (GO) terms like anatomical structure morphogenesis and blood vessel morphogenesis (Supplementary Data 2), while positively correlated BRGs were significantly enriched for functions like anterograde trans-synaptic signaling and nervous system development (BH-corrected $P < 0.05$, DAVID; Supplementary Data 3).

Since the expression patterns of the 960 BRGs were observed in only six non-neurological brains from the AHBA, we used two independent datasets from non-neurological controls for validation. For each dataset we assessed whether BRGs were also differentially expressed

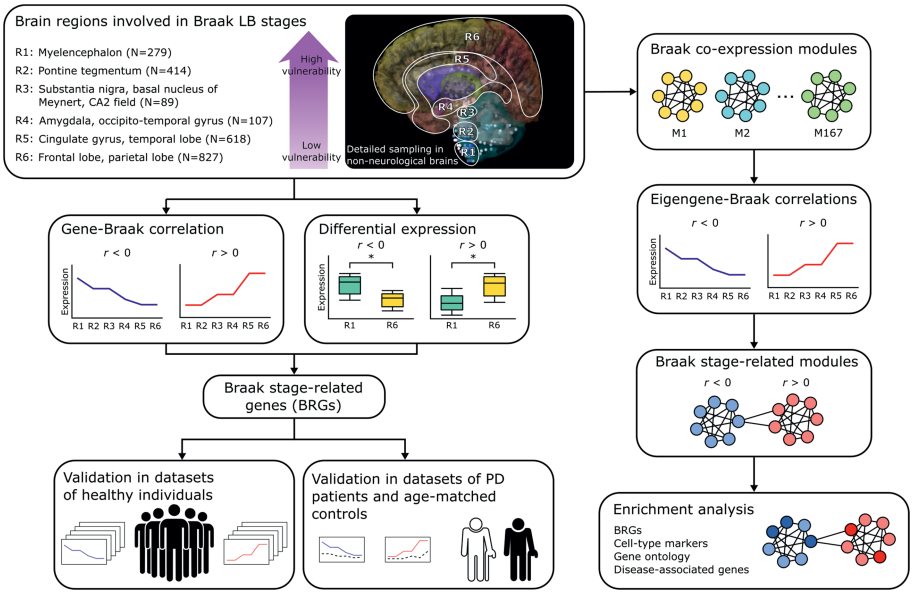


Figure 6.1 Study overview. Differential vulnerability to Parkinson's disease (PD) was examined across brain regions R1-R6 (image credit: Allen Institute). N is the number of samples across all six non-neurological donors from the Allen Human Brain Atlas (AHBA), which are involved in the six PD Braak stages as they sequentially accumulate Lewy bodies during disease progression (Supplementary Table 1 and Supplementary Figure 1). Through correlation and differential expression analysis, we identified Braak stage-related genes (BRGs) with expression patterns that are either positively ($r > 0$) or negatively ($r < 0$) correlated with Braak stages in the non-neurological brain. These were validated in cohorts of non-neurological individuals and subsequently in PD patients and age-matched controls. To obtain a more global view of BRG expression signatures, we focused on co-expression modules of all genes and correlated the module eigengene expression with Braak stages. The resulting modules of genes were subsequently analyzed to detect common biologically meaningful pathways.

between regions related to the most distant Braak stages, and whether the decreasing or increasing expression patterns could be replicated. First, using microarray data from 134 individuals in the UKBEC [20], we selected brain samples corresponding to the myelencephalon (R1), substantia nigra (R3), temporal cortex (R5), and frontal cortex (R6). For the 885 BRGs present in UKBEC, 139 out of 314 (44.3%) negatively correlated BRGs and 400 out of 571 (70.1%) positively correlated BRGs were differentially expressed between R1 and R6 ($|FC_{R1\text{-}R6}| > 1$, BH-corrected $P < 0.05$, *t*-test). The mean expression of negatively and positively correlated BRGs showed indeed decreasing and increasing expression patterns, respectively, across regions R1, R3, R5, and R6 (Figure 6.2D). Second, we used RNA-sequencing (RNA-seq) data from 88-129 individuals in the GTEx consortium [19] and selected samples of the substantia nigra (R3), amygdala (R4), anterior cingulate cortex (R5), and frontal cortex (R6). For the 883 BRGs present in the GTEx consortium, 204 out of 318 (64.2%) negatively correlated BRGs and 475 out of 565 (84.1%) positively correlated BRGs were differentially expressed between the two most distant regions R3 and R6 in this dataset ($FC_{R3\text{-}R6}| > 1$, BH-corrected $P < 0.05$, DESeq2). The mean expression of BRGs again showed decreasing and increasing patterns, here across regions R3-R6 (Figure 6.2E). Together, this indicates that the expression patterns of BRGs in the brain are consistent across non-neurological individuals.

We next hypothesized that if the identified BRGs are associated with vulnerability to PD, they are also indicative of vulnerability differences between PD patients and controls. To test this hypothesis, we used two datasets with transcriptomic measurements from brain regions covering most Braak stage-related regions sampled from PD and iLBD patients, and non-demented age-matched controls (microarray [11] (Supplementary Table 2 and Supplementary Data 4) and RNA-seq datasets (Supplementary Table 3 and Supplementary Data 5); see the "Methods" section). First, we found more differentially expressed genes between brain regions within the same group of individuals (PD, iLBD, and control) than between conditions within the same region (Supplementary Figure 4). This observation further highlights the importance of assessing expression patterns across regions rather than disease conditions [28]. Next, we validated the expression patterns of BRGs, which we identified in brains of non-neurological adults from the AHBA, in both the PD microarray and RNA-seq datasets. First, we observed (again) similar patterns in non-demented age-matched controls (Figure 6.2F and G). Interestingly, the increasing and decreasing expression patterns of BRGs were diminished in iLBD patients and totally disrupted in PD patients across regions involved in preclinical stages R1-R3 (Figure 6.2F). Across regions R3 and R4/R5 however, these expression patterns were preserved in PD patients (Figure 6.2G). In addition to the patterns across brain regions, we found that BRGs also captured patterns across conditions PD, iLBD, and control (Supplementary Figure 5). For both PD datasets, this is most apparent within the substantia nigra (R3), where negatively correlated BRGs that had higher expression in more vulnerable brain regions also had higher expression in PD patients compared to controls. Vice versa, positively correlated BRGs that had higher expression in less vulnerable brain regions also had higher expression in controls compared to PD patients.

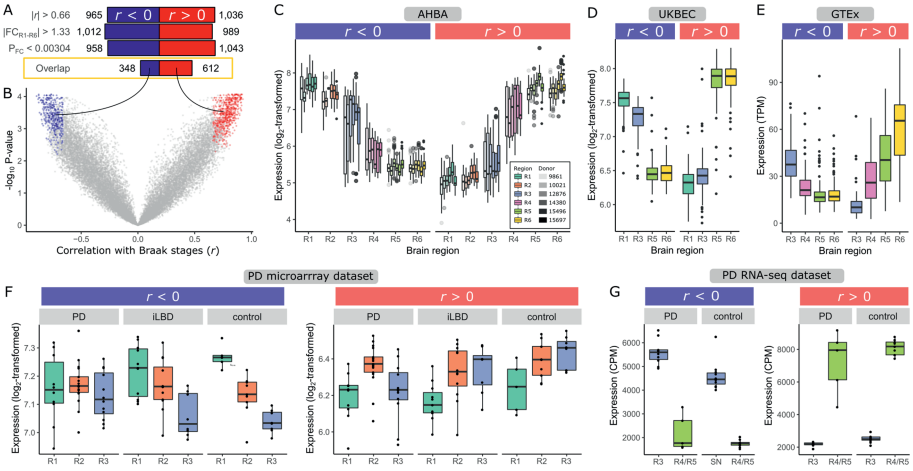


Figure 6.2 Expression patterns of Braak stage-related genes (BRGs) across brain regions of non-neurological, incidental Lewy body disease (iLBD) and Parkinson's disease (PD) brains. (A) Selection of BRGs that were either negatively (blue; $r < 0$) or positively (red; $r > 0$) correlated with Braak stages. Genes were selected based on (1) highest absolute correlation ($|r|$) of gene expression and Braak stage labels, (2) highest absolute fold-change (FC) between R1 and R6, and (3) lowest *P*-value of FC in the differential expression analysis (BH-corrected P_{FC}), for which the top 10% (2,001) genes resulted in the shown thresholds. The overlap between the three sets of top 10% genes resulted in 960 BRGs. (B) Correlation r of BRGs (red and blue points) with Braak stages (x-axis) and $-\log_{10}$ BH-corrected *P*-value (y-axis). (C) Mean expression of BRGs for each region (colors) and donor (opacity) in the AHBA (number of samples in Supplementary Table 1). (D) Validation across 134 non-neurological individuals in UK Brain Expression Consortium (UKBEC; R1: medulla, R3: substantia nigra, R5: temporal cortex, R6: frontal cortex), and (E) 88-129 non-neurological individuals in Genotype-Tissue Expression Consortium (GTEx; R3: substantia nigra, R4: amygdala, R5: anterior cingulate cortex, R6: frontal cortex). Each data point is a sample with the mean expression of negatively or positively correlated BRGs. (F) Validation in PD microarray dataset (R1: medulla oblongata, R2: locus coeruleus, R3: substantia nigra; number of samples in Supplementary Table 2) and (G) PD RNA-seq dataset (R3: substantia nigra, R4/R5: medial temporal gyrus; number of samples in Supplementary Table 3). Boxplots (f, g) are shown per patient group (PD, iLBD, and control) and per brain region (Supplementary Figure 5). The boxplots indicate the median and interquartile range (25th and 75th percentiles) with whiskers indicating 1.5 times the interquartile range; outliers beyond the whiskers are plotted individually.

To summarize, these validations showed that the expression patterns of the detected BRGs are replicated in independent datasets (UKBEC and GTEx) and indeed showed progressively disturbed patterns in iLBD and PD patients (PD microarray and PD RNA-seq datasets). This was shown for the mean of both BRGs groups (increasing and decreasing), but is also shown for individual BRGs (Supplementary Figure 6). These findings support the relationship of BRGs with PD vulnerability encountered in brain regions of non-neurological individuals and show how their expression may influence the vulnerability at a region-specific level as well as between patients and controls.

6.2.3 BRAAK STAGE-RELATED CO-EXPRESSION MODULES

In addition to the expression of individual genes, we analyzed non-neurological brains from the AHBA to examine the expression of gene sets that may jointly affect the vulnerability of brain regions to PD. To study genetic coherence in vulnerable brain regions, we clustered all 20,017 genes into 167 modules based on their pairwise co-expression across regions R1-R6. The module eigengene, which summarizes the overall expression of genes within a module, was correlated with the labels of regions R1-R6 as defined by Braak stages (Figure 6.3A and Supplementary Data 6). Whether or not the modules showed expression patterns that correlated with Braak stages, their expression in the arcuate nucleus of medulla, locus coeruleus and CA2-field was consistently low (Figure 6.3B and Supplementary Figure 7). For the CA2-field this might be explained by the presence of Lewy neurites rather than LBs [18]. Correlations with Braak stages were mostly driven by the expression change between regions involved in preclinical stages (R1-R3) and clinical stages (R4-R6). In addition, regions R1-R3 showed more extreme expression values (high and low) than in regions R4-R6.

We selected 23 co-expression modules for which the eigengene was significantly correlated with Braak stages (BH-corrected $P < 0.0001$, *t*-test). These modules have distinct expression patterns (Supplementary Figure 8) and those that were negatively correlated with Braak stages showed more distinct expression patterns than the positively correlated modules (Supplementary Figure 9). Module M39 showed the lowest correlation with Braak stages ($r = -0.87$, BH-corrected $P = 3.65e\text{-}7$, *t*-test), while M50 showed the highest correlation ($r = 0.92$, BH-corrected $P = 4.42e\text{-}7$, *t*-test). Most modules were significantly enriched for BRGs that were similarly correlated with Braak stages (BH-corrected $P < 0.05$, hypergeometric test; Figure 6.3C). For functional characterization, modules were further assessed for enrichment of cell-type markers [26], and gene sets associated with functional GO terms or diseases. A full version of the table in Figure 6.3C showing all significant associations is given in Supplementary Figure 10.

We found that modules that were negatively correlated with Braak stages were enriched for markers for all different cell-types, and linked to various functions and diseases. M39 was enriched for markers of astrocytes and endothelial cells, and the function membrane raft which plays a role in neurotransmitter signaling. M127 was enriched for microglia and neurons, and associated with functional GO terms such as locomotory behavior and dopamine biosynthetic process, as well as diseases including dopa–responsive dystonia, dystonia–limb, and tremor, highlighting their role in motor circuitry. M47 was enriched for endothelial cell markers and genes involved in immune response, blood coagulation, interferon-gamma-mediated signaling pathway, and oxygen transport. This module was also enriched for genes involved in auto-inflammatory or auto-immunity disorders, e.g., hypersensitivity, infection, and inflammation. These modules and their associated pathways were associated with the preclinical stages of PD, because of their higher expression in regions R1-R3.

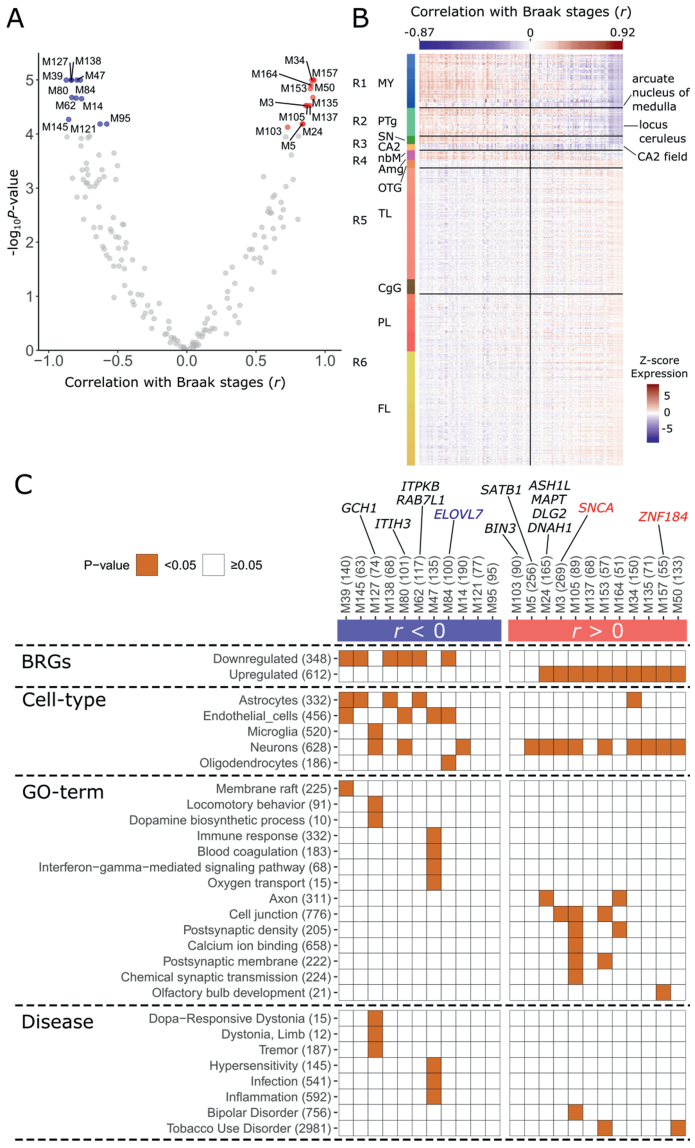


Figure 6.3 Braak co-expression modules. Genes were analyzed for co-expression across regions R1-R6 in the Allen Human Brain Atlas. (A) Module eigengene correlation with Braak stages. Each point reflects a module showing its correlation r with Braak stages (x-axis) and -$\log_{10}$-transformed *P*-values (BH-corrected; y-axis); 23 significant modules (BH-corrected $P < 0.0001$, *t*-test) were selected for further analysis (blue and red points). (B) Eigengene expression of all 167 modules across brain regions (rows) of donor 9861 sorted by their correlation with Braak stages (column colors). The vertical line separates negatively and positively correlated modules, and correlations are shown for two modules

with the lowest and highest correlation. Brain regions involved in Braak includes the following anatomical structures: myelencephalon (MY), pontine tegmentum (PTg), substantia nigra (SN), CA2-field (CA2), basal nucleus of Meynert (nbM), amygdala (Amg), occipito-temporal gyrus (OTG), temporal lobe (TL), cingulate gyrus (CgG), parietal lobe (PL), and frontal lobe (FL). Modules were low expressed in the arcuate nucleus of medulla, locus coeruleus and CA2-field, independently of their correlation with Braak stages (Supplementary Figure 7). (C) Significant modules were sorted based on their correlation with Braak stages (columns) and assessed for significant overlap with Braak stage-related genes (BRGs), cell-type markers, and gene sets associated with functional GO terms or diseases (brown squares, BH-corrected P < 0.05, hypergeometric test). The number of genes within each module and tested gene set is given between brackets. Additionally, these modules revealed the presence of genes associated with Parkinson's disease variants (annotated at the top) that have (blue and red) or have not (black) been identified as BRGs. A full version of this table showing all significant associations is given in Supplementary Figure 10.

Modules that were positively correlated with Braak stages were specifically enriched for neuronal markers and related functions (e.g., axon, cell junction, and chemical synaptic transmission) reflecting higher expression of these modules in the synapse-dense cerebral cortex. M157 was enriched for the function olfactory bulb development, M105 for functions such as cell junction, postsynaptic density, calcium ion binding, and genes linked to bipolar disorder, M153 for functions cell junction and postsynaptic membrane, and both M153 and M50 were linked to tobacco use disorder. Overall, gene co-expression across Braak stage-related regions R1-R6 revealed interesting modules that highlight pathways and potential gene interactions involved in the preclinical or clinical stage of PD.

6.2.4 BRGs ARE NOT FULLY CONFOUNDED BY CELLULAR COMPOSITION

We validated whether the identification of BRGs was confounded by variations in cellular compositions across the six Braak stage-related regions R1-R6. We applied population-specific expression analysis (PSEA) [25] to the AHBA to validate the cell-type specificity of each of the 960 BRGs. We found all 960 BRGs to be differentially expressed (BH-corrected $P < 0.05$, *t*-test) between regions R1 and R6 after correcting for five major cell-types (neurons, astrocytes, oligodendrocytes, microglia, and endothelial cells). For example, the neuronal marker *ADCY1* which was identified as a BRG remains differentially expressed between regions R1 and R6 when corrected for neurons or other cell-types (Figure 6.4). Similarly as for BRGs, PSEA analysis on all 23 Braak stage-related co-expression modules showed significant differential expression between regions R1 and R6 which cannot be fully explained by differences in cellular composition.

In the PD datasets, not all BRGs were found significant after correction for cellular composition, however smaller changes can be expected when comparing regions that are less distant (R1-R3 and R3-R4/R5). Similar to the differential expression analysis without correction for cellular composition (Supplementary Figure 4), PSEA revealed more changes between brain regions than between patients and controls (Supplementary Figure 11).

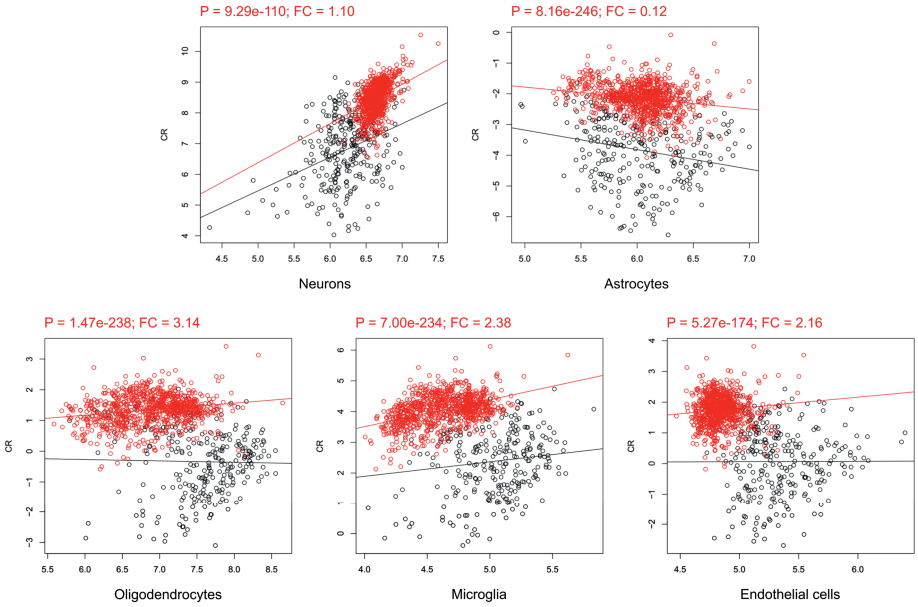


Figure 6.4 Differential expression of neuronal marker *ADCY1* in the AHBA corrected for cell-type abundance. *ADCY1* is a neuronal marker identified as one of the 960 Braak stage-related genes (BRGs). We found it was still significantly differentially expressed between samples from region R1 (black) and R6 (red) when correcting for one of the five main cell-types with PSEA (BH-corrected $P < 0.05$, *t*-test). Significant BH-corrected *P*-values are highlighted in red text together with cell-type specific fold-changes (FC; slope change of red line).

6.2.5 Expression of PD-implicated genes is related to Braak staging

We found that the expression patterns of several PD-implicated genes, identified in the two most recent genome-wide association studies [7,8], were correlated with the Braak LB staging scheme. These included BRGs (*SCARB2*, *ELOVL7*, *SH3GL2*, *SNCA*, *BAP1*, and *ZNF184*; Table 6.1 and Supplementary Figure 12) or genes present in Braak stage-related co-expression modules (*GCH1*, *ITIH3*, *ITPKB*, *RAB7L1*, *BIN3*, *SATB1*, *ASHL1*, *MAPT*, *DLG2*, and *DNAH1*; Figure 6.3C).

We further explored the relationship between *SNCA* expression and PD vulnerability in more detail. *SNCA* was positively correlated with Braak stages in non-neurological brains from the AHBA, with a lower expression in regions R1-R2 and higher expression in R3-R6 (Figure 6.5A-C), which was replicated in larger cohorts of non-neurological individuals (Figure 6.5D and E). This observation suggests that lower *SNCA* expression indicates high vulnerability of brain regions to develop LB pathology. We further validated this concept in two cohorts of PD patients in which *SNCA* expression similarly increased across the medulla

oblongata (R1), locus coeruleus (R2), and substantia nigra (R3) of PD and iLBD patients, and age-matched controls. *SNCA* was significantly lower expressed in region R1 compared to R2 and R3 in PD and iLBD patients, but not in controls (BH-corrected $P < 0.05$, *t*-test; Figure 6.5F). In the PD RNA-seq dataset, *SNCA* was significantly lower expressed in the substantia nigra (R3) compared to the medial temporal gyrus (R4/R5) in PD patients, but again not in controls (BH-corrected $P < 0.05$, DESeq2; Figure 6.5G). Altogether, *SNCA* expression patterns could be replicated in brain regions of age-matched controls, however changes were larger between brain regions in PD and iLBD cases. We further assessed *SNCA* expression using PSEA in the AHBA (Figure 6.5H) and found that changes were independent of neuronal or other cell-type densities when comparing different brain regions. In the PD datasets, PSEA results were scattered and did not align between the microarray and RNA-seq dataset, which might be caused by the small sample sizes and the comparison of different brain regions (Supplementary Figure 13).

Co-expression analysis in non-neurological brains from the AHBA revealed several dopaminergic genes present in module M127. Their expression patterns were further investigated together with *SNCA* which is also known to regulate dopamine homeostasis [29]. *GCH1*, *TH*, and *SLC6A3* (also known as *DAT*) were related to the functional term dopamine biosynthetic process, and *SLC18A2* (also known as *VMAT2*) is known to store dopamine into synaptic vesicles [30]. Unlike *SNCA*, the expression of *GCH1*, *TH*, *SLC6A3*, and *SLC18A2* was higher expressed in regions involved at preclinical stages than those involved at clinical stages (Figure 6.6 and Supplementary Figure 14). Furthermore, all these dopaminergic genes and *SNCA* showed a clear peak of expression in region R3 which includes the substantia nigra, basal nucleus of Meynert, and CA2-field.

Table 6.1 Braak stage-related genes that previously have been associated with Parkinson's disease. Several Parkinson's disease variant-associated genes showed expression profiles that are correlated with Braak stages. The correlation r with Braak stages, fold-change (FC), and *P*-value of FC (*t*-test, BH-corrected) are within the selection thresholds for BRGs.

Gene symbol	Entrez ID	Correlation with Braak (*r*)	*P*-value of correlation with Braak (BH-corrected)	FC between R1 and R6	*P*-value of FC between R1 and R6 (BH-corrected)	Module mem-ber	Reference
SCARB2	950	-0.78	4.4e-04	-1.44	1.7e-03	M130	Nalls et al. 2014
ELOVL7	79993	-0.67	7.2e-04	-1.35	1.4e-03	M84	Chang et al. 2017
SH3GL2	6456	0.70	4.5e-04	1.4	2.3e-03	-	Chang et al. 2017
SNCA	6622	0.70	3.3e-04	1.75	4.3e-04	M3	Nalls et al. 2014, Chang et al. 2017
BAP1	8314	0.77	3.2e-03	1.99	1.6e-03	M85	Chang et al. 2017
ZNF184	7738	0.81	4.6e-04	2.34	2.9e-03	M157	Chang et al. 2017

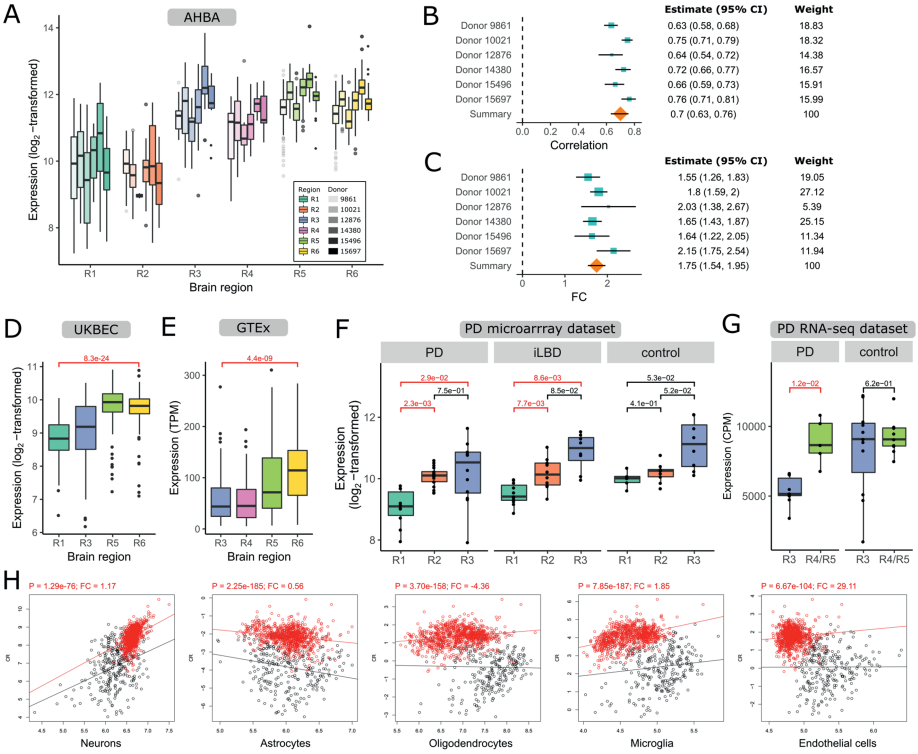


Figure 6.5 *SNCA* expression in Braak stage-related regions R1-R6 of non-neurological individuals and Parkinson's disease (PD) patients. (A) Boxplots of *SNCA* expression in regions R1-R6 (colored) for each donor (opacity) in the AHBA (number of samples in Supplementary Table 1). Meta-analysis of (B) *SNCA* correlation with Braak stages and (C) *SNCA* expression fold-change (FC) between region R1 and R6 across the six donors in the AHBA. To calculate the summary effect size (orange diamonds) from the individual effect sizes (turquoise squares), the 95% confidence intervals (CI) and weights are taken into account. The positive correlation with Braak stages was validated in datasets from two healthy cohorts, (D) UK Brain Expression Consortium (UKBEC; 134 donors) and (E) Genotype-Tissue Expression Consortium (GTEx; 88-129 donors), and (f, g) two PD cohorts with PD patients, incidental Lewy body disease (iLBD) patients, and non-demented age-matched controls (number of samples in Supplementary Table 2 and 3). In the PD datasets, *SNCA* expression was tested for differential expression between regions and conditions (red, BH-corrected $P < 0.05$, *t*-test and DESeq2, respectively). The boxplots indicate the median and interquartile range (25th and 75th percentiles) with whiskers indicating 1.5 times the interquartile range; outliers beyond the whiskers are plotted individually. (H) *SNCA* was still significantly differentially expressed between region R1 (black) and R6 (red) when correcting for five main cell-types with PSEA in the AHBA (BH-corrected $P < 0.05$, *t*-test). Significant BH-corrected *P*-values are highlighted in red together with cell-type specific fold-changes (slope change of red line). PSEA results for PD data are shown in Supplementary Figure 13.

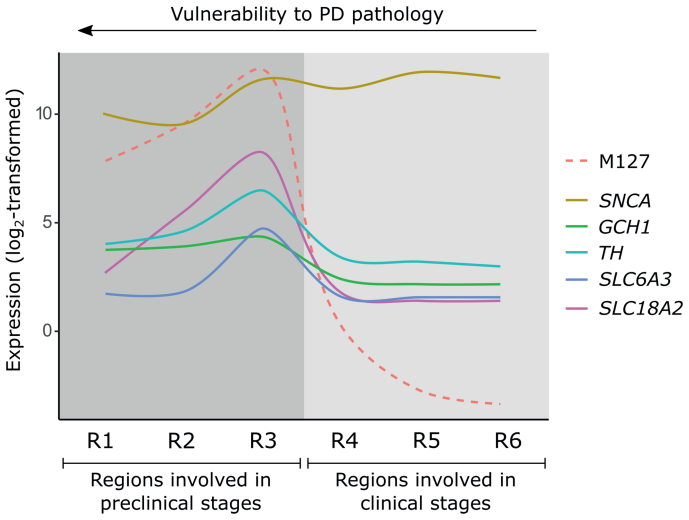


Figure 6.6 Schematic overview of molecular activity of dopaminergic genes in module M127 and *SNCA* across brain regions of the Braak staging scheme. Lines across regions R1-R6 were based on transcriptomic data from the Allen Human Brain Atlas (Figure 6.1 and Supplementary Figure 14). Expression of module M127 is in the eigengene space. Genes showed peak activity in region R3 that includes the substantia nigra, basal nucleus of Meynert, and CA2-field. While *SNCA* was generally high expressed in all regions, dopaminergic genes in M127 were low or not expressed in other regions than R3. *SNCA*: responsible for dopamine release, *GCH1*: together with *TH* required for production of dopamine, *TH*: catalyzes tyrosine to the dopamine precursor L-3,4-dihydroxyphenylalanine (L-DOPA), *SLC6A3* (also known as *DAT*): transports dopamine from the synaptic cleft back to the cytosol, *SLC18A2* (also known as *VMAT2*): stores dopamine into synaptic vesicles.

6.3 Discussion

In PD, the progressive accumulation of LB pathology across the brain follows a characteristic pattern, which starts in the brainstem and subsequently evolves to more rostral sites of the brain (Braak ascending scheme) [1]. Using transcriptomic data of non-neurological brains, we identified genes (e.g., *SNCA*, *SCARB2*, and *ZNF184*) and modules of co-expressed genes for which the expression decreased or increased across brain regions defined by the Braak ascending scheme. Interestingly, these patterns were disrupted in brains of PD patients across regions that are preclinically involved in the pathophysiology of PD. One gene co-expression module that showed higher expression in preclinically involved regions was related to dopamine synthesis, locomotory behavior, and microglial and neuronal activity. Another module was related to blood-oxygen transport, the immune system, and may involve endothelial cells. Our results highlight the complex genetic architecture of PD in which the combined effects of genetic variants and co-expressed genes may underlie the selective regional vulnerability of the brain.

Multiple studies suggests that a cytotoxic role and prion-like transfer of α-synuclein may contribute to its progressive spread across the brain in PD, assuming a gain-of-function [3,31,32]. In line with this assumption are reports of familial PD caused by *SNCA* multiplications, suggesting a *SNCA* dosage effect in causing PD [4,5]. Interestingly, in contrast to the temporal and spatial pattern of the α-synuclein distribution associated with the ascending Braak scheme in PD, the *SNCA* expression signature across brain regions R1-R6 in non-neurological brains followed a reverse pattern with lowest expression in preclinically involved regions (brainstem) and highest expression in clinically involved regions (limbic system and cortex). Expression changes between regions were larger in PD and iLBD brains, because of lower expression in preclinically involved regions compared to age-matched controls. The abundance of physiological *SNCA* in non-neurological brains suggests a protective role, while at the same time it may impact vulnerability to LB pathology in PD brains as demonstrated in earlier studies detecting both proteins and mRNA levels (literature overview in Supplementary Table 4). Cell lines or animal models without *SNCA* showed a synaptic deficit, increased susceptibility to viruses, sensitivity to reward, and resulted in nigrostriatal neurodegeneration underscoring the importance of the presence of α-synuclein for neuronal function. Mutant α-synuclein accelerated cell death induced by various stimuli (staurosporine, serum deprivation, trypsin, or oxidative stress by H_2O_2), while wild-type α-synuclein exerted anti-apoptotic effects. In contrast to the suggested neuroprotective role of α-synuclein, other studies suggest a deleterious effect when overexpressed and that removing *SNCA* mediates resistance to LB pathology. Collectively, our findings suggest that low *SNCA* expression in preclinically involved regions may increase the vulnerability of brain regions to LB pathology.

Next to *SNCA*, the expression of several other genes known as genetic risk factors for PD [7,8] were related to the Braak staging scheme. Two genes *ZNF184* (zinc finger protein 184) and *ELOVL7* (fatty acid elongase 7) have recently been associated with early onset PD in a Chinese population [33]. *SCARB2* (scavenger receptor class B member 2) encodes for the lysosomal integral membrane protein-2 (*LIMP2*), the specific receptor for glucocerebrosidase (GCase), and is important for transport of GCase from the endoplasmic reticulum via Golgi to lysosomes [34]. *SCARB2*-deficiency in mice brains led to α-synuclein accumulation mediating neurotoxicity in dopaminergic neurons [34]. Overexpression in murine and human cell lines improved lysosomal activity of this enzyme and enhanced α-synuclein clearance [34]. *SH3GL2* (SH3 Domain Containing GRB2 Like 2, Endophilin A1) is thought to act downstream of *LRRK2* to induce synaptic autophagosome formation and may be deregulated in PD [35]. *BAP1* (ubiquitin carboxyl-terminal hydrolase) is a deubiquitinase that acts as a tumor suppressor. Cancer-associated mutations within this gene were found to destabilize protein structure promoting amyloid-β aggregation in vitro, which is the pathological hallmark in Alzheimer's disease [36].

A number of functional pathways have been suggested to play a role in the pathogenesis of PD, such as lysosomal function, immune system response, and neuroinflammation [6–8,37]. We identified modules of genes that co-expressed across the six Braak stage-related

regions and found they were enriched for genes related to molecular processes that have been linked to the (pre)clinical symptoms and functional deficits in PD.

One negatively correlated module M127 was enriched for genes related to functions and diseases involving dopamine synthesis and motor functions. This module also contained the PD variant-associated gene *GCH1* (GTP cyclohydrolase 1) that is known to co-express with *TH* (tyrosine hydroxylase, the enzyme responsible for converting tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA) in the dopamine synthesis pathway) to enhance dopamine production and enable recovery of motor function in rat models of PD [38]. In this study, both *GCH1* and *TH* occur in M127 and thus were co-expressed across brain regions involved in Braak stages supporting their interaction. The higher expression in more vulnerable brain regions R1-R3 indicates that *GCH1*, *TH*, and possibly other genes within module M127 are essential to maintain dopamine synthesis that is affected in the early Braak stages of PD. Indeed, by inhibiting *TH* activity, α-synuclein can act as a negative regulator of dopamine release [30,39]. In this module, *SLC18A2* (vesicular monoamine transporter 2) and *SLC6A3* (dopamine transporter) were also present, which are important for dopamine storage and transport in the cell [30]. Interestingly, dopamine may increase neuronal vulnerability, as was suggested by an earlier study showing that α-synuclein is selectively toxic in dopaminergic neurons, and neuroprotective in non-dopaminergic cortical neurons [40]. Cell-type marker enrichment showed that module M127 was enriched for microglia- and neuronal markers, suggesting a role in neuroinflammation. α-Synuclein aggregates evoke microglia activation which in turn promotes aggregated protein propagation to other brain regions, possibly even from the gut or periphery to the brain [31,37]. The higher expression of microglial genes within module M127 may contribute to the higher vulnerability of brain regions affected during preclinical stages to form protein aggregates. Further investigation of genes within module M127 will provide a better understanding of the molecular mechanisms underlying microglia activation, dopaminergic pathways and motor functions.

Another negatively correlated module M47 was enriched for endothelial cell markers and genes involved in functions and disorders that relate to the immune response and oxygen transport in blood. One previous case-control study showed that anemia or low hemoglobin levels may precede the onset of PD [41]. Several studies using blood transcriptomic meta-analysis revealed genes associated with hemoglobin and iron metabolism were downregulated in PD patients compared to controls [42–44]. In our study, several hemoglobin genes (*HBD, HBB, HBA1*, *HBA2*, and *OASL*) were also present in module M47 of which *HBD* and *HBB* have been described to be highly interconnected with *SNCA* [44]. We also found an association between the interferon-gamma-mediated signaling pathway and M47 in which *OASL* also plays a role. Module M47 was negatively co-expressed with *SNCA*. Notably, a significant loss of negative co-expression between *SNCA* and interferon-gamma genes in the substantia nigra has been demonstrated in PD patients as compared to controls [45]. This loss may result from a downregulation of genes within M47 in the substantia nigra of PD patients, similarly as was observed in blood transcriptomics of PD [42–44]. This could be confirmed for *ATXN3* in the substantia nigra of PD patients [46]. Therefore, these genes

have the potential to serve as blood biomarkers for PD vulnerability. Overall, these studies suggest that dysregulation of genes within module M47 involved in blood-oxygen transport and the immune system influence brain regions to be selectively vulnerable to PD.

Identification of transcriptomic features in regions or disease conditions may be confounded by changes in cell-type composition. We used PSEA [25] to examine the impact of this confounding factor and found that all 960 BRGs remained differentially expressed between regions R1 and R6 in the AHBA. We also applied PSEA in the two PD datasets that allowed us to examine cell-type specificity between regions as well as between disease conditions. Although it is known that gene expression varies more between regions than between disease conditions [28], it is less clear how cell-type composition contributes to this variation. Here, we found that regional comparisons yielded more significant results than when comparing disease conditions. Therefore, BRGs also captured expression changes between patients and controls, but changes were less dependent on cell-type abundance between regions than between patients and controls.

To get a full understanding of how cell-types affect PD progression, we would need single cell data that map to all six Braak stage-related regions. This would allow us to assess which cell-types might influence regional vulnerability and which genes or modules relate to these cell-type specific processes. Currently, there are many single cell datasets available, but they mainly cover limbic and cortical regions [47,48], which only includes regions involved in the clinical stages of PD. For the substantia nigra, the hallmark region of PD, there is currently only one human single cell dataset available derived from archived samples [49]. Thus, the single cell datasets that are available now do not cover all regions in the Braak staging scheme. However, as the field of single cell analysis is growing at a fast pace and more datasets are being published, we expect that studying the role of cell-types with respect to regional vulnerability to PD in a similar way as we presented here will be possible in the near future.

Our findings on BRGs were based on regional expression differences that we analyzed using the AHBA. Although the number of AHBA donors is low, we confirmed these expression patterns in UKBEC and GTEx where the number of donors is high. Since most PD studies are limited by the availability of post-mortem brains of PD patients, the two PD datasets in our study had both low numbers of regional samples and donors. Thus, our findings on regional differences in PD patients are less reliable than our findings based on non-neurological controls. Nevertheless, they can still give an indication on how the expression of BRGs changes in brains of PD patients. This study showed that collecting more samples from multiple brain regions in post-mortem PD brains is valuable to get a better understanding of the vulnerability to PD.

In conclusion, we identified genes and pathways that may be important to maintain biological processes in specific brain regions, but may also contribute to a higher selective vulnerability to PD. Our results suggest that interactions between microglial genes and genes involved in dopamine synthesis and motor functions, as well as between genes

involved in blood-oxygen transport and the immune system may contribute to the early involvement of specific brain regions in PD progression. Our observations highlight a potential complex interplay of pathways in healthy brains and provide clues for future genetic targets concerning the pathobiology in PD brains.

6.4 Acknowledgements

We thank S.M.H. Huisman and prof. J. J. Goeman for their support in the statistical analyses. We also want to thank G. Bonvicini for her help in accessing the RNA-seq data from PD patients and V. Bonifati for critical discussions during the preparation of the manuscript. This research received funding from The Netherlands Technology Foundation (STW), as part of the STW project 12721 (Genes in Space, PI Lelieveldt). WvB and MCCH received funding from Alzheimer Netherlands and LECMA/Vaincre Alzheimer to collect the RNA-sequencing datasets that are used in this study. WvB was financially supported by grants from Amsterdam Neuroscience, Dutch Research council (ZonMW), Stichting Parkinson Fonds, Alzheimer association, and Rotary Aalsmeer-Uithoorn. WvB performed contract research and consultancy for Roche Pharma, Lysosomal Therapeutics, CHDR, Cross beta Sciences and received research consumables from Roche and Prothena. MCCH was financially supported by grants from INSERM, CHU de Lille, Université de Lille, BiLille, LECMA/Vaincre Alzheimer, and French Health Ministry for the PHRCs.

6.5 Methods

6.5.1 Allen Human Brain Atlas

To examine gene expression patterns across brain regions involved in PD, we used normalized gene expression data from the AHBA [15], a human post-mortem microarray dataset of 3,702 anatomical brain regions from six non-neurological individuals (five males and one female, mean age 42, range 24–57 years). We downloaded the data from http://human.brain-map.org/. To filter and map probes to genes, the data were concatenated across the six donors. We removed 10,521 probes with missing Entrez IDs, and 6,068 probes with low presence as they were expressed above background in <1% of samples (PA-call containing presence/absence flag [15]). The remaining 44,072 probes were mapped to 20,017 genes with unique Entrez IDs using the *collapseRows*-function in R-package WGCNA v1.64.1 [22] as follows: (i) if there is one probe, that one probe is chosen, (ii) if there are two probes, the one with maximum variance across all samples is chosen (method=maxRowVariance), (iii) if there are more than two probes, the probe with the highest connectivity (summed adjacency) is chosen (connectivityBasedCollapsing=TRUE). Based on the anatomical labels given in the AHBA, 2,334 out of 3,702 samples were mapped to Braak stage-related regions R1-R6 as defined by the BrainNet Europe protocol [18] and each region corresponds to one or multiple anatomical structures (Supplementary Table 1). The locus coeruleus and pontine raphe nucleus are both part of the pontine tegmentum in R2.

6.5.2 UK Brain Expression Consortium (UKBEC)

UKBEC [20] (http://www.braineac.org) contains microarray expression data from 10 brain regions of 134 non-neurological donors (74.5% males, mean age 59, range 16–102 years) for which their control status was confirmed by histology. We used the biomaRt R-package version 2.38 [23] to map Affymetrix probe IDs from UKBEC to gene Entrez IDs; 262,134 out of 318,197 probes could be mapped. Similar as with the AHBA, expression data for all probes and samples was concatenated across the 10 brain regions before mapping probes to 18,333 genes with unique Entrez IDs using the *collapseRows*-function.

6.5.3 Genotype-Tissue expression consortium (GTEx)

From GTEx [19] (https://gtexportal.org), we obtained RNA-sequencing (RNA-seq) samples from four brain tissues from multiple non-neurological subjects (65.7% males, range 20–79 years): substantia nigra (88 samples), amygdala (121 samples), anterior cingulate cortex (100 samples), and frontal cortex (129 samples). These brain regions corresponded to Braak stage-related regions R3-R6, respectively. We downloaded gene read counts (v7) for differential expression analysis and gene transcript per million (TPM) expression values (v7) for visualization. Out of 56,202 genes, we selected 19,820 protein coding genes and removed 405 genes with zero counts in one of the four regions of interest; 19,415 genes were left for analysis.

6.5.4 PD microarray dataset

In the PD microarray dataset, samples were collected from the medulla oblongata (R1), locus coeruleus (R2), and substantia nigra (R3) from PD patients (67.6% males, mean age 78, range 61–87 years), iLBD patients (42.4% males, mean age 80, range 56–98 years), and non-demented controls (54.5% males, mean age 77, range 60–91 years) (Supplementary Table 2 and Supplementary Data 4). The PD microarray data of the substantia nigra (R3) was previously published in Dijkstra et al. [11] (GEO accession number GSE49036). Based on pathological examination, PD patients in the microarray dataset revealed LB pathology in accordance with Braak stages 4-6, and iLBD patients showed LB pathology in the brainstem (Braak stages 1-3), and therefore represent the early stages of PD. Additional samples from the medulla oblongata (R1) and locus coeruleus (R2) were collected and processed of the same cohort in the same manner for hybridization on GeneChip® Human Genome U 133 Plus 2.0 arrays. Probe IDs were mapped to Entrez IDs with the *mapIds*-function in the hgu133plus2.db R-package v3.2.3. We removed 10,324 out of 54,675 probes with missing Entrez IDs. The remaining 44,351 probes were mapped to 20,988 genes with unique Entrez IDs using the *collapseRows*-function similarly as was done for the AHBA.

6.5.5 PD RNA-sequencing dataset

In the PD RNA-seq dataset, samples from the substantia nigra (R3) and medial temporal gyrus (R4/R5) were collected from PD patients (61.1% males, mean age 79, range 57–88 years), and non-demented age-matched controls (48.0% males, mean age 78, range 59–93 years) (Supplementary Table 3 and Supplementary Data 5). The extracted RNA was quantified using an Ozyme NanoDrop System, of which 500 ng of total RNA from each sample was further processed for purification of ribosomal RNA (rRNA) using human

Illumina Ribo-Zero™ rRNA Removal Kit. Then the Illumina TruSeq stranded total RNA protocol was used for library preparation. The library was sequenced on a Hiseq4000. RNA-seq reads were aligned to human genome (GRCh 38) with TopHat software (version: 2.1.1) using reference gene annotations (Ensembl GRCh38.p3) to guide the alignment. The count of reads per gene were determined from the alignment file (bam) and reference gene annotations (Ensembl) using FeatureCounts software (version: 1.5.3), resulting in 52,411 transcripts with Ensembl IDs. Entrez IDs of 20,017 genes in the AHBA were mapped to Ensembl IDs using biomaRt R-package version 2.38.

The brain samples for the PD microarray and RNA-seq analysis were obtained from The Netherlands Brain Bank (NBB), Netherlands Institute for Neuroscience, Amsterdam (open access: www.brainbank.nl). All Material has been collected from donors for or from whom a written informed consent for a brain autopsy and the use of the material and clinical information for research purposes had been obtained by the NBB. All procedures performed in studies involving human participants were in accordance with the ethical standards of the VU University Medical Center (VUmc, Amsterdam) and local Medical Ethics Committee (METC VUmc, reference number 2009/148) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

6.5.6 Braak stage-related genes (BRGs)

Two analysis methods were used to find BRGs for which the spatial expression in the AHBA is related to the progression of the disease: (i) Pearson's correlation between gene expression and labels 1-6 according to their assignment to one of the Braak stage-related regions R1-R6, and (ii) differential expression between Braak stage-related regions R1 and R6. As the expression values were $\log_2$-transformed, the mean difference between two regions was interpreted as the FC. Genes were assigned as BRGs based on the overlap of the top 10% (2,001) ranked genes with: (i) highest absolute correlation between gene expression and Braak stage labels, (ii) highest absolute FC between R1 and R6, and (iii) lowest Benjamini-Hochberg (BH) corrected *P*-value of the FC.

To avoid capturing donor-specific changes, we applied correlation and differential expression analyses for each of the six brain donors separately, and effect sizes were then combined by meta-analysis (metafor R-package 2.0). A random effects model was applied which assumes that each brain is considered to be from a larger population of brains and therefore takes the within-brain and between-brain variance into account. The between-brain variance (tau^2) was estimated with the Dersimonian-Delaird model. Variances and confidence intervals were obtained using the *escalc*-function. Correlations were Fisher-transformed ($z = \text{arctanh}(r)$) to obtain summary estimates, which were then back-transformed to correlation values ranging between -1 and +1. The significance of summary effect sizes (correlations and FCs) was assessed through a two-sided *t*-test with 5 degrees of freedom (H_0: FC=0). *P*-values were BH-corrected for all 20,017 genes. The weights used in the meta-analysis are based on the non-pooled expression variance in R1-R6.

The negatively and positively correlated BRGs were assessed for enrichment of functional GO terms using RDAVIDWebService R-package 1.20. The 20,017 genes from the AHBA were used as background genes. Functional GO terms were selected when BH-corrected *P*-value < 0.05 and gene count was at least 20.

6.5.7 Differential gene expression in validation datasets

A two-sided unpaired *t*-test was used to assess expression differences between conditions (PD, iLBD, and age-matched controls) and brain regions (R1-R6) in the AHBA, UKBEC, and PD microarray dataset. For GTEx, we used DESeq2 version 1.22.2 [24]. For the PD RNA-seq dataset, normalization and differential expression was done with 'DESeq2' R-package version 1.10.1, with age and sex introduced in the statistical model to take into account possible biases. Each analysis done with DESeq2 used a two-sided Wald test. The cut-off for differentially expressed genes was $P < 0.05$ (BH-corrected). For microarray experiments, the FC was interpreted as the difference in mean expression μ_B-μ_A, with μ as the mean expression in either group A and B. For RNA-seq experiments, FC is the $\log_2$ FC obtained from DESeq2.

6.5.8 Cell-type specific analysis

To assess whether results were confounded by cell-type composition in different brain regions and conditions, we applied PSEA [25] in the AHBA, PD microarray dataset, and PD RNA-seq datasets. Data from the AHBA were first concatenated across the six donors before applying PSEA. This method applies linear regression to examine whether the expression between two groups of samples is different (two-tailed *t*-test) while correcting for cell-type composition estimated from cell-type markers. To define cell-type markers, we used gene expression data from sorted cells of the mouse cerebral cortex [26]. Genes were selected as markers when they had a 20-fold higher expression compared to the mean of the other cell-types. All genes were analyzed while correcting for five main cell-types for which the cell-type signal was estimated by taking the mean expression of markers: 628 neurons, 332 astrocytes, 186 oligodendrocytes, 520 microglia, and 456 endothelial cells. *P*-values were BH-corrected across all genes in a dataset and significant when <0.05.

6.5.9 Gene co-expression modules in Braak stage-related regions

Gene co-expression matrices (pairwise Pearson's correlation, *r*, across Braak stage-related regions R1-R6) were calculated for each one of the six brain donors in the AHBA separately, and then combined into one consensus matrix based on the element-wise mean across all donors. Co-expression was converted to dissimilarity based on $1 - r$; in this way only positively co-expressed genes were taken into account. All genes were hierarchically clustered using single, complete, and average linkage and co-expression modules were obtained with the *cutreeDynamicTree*-function in the dynamicTreeCut R-package 1.63; minimum module size was set to 50 by default. The weighted correlation network analysis (WGCNA) R-package version 1.64.1 [22] was used for further analysis of the modules. Hierarchical clustering by average linkage resulted in an acceptable number of missing genes while retaining the maximum number of modules (Supplementary Figure 15; 167 modules with sizes up to 297 genes). For each module, the eigengene was obtained based on

the first principle component and thus summarizes the expression of all genes within a module across all samples in Braak stage-related regions R1-R6. This was done for each brain donor separately. The sign of the eigengene expression was corrected based on the sign of its Pearson's correlation with the mean expression of all genes within the module. Similar to the BRGs, the eigengene of each module was correlated with Braak stage labels for each donor separately and correlations were combined across donors using meta-analysis.

6.5.10 Gene set enrichment analysis of Braak stage-related modules

The one-sided hypergeometric test was used to identify modules that are significantly enriched for BRGs, cell-type markers [26], gene ontology- (GO), and disease-associated genes from DisGeNET [27]. A table of 561,119 gene-disease associations were obtained from DisGeNET version 5.0 (May, 2017) from http://www.disgenet.org/. Gene sets associated with 17,857 GO terms were obtained from the Ensembl dataset *hsapiens_gene_ensembl* version 92 through biomaRt R-package version 2.38. All gene sets were filtered to contain only genes matching the 20,017 genes in the AHBA and at least 10 genes. Modules were significantly enriched when $P < 0.05$ (BH-corrected for number of modules and gene sets) using all 20,017 genes from the AHBA as background genes.

6.5.11 Data and code availability

Gene expression data from healthy subjects used in this study are publicly available at brain-map.org, braineac.org, and gtexportal.org. Microarray data from PD, and iLBD patients, and controls were collected and shared by Amsterdam University Medical Center, the Netherlands. More details are described in the above sections. Scripts to run all analyses can be found online: https://github.com/arlinkeo/pd_braak. Scripts to analyze the microarray and RNA-seq datasets of PD patients were run in R version 3.4. The DESeq2 analysis of the PD RNA-seq data is shared on https://gitlab.univ-lille.fr/bilille/2017-mc-chartier-rna-seq.

References

[1] H. Braak, K. Del Tredici, U. Rüb, R. A. I. De Vos, E. N. H. Jansen Steur, and E. Braak, "Staging of brain pathology related to sporadic Parkinson's disease," *Neurobiol. Aging* 24, 197–211 (2003).

[2] D. J. Surmeier, J. A. Obeso, and G. M. Halliday, "Selective neuronal vulnerability in Parkinson disease," *Nat. Rev. Neurosci.* 3, 973–982 (2016).

[3] P. Borghammer, "How does parkinson's disease begin? Perspectives on neuroanatomical pathways, prions, and histology," *Mov. Disord.* 33, 48–57 (2018).

[4] M.-C. Chartier-Harlin, J. Kachergus, C. Roumier, V. Mouroux, X. Douay, S. Lincoln, C. Levecque, L. Larvor, J. Andrieux, et al., "α-synuclein locus duplication as a cause of familial Parkinson's disease," *Lancet* 364, 1167–1169 (2004).

[5] A. B. Singleton, M. Farrer, J. Johnson, A. Singleton, S. Hague, J. Kachergus, M. Hulihan, T. Peuralinna, A. Dutra, et al., "α-Synuclein locus triplication causes Parkinson's disease," *Science* 302, 841 (2003).

[6] V. Bonifati, "Genetics of Parkinson's disease – state of the art, 2013," *Parkinsonism Relat. Disord.* 20, S23–S28 (2014).

[7] D. Chang, M. A. Nalls, I. B. Hallgrímsdóttir, J. Hunkapiller, M. van der Brug, F. Cai, G. A. Kerchner, G. Ayalon, B. Bingol, et al., "A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci," *Nat. Genet.* 49, 1511–1516 (2017).

[8] M. A. Nalls, N. Pankratz, C. M. Lill, C. B. Do, D. G. Hernandez, M. Saad, A. L. Destefano, E. Kara, J. Bras, et al., "Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease," *Nat. Publ. Gr.* 46, 989–993 (2014).

[9] E. Glaab and R. Schneider, "Neurobiology of Disease Comparative pathway and network analysis of brain transcriptome changes during adult aging and in Parkinson's disease," *Neurobiol. Dis.* 74, 1–13 (2015).

[10] B. E. Riley, S. J. Gardai, D. Emig-Agius, M. Bessarabova, A. E. Ivliev, B. Schüle, J. Alexander, W. Wallace, G. M. Halliday, et al., "Systems-based analyses of brain regions functionally impacted in Parkinson's disease reveals underlying causal mechanisms," *PLoS One* 9, 1–14 (2014).

[11] A. A. Dijkstra, A. Ingrassia, R. X. De Menezes, R. E. van Kesteren, A. J. M. Rozemuller, P. Heutink, and W. D. J. van de Berg, "Evidence for immune response, axonal dysfunction and reduced endocytosis in the substantia nigra in early stage Parkinson's disease," *PLoS One* 10, 1–21 (2015).

[12] E. Oerton and A. Bender, "Concordance analysis of microarray studies identifies representative gene expression changes in Parkinson's disease: a comparison of 33 human and animal studies," *BMC Neurol.* 17, 1–14 (2017).

[13] R. Freer, P. Sormanni, G. Vecchi, P. Ciryam, C. M. Dobson, and M. Vendruscolo, "A protein homeostasis signature in healthy brains recapitulates tissue vulnerability to Alzheimer's disease," *Sci. Adv.* 2, 1–8 (2016).

[14] J. Sepulcre, M. J. Grothe, O. Uquillas, L. Ortiz-terán, I. Diez, H. Yang, H. I. L. Jacobs, B. J. Hanseeuw, Q. Li, et al., "Neurogenetic contributions to amyloid beta and tau spreading in the human cortex," *Nat. Med.* 24, 1910–1918 (2018).

[15] M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C.-K. K. Lee, et al., "Canonical genetic signatures of the adult human brain," *Nat. Neurosci.* 18, 1832–1844 (2015).

[16] T. Rittman, M. Rubinov, P. E. Vértes, A. X. Patel, C. E. Ginestet, B. C. P. Ghosh, R. A. Barker, M. G. Spillantini, E. T. Bullmore, et al., "Regional expression of the *MAPT* gene is associated with loss of hubs in brain networks and cognitive impairment in Parkinson disease and progressive supranuclear palsy," *Neurobiol. Aging* 48, 153–160 (2016).

[17] B. S. Freeze, D. Acosta, S. Pandya, Y. Zhao, and A. Raj, "Regional expression of genes mediating trans-synaptic alpha-synuclein transfer predicts regional atrophy in Parkinson disease," *NeuroImage Clin.* 18, 456–466 (2018).

[18] I. Alafuzoff, P. G. Ince, T. Arzberger, S. Al-Sarraj, J. Bell, I. Bodi, N. Bogdanovic, O. Bugiani, I. Ferrer, et al., "Staging/typing of Lewy body related α-synuclein pathology: A study of the BrainNet Europe Consortium," *Acta Neuropathol.* 117, 635–652 (2009).

[19] L. J. Carithers, K. Ardlie, M. Barcus, P. A. Branton, A. Britton, S. A. Buia, C. C. Compton, D. S. Deluca, J. Peter-demchok, et al., "A novel approach to high-quality postmortem tissue procurement: The GTEx project," *Biopreserv Biobank* 13, 311–319 (2015).

[20] D. Trabzuni, M. Ryten, R. Walker, C. Smith, S. Imran, A. Ramasamy, M. E. Weale, and J. Hardy, "Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies," *J. Neurochem.* 119, 275–282 (2011).

[21] M. J. Hurley, P. F. Durrenberger, S. M. Gentleman, A. F. Walls, and D. T. Dexter, "Altered expression of brain proteinase-activated receptor-2, trypsin-2 and serpin proteinase inhibitors in Parkinson's disease," *J. Mol. Neurosci.* 57, 48–62 (2015).

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

[23] D. Smedley, S. Haider, S. Durinck, L. Pandini, P. Provero, J. Allen, O. Arnaiz, M. H. Awedh, R. Baldock, et al., "The BioMart community portal: An innovative alternative to large, centralized data repositories," *Nucleic Acids Res.* 43, W589–W598 (2015).

[24] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol.* 15, 1–21 (2014).

[25] A. Kuhn, D. Thu, H. J. Waldvogel, R. L. M. Faull, and R. Luthi-Carter, "Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain," *Nat. Methods* 8, 945–947 (2011).

[26] Y. Zhang, K. Chen, S. A. Sloan, M. L. Bennett, A. R. Scholze, S. O'Keeffe, H. P. Phatnani, P. Guarnieri, C. Caneda, et al., "An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex," *J. Neurosci.* 34, 11929–11947 (2014).

[27] J. Piñero, À. Bravo, N. Queralt-Rosinach, A. Gutiérrez-Sacristán, J. Deu-Pons, E. Centeno, J. García-García, F. Sanz, and L. I. Furlong, "DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Res.* 45, D833–D839 (2017).

[28] 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).

[29] H. Deng and L. Yuan, "Genetic variants and animal models in *SNCA* and Parkinson disease," *Ageing Res. Rev.* 15, 161–176 (2014).

[30] L. L. Venda, S. J. Cragg, V. L. Buchman, and R. Wade-, "α-Synuclein and dopamine at the crossroads of Parkinson's disease," *Trends Neurosci.* 33, 559–568 (2013).

[31] J. A. Steiner, E. Quansah, and P. Brundin, "The concept of alpha-synuclein as a prion-like protein: ten years after," *Cell Tissue Res.* 373, 161–173 (2018).

[32] D. Sulzer, R. N. Alcalay, F. Garretti, L. Cote, E. Kanter, J. Agin-Liebes, C. Liong, C. McMurtrey, W. H. Hildebrand, et al., "T cells from patients with Parkinson's disease recognize α-synuclein peptides," *Nature* 546, 656–661 (2017).

[33] G. Li, S. Cui, J. Du, J. Liu, P. Zhang, Y. Fu, and Y. He, "Association of GALC, ZNF184, IL1R2 and ELOVL7 with Parkinson's disease in southern chinese," *Front. Aging Neurosci.* 10, 1–6 (2018).

[34] M. Rothaug, F. Zunke, J. R. Mazzulli, M. Schweizer, and H. Altmeppen, "LIMP-2 expression is critical for β-glucocerebrosidase activity and α-synuclein clearance," *Proc. Natl. Acad. Sci.* 111, 15573–15578 (2014).

[35] S. Soukup and P. Verstreken, "EndoA/Endophilin-A creates docking stations for autophagic proteins at synapses," *Autophagy* 13, 971–972 (2017).

[36] S. Bhattacharya, P. Hanpude, and T. K. Maiti, "Cancer associated missense mutations in BAP1 catalytic domain induce amyloidogenic aggregation: A new insight in enzymatic inactivation," *Sci. Rep.* 5, 18462 (2015).

[37] T. R. Sampson, J. W. Debelius, T. Thron, P. Wittung-stafshede, R. Knight, and S. K. Mazmanian, "Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease," *Cell* 167, 1469–1480 (2016).

[38] E. Cederfjäll, G. Sahin, D. Kirik, and T. Björklund, "Design of a single AAV vector for coexpression of TH and GCH1 to establish continuous DOPA synthesis in a rat model of Parkinson's disease," *Mol. Ther.* 20, 1315–1326 (2012).

[39] A. Abeliovich, Y. Schmitz, I. Farin, D. Choi-lundberg, W. Ho, P. E. Castillo, N. Shinsky, J. Manuel, G. Verdugo, et al., "Mice lacking α-synuclein display functional deficits in the nigrostriatal dopamine system," *Cell Neuron* 25, 239–252 (2000).

[40] J. Xu, K. Shyan-Yuan, F. J. . Lee, W. Song, L.-W. Jin, and B. A. Yankner, "Dopamine-dependent neurotoxicity of α-synuclein: A mechanism for selective neurodegeneration in Parkinson disease," *Nat. Med.* 8, 600–606 (2002).

[41] R. Savica, "Anemia or low hemoglobin levels preceding Parkinson disease," *Neurology* 73, 1381–1388 (2009).

[42] E. Mutez, L. Larvor, F. Leprêtre, V. Mouroux, D. Hamalek, J.-P. Kerckaert, J. Pérez-Tur, N. Waucquier, C. Vanbesien-Mailliot, et al., "Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation," *NBA* 32, 1839–1848 (2011).

[43] E. Mutez, A. Nkiliza, K. Belarbi, A. De Broucker, C. Vanbesien-mailliot, S. Bleuse, A. Du, T. Comptdaer, P. Semaille, et al., "Neurobiology of Disease Involvement of the immune system, endocytosis and EIF2 signaling in both genetically determined and sporadic forms of Parkinson's disease," *Neurobiol. Dis.* 63, 165–170 (2014).

[44] J. A. Santiago and J. A. Potashkin, "Blood transcriptomic meta-analysis identifies dysregulation of hemoglobin and iron metabolism in Parkinson' Disease," *Front. Aging Neurosci.* 9, 1–8 (2017).

[45] N. Liscovitch and L. French, "Differential co-expression between α-synuclein and IFN-γ signaling genes across development and in Parkinson's disease," *PLoS One* 9, 1–13 (2014).

[46] J. Kelly, R. Moyeed, C. Carroll, D. Albani, and X. Li, "Gene expression meta-analysis of Parkinson's disease and its relationship with Alzheimer's disease," *Mol. Brain* 12, 1–10 (2019).

[47] Q. Mu, Y. Chen, and J. Wang, "Deciphering brain complexity using single-cell sequencing," *Genomics, Proteomics Bioinforma.* 17, 344–366 (2019).

[48] M. Maitra, C. Nagy, and G. Turecki, "Sequencing the human brain at single-cell resolution," *Curr. Behav. Neurosci. Rep.* 6, 197–208 (2019).

[49] J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, and E. Z. Macosko, "Single-cell multi-omic integration compares and contrasts features of brain cell identity," *Cell* 177, 1873-1887.e17 (2019).

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).

Summary

Neurodegenerative diseases are hallmarked by protein inclusions and cell loss in disease-related brain regions. Many studies have tried to understand the molecular mechanisms that lead to the pathological and symptomatic hallmarks of neurodegeneration. Although studies highlighted important genes and biological pathways, the exact disease mechanisms are still not fully understood.

In this thesis, we make use of bioinformatics approaches to analyze spatial molecular data from the human brain. Using different computational methods, we mainly focused our research on Parkinson's disease (PD) that is characterized by the loss of dopaminergic neurons in the substantia nigra and the progressive deposition of protein inclusions, called Lewy bodies, across the brain. These pathological findings are associated with symptoms including slowing of movements, tremor, and cognitive impairment, but the exact cause for PD remains unknown. To better understand the molecular mechanisms within brain regions associated with PD, and more general neurodegenerative diseases, we exploited a high-resolution spatial gene expression atlas of the healthy human brain generated by the Allen Institute of Brain Science. Spatial transcriptomics allows examining the molecular and functional organization of the human brain and can be combined with neuroimaging data to identify brain regions and anatomical structures that are vulnerable to cell loss in neurodegenerative diseases. By combining both data modalities, we examined healthy molecular functions in brain regions associated with disease vulnerability based on neuroimaging features, namely gray matter loss within brain networks in individuals with Parkinson's disease, Huntington's disease, and individuals at risk of schizophrenia. Here, the analyses were based on gene expression differences between regions associated with a disease in neuroimaging studies and a region that is considered unaffected (non-susceptible). While the consecutive analyses of both data modalities revealed interesting associations, integrated analysis of both data modalities revealed possible new relationships between gene expression levels and disease-related changes measured with neuroimaging. Since our main focus is on Parkinson's disease, we were also interested in gene expression patterns across Braak stage-related regions and our analyses revealed genes that may play a role in the progression and the pathological spreading of Lewy bodies in PD.

With this thesis, we have shown that by applying data-driven computational methods we can explore the whole genome and find gene expression patterns informative of regional brain vulnerability in neurodegenerative diseases. Our methods can similarly be applied to unravel the molecular mechanisms in other neurodegenerative diseases, and potentially even reveal shared mechanisms between neurological disorders.

Samenvatting

Neurodegeneratieve ziekten worden gekenmerkt door eiwitinsluitingen en het verlies van cellen in hersengebieden die verband houden met deze ziekten. Meerdere onderzoeken hebben geprobeerd de moleculaire mechanismen te begrijpen die leiden tot de pathologische en symptomatische kenmerken van neurodegeneratie. Hoewel de onderzoeken belangrijke genen en biologische processen hebben blootgelegd, zijn de exacte mechanismen van deze ziekten nog steeds niet volledig begrepen.

In dit proefschrift gebruiken we bioinformatica benaderingen om de ruimtelijke moleculaire gegevens van het menselijk brein te analyseren. Met behulp van verschillende berekeningsmethoden hebben we ons onderzoek voornamelijk gericht op de ziekte van Parkinson (ZvP). Parkinson wordt gekenmerkt door het verlies van dopaminerge neuronen in de *substantia nigra* en de progressieve afzetting van eiwitinsluitingen, Lewy-lichaampjes genaamd, door het gehele brein. Deze pathologische bevindingen gaan gepaard met allerlei symptomen, waaronder vertraagde bewegingen, trillingen en cognitieve stoornissen. Maar de exacte oorzaak van de ZvP blijft vooralsnog onbekend. Om de moleculaire mechanismen in hersengebieden geassocieerd met de ZvP en andere neurodegeneratieve ziekten beter te begrijpen, hebben we gebruik gemaakt van een hoge resolutie ruimtelijke atlas van de expressie van alle genen in breinen van gezonde mensen, gegenereerd door het Allen Institute of Brain Science. Ruimtelijke genexpressiepatronen stellen ons in staat de moleculaire en functionele organisatie van het menselijk brein te bestuderen en deze data kan gecombineerd worden met medische beelddata voor het identificeren van hersengebieden en anatomische structuren die kwetsbaar zijn voor celverlies bij neurodegeneratieve ziekten. Door de twee datamodaliteiten te combineren, hebben we de gezonde moleculaire functies in hersengebieden onderzocht die verband houden met de kwetsbaarheid voor ziekten op basis van kenmerken uit beelddata, namelijk het verlies van grijze materie in hersennetwerken bij mensen met de ziekte van Parkinson, de ziekte van Huntington en mensen met een verhoogde kans op schizofrenie. Hierbij waren de analyses gebaseerd op verschillen in genexpressie tussen de gebieden die geassocieerd zijn met een ziekte in medische beelddata studies en gebieden die beschouwd wordt als onaangetast. Hoewel de opeenvolgende analyses van beide datamodaliteiten interessante associaties opleverden, onthulde de geintegreerde analyse van de twee datamodaliteiten nieuwe mogelijke relaties tussen niveaus van genexpressie en ziektegerelateerde veranderingen die zijn waargenomen met beelddata. Aangezien we ons voornamelijk op de ZvP richten, hebben we ook gekeken naar genexpressiepatronen in regio's die verband houden met het Braak-stadium en uit onze analyses bleek dat genen een rol zouden kunnen spelen in de progressie en de pathologische verspreiding van Lewy-lichaampjes bij de ZvP.

Met dit proefschrift hebben we aangetoond dat we door het toepassen van data-gedreven computationele methoden het hele genoom kunnen verkennen en dat we

genexpressiepatronen kunnen vinden die indicatie verschaffen over de regionale kwetsbaarheid van de hersenen bij neurogeneratieve ziekten. Onze methoden kunnen ook worden toegepast om de moleculaire mechanismen bij andere neurodegeneratieve ziekten te ontrafelen en mogelijk zelfs gedeelde mechanismen tussen neurologische aandoeningen te onthullen.

Publications

A. Keo, O. Dzyubachyk, J. van der Grond, J. J. van Hilten, M. J. T. Reinders, and A. Mahfouz, "Transcriptomic signatures associated with regional thickness changes in Parkinson's disease," *bioRxiv* (2020).

M. Meijer, **A. Keo**, J. M. C. van Leeuwen, O. Dzyubachyk, O. C. Meijer, C. H. Vinkers, and A. Mahfouz, "Molecular characterization of the stress network in the human brain," *bioRxiv* (2020).

A. Keo, A. Mahfouz, A. M. T. Ingrassia, J. P. Meneboo, C. Villenet, E. Mutez, T. Comptdaer, B. P. F. Lelieveldt, M. Figeac, et al., "Transcriptomic signatures of brain regional vulnerability to Parkinson's disease," *Commun. Biol.* 3, 1–12 (2020).

C. Anyansi, **A. Keo**, B. J. Walker, T. J. Straub, A. L. Manson, A. M. Earl, and T. Abeel, "QuantTB- A method to classify mixed Mycobacterium tuberculosis infections within whole genome sequencing data," *BMC Genomics* 21, 1–16 (2020).

A. Keo, O. Dzyubachyk, J. van der Grond, A. Hafkemeijer, W. D. J. van de Berg, J. J. van Hilten, M. J. T. Reinders, and A. Mahfouz, "Cholinergic circuit genes in the healthy brain are differentially expressed in regions that exhibit gray matter loss in Parkinson's disease," *bioRxiv* (2019).

A. Keo, N. A. Aziz, O. Dzyubachyk, J. van der Grond, W. M. C. van Roon-Mom, B. P. F. Lelieveldt, M. J. T. Reinders, and A. Mahfouz, "Co-expression Patterns between *ATN1* and *ATXN2* Coincide with Brain Regions Affected in Huntington's Disease," *Front. Mol. Neurosci.* 10, 1–13 (2017).

Acknowledgements

The PhD time has been a great journey, where family, friends, and colleagues have stood by my side and have supported me all the way. Your support and encouraging words have lifted me and helped me to push my own boundaries. Getting to the place where I am now as a person and as a scientist would have not been possible without you.

First, I want to thank the promoters and supervisors for giving me the opportunity to pursue my interest in Bioinformatics. Ahmed, you have been a great supervisor and I also appreciate you as a friend. Although I have been stubborn sometimes, your valuable ideas helped me grow and improve. Marcel, thank you for your trust and the chance to pursue a PhD. I enjoyed working with you and that is because I appreciate your honesty but also the teasing jokes. Boudewijn, I am honored to have been part of your team. Your enthusiasm for research is great to see and is uplifting to everyone. To my dear colleagues at the Division of Image Processing (Oleh, Baldur, Denis, Michèle, Leo, Els, Anna-Carien, Thomas, Marius, Mohamed, Niels, Jeroen, Antonis, Zhiwei), Leiden Computational Biology Center (Erik, Daniele, Indu, Thies, Thomas, Lieke, Martijn), Delft Bioinformatics Lab (Amin, Sjoerd, Jasper, Alexey, Christian, Soufiane, Alex, Tom, Stavros, Christine, Aysun, Joana, Ramin, Tamim, Meng, Mostafa, Mohammed, Mark, Sven, Henne, Thomas, Niccolo, Lucas), and the Pattern Recognition group (Gorkum, Seyran, Silvia, Taygun, Saskia, Bart, Ruud, Tom, Laura, Ekin, Wouter, Stephanie, Yancong, Alex, Marco, David, Chirag, Ziqi, Jose, Jan, Hamdi, Yunqiang, Yeswanth, Osman), thank you for the fun times, your helpfulness, and for accepting me weird as I am. I enjoyed the talks, conferences, social sport events, lab days, retreats, concerts, drinks, and cakes and I am happy to have shared this experience with you. To my friends and family, thank you for patiently listening to me when I talked about my work and for the good times that I needed in between.

To my parents, thank you for the unconditional love and pride. អរគុណម៉ែឪអោយខ្ញុំមិនឃ្លាតឆ្ងាយពីឈាមខ្មែរ Cheerkim, I know you always have my back. Cheerted, thank you for using your talent to create this awesome cover and I am proud that we have put this together as brother and sister. Niels, my rebellious husband, I'm very lucky to have you by my side. Your humor, creativity, intelligence, and unlimited skills inspire me in every aspect of life. You keep me grounded and down to earth.

Curriculum Vitæ

Arlin Keo was born on November 8th, 1988 in Oss, the Netherlands. She attended her pre-university education (VWO) in 2000 at the Prisma College Graaf Engelbrecht in Breda and graduated in 2007. In 2011, she obtained a Bachelor's degree in Biology and Applied Medical Laboratory Technology at Avans University of Applied Sciences in Breda, with a specialization in Biotechnology and Bio-informatics. After working as a chemical analyst at Purac BioChem in Gorinchem for a short while, she pursued her interest in bioinformatics and started studying Bioinformatics at TU Delft in 2012. In 2016, she obtained her Master's degree in Computer Science with a specialization in Bioinformatics. Thereafter, in 2016 until 2020, she continued with a PhD in Computational Biology at Leiden University Medical Center, where she started at the Radiology department and later at the new department Leiden Computational Biology Center. Since September 2020, Arlin is working as a postdoctoral researcher at the Cancer Computational Biology Center at the Erasmus Medical Center in Rotterdam doing research with the Department of Urology, Department of Medical Oncology, and Department of Gastroenterology and Hepatology.

ILLUSTRATION: Cheerted Keo