bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Transcriptomic signatures of brain regional vulnerability to Parkinson's disease

Arlin Keo[1,2], Ahmed Mahfouz[1,2], Angela M.T. Ingrassia[3], Jean-Pascal Meneboo[4,5], Celine Villenet[4], Eugénie Mutez[6,7,8], Thomas Comptdaer[6,7], Boudewijn P.F. Lelieveldt[1,2,9], Martin Figeac[4,5], Marie-Christine Chartier-Harlin[6,7]*, Wilma D.J. van de Berg[3]*, Jacobus J. van Hilten[10]*, and Marcel J.T. Reinders[1,2]*.

1. Leiden Computational Biology Center, Leiden University Medical Center, Leiden, The Netherlands.
2. Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands.
3. Department of Anatomy and Neurosciences, Amsterdam Neuroscience, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands.
4. Univ. Lille, Plate-forme de génomique fonctionnelle et Structurale, F-59000 Lille, France.
5. Univ. lille. Bilille, F-59000 Lille, France
6. Univ. Lille, Inserm, CHU Lille, UMR-S 1172 - JPArc - Centre de Recherche Jean-Pierre AUBERT Neurosciences et Cancer, F-59000 Lille, France
7. Inserm, UMR-S 1172, Early Stages of Parkinson's Disease, F-59000 Lille, France.
8. Univ. Lille, Service de Neurologie et Pathologie du mouvement, centre expert Parkinson, F-59000 Lille, France
9. Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands.
10. Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands

*Shared corresponding authors. E-mail: marie-christine.chartier-harlin@inserm.fr; wdj.vandeberg@amsterdamumc.nl; j.j.van_hilten@lumc.nl; m.j.t.reinders@tudelft.nl.

Keywords: Parkinson's disease, Braak Lewy body stages, transcriptomics, selective vulnerability, gene co-expression.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

The molecular mechanisms underlying the caudal-to-rostral progression of Lewy body pathology in Parkinson's disease (PD) remains poorly understood. Here, we aimed to unravel transcriptomic signatures across brain regions involved in Braak Lewy body stages in non-neurological controls and PD donors. Using human postmortem brain datasets of non-neurological adults from the Allen Human Brain Atlas, we identified expression patterns related to PD progression, including genes found in PD genome-wide associations studies: *SNCA*, *ZNF184*, *BAP1, SH3GL2, ELOVL7*, and *SCARB2*. We confirmed these patterns in two datasets of non-neurological subjects (Genotype-Tissue Expression project and UK Brain Expression Consortium) and found altered patterns in two datasets of PD patients. Additionally, co-expression analysis across vulnerable regions identified two modules associated with dopamine synthesis, the motor and immune system, blood-oxygen transport, and contained microglial and endothelial cell markers, respectively. Alterations in genes underlying these region-specific functions may contribute to the selective regional vulnerability in PD brains.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Background

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]. 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 remains 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 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]. 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

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 aim to unravel the molecular factors underlying selective vulnerability to LB pathology during PD progression. We used transcriptomic postmortem human brain data of non-neurological adult donors from the AHBA and analyzed brain regions involved in Braak LB stages[18]. We identified genes whose expression patterns increase or decrease across these regions, referred to as Braak stage-related genes (BRGs) and hypothesized that they might contribute to higher vulnerability to LBs in PD brains based on the sequence of events as postulated by Braak et al.[1] (Figure 1). We validated the relevance of the identified BRGs to PD progression with two datasets of non-neurological controls (Genotype-Tissue Expression project (GTEx)[19] and UK Brain Expression Consortium (UKBEC)[20]) and two datasets of PD donors. We found altered expression patterns in patients with PD and incidental Lewy body disease (iLBD), who are assumed to represent the pre-clinical stage of PD[11,21]. In non-neurological brains from the AHBA, we further identified modules of co-expressed genes across Braak LB stage involved regions, and characterized them functionally by assessing their enrichment for cell-type markers, gene ontology (GO)-terms, and disease-associated genes. Using our region-based approach, we identified genes known to harbor PD-associated variants (e.g. *SNCA*, *ZNF184*, *BAP1, SH3GL2, ELOVL7*, and *SCARB2*) and pathways linked to PD (e.g. dopamine biosynthetic process and immune response) with expression signatures associated with progressive LB pathology. In non-neurological controls, several dopaminergic genes were highly expressed in brain regions related to the preclinical stages of PD progression, highlighting their importance in maintaining motor functions. These observed transcriptomic signatures provide insights into the molecular mechanisms underlying brain regional vulnerability to PD, enabling the development of new and improved methods for diagnosis and treatment of PD.

Results

Parkinson's disease Braak stage-related genes

The PD Braak staging scheme defines a temporal order of brain regions affected during the progression of the disease18. We analyzed these brain regions using a microarray data set of 3,702 anatomical brain

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

regions from six individuals without any known neuropsychiatric or neurological background from the Allen Human Brain Atlas (AHBA)[15]. We first assigned brain samples to Braak stage-related regions R1-R618: 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) (Figure 1, Supplementary Table 1, and Supplementary Figure 1). To identify genes with expression patterns that are associated with selective vulnerability to PD, Braak stage-related genes (BRGs), we correlated gene expression with the label of these vulnerable regions as defined by Braak stage. To focus on genes with 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 (|fold-change (FC)| > 1; Benjamini-Hochberg (BH)-corrected $P < 0.05$; Supplementary Figure 2).

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}; two-sided t-test; Figure 2a and b). The top 10% (2,001) ranked genes for each criterion resulted in genes with $|r| > 0.66$, $|FC_{R1\text{-}R6}|>1.33$, and $P_{FC} < 0.00304$. 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 2c and Supplementary Table 2). Negatively correlated BRGs were significantly enriched for GO-terms like anatomical structure morphogenesis and blood vessel morphogenesis (Supplementary Table 3), while positively correlated BRGs were significantly enriched for functions like anterograde trans-synaptic signaling and nervous system development (Supplementary Table 4).

To validate the decreasing or increasing expression patterns of 960 BRGs observed in six non-neurological brains from the AHBA, we used two independent datasets from non-neurological controls and assessed differential expression between the two most distant regions in each dataset. 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

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

BRGs present in UKBEC, 139 out of 317 (43.8%) 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 $P < 0.05$, and Figure 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 R3 and R6 in this dataset ($FC_{R3\text{-}R6}| > 1$, BH-corrected $P < 0.05$, and Figure 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 age-matched controls. To test this hypothesis, we used two datasets with transcriptomic measurements from brain regions covering most Braak stage-related regions sampled from PD, iLBD and non-demented age-matched controls (PD microarray[11] (Supplementary Table 5 and 6) and PD RNA-seq datasets (Supplementary Table 7 and 8); see Methods). First, we found larger differences between brain regions within the same group of individuals (PD, iLBD, and control) than between conditions within the same region (number of differentially expressed genes in Supplementary Figure 3). This observation further highlights the importance of assessing expression patterns across regions rather than disease conditions[22]. Next, we validated the expression patterns of BRGs, which we identified in non-neurological adults, in non-demented age-matched controls in both the PD microarray and PD RNA-seq datasets and observed similar patterns (Figure 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 2f). Across regions R3 and R4/R5 however, these expression patterns were preserved in PD patients (Figure 2g). In addition to the changes across brain regions, we found that BRGs also captured changes across conditions PD, iLBD, and control within the substantia nigra (R3) for both PD datasets (Supplementary Figure 4). Negatively correlated BRGs that had higher expression in the most vulnerable brain regions R1-R3 also had higher expression in PD patients compared to controls. Vice versa, positively correlated BRGs that have higher expression in the least vulnerable brain regions

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

R4-R6 also have higher expression in controls compared to PD patients. These findings thus support the relation 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.

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 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 as defined by Braak stages (Figure 3a and Supplementary Table 9). 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 3b and Supplementary Figure 5). 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), which were most evidently reflected in R3 or R4. 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 ($P < 0.0001$, BH-corrected). Module M39 showed the lowest correlation with Braak stages ($r = -0.87$ and BH-corrected $P = 3.65e-7$), while M50 showed the highest correlation ($r = 0.92$ and BH-corrected $P = 4.42e-7$). Most modules were significantly enriched for BRGs that were similarly correlated with Braak stages (Figure 3c). For functional characterization, modules were further assessed for enrichment of cell-type markers[23], and gene sets associated with functional GO-terms or diseases.

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

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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.

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. Furthermore, 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 preclinical and clinical stages of PD.

Variable expression of BRGs is not fully explained by variations in 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)[24] 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 $P < 0.05$ and $\beta \neq 0$) 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 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.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 3), PSEA revealed most changes between brain regions than between patients and controls (Supplementary Figure 6).

Expression of several Parkinson's disease-implicated genes are related to Braak staging

We found that the expression patterns of several PD-implicated genes, identified in the two most recent PD 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 1 and Supplementary Figure 7) or genes present in Braak stage-related co-expression modules (*GCH1*, *ITIH3*, *ITPKB*, *RAB7L1*, *BIN3*, *SATB1*, *ASHL1*, *MAPT*, *DLG2*, and *DNAH1*; Figure 3).

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 5a, Figure 5b, and Figure 5c), which was replicated in larger cohorts of non-neurological individuals (Figure 5d and Figure 5e). 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 ceruleus (R2), and substantia nigra (R3) of PD-, 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 (Figure 5f). In the 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 (Figure 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 5h) and found that changes were independent of neuronal or other cell-type densities when comparing different brain regions. In PD datasets results were scattered and did not align between both PD

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

microarray and RNAseq datasets because of the small sample sizes and the comparison of different brain regions (Supplementary Figure 8).

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[25]. *GCH1*, *TH*, and *SLC6A3* (*DAT*) are related to the functional term dopamine biosynthetic process, and *SLC18A2* (*VMAT2*) is known to store dopamine into synaptic vesicles[26]. 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. 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 (Figure 6 and Supplementary Figure 9).

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 patients with PD 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[3,27,28]. In line with this assumption are reports of familial PD caused by *SNCA* multiplications, suggesting a *SCNA* dosage effect in causing

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

PD[5,29]. 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 10). 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 (Table 1 and Supplementary Figure 7). 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[30]. *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[31]. *SCARB2*-deficiency in mice brains led to α-synuclein accumulation mediating neurotoxicity in dopaminergic neurons[31]. Overexpression in murine and human cell lines improved lysosomal activity of this enzyme and enhanced α-synuclein clearance[31]. *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[32]. *BAP1* (ubiquitin carboxyl-terminal hydrolase) is a deubiquitinase that acts as a tumor

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

suppressor. Cancer-associated mutations within this gene were found to destabilize protein structure promoting β-amyloid aggregation in vitro, which is the pathological hallmark in Alzheimer's disease[33].

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[6–8,34]. We identified modules of genes that co-express 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 contains 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[35]. 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 (Figure 6 and Supplementary Figure 9). The higher expression in the 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[26,36]. In this module, *SLC18A2* (*VMAT2*; vesicular monoamine transporter 2) and *SLC6A3* (*DAT*; dopamine transporter) were also present, which are important for storage of dopamine and transport in the cell[26]. 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[37]. 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[27,34]. 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

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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[38]. Several studies using blood transcriptomic meta-analysis revealed genes associated with hemoglobin and iron metabolism were downregulated in PD patients compared to controls[39–41]. 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*[41]. 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[42]. This loss may result from a downregulation of genes within M47 in the substantia nigra of PD patients, similarly as was observed in PD blood transcriptomics[39–41]. 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[24] to examine the impact of this confounding factor and found that all 960 BRGs remained differentially expressed between regions R1 and R6 in 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[22], 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

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

changes were less dependent on cell-type abundance between regions than between patients and controls.

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.

Methods

Methods and any associated references are available in the online version of the paper.

References

1. Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24**, 197–211 (2003).

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

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

4. Chartier-Harlin, M.-C. *et al.* α-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169 (2004).

5. Singleton, A. B. *et al.* α-Synuclein Locus Triplication Causes Parkinson's Disease. *Science* **302**, 841 (2003).

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

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

7. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* 1–6 (2017). doi:10.1038/ng.3955

8. Nalls, M. A. *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. Glaab, E. & Schneider, R. 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. Riley, B. E. *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. Dijkstra, A. A. *et al.* 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. Oerton, E. & Bender, A. 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. Freer, R. *et al.* A protein homeostasis signature in healthy brains recapitulates tissue vulnerability to Alzheimer's disease. *Sci. Adv.* **2**, 1–8 (2016).

14. Sepulcre, J. *et al.* Neurogenetic contributions to amyloid beta and tau spreading in the human cortex. *Nat. Med.* **24**, 1910–1918 (2018).

15. Hawrylycz, M. *et al.* Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).

16. Rittman, T. *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. Freeze, B. S., Acosta, D., Pandya, S., Zhao, Y. & Raj, A. Regional expression of genes mediating

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

trans-synaptic alpha-synuclein transfer predicts regional atrophy in Parkinson disease. *NeuroImage Clin.* **18**, 456–466 (2018).

18. Alafuzoff, I. *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. Carithers, L. J. *et al.* A novel approach to high-quality postmortem tissue procurement: The GTEx project. *Biopreserv Biobank* **13**, 311–319 (2015).

20. Trabzuni, D. *et al.* 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. Hurley, M. J., Durrenberger, P. F., Gentleman, S. M., Walls, A. F. & Dexter, D. T. 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. Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).

23. Zhang, Y. *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).

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

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

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

27. Steiner, J. A., Quansah, E. & Brundin, P. The concept of alpha-synuclein as a prion-like protein : ten years after. *Cell Tissue Res.* (2018). doi:10.1007/s00441-018-2814-1

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

28. Sulzer, D. *et al.* T cells from patients with Parkinson's disease recognize α-synuclein peptides. *Nature* **546**, 656–661 (2017).

29. Chartier-Harlin, M.-C. *et al.* α-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169 (2004).

30. Li, G. *et al.* Association of *GALC*, *ZNF184*, *IL1R2* and *ELOVL7* With Parkinson's Disease in Southern Chinese. *Front. Aging Neurosci.* **10**, 1–6 (2018).

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

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

33. Bhattacharya, S., Hanpude, P. & Maiti, T. K. Cancer associated missense mutations in BAP1 catalytic domain induce amyloidogenic aggregation: A new insight in enzymatic inactivation. *Sci. Rep.* **5**, doi: 10.1038/ srep18462 (2015).

34. Sampson, T. R. *et al.* Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell* **167**, 1469–1480 (2016).

35. Cederfjäll, E., Sahin, G., Kirik, D. & Björklund, T. 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).

36. Abeliovich, A. *et al.* Mice Lacking α-Synuclein Display Functional Deficits in the Nigrostriatal Dopamine System. *Cell Neuron* **25**, 239–252 (2000).

37. Xu, J. *et al.* Dopamine-dependent neurotoxicity of α-synuclein : A mechanism for selective neurodegeneration in Parkinson disease. *Nat. Med.* **8**, 600–606 (2002).

38. Savica, R. Anemia or low hemoglobin levels preceding Parkinson disease. *Neurology* **73**, 1381–

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1388 (2009).

39. Mutez, E. *et al.* Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation. *NBA* **32**, 1839–1848 (2011).

40. Mutez, E. *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).

41. Santiago, J. A. & Potashkin, J. A. Blood Transcriptomic Meta-analysis Identifies Dysregulation of Hemoglobin and Iron Metabolism in Parkinson' Disease. *Front. Aging Neurosci.* **9**, 1–8 (2017).

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

43. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, (2008).

44. Smedley, D. *et al.* The BioMart community portal: An innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589–W598 (2015).

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

46. Piñero, J. *et al.* DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2017).

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.

Author contributions

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

AK, AM, BL, JvH, and MR designed the study. AK wrote the scripts and performed the data analyses. EM, MCCH, MF and WvB designed the RNASeq experiments, designed, and realized the RNAseq sample comparisons and validations. WvB selected brain tissue samples for RNAseq and AI processed the brain tissue samples. TC and CV prepared the sequencing libraries. JPM performed the analysis of the PD RNA-seq data (quality control, alignment, normalization) under the supervision of MF. AK, AM, MCCH, WvB, JvH, and MR interpreted the data and wrote the manuscript with input from all authors. AM and MR supervised the overall project. The manuscript was read and approved by all authors.

Grants

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 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, Vaincre Alzheimer, French Health Ministry for the PHRCs, French National Research Agency, and the Michael J. Fox Foundation for Parkinson's Research.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figures & Tables

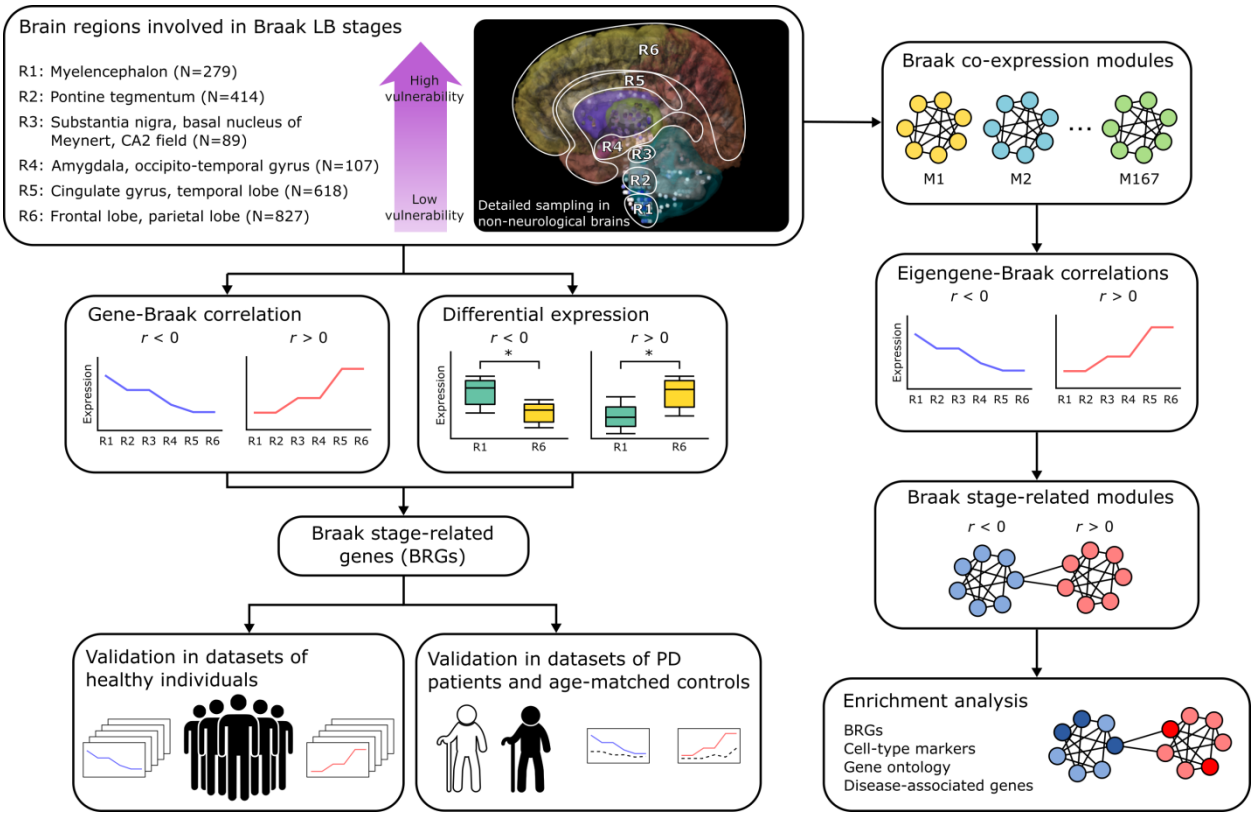

Figure 1 Study overview. Differential vulnerability to PD was examined across brain regions R1-R6. N is the tissue sample size across all six non-neurological donors from the Allen Human Brain Atlas (AHBA), which are involved in the six PD Braak stages as they sequentially accumulate Lewy bodies during 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 genes and correlated the module eigengene expression with Braak stages. The resulting modules of genes were subsequently analyzed to detect common biologically meaningful pathways.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

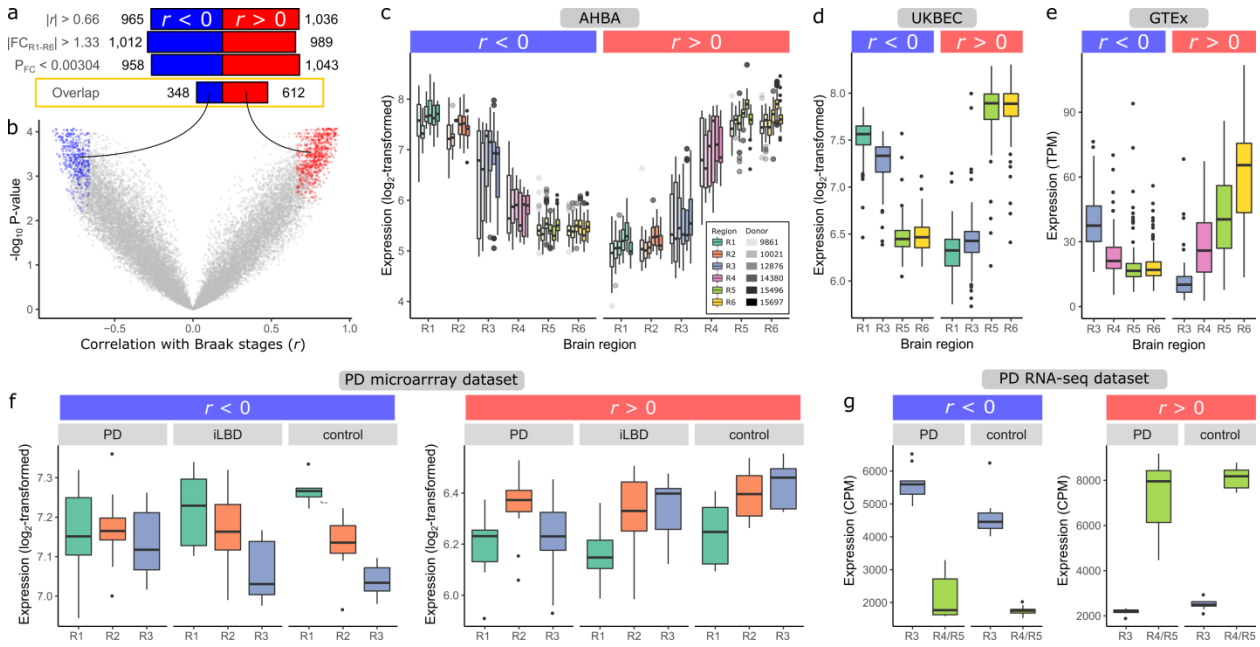

Figure 2 Expression patterns of Braak stage-related genes across brain regions of non-neurological, iLBD and PD brains. (a) Selection of BRGs that were either negatively (blue) or positively (red) correlated with Braak stages. Genes were selected based on 1) highest absolute correlation of gene expression 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 (P_{FC}; BH-corrected), 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 –log10 BH-corrected *P*-value (y-axis). (c) Mean expression of BRGs for each region (colors) and donor (opacity) in 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 the mean expression of BRGs across brain donors in one brain region for a BRG. (f) Validation in PD microarray dataset (R1: medulla oblongata, R2: locus ceruleus, R3: substantia nigra; number of samples in Supplementary Table 5) and (g) PD RNA-seq dataset (R3: substantia nigra, R4/R5: medial temporal gyrus; number of samples in Supplementary Table 7). Boxplots are shown (f, g) per patient group (PD, incidental Lewy body disease (iLBD), and control) and (Supplementary Figure 4) per brain region.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

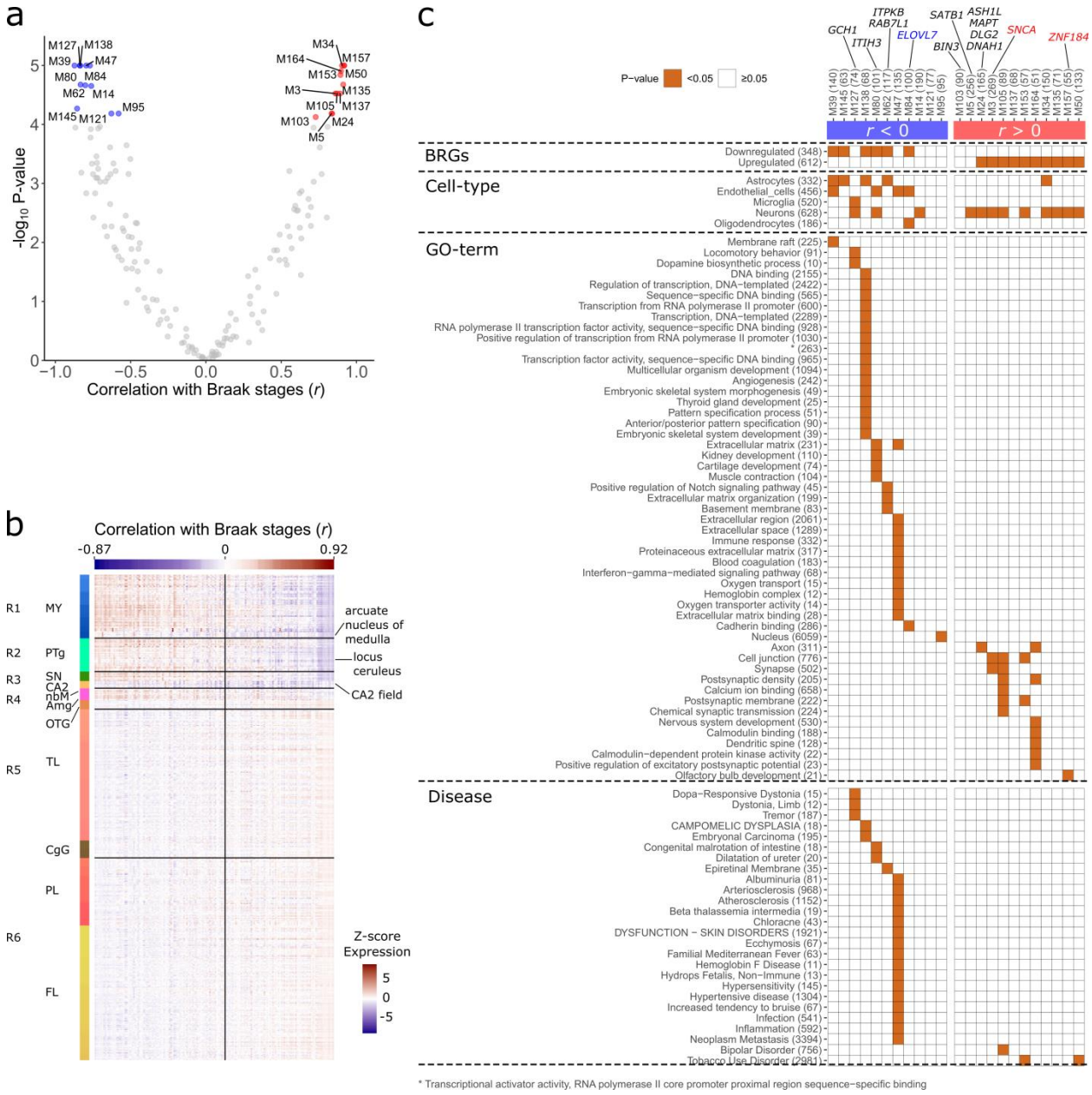

Figure 3 Braak co-expression modules. Genes were analyzed for co-expression across regions R1-R6 in the Allen Human Brain Atlas. (a) Module eigengene correlation with Braak. Each point reflects a module showing its correlation *r* with Braak stages (x-axis) and $\log_{10}$-transformed *P*-values (BH-corrected; y-axis); 23 significant modules (BH-corrected $P < 0.0001$; H_0: $r = 0$) were selected for further analysis (blue and red). (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 the 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

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(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 ceruleus and CA2-field, independently of their correlation with Braak stages. (c) Significant modules were sorted based on their correlation with Braak stages (columns) and assessed for significant overlap with BRGs, cell-type markers, and gene sets associated with functional GO-terms or diseases using a hypergeometric test (brown squares; $P < 0.05$; BH-corrected). Additionally, these modules reveal the presence of genes associated with PD variants (annotated at the top) that have (blue and red) and have not (black) been identified as BRGs.

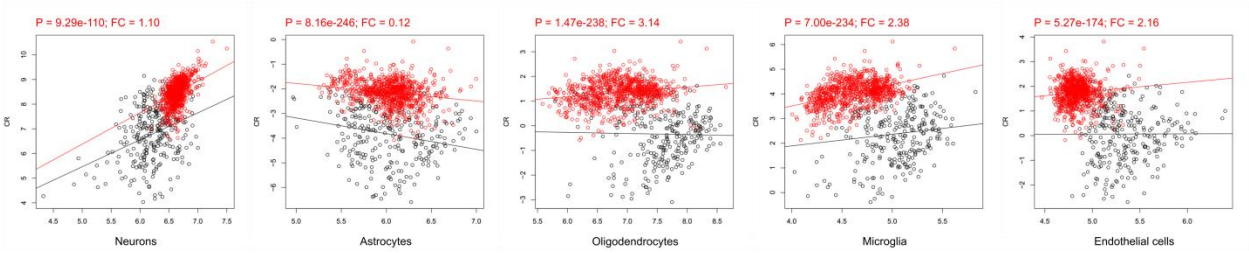

Figure 4 Differential expression of neuronal marker *ADCY1* in AHBA corrected for cell-type abundance. *ADCY1* is a neuronal marker identified as one of the 960 Braak stage-related genes (BRGs). We found it is still significantly differentially expressed between samples from region R1 (black) and R6 (red) when correcting for one of the five main cell-types with PSEA[24] (BH-corrected $P < 0.05$ for 960 BRGs). Significant BH-corrected *P*-values are highlighted in red text together with cell-type specific fold-changes (FC; slope change of red line).

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

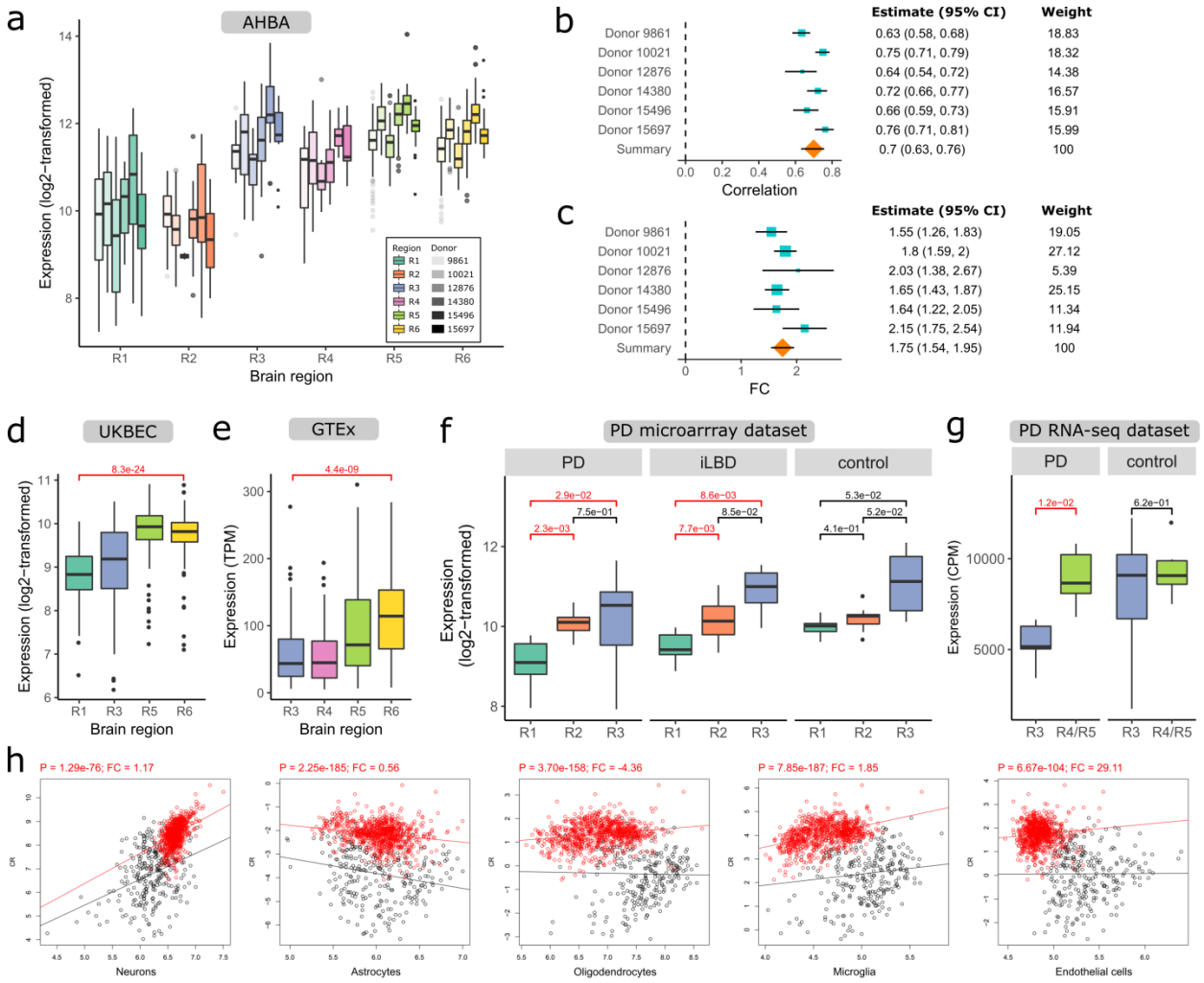

Figure 5 *SNCA* expression in brain regions R1-R6 of non-neurological individuals (a-e, h) and PD patients (f,g). (a) Boxplots of *SNCA* expression in regions R1-R6 (colored) for each donor (opacity) in AHBA (number of samples in Supplementary Table 1). Meta-analysis of (b) *SNCA* correlation with Braak across the six donors in AHBA and (c) *SNCA* expression fold-change between region R1 and R6. To calculate the summary effect size, the 95% confidence intervals (CI) and weights are taken into account. The positive correlation with Braak 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) and non-demented age-matched controls (number of samples in Supplementary Table 5 and 7). In the PD datasets, *SNCA* expression was tested for differential expression between (f, g) regions and conditions (Supplementary Figure 6) and was considered significant when BH-corrected $P < 0.05$ (red). (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 (BH-corrected $P < 0.05$ for 960 BRGs). Significant BH-corrected *P*-values are highlighted in red text

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

together with cell-type specific fold-changes (FC; slope change of red line). PSEA results for PD data are shown in Supplementary Figure 8.

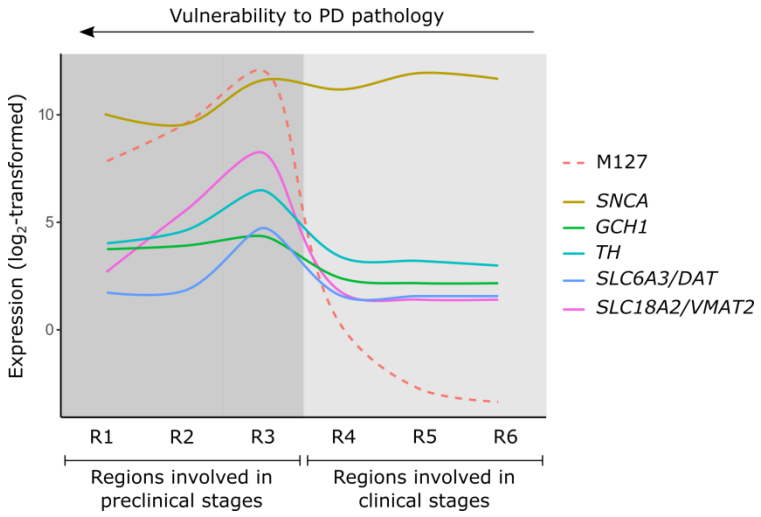

Figure 6 Schematic overview of molecular activity of dopaminergic genes in module M127 and *SNCA* across brain regions of the Braak staging scheme. Lines across regions R1-R6 were based on transcriptomic data from the Allen Human Brain Atlas (Figure 1 and Supplementary Figure 9). Expression of module 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, other genes 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 of L-3,4-dihydroxyphenylalanine (L-DOPA), *SLC6A3/DAT*: dopamine transporter: transports dopamine from the synaptic cleft back to the cytosol, *SLC18A2/VMAT2*: stores dopamine into synaptic vesicles.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 1 Braak stage-related genes that have previously been associated with PD. Several PD variant-associated genes showed expression profiles that are correlated with Braak stages. The correlation *r* with Braak, fold-change, and *P*-value of fold-change (t-test, BH-corrected) are within the selection thresholds for BRGs.

Gene symbol	Entrez ID	Correlation with Braak (*r*)	*P*-value (BH-corrected)	Fold-change	*P*-value (BH-corrected)	Module member	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

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Online methods

Allen Human Brain Atlas. To examine gene expression patterns across brain regions involved in Parkinson's disease (PD), we used normalized gene expression data from the Allen Human Brain Atlas (AHBA), 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[15]). We downloaded the data 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[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[43] 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 AHBA, 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. The locus ceruleus and pontine raphe nucleus are both part of the pontine tegmentum in R2.

Braak stage-related genes (BRGs). Two analysis methods were used to find genes for which the spatial expression in AHBA is related to the spread of the disease: 1) correlations between gene expression and labels 1-6 according to their assignment to one of the Braak stage-related regions R1-R6, and 2) 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 fold-change (FC). Genes were assigned as BRGs based on the overlap of the top 10% (2,001) ranked genes with: 1) highest absolute correlation of gene expression and Braak stage labels, 2) highest absolute FC between R1 and R6, and 3) lowest Benjamini-Hochberg (BH) corrected *P*-value of the FC.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To avoid capturing donor-specific changes, we applied the 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 = arctanh(*r*)) to obtain summary estimates, which were then back-transformed to correlation values ranging between -1 and +1. *P*-values were BH-corrected for all 20,017 genes. The significance of summary effect sizes (correlations and FCs) was assessed through a two-sided t-test (H_0: FC=0; unequal variances). The weights used in the meta-analysis are based on the non-pooled expression variance in R1-R6.

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[44] to map Affymetrix probe IDs from UKBEC to gene Entrez IDs; 262,134 out of 318,197 probes could be mapped. Similar as for 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.

Genotype-Tissue expression consortium (GTEx). From GTEx[19] (https://gtexportal.org), we obtained 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. From 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.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

PD microarray dataset. In the PD microarray dataset, samples were collected from the medulla oblongata (R1), locus ceruleus (R2), and substantia nigra (R3) from PD- (67.6% males, mean age 78, range 61–87 years) and, incidental Lewy body disease (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 5 and 6). The PD microarray data of the substantia nigra (R3) was previously published in Dijkstra et al.[11]. Based on pathological examination, PD patients in the microarray dataset revealed Lewy body (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 ceruleus (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 AHBA.

PD RNA-sequencing (RNA-seq) 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 7 and 8). 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 Illumina TruSeq stranded total RNA protocol was used for library preparation. The library was sequenced on a Hiseq4000. RNA-seq reads were aligned to the 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 AHBA were mapped to Ensembl IDs using biomaRt R-package version 2.38.

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

For both the PD microarray and RNA-seq dataset, a written informed consent for a brain autopsy and the use of the material and clinical information for research purposes had been obtained from all donors by The Netherlands Brain bank (NBB), Amsterdam. Autopsy was performed using a standardized protocol by NBB (open access: www.brainbank.nl). All procedures of NBB were approved by the local VUmc medical ethics committee.

Differential gene expression between brain regions and conditions. A two-sided unpaired t-test was used to assess expression differences between conditions (PD, iLBD, and controls) and brain regions (R1-R6) in the AHBA, UKBEC, and PD microarray dataset. For GTEx, we used DESeq2 version 1.22.2[45]. 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. *P*-values were BH-corrected across all genes. 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 $\log_2$ fold-change obtained from DESeq2.

Cell type-specific analysis. To assess whether results were confounded by cell-type composition in different brain regions and conditions, we applied population-specific expression analysis (PSEA)[24] in the AHBA, PD microarray, and PD RNA-seq datasets. Data from 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 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[23]. 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.

Gene co-expression modules in Braak stage-related regions R1-R6. Co-expression matrices (pairwise Pearson's correlation, *r*) were calculated for each one of the six brain donors of the AHBA

bioRxiv preprint doi: https://doi.org/10.1101/664771; this version posted June 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 are taken into account. All genes were hierarchically clustered using either single, complete, average linkage and co-expression modules were obtained with the *cutreeDynamicTree*-function in the WGCNA R-package version 1.64.1[43]; minimum module size was set to 50 by default. Hierarchical clustering by average linkage resulted in an acceptable number of missing genes while retaining the maximum number of modules (Supplementary Figure 10; 167 modules with sizes up to 297 genes). For each module, the eigengene was obtained for each brain donor separately 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. The sign of the eigengene expression was corrected based on the sign of its correlation with the mean expression of all genes within the module. Similar to the BRGs, the eigengene of each module was correlated with Braak labels for each donor separately and correlations were combined across donors using meta-analysis.

Gene set enrichment analysis of Braak stage-related modules. The hypergeometric test was used to identify modules that are significantly enriched for BRGs, cell-type markers[23], gene ontology- (GO), and disease-associated genes from DisGeNET[46]. A table of 561,119 gene-disease associations were obtained from DisGeNET version 5.0 (May, 2017) from http://www.disgenet.org/. Genes associated with 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 AHBA and at least 10 genes. Modules were significantly enriched when *P*-value < 0.05 (BH-corrected for number of modules) using all 20,017 genes from AHBA as background genes.

Data availability. 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.

Code availability. Scripts were run in R version 3.5 and can be found online: https://github.com/arlinkeo/PD. Scripts to analyze the PD microarray and PD RNA-seq dataset were run in R version 3.4.