Title
Regional transcriptional architecture of Parkinson's disease pathogenesis and network spread.

Permalink
https://escholarship.org/uc/item/9w66v2hh

Authors
Freeze, Benjamin
Pandya, Sneha
Zeighami, Yashar
et al.

Publication Date
2019-07-29

DOI
10.1093/brain/awz223

Peer reviewed
Regional transcriptional architecture of Parkinson’s disease pathogenesis and network spread

Benjamin Freeze,¹ Sneha Pandya,¹ Yashar Zeighami² and Ashish Raj¹,³

Although a significant genetic contribution to the risk of developing sporadic Parkinson’s disease has been well described, the relationship between local genetic factors, pathogenesis, and subsequent spread of pathology throughout the brain has been largely unexplained in humans. To address this question, we use network diffusion modelling to infer probable pathology seed regions and patterns of disease spread from MRI atrophy maps derived from 232 de novo subjects in the Parkinson’s Progression Markers Initiative study. Allen Brain Atlas regional transcriptional profiles of 67 Parkinson’s disease risk factor genes were mapped to the inferred seed regions to determine the local influence of genetic risk factors. We used hierarchical clustering and L₁ regularized regression analysis to show that transcriptional profiles of immune-related and lysosomal risk factor genes predict seed region location and the pattern of disease propagation from the most likely seed region, substantia nigra. By leveraging recent advances in transcriptomics, we show that regional microglial abundance quantified by high fidelity gene expression also predicts seed region location. These findings suggest that early disease sites are genetically susceptible to dysfunctional lysosomal α-synuclein processing and microglia-mediated neuroinflammation, which may initiate the disease process and contribute to spread of pathology along neural connectivity pathways.

Introduction

Parkinson’s disease is the second most common neurodegenerative disease (Kowal et al., 2013), causing devastating motor, neuropsychiatric and visuospatial dysfunction (Chaudhuri et al., 2006). Numerous genetic factors have been linked to both hereditary and sporadic forms of Parkinson’s disease, but the role of genetics in initiating region-specific pathology and controlling spread of disease across brain regions is poorly understood. Recent work has identified a role for genes involved in trans-synaptic transfer of α-synuclein in influencing the regional pattern of
brain atrophy (Freeze et al., 2018) but because of experimental limitations in humans, most work in this area has been limited to animal models (Mao et al., 2016). A better understanding of how genetic factors contribute to the region-specific risk of developing Parkinson’s disease pathology is crucial to identify promising genetic pathways for therapeutic intervention.

In humans, MRI-derived regional brain atrophy is an experimentally accessible measure of local pathology in vivo. Both cortical and subcortical brain regions have been shown to exhibit atrophy in Parkinson’s disease (Tinaz et al., 2011; Jia et al., 2015; Yau et al., 2018), including basal ganglia regions that contribute to cardinal motor symptoms. Multiple lines of evidence have suggested that pathology propagates from a small number of early disease sites along neural connectivity pathways (Desplats et al., 2009; Luk et al., 2012), ultimately affecting most of the brain. Previous work using network diffusion modelling (NDM) has shown that this process can be described by considering the spread of pathology as a diffusion process on the structural connectome (Raj et al., 2012; Pandya et al., 2019), with grey matter atrophy serving as a measure of local pathology. By seeding NDM pathology at every possible brain region, and comparing NDM predictions of atrophy with the true atrophy map, the regional origin of pathology can be inferred (Torok et al., 2018; Pandya et al., 2019). In Parkinson’s disease, NDM predicts that likely seed regions are located in the midbrain and other subcortical regions, with substantia nigra representing the most likely seed region (Pandya et al., 2019).

Here we ask whether regional transcriptional profiles of Parkinson’s disease risk factor genes can predict the pathology seed region and pattern of spread derived from the atrophy map and NDM. Genes that predict the regional origin of pathology may represent therapeutic targets that are particularly important in the early disease process and may therefore be more likely to significantly affect disease progression. We find that transcriptional profiles of immune-related and lysosomal Parkinson’s disease risk factor genes predict the pathology seed region, as well as the pattern of early disease spread. By exploiting recent advances in transcriptomics (Kelley et al., 2018), we also show that regional microglial abundance predicts the regional origin of pathology, suggesting that vulnerable regions may be susceptible to microglial-mediated pathology in the early disease state.

### Materials and methods

#### Parkinson’s disease regional atrophy

The regional Parkinson’s disease atrophy map was calculated as in Zeighami et al. (2015). Briefly, high resolution $T_1$-weighted MRIs were obtained from the prospective multicentre Parkinson’s Progression Markers Initiative (PPMI, ClinicalTrials.gov identifier: NCT0114102; Marek et al., 2011) for healthy control subjects ($n=117$ total; $n=74$ male, $n=43$ female; mean age 59.7 years) and de novo Parkinson’s disease subjects ($n=232$ total; $n=155$ male, $n=77$ female; mean age 61.2 years) enrolled between 2010 and 2013. Informed consent and Institutional Review Board approval were obtained, and HIPAA (Health Insurance Portability and Accountability Act) compliance was maintained. Additional study information is available at www.ppmi-info.org. Images were pre-processed, which included normalization, correction of intensity non-uniformity and denoising. Following linear and non-linear registration to the MNICBM 152 template, regional atrophy was calculated for each region in a 112 region parcellation using deformation-based morphometry (Ashburner et al., 1998). Resultant atrophy was re-scaled with regional atrophy represented as a z-score. Thirty-four cerebellar regions were excluded, and further analysis was performed on the remaining 78 cortical and subcortical regions.

#### Connectome construction

Structural connectivity was derived from high resolution diffusion-weighted MRI from the IIT Human Brain Atlas v.3 ($n=72$ healthy control subjects). Connectivity between two grey matter regions was defined as: anatomical connection density (ACD) – the fraction of superficial connected nodes with respect to total number of superficial nodes (Iturria-Medina et al., 2007). The connectivity matrix $C = [c_{ij}]$ is defined as all pairwise values of ACD such that $c_{ij}$ represents the connection strength of white matter fibre pathways between the $i^{th}$ and $j^{th}$ grey matter regions. This measure intrinsically corrects for variation in brain size as raw connection strength is divided by the sum of region-pair surface areas. Bidirectional connections are assumed (Pandya et al., 2019).

#### Network diffusion modelling

Parkinson’s disease progression is considered as a diffusion process on the connectivity matrix $C$ (Raj et al., 2012). A detailed description of Parkinson’s disease progression as predicted by NDM is available in Pandya et al. (2019). Briefly, the transmission of pathology from region 1 to region 2 is asserted to satisfy the equation

$$\frac{dx_1}{dt} = \beta c_{12} (x_2 - x_1)$$

where $x_1$ and $x_2$ are pathology concentrations in each region, and $\beta$ is a global diffusivity constant. Extending this relationship to all regions $i$ we define the regional pathology vector $x(t) = [x_i(t)]$, and the above equation become: $\frac{dx_i}{dt} = -\beta H x_i(t)$, where $H$ is the graph Laplacian $H = I - D^{-1}C D^{-1}$.

The regional atrophy vector $x$ at time point $t$ may then be adequately modelled by the equation $x(t) = e^{\gamma t} x_0$ where $x_0$ is the initial state of the model at $t=0$ (e.g. initial atrophy or pathological insult). As in Pandya et al. (2019), the NDM was separately initialized at all possible seed regions such that $x_0$ is 1 at each of the bilateral seed regions (e.g. left and right substantia nigra) and 0 at all other regions. The Pearson correlation coefficient (R) of the actual measured atrophy map with the NDM-predicted regional atrophy vector $x(t)$ was calculated for each timepoint $t$ and for each seed region. The
maximum $R$ occurring over all model time points $t$ for each seed region was determined and is considered a measure of seed region likelihood (SR). Following model initialization, we also calculate arrival time for each non-seed region, defined as the time $t$ at which 98% of the maximum theoretical $x(t)$ value in that region is achieved. As SR approaches zero for unlikely seed regions, the top 20 bilateral seed regions were considered for joint regional transcriptional analysis. Similarly, the top 20 arrival time regions were analysed to capture the pattern of early pathology spread.

**Regional transcriptional analysis**

Sixty-seven putative Parkinson’s disease risk factor genes from the recent Parkinson’s disease genome-wide association study (GWAS) meta-analysis by Chang et al. (2017) were identified for regional transcriptional analysis. For each gene, data were obtained from the publicly available human Allen Brain Atlas (ABA) (Hawrylycz et al., 2012). Human ABA data for the listed gene ‘CDC71’ could not be located. The ABA includes 926 brain regions, with each region having microarray expression levels from a set of 58,692 probes that correspond to 29,181 distinct genes. All ABA regions were mapped to the same 78 region atlas used for NDM. Semantic matching was used as the initial mapping strategy for all regions (i.e. ABA samples mapping to ‘hippocampus’ on the ABA atlas are mapped to ‘hippocampus’ in the 78 region parcellation used in this study). In cases for which semantic matching could not be used because of differences in region demarcation, atlases were visually compared, and ABA samples were mapped to the closest corresponding grey matter region in our 78 region parcellation. ABA samples spanning more than one region in our 78 region parcellation were excluded from analysis. All samples for all probes within the same region were then averaged for each gene. White matter tracts were excluded from analysis. Expression for each gene was averaged for six subject brains (which comprises data for six left hemispheres and two right hemispheres; more information can be found at help.brain-map.org/download/attachments/2818165/Normalization_WhitePaper.pdf). The top 10 high differential expressions were defined using the morgenstemning function (M. Geissbuelher).

**Gene functional classification**

Lyosomal, mitochondrial and autophagy-related genes were initially classified as described previously (Chang et al., 2017). Immune-related genes were initially classified according to the ImmPort database (Bhattacharyya et al., 2018). Gene classification was manually reviewed, and additional genes were added to the relevant functional class in cases in which literature evidence supports a role in the class.

**Hierarchical clustering of regional transcription**

Unweighted pair group method with arithmetic mean (UPGMA) agglomerative hierarchical clustering of expression data across genes and regions was performed using the MATLAB routine clustergram (Bar-Joseph et al., 2001). The Euclidean distance metric was used. Colour maps were defined using the morgenstemning function (M. Geissbuelher).

**L_1 regularized regression analysis**

$L_1$ regularized regression was performed using the MATLAB routine lasso (Tibshirani, 2011). Five-fold cross-validation was used in all cases. Display of regression coefficient and mean squared error (MSE) curves was performed using the routine lassoplot.

**Permutation analysis**

Genetic expression values were permuted across brain regions $10^3$ times for each tested gene set as described in the ‘Results’ section. Correlation coefficients for average permuted gene class expression and either SR or arrival time were then calculated. Cumulative probability curves were created using the MATLAB routine cdfplot.

**Data availability**

All data used in this study will be made available upon reasonable request.

**Results**

**Network diffusion modelling of Parkinson’s disease pathology**

Previous work used NDM as a mechanistic framework to dissect Parkinson’s disease pathogenesis and network propagation of pathology. In this context, brain region-specific pathology is considered to be indirectly but adequately measurable from the far more accessible surrogate measure of a deformation based morphometry (DBM) atrophy map, which is derived from T1-weighted volumetric MRI studies from 232 PPMI Parkinson’s disease and 117 healthy control subjects (Zeighami et al., 2015). Spread of pathology occurs via neural pathways, which are defined by diffusion MRI-derived structural connectivity. As in a recent article by Pandya et al., NDM was employed using each region in the atlas as an initial pathology seed and the likelihood of each seed was quantified by the correlation of the NDM regional atrophy prediction with the true atrophy map (Pandya et al., 2019). As every atlas region is simulated as a seed region, this method makes no a priori assumptions about the importance of seeds; and the correlation of the NDM prediction with the atrophy map (referred to as seed region likelihood) thereby represents an unbiased estimate of the propensity of each seed to explain the atrophy map.

Using this approach, the substantia nigra was determined to be the most likely seed region (SR = 0.65). Figure 1 illustrates the process by which the NDM prediction for substantia nigra seeding was derived and the similarity of
the substantia nigra-seeded NDM prediction to the atrophy map, which demonstrates extensive involvement of subcortical brain regions. The top 20 seed regions with the most accurate atrophy predictions as quantified by SR are shown in Table 1. Top seed regions tend to be very atrophic, suggesting that the role as an early disease site also allows for time-dependent accumulation of cellular pathology, dysfunction, and atrophy. However, there are regions such as putamen, which despite being more atrophic than subthalamic nucleus (atrophy z-score of 2.8 versus 2.0), is a much poorer seed region (SR 0.30 versus 0.53), which demonstrates the crucial role of interregional connectivity in determining SR. The spread of pathology from the most likely substantia nigra seed to subsequent regions can be quantified by arrival time, which is defined as the time to achieve 98% of the maximum atrophy. The top 20 regions with the smallest arrival times from the substantia nigra seed are also shown (Table 2). In accordance with actual atrophy progression, NDM predicts that early pathology spreads from the substantia nigra to connected subcortical regions in the basal ganglia and limbic system.

### Increased transcription of Parkinson’s disease risk factor genes in early disease sites

The set of putative Parkinson’s disease risk factor genes from the recent Chang et al. (2017) GWAS meta-analysis was mapped to the same atlas as that used for network diffusion analysis. Hierarchical clustering was performed across all genes (n = 67) and brain regions (n = 39 bilateral

<table>
<thead>
<tr>
<th>Region</th>
<th>SR (R)</th>
<th>Atrophy (z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substantia nigra</td>
<td>0.65</td>
<td>4.1</td>
</tr>
<tr>
<td>Red nucleus</td>
<td>0.58</td>
<td>3.0</td>
</tr>
<tr>
<td>Amygdala</td>
<td>0.56</td>
<td>3.2</td>
</tr>
<tr>
<td>Subthalamic nucleus</td>
<td>0.53</td>
<td>2.0</td>
</tr>
<tr>
<td>Parahippocampal gyrus</td>
<td>0.44</td>
<td>1.9</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>0.44</td>
<td>1.5</td>
</tr>
<tr>
<td>Anterior middle temporal lobe</td>
<td>0.37</td>
<td>1.4</td>
</tr>
<tr>
<td>Lateral occipitotemporal gyrus</td>
<td>0.37</td>
<td>1.0</td>
</tr>
<tr>
<td>Anterior inferolateral temporal lobe</td>
<td>0.35</td>
<td>1.0</td>
</tr>
<tr>
<td>Anterior superior temporal gyrus</td>
<td>0.34</td>
<td>1.0</td>
</tr>
<tr>
<td>Inferior middle temporal gyrus</td>
<td>0.32</td>
<td>1.0</td>
</tr>
<tr>
<td>Putamen</td>
<td>0.30</td>
<td>2.8</td>
</tr>
<tr>
<td>Thalamus</td>
<td>0.30</td>
<td>0.6</td>
</tr>
<tr>
<td>Pallidum</td>
<td>0.28</td>
<td>1.1</td>
</tr>
<tr>
<td>Superior central temporal gyrus</td>
<td>0.27</td>
<td>0.8</td>
</tr>
<tr>
<td>Insula</td>
<td>0.24</td>
<td>0.5</td>
</tr>
<tr>
<td>Nucleus accumbens</td>
<td>0.24</td>
<td>–0.4</td>
</tr>
<tr>
<td>Subcallosal area</td>
<td>0.24</td>
<td>–2.6</td>
</tr>
<tr>
<td>Anterior orbital gyrus</td>
<td>0.24</td>
<td>2.1</td>
</tr>
<tr>
<td>Caudate nucleus</td>
<td>0.24</td>
<td>0.7</td>
</tr>
</tbody>
</table>

R is shown for NDM seeded at the top 20 bilateral regions. Average atrophy (z-score) is also shown for the top 20 seed regions.
averaged regions) to determine whether discrete patterns of transcript abundance are present. The resultant dendrogram is shown in Fig. 2A. A striking pattern of differential expression of Parkinson’s disease genes is observed, with largely reversed patterns of transcription between primarily subcortical regions (left) and cortical regions (right). At the top level of region clustering (two clusters), one cluster contains only five brain regions, while the other contains 34 regions. Despite containing only five regions, the first cluster is composed of subcortical regions with SR rank 1 (substantia nigra), 2 (red nucleus) and 4 (subthalamic nucleus), suggesting that there is a relationship between expression of Parkinson’s disease genes and likelihood of a region representing an initiation site for pathology as inferred by NDM. A similar observation is made when considering the arrival time of pathology to other regions from the top ranked substantia nigra seed. The same five-region cluster is composed of the substantia nigra seed itself, and regions with arrival time rank 1 (red nucleus), 2 (subthalamic nucleus) and 4 (pallidum). The remaining 34-region cluster contains only two of the top five NDM seeds and two of the top five arrival time regions, despite containing more than 6-fold more brain regions. This regional distribution of top seed regions and top arrival time regions is unlikely to be due to chance (Fisher’s exact test, $P = 0.01$ in both cases). Further inspection of the larger 34 region cluster reveals that it is composed of a five-region cluster including the limbic system structures amygdala (seed rank 3) and hippocampus (arrival time rank 5). The larger 29-region cluster primarily contains unranked cortical structures, with the exception of the parahippocampal gyrus (seed rank 5, arrival time rank 3). As expected from visual inspection of the dendrogram, top five seed regions exhibit significantly more upregulated and downregulated genes than all other regions, with relatively more upregulated genes (Fig. 2B).

### Expression of immune-related and lysosomal gene classes predicts pathology seed region

Consistent with the overall pattern of gene expression visualized in Fig. 2A, average expression across all 67 Parkinson’s disease-related genes is positively correlated with SR ($r = 0.36$, $P = 0.02$; Fig. 3A). As a control, we performed a similar analysis using the expression of the top 10 high differential stability genes (Hawrylycz et al., 2015). These genes have differential regional expression patterns that are highly conserved across individuals. In contrast to the Parkinson’s disease risk factor genes, average expression of high differential stability genes is not significantly correlated with SR, demonstrating that this effect is not a general feature of transcription in high likelihood seed regions ($r = -0.30$, $P = 0.059$, Supplementary Fig. 1).

The set of Parkinson’s disease risk factor genes was subsequently subdivided according to functional classes implicated in Parkinson’s disease: immune-related, lysosomal, autophagy-related, and mitochondrial. Average expression of immune-related genes is highly correlated with SR ($r = 0.72$, $P = 3.5 \times 10^{-4}$), while average lysosomal gene expression demonstrates correlation similar to that of the Parkinson’s disease gene set as a whole ($r = 0.40$, $P = 0.01$). Average expression of autophagy-related and mitochondrial genes was not correlated with SR (Fig. 3D and E). Interestingly, expression of SNCA, which encodes $\alpha$-synuclein, is also uncorrelated with SR ($r = -0.14$, $P = 0.38$, Supplementary Fig. 2).

To determine whether the observed significant correlation coefficients were likely to be observed because of random structure within the set of gene expression values, we performed $10^4$ permutations of expression values across brain regions and calculated the resultant correlation coefficients for SR and average permuted gene expression. As shown in Fig. 3F, the actual correlation coefficients obtained for the entire gene set, and the lysosomal and immune-related gene subsets, are located in the rightmost region of the cumulative probability curve, distant from the mean $r$ of $\sim0$ for permuted datasets. Together, these results suggest that likely seed regions are enriched in transcripts for Parkinson’s disease risk factor genes, and that immune-related and lysosomal genes are specifically related to this effect.

### Table 2: Top 20 arrival time regions for NDM seeded at the top seed region, substantia nigra

<table>
<thead>
<tr>
<th>Region</th>
<th>Arrival time, substantia nigra (AU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red nucleus, R</td>
<td>10.1</td>
</tr>
<tr>
<td>Red nucleus, L</td>
<td>12.1</td>
</tr>
<tr>
<td>Subthalamic nucleus, L</td>
<td>12.1</td>
</tr>
<tr>
<td>Subthalamic nucleus, R</td>
<td>12.1</td>
</tr>
<tr>
<td>Parahippocampal gyrus, R</td>
<td>14.1</td>
</tr>
<tr>
<td>Pallidum, R</td>
<td>14.1</td>
</tr>
<tr>
<td>Hippocampus, R</td>
<td>16.1</td>
</tr>
<tr>
<td>Parahippocampal gyrus, L</td>
<td>18.1</td>
</tr>
<tr>
<td>Thalamus, L</td>
<td>18.1</td>
</tr>
<tr>
<td>Hippocampus, L</td>
<td>20.2</td>
</tr>
<tr>
<td>Amygdala, L</td>
<td>20.2</td>
</tr>
<tr>
<td>Thalamus, R</td>
<td>20.2</td>
</tr>
<tr>
<td>Nucleus accumbens, R</td>
<td>24.2</td>
</tr>
<tr>
<td>Putamen, R</td>
<td>24.2</td>
</tr>
<tr>
<td>Pallidum, L</td>
<td>24.2</td>
</tr>
<tr>
<td>Amygdala, R</td>
<td>28.2</td>
</tr>
<tr>
<td>Caudate nucleus, R</td>
<td>28.2</td>
</tr>
<tr>
<td>Nucleus accumbens, L</td>
<td>28.2</td>
</tr>
<tr>
<td>Subcallosal area, R</td>
<td>28.2</td>
</tr>
<tr>
<td>Caudate nucleus, L</td>
<td>32.3</td>
</tr>
</tbody>
</table>

Smaller arrival times reflect earlier involvement in Parkinson disease pathology. AU = arbitrary units; L = left; R = right.
Transcription of specific lysosomal and immune-related genes predicts pathology seed region

To determine whether transcriptional profiles of individual lysosomal genes predict SR we performed cross-validated L₁ regularized regression with expression profiles of the classified lysosomal genes as independent variables and SR as the dependent variable. At the value of the tuning parameter λ commonly used to choose a sparse regression model [i.e. that which achieves a MSE one standard error larger than that for the minimum MSE (Hastie, 2015)], coefficients for GBA and TMEM175 expression profiles remain non-zero, indicating that these are the most important predictor variables (Fig. 3G). The remaining expression
Figure 3 Prediction of seed region by immune-related and lysosomal gene transcription. (A–E) Scatterplots of seed region likelihood (R) and average regional expression (n = top 20 bilateral seed regions) across gene classes. Expression across all Parkinson’s disease-risk factor genes (n = 67) (A), as well as immune-related (n = 8) (B) and lysosomal (n = 5) (C) subsets is positively correlated with SR. Autophagy-related gene expression (n = 3) (D) and mitochondrial gene (n = 5) (E) expression is not significantly correlated with SR. (F) Cumulative probability distributions of correlation coefficients for expression and SR (Seed R scrambled) obtained by permutation of expression values across regions (n = 10⁴ permutations per tested gene class). (G and J) Cross-validated L₁ regularized regression coefficient curves for individual lysosomal (G) and immune-related (J) gene expression plotted against the logarithm of the tuning parameter λ. Genes that maintain non-zero coefficients at the value of λ one standard error greater than the value of λ that achieves minimum cross-validation MSE are considered important predictor variables for SR. Lysosomal genes GBA and TMEM175; and immune-related genes HLA-DQA1, IL1R2 and HLA-DRB6 are important predictor variables. (H and K) Mean squared error (MSE) curves corresponding to coefficient curves in (G) and (J), respectively. (I and L) Cumulative probability curves for regional expression of important lysosomal (I) and immune-related (L) predictor genes. Each gene is highly expressed in the top seed region, substantia nigra, as indicated on the curve.
profile coefficients for GALC, CTSB, and ATP6V0A1 are zero at this value of λ. Analysis of the distribution of these genes across all brain regions shows that GBA and TMEM175 are highly expressed in substantia nigra (Fig. 3I), compatible with a causal role for lysosomal dysfunction in Parkinson’s disease pathogenesis in the most likely seed region. We performed a similar L1 regularized regression analysis for the set of immune-related genes. Notably, transcriptional profiles of two major histocompatibility complex class II genes, HLA-DQA1 and HLA-DRB6, are important predictors of pathology seed region. The other important predictor gene is IL1R2, a complex regulator of IL1 signalling (Fig. 3J). These genes are also highly expressed in substantia nigra (Fig. 3L).

Enrichment in risk factor transcripts is uncorrelated with the overall atrophy pattern

Although SR is correlated with regional atrophy ($r = 0.70$, $P = 4.0 \times 10^{-7}$, Fig. 4A), average expression across all Parkinson’s disease genes is not correlated with the magnitude of atrophy across top seed regions (Fig. 4B, $n = 40$ regions for top 20 bilateral NDM seeds, $r = 0.08$, $P = 0.61$) or across all regions in the full brain parcellation (Fig. 4C, $n = 78$ regions, $r = 0.07$, $P = 0.56$). This result suggests that enrichment in Parkinson’s disease-risk factor transcripts is more related to disease initiation than control of the overall regional atrophy pattern.

Transcription of immune-related and lysosomal genes predicts the early pattern of spread from the substantia nigra

We next sought to determine whether Parkinson’s disease risk factor transcription also predicts the pattern of pathology spread from substantia nigra, the most likely seed region. To this end, we calculated the correlation between regional average transcript abundance and NDM-predicted arrival time from substantia nigra (defined as the model time needed to achieve 98% of the maximum theoretical x(t) at that region, after seeding $x_0$ at the substantia nigra only) across regions. Given that pathology originates in the substantia nigra in this model, the substantia nigra itself has arrival time 0 by definition. Arrival times for subsequently affected brain regions become progressively greater as regions achieve near maximal $x(t)$. Average transcription of all genes is negatively correlated with arrival time (Fig. 5A, $r = -0.62$, $P = 3.5 \times 10^{-3}$, $n = 20$ smallest arrival time regions, excluding substantia nigra itself) suggesting that regions most likely to be affected early in the disease process are enriched in Parkinson’s disease risk factor transcripts. Expression of immune-related and lysosomal genes is negatively correlated with arrival time ($r = -0.73$, $P = 2.6 \times 10^{-4}$; and $r = -0.59$, $P = 5.7 \times 10^{-3}$, respectively), while expression of autophagy-related and mitochondrial genes is not significantly correlated with arrival time (Fig. 5B–E). L1 regularized regression analysis was again used to determine whether important lysosomal and immune-related genetic predictors of the spread pattern from substantia nigra overlap with those predictive of the pathology seed region. Indeed, all of the genes identified as predictive of the seed region were also identified as predictive of the early spread pattern as quantified by arrival time. Other genes were also identified as important predictors, including the lysosomal gene ATP6V0A1 and multiple immune-related genes including CRHR1 and MAPT (which encodes tau) (Fig. 5G–J).

Microglial abundance predicts seed region location

We observed that several of the immune-related genes identified as important predictors of the seed region and early

Figure 4 Regional transcription of Parkinson’s disease risk factor genes does not predict the overall atrophy pattern. (A) Scatterplot of SR and regional atrophy ($n = $ top 20 bilateral seed regions), showing positive correlation. (B) Scatterplot of regional atrophy and average Parkinson’s disease risk factor gene expression for top seed regions, demonstrating absence of significant correlation. (C) Regional atrophy and average Parkinson’s disease risk factor gene expression are also uncorrelated across all regions, including top seed and non-top seed regions ($n = 78$).
Figure 5 Prediction of spread pattern by regional transcription of immune-related and lysosomal genes. (A–E) Scatterplots of NDM arrival time from top seed region substantia nigra and average regional expression (n = top 20 bilateral seed regions) across gene classes. Expression across all Parkinson’s disease-risk factor genes (n = 67) (A), as well as immune-related (n = 8) (B) and lysosomal (n = 5) (C) subsets is positively correlated with arrival time. Autophagy-related gene expression (n = 3) (D) and mitochondrial gene (n = 5) (E) expression is not significantly correlated with arrival time. (F) Cumulative probability distributions of correlation coefficients for expression and arrival time (Arrival time \( R_{\text{scrambled}} \)) obtained by permutation of expression values across regions (n = 10^4 permutations per tested gene class). (G and I) Cross-validated L1 regularized regression coefficient curves for individual lysosomal (G) and immune-related (I) gene expression plotted against the logarithm of the tuning parameter \( \lambda \). Important predictor variables have non-zero coefficients at \( \lambda \) one standard error from that achieving minimum MSE. (H and J) Mean squared error (MSE) curves corresponding to coefficient curves in (G) and (I), respectively.
pattern of spread are selectively expressed by microglia, such as HLA-DQA1. Recent work has identified genes with regional transcriptional levels highly predictive of the regional abundance of the cell type. The first principal component of regional expression of even a small number of such ‘high-fidelity genes’ is nearly perfectly correlated ($r > 0.99$) with regional cell type abundance (Kelley et al., 2018). To determine whether regional microglial abundance predicts pathology seed region and pattern of spread, we mapped expression of the top five high fidelity microglial genes to the same 78 region atlas used before. Similar to our analysis of all Parkinson’s disease risk factor genes,
hierarchical clustering was applied to regional expression of the top five high fidelity microglial genes. At the top level of clustering (two clusters), a four-region cluster was identified that contained the regions with highest expression of the microglial high-fidelity genes (Fig. 6A). This region contained the pallidum, substantia nigra, subthalamic nucleus, and red nucleus. Three of the top five seed regions and three of the top five arrival time regions were represented in this cluster. The remaining 35-region cluster contained only two each of the top seed and arrival time regions. Enrichment of top five seed regions and arrival time regions is unlikely to be due to chance (Fisher exact test, $P = 4.2 \times 10^{-3}$).

Given the correspondence between high fidelity gene expression and cell type abundance, we computed the first principal component (PC1) of high fidelity gene expression as a proxy measure of regional microglial abundance (Kelley et al., 2018). Interestingly, microglial abundance was significantly correlated with SR (Fig. 6B, $r = 0.38$, $P = 0.016$), suggesting that regions enriched in microglia at baseline may be more susceptible to neuroinflammatory processes. Average expression was equally correlated with SR ($r = 0.38$, $P = 0.016$), with similarity between average expression and PC1 driven by highly correlated expression across these genes. A trend towards a negative correlation between microglial abundance and arrival time from the substantia nigra seed was not statistically significant ($r = -0.34$, $P = 0.14$). As a control, we calculated the correlation between PC1 of the top five neuronal high fidelity genes (SNF, SNAP25, TRIM37, APT2A2, and MAPK9) and SR. There was no correlation between these measures (Fig. 6C, $r = -0.02$, $P = 0.89$), suggesting that neuronal abundance is unrelated to Parkinson’s disease pathology initiation. Similarly, average neuronal high fidelity average expression was also uncorrelated ($r = -0.03$, $P = 0.86$).

**Discussion**

Despite great effort investigating the genetic features of Parkinson’s disease, the role of risk factor genes in determining sites of Parkinson’s disease pathogenesis and patterns of spread remains largely unknown in humans. Here we have uncovered some of the molecular mediators of Parkinson’s disease seeding propensity and quantity, for the first time, the contribution of innate molecular factors in mediating the regional vulnerability of certain brain regions to act as pathology seeding loci. By combining state of the art neuroimaging, computational modelling and transcriptional analysis we show that likely Parkinson’s disease initiation sites are enriched in risk factor transcripts, specifically related to immune and lysosomal function. High levels of expression in seed regions may make them more susceptible to the effects of genetic variation, thereby explaining the increased risk for developing Parkinson’s disease observed in GWAS studies. As we specifically examined genetic expression in healthy subjects, this suggests that there is a pattern of regional genetic vulnerability which promotes early Parkinson’s disease pathology in the presence of an inciting event.

Transcriptional profiles of lysosomal genes GBA and TMEM175 are important predictors of SR and pattern of spread, with high levels of expression in top seed regions such as substantia nigra. GBA encodes the lysosomal protein glucocerebrosidase, which has been implicated in Parkinson’s disease pathogenesis by multiple experimental approaches in addition to GWAS. Both levels and activity of glucocerebrosidase are decreased in Parkinson’s disease patients, with the most pronounced effects occurring in substantia nigra (Gegg et al., 2012). Loss of function mutations affecting GBA enzymatic activity interfere with lysosomal protein degradation, leading to $\alpha$-synuclein accumulation and neurodegeneration in animal models (Mazzulli et al., 2011). Decreased levels of GBA also enhance intercellular transmission of $\alpha$-synuclein (Bae et al., 2014), suggesting that transcription of GBA can contribute to both initiation and spread of Parkinson’s disease pathology.

TMEM175 is a lysosomal K$^+$ channel, serving as the primary lysosomal K$^+$ conductance, with major roles in maintaining the lysosomal pH gradient and regulating lysosomal fusion with other organelles (Cang et al., 2015). Mutations of TMEM175 have been shown to decrease glucocerebrosidase activity by destabilizing lysosomal pH, leading to $\alpha$-synuclein accumulation (Jinn et al., 2017). GBA and TMEM175 therefore represent two lysosomal proteins in the same potential mechanistic pathway, positioned to modulate early $\alpha$-synuclein-related pathology. Interestingly, SNCA expression itself does not predict SR or regional atrophy, in accordance with other results (Freeze et al., 2018). This may suggest that levels of $\alpha$-synuclein are less important than pathways that promote the stochastic production of even small amounts of pathologic species, which may then increase in concentration by prion-like templating and aggregation mechanisms (Iljina et al., 2016), possibly involving dysfunctional lysosomal processing.

Increased expression of Parkinson’s disease immune-related genes also predicts SR and spread pattern. One of these genes, IL1R2, is a complex regulator of IL1-mediated inflammation (McMahan et al., 1991; Lang et al., 1998). Intriguingly, two MHCII genes also predict SR, including HLA-DQA1, which plays an important role in antigen presentation. As microglia are the primary antigen presenting cells in the brain (Hickey and Kimura, 1988), we sought to determine whether microglia are enriched in likely seed regions. Indeed, microglial high fidelity gene expression shows increased microglia abundance in likely seed regions. This finding is concordant with neuropathological studies in normal subjects showing larger numbers of microglia in areas related to Parkinson’s disease pathology. For instance in humans, substantia nigra, putamen and pallidum contain ~2-fold more microglia than frontal grey matter (Mittelbronn et al., 2001). The relative...
distribution of microglia is similar in mice, suggesting that anatomic variability in microglial density may be conserved across species. As in humans, substantia nigra, putamen and pallidum also contain ~2-fold more microglia than cortical grey matter (Lawson et al., 1990). Although regional enrichment of microglia in normal brains does not prove that such regions are more susceptible to microglia mediated pathology in Parkinson’s disease, there is correspondence between areas of high microglial density in Parkinson’s disease patients. For example, histopathological analysis has shown that the substantia nigra in Parkinson’s disease subjects contains 4-fold more activated microglia than cortical regions (Imamura et al., 2003). Improved imaging of different microglial species in vivo may provide insight into the relationship between baseline numbers of microglia and activated subtypes in Parkinson’s disease. Further work on high fidelity gene expression may also allow resolution of these subtypes in Parkinson’s disease patients by transcriptomic methods, an area of active investigation.

It is possible that microglia-mediated pathology may require additional disease initiation events, such as accumulation of pathological α-synuclein species. Microglia have been shown to interact with α-synuclein in numerous studies. For instance, α-synuclein acts as a microglial chemoattractant (Wang et al., 2015) and α-synuclein aggregates trigger a microglial pro-inflammatory response that causes neuronal toxicity (Zhang et al., 2005). Together our findings suggest a model in which early pathology sites are vulnerable to dysfunctional lysosomal α-synuclein processing as well as microglia mediated neuroinflammation that may directly cause neuronal damage and also promote α-synuclein spread, similar to the spread of tau in Alzheimer disease (Asai et al., 2015).

We emphasize that our findings primarily relate to disease initiation and patterning of early disease spread, with connectivity patterns playing a critical role in determining the overall atrophy map as disease progresses throughout the brain. Other local factors may also play an important role in determining the regional pattern of disease, although findings from our own group suggest that trans-synaptic models of disease spread are more predictive than purely distance-based spread models (Pandya et al. 2019). As posited by other groups, there may be cell-type specific vulnerability to α-synuclein pathology, or other types of pathological insult mediated by the immune system. Downstream effects from cell death in connected regions may also be relevant. For example, the striatum is affected by substantia nigra pathology not only due to transmission of pathological α-synuclein, but also by dopamine deprivation itself.

There are several limitations of this study that warrant discussion. We use MRI-derived atrophy as a marker of regional neurodegeneration. This provides a measure of local effective pathology, as regional atrophy largely reflects neurodegeneration in the form of cellular atrophy and death. For instance, substantia nigra dopaminergic cell death has been shown to correlate with disease duration and progression of neurodegeneration (Damier et al., 1999). In mice and non-human primates, substantia nigra injection of human Lewy body extracts induces progressive death of substantia nigra dopaminergic neurons and concomitant reduction in striatal dopaminergic terminals (Recasens et al., 2014), suggesting that there is a close relationship between cell death and α-synuclein pathology. Indeed, highly atrophic top seed regions in this study, such as substantia nigra and amygdala, have been shown to exhibit Lewy pathology in all patients in a cohort studied by Jellinger (2003). However, we acknowledge that there are regions in which Lewy pathology can frequently occur without marked atrophy on MRI. For instance, cingulate cortex exhibits Lewy pathology in 34% of patients in the same study, and yet is not significantly atrophic in our dataset. Certain brain regions may be more resistant to α-synuclein pathology, possibly due to cell-type specific mechanisms. Further research in this area is likely to provide important insight into additional mechanisms that control the regional pattern of disease.

We used a DBM approach to quantifying regional atrophy on a 78-region parcellation of the brain. Although some other commonly used parcellations contain more brain regions, we chose to sacrifice some spatial resolution for improved signal to noise ratio in the atrophy map, which was derived from a large PPMI Parkinson’s disease subject cohort. The high precision of the atrophy map is relevant, as conflicting results in other studies may be in part due to lack of statistical power and precision. For example, while substantia nigra volumetric change in Parkinson’s disease has been analysed in several other studies, there have been discrepant results. While some groups have reported volume loss, others have reported no change (Pérant et al., 2010) or even increased volume of the substantia nigra (Kwon et al., 2012). Our study is notable for containing many more subjects (232) than the studies by Péran et al. (30 subjects) and Kwon et al. (10 subjects); suggesting that this may be an important consideration in addition to other methodological differences.

Although DBM has been shown to be more sensitive to subcortical atrophy in other work (Scanlon et al., 2011), it remains difficult to determine atrophy in the brainstem and certain other brain regions due to technical limitations of MRI including lack of spatial resolution and tissue contrast. Such regions include olfactory cortex and brainstem structures such as the dorsal motor nucleus of the vagus nerve and locus coeruleus, which have been posited to serve as early pathology sites in Parkinson’s disease (Braak et al., 2003). Olfactory cortex and the dorsal motor nucleus of the vagus nerve specifically have also been suggested to act as accumulation sites for extra-CNS α-synuclein transmitted by cranial nerve I and visceral afferents. While our findings cannot address these claims, in the future these structures may be more amenable to analysis with improved atlas construction (Bianciardi et al., 2015) and T2* imaging at high field strength. Even if there are additional early sites of disease, we suggest that the top seed...
regions in this study likely still serve as important mediators of the regional atrophy pattern; possibly by accumulating subthreshold levels of pathology from other connected early disease sites.

In this study we used gene expression data derived from a relatively small number of healthy control subjects. The creation of the Allen Brain Atlas was extremely resource intensive, and a similar dedicated gene expression atlas for Parkinson’s disease subjects is not yet available. There may be differences in Parkinson’s disease and healthy control regional transcription that we cannot address with our current approach, but this will likely be a fruitful area of future research. We also note that our analysis of genetic factors is limited to transcript abundance. Although this is an informative measure, protein levels may not scale linearly with transcript abundance. Importantly, protein levels and activity are known to often be more relevant to cellular function. Additional factors such as protein folding, post-translational modification, and subcellular localization and trafficking, are also highly relevant to the ultimate impact of gene expression on cellular dynamics. These reasons may explain why autophagy-related and mitochondrial gene expression does not predict SR and pattern of spread, despite evidence for their importance from GWAS. Advances in proteomics may provide further insight into these issues, although most of this work is currently being performed in model organisms (Sharma et al., 2015).

In summary, we demonstrate that lysosomal and immune-related molecular pathways are likely critical contributors to Parkinson’s disease initiation and early disease propagation, and may be particularly relevant in designing disease-modifying treatments that are not reliant on dopamine replacement or ablation/DBS-based circuit modulation. Importantly, we are able to deduce these genetic effects from non-invasive MRI and computational analysis, which may facilitate future development of combined genetic/imaging biomarkers. Further work on techniques to image various α-synuclein species, and microglial distribution and activity, may provide further insight into early disease states.

Acknowledgements

PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson’s Research funding partners AbbVie, Avid Radiopharmaceuticals, Biogen Idec, BioLegend, Bristol-Myers Squibb, Eli Lilly & Co., F. Hoffman-La Roche, Ltd., GE Healthcare, Genentech, GlaxoSmithKline, Llandber, Merck, MesoScale Discovery, Piramal, Pfizer, Sanofi Genzyme, Servier, Takeda, Teva, and UCB.

Funding

Radiological Society of North America (RSNA) resident research grant RR1813 (B.F.).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at Brain online.

References


