

SCIENTIFIC REPORTS



OPEN

Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology

Received: 20 April 2016

Accepted: 04 July 2016

Published: 27 July 2016

Simon G. Coetzee¹, Steven Pierce², Patrik Brundin², Lena Brundin², Dennis J. Hazelett¹ & Gerhard A. Coetzee²

Recent genome-wide association studies (GWAS) of Parkinson's disease (PD) revealed at least 26 risk loci, with associated single nucleotide polymorphisms (SNPs) located in non-coding DNA having unknown functions in risk. In order to explore in which cell types these SNPs (and their correlated surrogates at $r^2 \geq 0.8$) could alter cellular function, we assessed their location overlap with histone modification regions that indicate transcription regulation in 77 diverse cell types. We found statistically significant enrichment of risk SNPs at 12 loci in active enhancers or promoters. We investigated 4 risk loci in depth that were most significantly enriched ($-\log_{10}P > 14$) and contained 8 putative enhancers in the different cell types. These enriched loci, along with eQTL associations, were unexpectedly present in non-neuronal cell types. These included lymphocytes, mesendoderm, liver- and fat-cells, indicating that cell types outside the brain are involved in the genetic predisposition to PD. Annotating regulatory risk regions within specific cell types may unravel new putative risk mechanisms and molecular pathways that contribute to PD development.

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting about 1% of those over the age of 60. Classically, PD was considered a movement disorder with akinesia, rigidity, tremor and postural instability as the predominant motor features. More recently, it became clear that patients also experience a wide range of disabling nonmotor disturbances, e.g. olfactory deficits, sleep disorders, depression, cognitive decline, constipation and other autonomic changes¹. The neuropathology is characterized by severe degeneration of substantia nigra dopaminergic neurons (crucial for the motor deficits) and progressive development of intraneuronal α -synuclein aggregates throughout the central and peripheral nervous systems². Monogenetic forms of PD are rare and only account for 5–10% of cases³ (see also longer reviews^{4–6}); a total of 7 autosomal and recessive genes have been identified, exhibiting varying degrees of penetrance. Thus, the vast majority of cases of PD are sporadic and a combination of environmental factors and complex genetic loci likely play a causative role⁴. In addition to the rare Mendelian-inherited cases of PD, the genetic predisposition to PD includes at least 26 common single nucleotide polymorphism (SNP) variants ("index SNPs"), each imposing low but significant risk^{7,8}. These risk loci associated with PD were discovered by population-based genome-wide association studies (GWAS), published in 2014 in a comprehensive large-scale meta-analysis of reproducible hits⁷ and conveniently summarized in a Cell snapshot³. Importantly, most (>90%) of the identified index SNPs in PD GWAS^{7,8} are located in noncoding DNA regions, making the assignment of potential functionality or causality—even the identification of the specific genes associated with the risk-SNPs—quite challenging.

Since 2005, over 1,600 significant SNP risk variants have been identified by GWAS for more than 250 traits, many with complex genetic predisposition and unknown gene involvements (<https://www.genome.gov/26525384>). Most genetic epidemiological studies have attempted to characterize the risk SNPs found in non-coding DNA simply by considering the "nearest gene" as the one being involved in risk. However, many index SNPs and their surrogates [in linkage disequilibrium (LD)] reside in regulatory regions (mainly in

¹Bioinformatics and Computational Biology Research Center, Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA. ²Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, Michigan, USA. Correspondence and requests for materials should be addressed to D.J.H. (email: dennis.hazelett@csmc.edu) or G.A.C. (email: gerry.coetzee@vai.org)

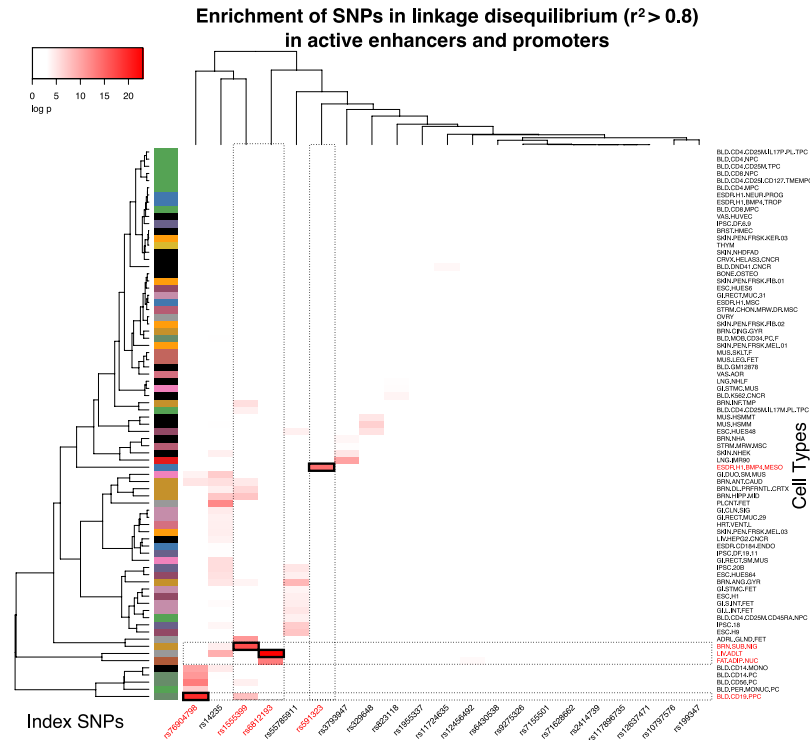


Figure 1. Tissue specific enrichment at PD GWAS loci. The relative enrichment of 21 index SNP (x axis) is shown by multiple hypothesis adjusted p -value (negative natural log) for each of the 77 REMC tissues (y axis). The data from Supplemental Table 1 are plotted; color from white to red indicates increasingly significant hits. SNPs and tissues selected for in-depth analysis are highlighted with red type on the axes and with guidelines on the plot. The specific SNP–tissue interactions that appear in subsequent figures are outlined in black.

enhancers) and the target genes of such regulation are most likely not the nearest genes⁹. Therefore, there is still a significant gap between the identification of most risk SNPs and an understanding of their biological function in human disease. Here, we have addressed this conundrum via the development and use of two Bioconductor software tools, FunciSNP and motifbreakR^{10,11}, to functionally annotate risk SNPs for PD. The approach allows the identification of risk SNP enrichment at active regulatory elements in non-coding DNA (promoters and enhancers, in this study). Enrichment is defined as the presence of risk SNPs nonrandomly distributed and at an increased prevalence within chromatin regions with putative functions. The two alleles of the SNP may each impose a differential risk by affecting the regulatory capacity of the region on target genes.

In the past, we and others have used bioinformatic tools, including FunciSNP and motifbreakR, to interrogate genetic studies in light of both public and in-house next generation sequencing data. Noncoding risk regions of many complex diseases have thus been investigated in recent years. However, in most cases, this was done only in one specific cell- or tissue type within databases. More recently, with the curation of epigenetic data within the Roadmap Epigenomics Mapping Consortium (REMC), the mapping of risk alleles in a variety of cell types¹² became possible. Thus, our present approach is based on the assessment of PD risk SNP enrichment at specific chromatin regulatory regions in multiple distinct cell- and tissue types. Twenty-one PD risk loci were assessed in 77 REMC cell types of diverse tissue- and lineage-origin. Eight enhancers at the four most significantly enriched PD risk loci were annotated in detail. They were found respectively in lymphocytes, mesendoderm, brain-, liver- and fat cells, allowing refined hypotheses to be formulated for the genetic predisposition to PD at these loci.

Results

We assessed 21 PD risk loci in 77 REMC tissues and cell types (supplemental Tables 1 and 2). One of the 77 REMC tissues was derived from human substantia nigra, which contains the dopaminergic neurons that undergo extensive degeneration in PD. The majority of the data in REMC are from tissues or cells that classically have not been considered to be affected by PD and have, so far, not been implicated in the disease development. We collected three data types. They were histone H3K4 mono and trimethylation (H3K4me1&3), and acetylation of H3K27 (H3K27ac). For the purposes of the present analyses, we restricted ourselves to active regulatory sequences in non-coding DNA (The data from Supplemental Table 1 are presented as a heat-map in Fig. 1). Regions with positive signals for the histone modification H3K4me3 were classified as promoters, and we considered them to be active if they overlapped with the active histone mark H3K27ac. On the other hand, we classified H3K4me1 regions as enhancers, and similarly we considered them to be active if there was overlap with H3K27ac. We compared enrichment of risk SNPs with all 1000 genomes SNPs within a window of 1 MB around each GWAS index SNP that we had found to occur within enhancer/promoter regions in the 77 REMC cell types. To this end, we counted correlated SNPs (at $r^2 \geq 0.8$; denoted as risk SNPs) at each risk locus in each of the 77 REMC cell types.

We statistically assessed them (risk SNPs in regulatory regions vs. non-risk 1000genomes SNPs in regulatory regions) for significant differences, using the hypergeometric test after locus-specific corrections for multiple hypotheses (Supplemental Table 1 and Fig. 1). A schematic presentation of the workflow is found in Supplemental Fig. 1.

Enrichment. We found enrichment of risk SNPs at 12 of the 21 loci, mainly at enhancers and not exclusively in nervous tissues, suggesting that the etiology of genetic risk is not restricted to the central nervous system. Based on very high significance ($-\log_{10}P > 14$) of enrichment, we chose to analyze eight putative enhancers in depth, as a proof-of-principle that our approach generates interesting and novel hypotheses about etiological and pathogenetic mechanisms of PD. The eight putative enhancers were at 4 most significantly enriched loci. They were at chromosome 4q21 (SNP enrichment in liver and fat cells, Fig. 2), chromosome 8p22 (SNP enrichment in BMP4 treated ES-derived mesendoderm, Fig. 3) and 12q12 (SNP enrichment in CD19-B lymphocytes, Fig. 4), and chromosome 14q24 (SNP enrichment in CD19-B lymphocytes and substantia nigra cells, Fig. 5).

Locus 4q21.1. SNP enrichment was found at active enhancers in both liver and fat cells, but not in substantia nigra cells at 4q21.1 (Fig. 2). In both active cases the regulatory regions are likely to be super-enhancers¹³ (a.k.a. stretch enhancers¹⁴) each spanning some 10 kb. One of the enriched SNPs disrupts topoisomerase I binding, an arginine/serine-rich, E3 ubiquitin protein ligase (TOPORS). TOPORS is a nuclear protein and contains a RING-type zinc finger domain. It functions as an ubiquitin-protein E3 ligase¹⁵ but its function as a DNA binding protein is unknown. The response element (binding motif) falls within an intron for five isoforms of FAM47E. FAM47E is a protein of unknown function, but it is likely involved in inflammation and within epithelial biology networks¹⁶.

Locus 8p22. Four putative active risk enhancers at chromosome 8p22 demonstrate risk-SNP response-element disruption in BMP4-treated ES-derived mesendoderm, but not in substantia nigra cells (Fig. 3A,B). Three of the enhancers (Fig. 3A) are either independent or form part of a super/stretch-enhancer^{13,14} spanning about 5 kb. Three SNPs (rs6991493, rs7001099, and rs62504203) disrupt response elements in DMRTA2, POU6F2, and FOXG1/IRX2,3&6, respectively. DMRTA2 (doublesex and mab-3 related transcription factor-like family A2) is mutated in novel cortical brain malformation¹⁷. POU6F2 (POU-domain homeobox transcription factor class 5) is associated with autism risk¹⁸. FOXG1 (forkhead box G1) is a transcription factor involved in the dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders¹⁹. IRX2,3&6 (Iroquois homeobox transcription factors) are involved in internal organ development such as the heart²⁰. The fourth SNP (rs1717289) right next to H3K27ac mark in BMP4 treated ES-derived mesendoderm, disrupts several transcription response elements including a CTCF site (Fig. 3B). The CTCF site may demarcate enhancer/promoter loops important in distal enhancer regulation. The transcription factors in question RXRA, BCL, PAX5, SP11, EP300, and MEF2 are all involved in cell type differentiation in some way or another. All four enhancers are located in a large intron of the gene RP11-13N12.1, which encodes for a lincRNA of unknown function.

Locus 12q12. A putative enhancer cluster at chromosome 12q12 is within an active enhancer region (H3K4me1 & H3K27ac) in CD19 B-lymphocytes cells but not brain cells. A strong promoter (H3K27ac) is also nearby. The enhancer is likely a super/stretch-enhancer^{13,14} spanning about 8 kb. A risk SNP (rs74324737) in the vicinity disrupts the response elements of four transcription factors with very similar response elements, namely TCF3, MYOD1, LMO2 and ZEB1 (Fig. 4). All four transcription factors are involved in tissue-specific signaling. The enhancer is within intron 2 of a multi-domain leucine-rich repeat kinase 2 (*LRRK2*) gene, which encodes a kinase regulating a subset of Rab GTPases and is known to give rise to autosomal dominant PD when mutated²¹.

Locus 14q24.1. Two very strong risk enhancers (as judged by the type, extent and sizes of the modified histone peaks) likely forming part of another super/stretch-enhancer^{13,14} (spanning about 5 kb), are located at chromosome 14q24.1 (Fig. 5). They are found specifically in substantia nigra and to a lesser extent in CD19 B-lymphocyte cells. Risk SNPs (rs2273596 and rs7160654) are respectively disrupting response elements for NHLH1 (Nescient helix-loop-helix) and NR3C1 (nuclear receptor subfamily 3, group C member 1). The NHLH1 transcription factor correlates with DNA methylation and gene expression in the brains of patients with bipolar disorder and schizophrenia²². NR3C1, also known as the glucocorticoid receptor, is involved in multiple aspects of the inflammatory response.

Expression Quantitative Trait Loci (eQTL) associations. Using data from the Genotype-Tissue Expression consortium (GTEx), we interrogated risk SNP genotypes and gene expression in different tissues. Since the tissue overlap between REMC and GTEx was only partial, risk polymorphisms at two loci (4q21.1 and 12q12) were significantly associated ($p < 2.5 \times 10^{-3}$) with gene expression within tissues where enhancer enrichment was observed. Thus we found that the risk alleles of rs6812193 and its surrogates were associated with decreased expression of CCDC158, FAM47E, NAAA and NUP54 in many tissues including liver and fat. The risk alleles of rs76904798 and its surrogates were associated with increased expression of LRRK2 in many tissues including blood.

Discussion

In this study we employed a novel approach to identify the potential functionality of risk-SNPs as identified by GWAS in PD. We first determined whether risk SNPs are enriched at regulatory regions in distinct cell types or tissues in the REMC database, followed by an assessment from known databases of whether and how risk SNPs disrupt response elements (transcription factor binding motifs). Although similar enrichment analyses have been employed in other studies (e.g., by Farh *et al.*¹²), this is the first time such an approach has been undertaken for

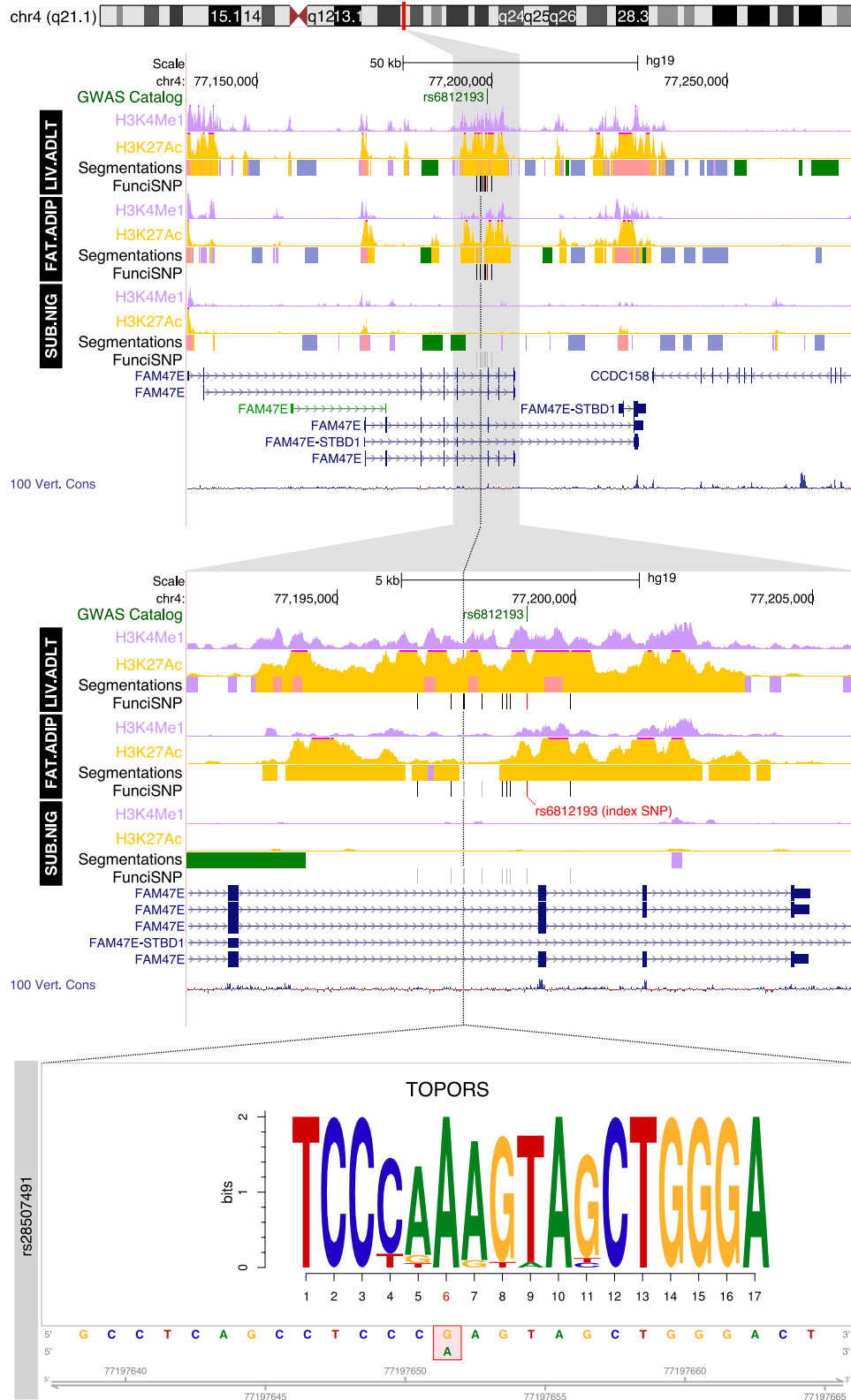


Figure 2. Functional enrichment at the 4q21 locus. Segmentation tracks summarize the functional annotations according to the REMC color scheme: orange for EAR (active enhancer), rose for PAR (active promoter), light purple for EPR and PPR (poised enhancers and promoters), grey for SCR (silenced regions), blue-grey for HET (heterochromatic) and green for TRS (transcribed). Wiggle tracks for H3K4me1 and H3K27Ac show the fine-grained epigenomics context in the regions surrounding each locus. The location of SNPs that disrupt transcription factor motifs are indicated with a dotted black line and an expanded view at the bottom. The motif is aligned with the genomic sequence surrounding the SNP, whose position is indicated in red; reference and alternate alleles are indicated in the red bounding box in the sequence below.

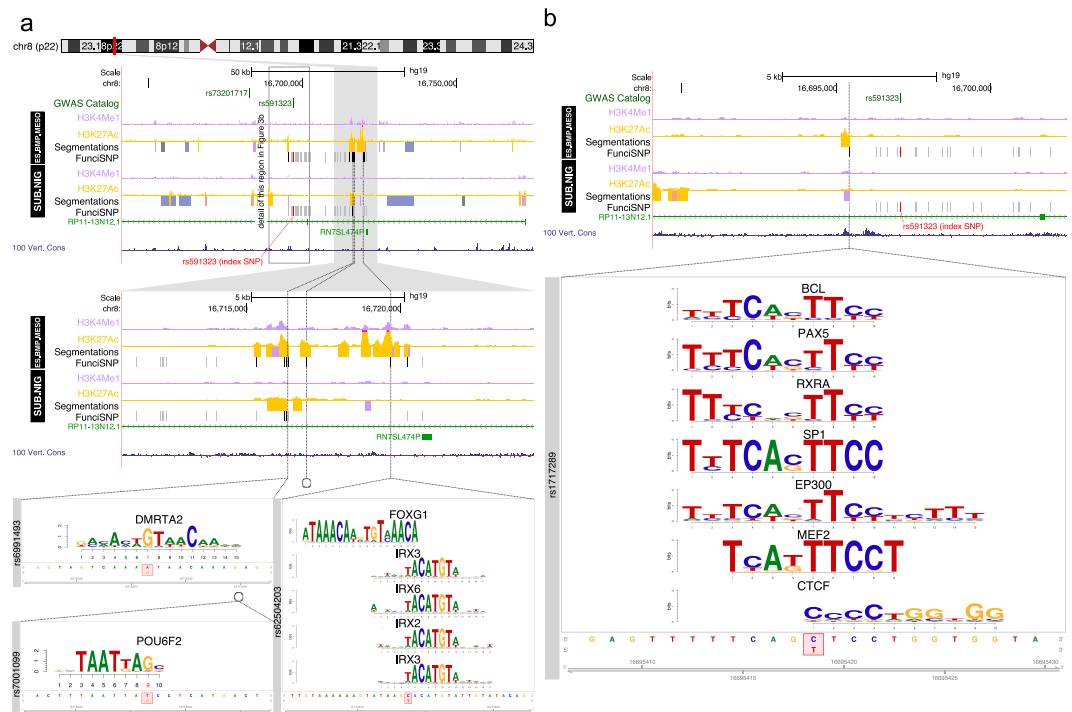


Figure 3. Functional enrichment at the 8p22 locus. (a,b) Two different regions at 8p22. Segmentation tracks summarize the functional annotations according to the REMC color scheme: orange for active enhancer, rose for active promoter, light purple for poised enhancers and promoters, grey for silenced regions, blue-grey for heterochromatin and green for transcribed regions. Wiggle tracks for H3K4me1 and H3K27ac show the fine-grained epigenomics context in the regions surrounding each locus. The location of SNPs that disrupt transcription factor motifs are indicated with a dotted black line and an expanded view at the bottom. The motif is aligned with the genomic sequence surrounding the SNP, whose position is indicated in red; reference and alternate alleles are indicated in the red bounding box in the sequence below.

PD or any neurodegenerative disease. We argue that if clusters of risk SNPs are found more often than by chance in regulatory regions (i.e., significant enrichment in that region), it is likely that alleles as defined by one or more SNPs are functional and modulate the risk of PD. In other words, alleles as defined by one or more SNP drivers carry the other SNP alleles as passengers within haplotype blocks. By this approach, we identified 12 loci significantly enriched in distinct cell types or tissues. Interestingly, only one locus with significant SNP enrichment in active regulatory regions was found in the substantia nigra. Our conclusion is that the risk SNPs identified by GWAS are likely involved in biological mechanisms conferred by cell types beyond the substantia nigra dopaminergic neurons that have traditionally been the focus of PD risk.

Despite a dramatic increase in knowledge regarding the genetics of PD during the past decade^{4–6}, the precise underlying pathogenetic mechanisms have remained elusive. Indeed, the contributing mechanisms might differ greatly among PD patients, because the clinical features, age at onset and rate of disease progression vary significantly¹. The emerging picture of PD is one of a clinical syndrome with different underlying etiologies and molecular mechanisms, rather than a single disease^{1,5,23}. This concept is consistent with the idea that changes in multiple independent molecular pathways, and the genes that control their activities, might influence the risk for PD. Several studies have implicated an impairment of the lysosomal autophagy system; mitochondrial deficits, oxidative stress, and neuroinflammation as key upstream events in the disease process²⁴. This concept also agrees with the emerging view that the initial molecular triggers for PD might not always occur in the cell types that eventually bear the brunt of the disease, i.e., the midbrain dopaminergic neurons and other neurons that show exhibit α -synuclein aggregates. The concept of such “non-cell autonomous” toxicity has gained recent support²⁵ and stresses the need to also study non-neuronal cells when searching for the causes of PD.

We observed that genetic variants in lymphocytes, mesendoderm, brain, liver- and fat-cells could all be involved in PD risk, which is consistent with the theory that PD risk is affected by changes in multiple independent molecular pathways and the genes that control their activities; in different cell and tissue types, even outside the brain. Indeed, only one of the risk loci we identified has risk SNP enrichment in substantia nigra. Despite the traditional research focus on nigral dopaminergic neurons, pathological alpha-synuclein aggregates are present throughout both the peripheral nervous system and brain²⁶ and they are believed to first appear in the olfactory system and in the nerves innervating the gastrointestinal tract and heart²⁷, possibly as early as 5–15 years before diagnosis^{28,29}. This phase is entitled “prodromal PD” and excludes overt motor symptoms, but it is associated with other functional deficits, e.g., anosmia, sleep disorder and constipation^{28,29}. Considering the length of this prodromal phase, genetic variations that influence the pathogenetic mechanisms, even in a minor way and outside the nervous system, could conceivably profoundly impact lifetime PD risk. Although this paper does not attempt

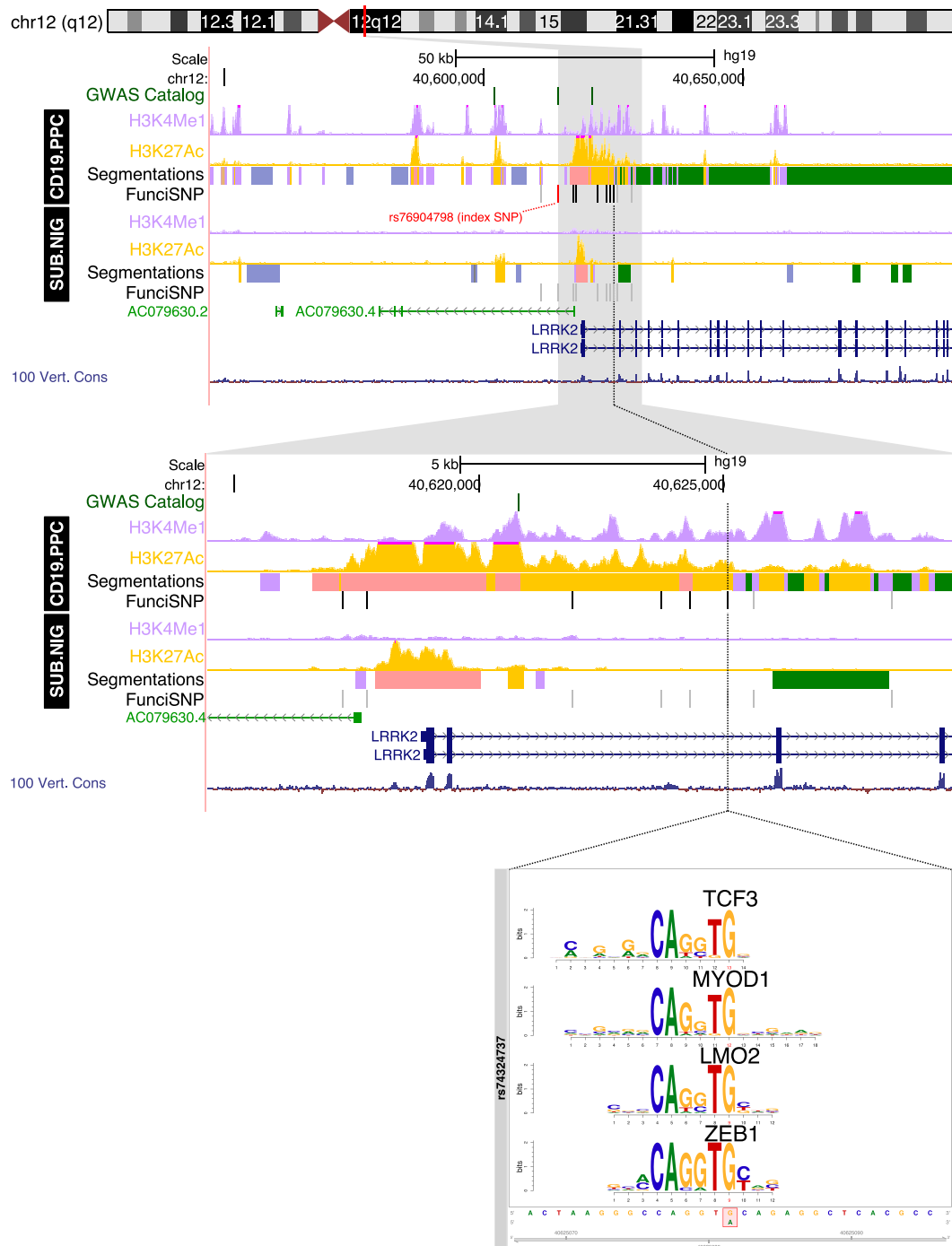


Figure 4. Functional enrichment at the 12q12 locus. Segmentation tracks summarize the functional annotations according to the REMC color scheme: orange for active enhancer, rose for active promoter, light purple for poised enhancers and promoters, grey for silenced regions, blue-grey for heterochromatin and green for transcribed regions. Wiggle tracks for H3K4me1 and H3K27ac show the fine-grained epigenomics context in the regions surrounding each locus. The location of SNPs that disrupt transcription factor motifs are indicated with a dotted black line and an expanded view at the bottom. The motif is aligned with the genomic sequence surrounding the SNP, whose position is indicated in red; reference and alternate alleles are indicated in the red bounding box in the sequence below.

to establish the exact mechanisms by which the identified loci confer risk, in the following sections we discuss hypothetically how they may be involved in the pathobiological mechanisms of PD.

Three of the risk loci we identified are related to the immune system. Neuroinflammation is an integral part of the disease process in PD^{30–32}, with ample evidence available for both activated microglia and increased levels

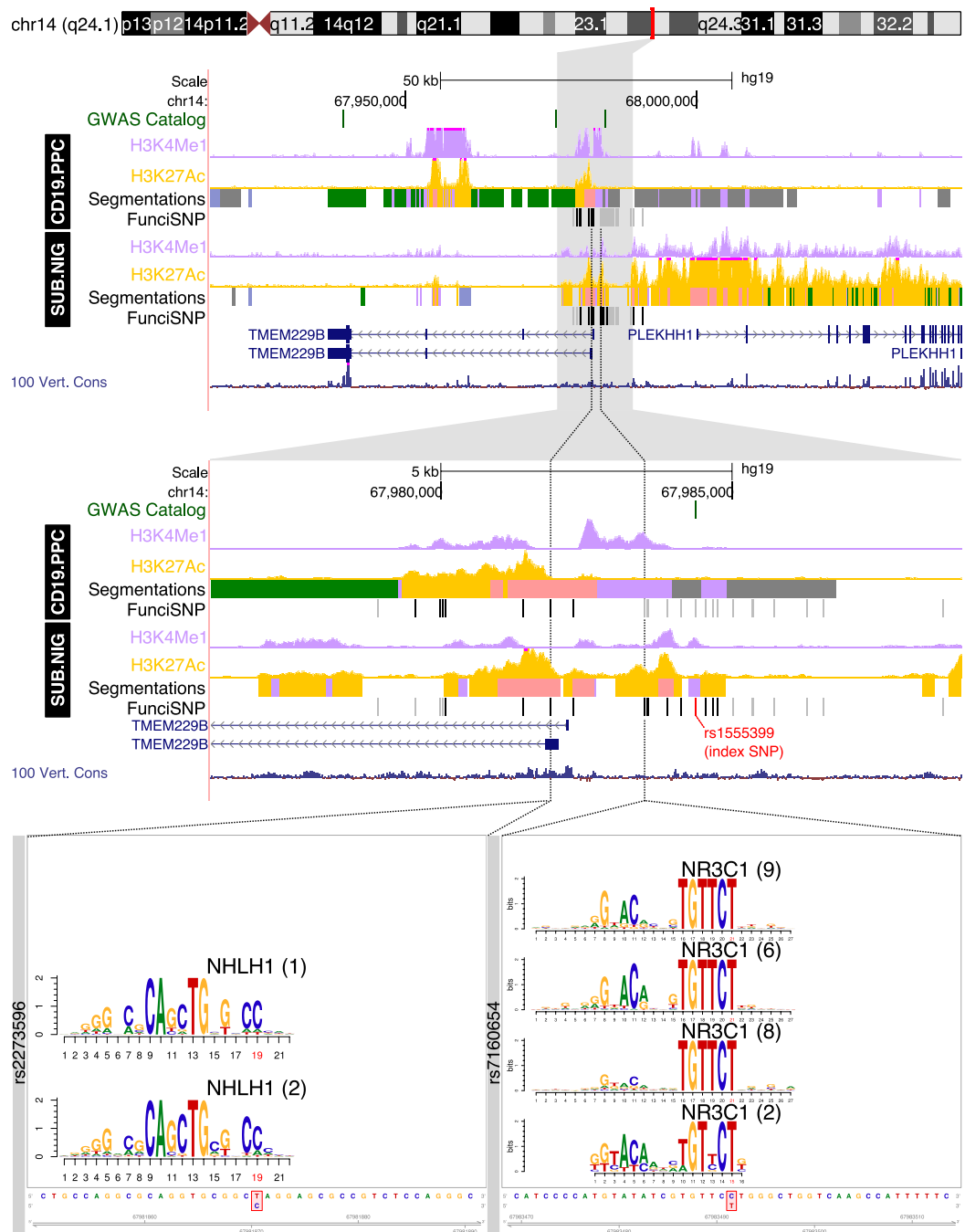


Figure 5. Functional enrichment at the 14q24 locus. Segmentation tracks summarize the functional annotations according to the REMC color scheme: orange for active enhancer, rose for active promoter, light purple for poised enhancers and promoters, grey for silenced regions, blue-grey for heterochromatin and green for transcribed regions. Wiggle tracks for H3K4me1 and H3K27ac show the fine-grained epigenomics context in the regions surrounding each locus. The location of SNPs that disrupt transcription factor motifs are indicated with a dotted black line and an expanded view at the bottom. The motif is aligned with the genomic sequence surrounding the SNP, whose position is indicated in red; reference and alternate alleles are indicated in the red bounding box in the sequence below.

of pro-inflammatory cytokines³³. Taken together, this brings changes in the genetic regulation of immune cells to the center stage. Two of the PD risk SNPs were enriched in CD19⁺ B-lymphocytes, at enhancers for LRRK2 (chromosome 12) and at another locus (chromosome 14) the disruption of the glucocorticoid receptor response element. The immune system might be relevant to PD in at least three fundamentally important pathogenetic aspects: (i) interactions of the immune system and α -synuclein aggregation; (ii) autoimmunity; and (iii) changes in host defenses against pathogens.

Inflammation has been suggested to be a potential trigger of Lewy pathology. Specifically, neuroinflammation promotes α -synuclein misfolding in the brain³⁴, and inflammation in the gut has been proposed to cause α -synuclein aggregation in enteric nerves starting a cascade of propagating Lewy pathology via the vagal nerve^{35–37}. As stated above, one of the risk loci in B-lymphocytes is localized at an enhancer region within *LRRK2*, a gene which can harbor mutations that collectively are the most common known genetic cause of both familial and sporadic PD. *LRRK2* is also associated with the inflammatory bowel disorder Crohn's disease, hinting that it is involved in regulating intestinal inflammation³⁸. The *LRRK2* protein possesses both kinase and GTPase activities. The precise function of *LRRK2* is not fully understood and expression of the protein in nigral dopamine neurons remains controversial. On the other hand, *LRRK2* is expressed in immune cells including monocytes, microglia and CD19 + B-lymphocytes, and is up-regulated following recognition of microbial structures³⁹. Therefore *LRRK2* has been suggested to regulate immune responses in the brain, and microglia lacking *LRRK2* exposed to proinflammatory stimuli have attenuated responses^{40,41}. Brain microglia can phagocytose extracellular α -synuclein, reduce neuron-to-neuron transfer of α -synuclein and thereby potentially play a protective role by limiting the propagation of α -synuclein pathology between brain regions^{42–44}. Changes in *LRRK2* mediated signaling have been suggested to alter clearance of misfolded alpha-synuclein from the brain and these effects might at least in part be mediated by microglia. Interestingly, where comparisons could be made between the GTEx and REMC databases, two of the enriched risk loci (rs6812193 and rs76904798) were associated with near-cis (+/– 1000 kb) gene expression in the same tissues, indicating tissue-specific functionality of the risk enhancers involved. It is unclear how the eQTL genes at the two loci participate in PD risk, with the notable exception of *LRRK2*. Thus, the positive direction of association between risk alleles of rs76904798 and *LRRK2* expression is in line with the known gain-of-function Mendelian mutations in this gene, which in turn lead to PD.

Despite that reactions in brain tissue are more tightly controlled than in the peripheral tissue, peripheral immune cells can readily access the brain⁴⁵ and expression of risk loci in B-lymphocytes might directly contribute to inflammatory mechanisms in PD pathogenesis. Under physiological conditions, memory- and regulatory T-cells, and most likely also regulatory B-cells, traffic and monitor the central nervous system tissue and react to damage and invading pathogens⁴⁶. Potential aberrations in these patrolling immune cells could lead to autoimmunity or unrestricted immune reactions. Co-stimulatory receptors expressed on B cells include CD19, a B-cell activating factor receptor and Toll-like receptor. Hyperactivity of these receptors may break B-cell tolerance in several autoimmune diseases. B-lymphocytes could also be important in PD pathogenesis considering their key role in the primary host defense against pathogens and the evidence that B-cell-related genes, including CD19, are significantly down-regulated in peripheral blood leukocytes from patients with PD^{47,48}. Moreover, the absolute numbers of circulating CD19 cells are reduced in PD patients⁴⁹. Another risk locus, possibly regulating *FAM47E*, was present in liver and adipocytes. In the liver, processes such as oxidation and the detoxification of metabolites and exogenous agents might contribute to PD. States with increased metabolic inflammation, such as diabetes, are coupled to an increased vulnerability of dopaminergic neurons to certain neurotoxins^{50–52}. Furthermore, changes in liver cell function could impact uric acid metabolism. This would be important because changes in uric acid have been suggested to protect against PD, via its antioxidant effects of urate, and changed liver function is associated with changes in urate levels⁵³. The liver harbors a large population of monocytes and macrophages that functions as phagocytes, breaking down pathogens and toxins, and that also presents antigens to lymphocytes⁵⁴.

Possible effects of neurodevelopmental changes cannot be overlooked. A diagnosis of PD requires motor symptoms that are consequential to severe degeneration of nigral dopaminergic neurons and their striatal terminals⁵⁵. The number of nigral dopaminergic neurons varies greatly among normal human subjects^{56,57} and between different inbred strains of mice⁵⁸. Thus, genetic factors might affect dopaminergic neurogenesis in humans and one can postulate that individuals born with more nigral dopaminergic neurons are less likely to be diagnosed with PD during their lifetime. Mesendoderm gives rise to, *inter alia*, mesodermal cells⁵⁹, which in turn gives rise to (again *inter alia*) the notochord. The notochord releases hedgehog signaling proteins which are necessary for the development of the floor plate^{60,61} – an ectodermal tissue that forms neural progenitors including those that differentiate into midbrain dopaminergic neurons⁶². Our analysis of the *8p22* PD risk locus revealed disruption in several transcription factor response elements that are active in BMP4-treated ES-derived mesendoderm. It is conceivable that they affect the differentiation of the notochord; thereby they might have an effect on lifetime PD risk by reducing the number of midbrain dopaminergic neurons generated by the floor plate.

Concluding remarks

Genetic studies have led to novel insights in risk mechanisms of complex diseases, especially in studies of the genetic predisposition to cancer⁶³. Despite this, many identified risk loci still remain challenging to functionally assess, due to the presence of too many potentially functional SNPs and lack of knowledge of the 3D structure of chromatin, making it difficult to link enhancers with their target genes. Annotating regulatory risk regions within specific cell types, as reported here, will go a long way toward understanding the risk mechanisms of complex genetic diseases. The putative risk mechanisms for PD are summarized in Fig. 6, in which we have considered developmental, trigger, and adaptive response mechanisms. The work reported here represents a novel strategy for identifying biological mechanisms that might influence PD risk and, importantly, provides new insight into the etiology of this devastating and progressive neurological disorder. We are currently developing laboratory models to investigate the mechanisms of these genetic associations within their appropriate organ and cellular contexts as informed by our enrichment analysis.

Methods

Functional annotation of GWAS. Twenty-six index SNPs associated with PD were downloaded from http://www.pdgene.org/top_results. Five SNPs were removed from further analysis. Rs17649553 was removed

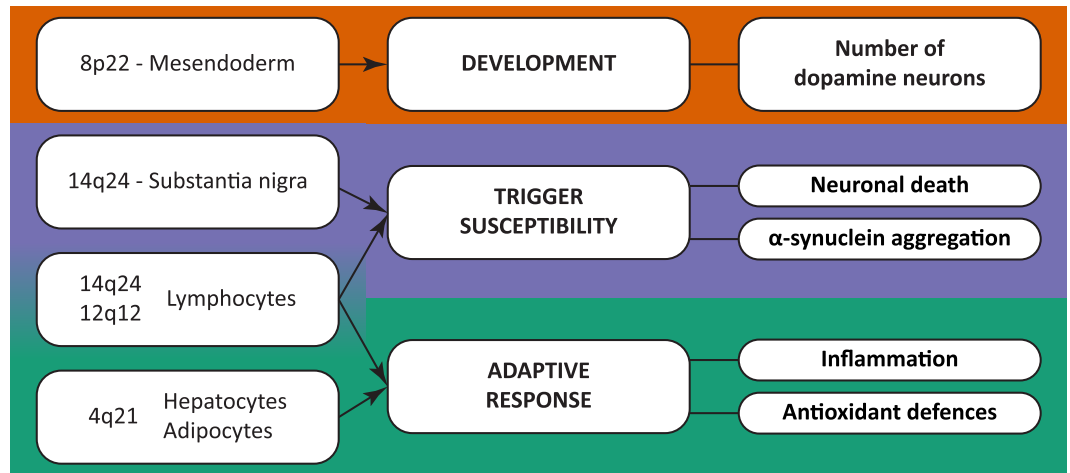


Figure 6. Outline of the novel connections made in this study.

because it was in an assembly exception. Rs11060180, rs34311866, rs356182 and rs62120679 were removed because they had no SNPs with $r^2 \geq 0.8$. This left 21 SNPs for further enrichment analysis. These SNPs were integrated with comprehensive data from the Roadmap Epigenomics Mapping Consortium (REMC) for 77 tissues and cell types and 1000 genomes using FucisNP software¹¹. For each of the cells and tissues from REMC, consortium IDR-processed data⁶⁴ were obtained for each of three data types, H3K4me1 and 3, and H3K27Ac. SNPs in LD at $r^2 \geq 0.8$ and overlapping these features were identified using FucisNP software. In order to functionally categorize the SNPs, we used a rules-based segmentation algorithm to classify regions in every listed cell type based on histone marks. The following describes the rules in order of decreasing precedence. Regions with a positive signal for the histone modification H3K4me3 were classified as promoters, and as active or poised on the basis of overlap with the active histone mark H3K27ac (promoter active region PAR). H3K4me1 regions were classified as enhancers, and similarly as active or poised on the basis of overlap with H3K27ac (enhancer active region EAR). Each region had a single annotation assigned such that SNPs could be assigned one of three unambiguous functions based on the combinations of histone marks. The 77 REMC cell types used for the study and numerical significance of enrichment (see below) are listed in Supplemental Table 1. Cell type abbreviations are listed in Supplemental Table 2. The enrichment is also presented as a heatmap in Fig. 1 for visual simplicity.

Enrichment analyses. We used the hypergeometric distribution to calculate the probability of the overlaps against the background of 1000 genomes (within 1 MB windows around the index SNPs inclusive) and adjusted for multiple comparisons for each cell type with FDR < 0.05. This was repeated for the pooled categories EAR & PAR (active enhancers and promoters). Enriched regions were evaluated in the UCSC genome browser with wiggle tracks (.wig) for H3K4me1 and H3K27ac, our genome segmentations and tissue-specific FucisNP results. We calculated a Euclidean distance metric for log-scaled p values for each index SNP in each tissue and used average linkage hierarchical clustering with the heatmap. 2 package in R (Fig. 1).

Analysis of Transcription Factor binding sites. SNPs located within enhancers or promoters of any kind were analyzed using the bioconductor motifbreakR package using default method settings (weighted sum) with a p-value cutoff at 5×10^{-5} (ref. 10). Figures were prepared from screenshots of the UCSC genome browser and MotifbreakR plotting functions for SNPs at $p < 10^{-5}$.

Expression Quantitative Trait Loci (eQTL). All polymorphisms within ± 50 kb of the FAM47E or LRRK2 locus which were included the meta-analysis by Nalls *et al.*⁷ and associated with a p-value of at least 0.05 were downloaded from pdgene.org. To these were added unique SNPs in LD ($r^2 \geq 0.8$) found via the Broad Institute SNAP program (<https://www.broadinstitute.org/mpg/snap/>). This enlarged set was used to subset the significant eQTL SNP-gene associations available from the GTEx consortium (<http://www.gtexportal.org/home/>). This subsetting produced a tissue specific collection of SNP-gene expression associations linked to PD.

Computers and Software. Genome segmentations, FucisNP, clustering, motifbreakR, and SNP-tissue enrichment analyses were carried out in R (<https://cran.r-project.org>) on an iMac Retina 5k with a 3.5-GHz processor and 24 GB of memory.

References

1. Berg, D. *et al.* Time to redefine PD? Introductory statement of the MDS Task Force on the definition of Parkinson's disease. *Mov Disord* **29**, 454–462 (2014).
2. Braak, H. *et al.* Stanley Fahn Lecture 2005: The staging procedure for the inclusion body pathology associated with sporadic Parkinson's disease reconsidered. *Mov Disord* **21**, 2042–2051 (2006).
3. Bras, J., Guerreiro, R. & Hardy, J. SnapShot: Genetics of Parkinson's disease. *Cell* **160**, 570–570 e571 (2015).
4. Kalinderi, K., Bostantjopoulou, S. & Fidani, L. The genetic background of Parkinson's disease: current progress and future prospects. *Acta Neurol Scand* (2016).

5. Langston, J. W., Schule, B., Rees, L., Nichols, R. J. & Barlow, C. Multisystem Lewy body disease and the other parkinsonian disorders. *Nat Genet* **47**, 1378–1384 (2015).
6. Gasser, T. Usefulness of Genetic Testing in PD and PD Trials: A Balanced Review. *J Parkinsons Dis* **5**, 209–215 (2015).
7. Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**, 989–993 (2014).
8. Verstraeten, A., Theuns, J. & Van Broeckhoven, C. Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. *Trends Genet* **31**, 140–149 (2015).
9. Yao, L., Shen, H., Laird, P. W., Farnham, P. J. & Berman, B. P. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol* **16**, 105 (2015).
10. Coetzee, S. G., Coetzee, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847–3849 (2015).
11. Coetzee, S. G., Rhie, S. K., Berman, B. P., Coetzee, G. A. & Noushmehr, H. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. *Nucleic Acids Res* **40**, e139 (2012).
12. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
13. Hnisz, D. *et al.* Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell* **58**, 362–370 (2015).
14. Parker, S. C. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci USA* **110**, 17921–17926 (2013).
15. Denuc, A. & Marfany, G. SUMO and ubiquitin paths converge. *Biochem Soc Trans* **38**, 34–39 (2010).
16. Ledo, N. *et al.* Functional genomic annotation of genetic risk loci highlights inflammation and epithelial biology networks in CKD. *J Am Soc Nephrol* **26**, 692–714 (2015).
17. Urquhart, J. E. *et al.* DMRTA2 (DMRT5) is mutated in a novel cortical brain malformation. *Clin Genet* (2016).
18. Anney, R. *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* **19**, 4072–4082 (2010).
19. Mariani, J. *et al.* FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375–390 (2015).
20. Kim, K. H., Rosen, A., Bruneau, B. G., Hui, C. C. & Backx, P. H. Iroquois homeodomain transcription factors in heart development and function. *Circulation research* **110**, 1513–1524 (2012).
21. Steger, M. *et al.* Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife* **5** (2016).
22. Chen, C. *et al.* Correlation between DNA methylation and gene expression in the brains of patients with bipolar disorder and schizophrenia. *Bipolar Disord* **16**, 790–799 (2014).
23. Koga, S. *et al.* When DLB, PD, and PSP masquerade as MSA: an autopsy study of 134 patients. *Neurology* **85**, 404–412 (2015).
24. Hirsch, E. C., Jenner, P. & Przedborski, S. Pathogenesis of Parkinson's disease. *Mov Disord* **28**, 24–30 (2013).
25. Ilieva, H., Polymenidou, M. & Cleveland, D. W. Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. *J Cell Biol* **187**, 761–772 (2009).
26. Halliday, G. M., Holtson, J. L., Revesz, T. & Dickson, D. W. Neuropathology underlying clinical variability in patients with synucleinopathies. *Acta Neuropathol* **122**, 187–204 (2011).
27. Braak, H., de Vos, R. A., Bohl, J. & Del Tredici, K. Gastric alpha-synuclein immunoreactive inclusions in Meissner's and Auerbach's plexuses in cases staged for Parkinson's disease-related brain pathology. *Neurosci Lett* **396**, 67–72 (2006).
28. Mahlknecht, P., Seppi, K. & Poewe, W. The Concept of Prodromal Parkinson's Disease. *J Parkinsons Dis* **5**, 681–697 (2015).
29. Berg, D. *et al.* MDS research criteria for prodromal Parkinson's disease. *Mov Disord* **30**, 1600–1611 (2015).
30. Hirsch, E. C. & Hunot, S. Neuroinflammation in Parkinson's disease: a target for neuroprotection? *Lancet Neurol* **8**, 382–397 (2009).
31. Hirsch, E. C., Vyas, S. & Hunot, S. Neuroinflammation in Parkinson's disease. *Parkinsonism Relat Disord* **18** Suppl 1, S210–S212 (2012).
32. Kannarkat, G. T., Boss, J. M. & Tansey, M. G. The role of innate and adaptive immunity in Parkinson's disease. *J Parkinsons Dis* **3**, 493–514 (2013).
33. Collins, L. M., Toulouse, A., Connor, T. J. & Nolan, Y. M. Contributions of central and systemic inflammation to the pathophysiology of Parkinson's disease. *Neuropharmacology* **62**, 2154–2168 (2012).
34. Gao, H. M. *et al.* Neuroinflammation and oxidation/nitration of alpha-synuclein linked to dopaminergic neurodegeneration. *J Neurosci* **28**, 7687–7698 (2008).
35. Angot, E., Steiner, J. A., Hansen, C., Li, J. Y. & Brundin, P. Are synucleinopathies prion-like disorders? *Lancet Neurol* **9**, 1128–1138 (2010).
36. Lebovitz, T. *et al.* The second brain and Parkinson's disease. *Eur J Neurosci* **30**, 735–741 (2009).
37. Svensson, E. *et al.* Vagotomy and subsequent risk of Parkinson's disease. *Ann Neurol* **78**, 522–529 (2015).
38. Nalls, M. A. *et al.* Genetic comorbidities in Parkinson's disease. *Hum Mol Genet* **23**, 831–841 (2014).
39. Hakimi, M. *et al.* Parkinson's disease-linked LRRK2 is expressed in circulating and tissue immune cells and upregulated following recognition of microbial structures. *J Neural Transm (Vienna)* **118**, 795–808 (2011).
40. Puccini, J. M. *et al.* Leucine-rich repeat kinase 2 modulates neuroinflammation and neurotoxicity in models of human immunodeficiency virus 1-associated neurocognitive disorders. *J Neurosci* **35**, 5271–5283 (2015).
41. Moehle, M. S. *et al.* LRRK2 inhibition attenuates microglial inflammatory responses. *J Neurosci* **32**, 1602–1611 (2012).
42. Allen Reish, H. E. & Standaert, D. G. Role of alpha-synuclein in inducing innate and adaptive immunity in Parkinson disease. *J Parkinsons Dis* **5**, 1–19 (2015).
43. Bae, E. J. *et al.* Antibody-aided clearance of extracellular alpha-synuclein prevents cell-to-cell aggregate transmission. *J Neurosci* **32**, 13454–13469 (2012).
44. George, S. & Brundin, P. Immunotherapy in Parkinson's Disease: Micromanaging Alpha-Synuclein Aggregation. *J Parkinsons Dis* **5**, 413–424 (2015).
45. Rezai-Zadeh, K., Gate, D., Szekely, C. A. & Town, T. Can peripheral leukocytes be used as Alzheimer's disease biomarkers? *Expert Rev Neurother* **9**, 1623–1633 (2009).
46. Mutnal, M. B., Hu, S., Schachtele, S. J. & Lokensgard, J. R. Infiltrating regulatory B cells control neuroinflammation following viral brain infection. *J Immunol* **193**, 6070–6080 (2014).
47. Kedmi, M., Bar-Shira, A., Gurevich, T., Giladi, N. & Orr-Urtreger, A. Decreased expression of B cell related genes in leukocytes of women with Parkinson's disease. *Mol Neurodegener* **6**, 66 (2011).
48. Kobo, H. *et al.* Down-regulation of B cell-related genes in peripheral blood leukocytes of Parkinson's disease patients with and without GBA mutations. *Mol Genet Metab* **117**, 179–185 (2016).
49. Stevens, C. H. *et al.* Reduced T helper and B lymphocytes in Parkinson's disease. *J Neuroimmunol* **252**, 95–99 (2012).
50. Wang, L. *et al.* Metabolic inflammation exacerbates dopaminergic neuronal degeneration in response to acute MPTP challenge in type 2 diabetes mice. *Exp Neurol* **251**, 22–29 (2014).
51. Wang, L., Li, C., Zhang, Y., Qiao, C. & Ye, Y. Synthesis and biological evaluation of benzofuroxan derivatives as fungicides against phytopathogenic fungi. *J Agric Food Chem* **61**, 8632–8640 (2013).
52. Santiago, J. A. & Potashkin, J. A. Shared dysregulated pathways lead to Parkinson's disease and diabetes. *Trends Mol Med* **19**, 176–186 (2013).

53. Gao, X., O'Reilly, E. J., Schwarzschild, M. A. & Ascherio, A. Prospective study of plasma urate and risk of Parkinson disease in men and women. *Neurology* **86**, 520–526 (2016).
54. Heymann, F. & Tacke, F. Immunology in the liver—from homeostasis to disease. *Nat Rev Gastroenterol Hepatol* **13**, 88–110 (2016).
55. Kordower, J. H. *et al.* Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. *Brain* **136**, 2419–2431 (2013).
56. Pakkenberg, B. *et al.* The absolute number of nerve cells in substantia nigra in normal subjects and in patients with Parkinson's disease estimated with an unbiased stereological method. *J Neurol Neurosurg Psychiatry* **54**, 30–33 (1991).
57. Iacono, D. *et al.* Parkinson disease and incidental Lewy body disease: Just a question of time? *Neurology* **85**, 1670–1679 (2015).
58. Ross, R. A., Judd, A. B., Pickel, V. M., Joh, T. H. & Reis, D. J. Strain-dependent variations in number of midbrain dopaminergic neurones. *Nature* **264**, 654–656 (1976).
59. Tada, S. *et al.* Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture. *Development* **132**, 4363–4374 (2005).
60. Hynes, M., Poulsen, K., Tessier-Lavigne, M. & Rosenthal, A. Control of neuronal diversity by the floor plate: contact-mediated induction of midbrain dopaminergic neurons. *Cell* **80**, 95–101 (1995).
61. Placzek, M., Yamada, T., Tessier-Lavigne, M., Jessell, T. & Dodd, J. Control of dorsoventral pattern in vertebrate neural development: induction and polarizing properties of the floor plate. *Development Suppl* **2**, 105–122 (1991).
62. Wang, M. Z. *et al.* Induction of dopaminergic neuron phenotype in the midbrain by Sonic hedgehog protein. *Nat Med* **1**, 1184–1188 (1995).
63. Freedman, M. L. *et al.* Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* **43**, 513–518 (2011).
64. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

Acknowledgements

The work presented here was supported by Van Andel Research Institute and Cedar-Sinai funds. We acknowledge edits by Dr. David Nadziejka (VARI).

Author Contributions

G.A.C., S.G.C. and D.J.H. conceived the study and acquired and interpreted the data. S.G.C., S.P. and D.J.H. performed the bioinformatics analyses. P.B. and L.B. provided Parkinson's disease expertise and in large part wrote the discussion. All authors contributed to the writing of, and concurred with the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Coetzee, S. G. *et al.* Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology. *Sci. Rep.* **6**, 30509; doi: 10.1038/srep30509 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016