Genetic studies of sporadic Parkinson’s disease

On the identification of genetic risk factors and the path towards a better understanding of underlying pathogenic mechanisms

Victoria Berge-Seidl

Department of Neurology, Oslo University Hospital
Faculty of Medicine, University of Oslo

2021

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-964-6

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Acknowledgements

This thesis was carried out at the Department of Neurology, Oslo University Hospital, and the Institute of Clinical Medicine at the University of Oslo. I am grateful to the South-Eastern Norway Regional Health Authority for their funding of this PhD project.

During my medical studies, I took part in research related to the Parkinson’s disease protein alpha-synuclein. Motivated for further scientific work and with a special interest in neurology, I was fortunate to get in contact with Mathias Toft and his research group. I am deeply grateful to my supervisor Mathias Toft for giving me the opportunity to do my PhD in such an exciting academic field as Parkinson’s disease genetics. Thank you for invaluable support and all the knowledge you have shared. You have given me trust and freedom in my work, inspired me to take on challenging tasks, provided guidance along the way, as well as recognition and encouragement. I also want to express my gratitude to my co-supervisor Espen Dietrichs for his constructive comments and valuable advice.

I am grateful to be part of a research group with skillful and nice colleagues. Deep thanks to Lasse Pihlstrøm for the important contributions he has made to the scientific work presented in this thesis. I have knocked more than once or twice on your office door with questions regarding everything from specific software commands to larger genetic principles. Thank you for your patience, instructive answers and for the opportunity to learn from you. I am grateful to Sandra Pilar Henriksen and my former colleague Aina Rengmark for their help and guidance, especially related to laboratory work. Thank you Margrete Langmyhr for the conversations and work we have shared. Research and life as a PhD student inevitably has its ups and downs, and it has meant a lot to me to take part in this together with you. Also, warm thanks to my colleagues Zafar Iqbal, Maren Stolp Andersen, Chiara Cappelletti, Silje Bjerknes and Jon-Anders Tunold.

I have very much appreciated working in a lively research environment at Domus Medica 4. Special thanks to Tone Berge and my other colleagues in the MS research group for valuable scientific input and support, as well as pleasant coffee breaks and luncheons.

I would like to express my warmest thanks to my family. I am truly grateful to my parents Elisabeth and Viktor for continuous care and support, the interest they show in my work and for sheering me on in life. And to my husband Sebastian, thank you for making me laugh and relax, for keeping the house and garden in order, and for sharing your thoughts, enthusiasm and life with me. Your experience with research and molecular biology has been very helpful, allowing me to discuss my work with you in ways that most PhD students cannot with their partner. Our children Josefine and Gustav are such a wonderful distraction to work and bring so much joy into my life. While they are too young to understand the scope of a PhD, they definitely do their own experiments at the playground, in the bathtub and out in nature.

Finally, the work presented in this thesis would not have been possible without the consent given and blood samples donated by study participants. Thank you so much for your contributions!

Table of Contents

Abbreviations

AAO	Age at onset
ATAC-seq	Assay for Transposase Accessible Chromatin followed by sequencing
bHLH	Basic helix-loop-helix
ChIP-seq	Chromatin immunoprecipitation sequencing
dbGaP	Database of Genotypes and Phenotypes
ddNTP	Di-deoxynucleotide
DREME	Discriminative regular expression motif elicitation
DNase-seq	DNase I hypersensitive sites sequencing
eQTL	Expression quantitative trait locus
FANS	Fluorescence-activated nuclear sorting
FIMO	Find Individual Motif Occurrences
GBA	Glucocerebrosidase gene
GCase	Glucocerebrosidase
GD	Gaucher’s disease
GWAS	Genome-wide association study
HTS	High-throughput sequencing
HWE	Hardy-Weinberg equilibrium
LD	Linkage disequilibrium
MEME	Multiple EM for motif elicitation
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
OCR	Open chromatin region
OR	Odds ratio
PCR	Polymerase chain reaction
PD	Parkinson’s disease
PPMI	Parkinson's Progression Markers Initiative
PRS	Polygenetic risk scores
PWM	Position weight matrix
RBD	Rapid eye movement sleep behavior disorder

Sammendrag på norsk

Mer enn seks millioner mennesker lever med Parkinsons sykdom i verden i dag og de fleste som rammes er i høyere aldersgrupper. Parkinsons sykdom er en nevrologisk lidelse som diagnostiseres basert på tilstedeværelse av motoriske symptomer med skjelvinger, langsomme bevegelser og stivhet. Pasienter med Parkinsons sykdom har ofte i tillegg en rekke ikke-motoriske symptomer som demens, søvnforstyrrelse, og plager relatert til dysfunksjon av det autonome nervesystemet. Ved Parkinsons sykdom degenererer og dør nerveceller, men man har en manglende forståelse av de underliggende patologiske prosessene. Det finnes medikamenter som kan lindre symptomer, men ingen behandling som er kurativ eller som kan hindre at sykdommen progredierer.

De siste tiårene har forskning vist at genetikken spiller en betydelig rolle i utviklingen av Parkinsons sykdom. Familiebaserte studier har identifisert sjeldne sykdomsforårsakende mutasjoner, mens storskala populasjonsbaserte genetiske studier har oppdaget en rekke genområder som påvirker risiko for sporadisk Parkinsons sykdom der det ikke er noen familiehistorie for denne tilstanden. Kartlegging av genetiske risikofaktorer er viktig fordi det kan bidra til en bedre forståelse av sykdomsforårsakende molekylære mekanismer og identifisering av mulige angrepspunkter for utvikling av ny effektiv terapi. Det kan i tillegg bidra til verdifull prognostisk informasjon, samt identifisere individer som vil ha nytte av fremtidige behandlingsmuligheter.

I denne avhandlingen presenteres og diskuteres tre studier av genetiske risikofaktorer for Parkinsons sykdom. I den første studien analyserte vi genetiske varianter i *GBA* genet i skandinaviske pasienter med Parkinsons sykdom og friske kontrollpersoner. *GBA* har blitt identifisert som et av de viktigste risikogenene for Parkinsons sykdom og medikamenter rettet mot nettopp dette genet testes nå i kliniske studier av sykdommen. Vi fant at den kodende varianten E326K er assosiert med Parkinsons sykdom. Våre resultater viser også at E326K i høy grad forekommer sammen med et nærliggende assosiasjonssignal fra helgenomstudier og derfor virker å være den underliggende kausale varianten bak dette signalet. I tillegg til å påvirke risikoen for å få sykdom, så bidrar genetiske varianter også til den betydelige variasjonen i utvikling av symptomer ved Parkinsons sykdom. I den andre studien analyserte vi om genetisk variabilitet i genet *DNM3* påvirker alderen for debut av symptomer ved Parkinsons sykdom. *DNM3* har tidligere blitt rapportert å

påvirke alder for symptomdebut i en liten gruppe av Parkinson pasienter med mutasjon i *LRRK2* genet, men vi fant ingen evidens for at denne effekten var overførbar til den store gruppen av pasienter med sporadisk Parkinsons sykdom.
En økende mengde tilgjengelig epigenomiske data i relevante celletyper har åpnet opp for nye mulige strategier for å utforske biologien bak de genetiske signalene. I den tredje studien kombinerte vi assosiasjonssignaler fra helgenomstudier med epigenomiske data for å identifisere nettverk av transkripsjonsfaktorer involvert i risikomekanismer. Vi fant at risikovarianter for Parkinsons sykdom var overrepresentert i åpent kromatin med bindingsseter for transkripsjonsfaktorer tilhørende bHLH-familien. Dette indikerer at bHLH-transkripsjonsfaktorer kan være involvert i patogene mekanismer ved sykdommen.

De tre studiene som presenteres i denne avhandlingen benytter ulike metoder for å utforske forskjellige aspekter av genetikkens rolle ved Parkinsons sykdom. Videre kartlegging og forståelse av genetiske risikofaktorer forventes å ha en sentral rolle i utviklingen av nye former for terapi som kan bedre livene til individer med Parkinsons sykdom.

Publications included

Paper 1

Berge-Seidl V, Pihlstrøm L, Maple-Grødem J, Forsgren L, Linder J, Larsen JP, Tysnes OB, Toft M

«The *GBA* variant E326K is associated with Parkinson's disease and explains a genome-wide association signal»

Neurosci Lett. 2017 Sep 29;658:48-52. doi: 10.1016/j.neulet.2017.08.040.

Paper 2

Berge-Seidl V, Pihlstrøm L, Wszolek ZK, Ross OA, Toft M

«No evidence for *DNM3* as genetic modifier of age at onset in idiopathic Parkinson's disease»

Neurobiol Aging. 2019 Feb;74:236.e1-236.e5. doi: 10.1016/j.neurobiolaging.2018.09.022.

Paper 3

Berge-Seidl V, Pihlstrøm L, Toft M

«Integrative analysis identifies bHLH transcription factors as contributors to Parkinson's disease risk mechanisms»

Sci Rep. 2021 Feb 10;11(1):3502. doi: 10.1038/s41598-021-83087-2.

1. Introduction

Parkinson’s disease (PD) is a neurodegenerative movement disorder that affects more than 6 million people worldwide (GBD 2016 Parkinson's Disease Collaborators, 2018). This progressive condition is characterized by the loss of dopaminergic neurons in a nucleus of the midbrain called the substantia nigra and manifests clinically as slowing of movement, tremor at rest and muscle rigidity. In addition to symptoms from the movement apparatus, PD patients also suffer from a range of non-motor symptoms such as cognitive impairment, mood- and sleep disorders, and symptoms related to dysfunction of the autonomic nervous system. Symptomatic treatment may alleviate some symptoms, but there is no preventive- or disease-modifying therapy that can affect the progressive nature of the disease. The presentation of symptoms and rate of progression vary between patients, but it will eventually lead to a debilitating stage where the patient requires extensive help with activities of daily living.

Who are likely to develop PD? And what are the pathological processes causing neuronal death in PD patients? These are crucial questions that need to be addressed for improved therapy to become available. In recent decades, our understanding of how genetics contribute to the development of PD has been substantially enhanced. This opens up new routes to uncover molecular targets for neuroprotective treatment and may aid in the identification of individuals that would benefit from such therapy.

In this thesis, I will present and discuss three studies that all explore the contribution of genetics in PD, however from different angles. In the first study, we analyzed genetic variability within the glucocerebrosidase gene (*GBA*). *GBA* has emerged as a candidate gene for targeted therapies in PD. We analyzed *GBA* variants in PD patients and controls to explore how these variants relate to a nascent genome-wide association signal. In the next study, we analyzed how genetic variability may affect the age at onset (AAO) of PD. We tested whether a genetic variant reported to affect the AAO in a genetic subgroup of PD, also affects the disease onset in the majority of PD patients having idiopathic PD. In the third study, our aim was to contribute to the challenging task of translating the growing number of genetic association signals into biological meaningful information.

We combined genome-wide association signals with epigenomic data to identify transcriptional networks potentially involved in PD risk mechanisms.

1.1 Parkinson’s disease

1.1.1 Neuropathology

In 1817, James Parkinson published “An essay on the shaking palsy”, where he described six individuals sharing a number of characteristic symptoms (Parkinson, 1817). This clinical syndrome would later be defined in greater detail by another famous neurologist, Jean-Martin Charcot, who named it after Parkinson. In his essay, James Parkinson captured many of the clinical features of PD and also noted the degenerative nature of the disease. Almost a century later, in 1912, Fritz Heinrich Lewy identified characteristic inclusions in neurons of certain brain nuclei in PD (Lewy, 1912). Shortly after, Konstantin Nikolaevich Tretiakoff described similar inclusions in the substantia nigra of PD patients that he named after Lewy (Tretiakoff, 1919). He also showed degeneration of the substantia nigra and suggested that there was a link between the cell loss and parkinsonian symptoms, an observation that was confirmed by Rolf Hassler in 1938 (Hassler, 1938, Goedert et al., 2013). The biochemical composition of Lewy bodies would however remain unknown until the late 1990s when immunohistochemical staining identified the protein alpha-synuclein as the main component (Spillantini et al., 1997).

These groundbreaking discoveries have been vital in defining what we still regard as the pathological hallmark of PD: abnormal aggregation of alpha-synuclein into Lewy bodies and the loss of dopaminergic neurons in the substantia nigra. At the time of clinical presentation of PD, about 50 % of the nigral dopaminergic cell bodies and their axon terminals in the putamen are lost. Five years after diagnosis, the loss is almost complete (Kordower et al., 2013). PD is however not a disease limited to dopaminergic neurons of the substantia nigra, but is instead a multisystem disorder affecting many different regions of the nervous system. Lewy body pathology has been reported in multiple regions of the brain, and also in the peripheral- and enteric nervous system, often accompanied by neuronal cell loss (Giguere et al., 2018).

In 2003, Braak and colleagues introduced a neuropathological staging model for PD. According to this model, Lewy pathology first occur in the dorsal motor nucleus of the vagal nerve and anterior olfactory nucleus, advances from there to subcortical nuclei, and at later stages reach the cerebral cortex (Braak et al., 2003). Consistent with this staging model the dual-hit hypothesis proposes that an unidentified neurotropic pathogen comes into contact with enteric- and olfactory neurons, initiating Lewy pathology and the following spread to the central nervous system (Hawkes et al., 2007). Misfolding and aggregation of alpha-synuclein into amyloid fibrils within Lewy bodies is considered a major pathogenic event in PD. However, the exact mechanisms through which alpha-synuclein aggregate and how this process relates to neuronal impairment are far from fully understood. Interestingly, experimental evidence suggests that alpha-synuclein pathology may spread from cell to cell through prion-like mechanisms, which could explain Braak's pathological findings (Steiner et al., 2018).

1.1.2 Clinical features and diagnostics of Parkinson's disease

The diagnosis of PD is based on clinical features and confirmation of the diagnosis can only be obtained postmortem based on neuropathological findings. The diagnostic accuracy of PD has been estimated at about 80% after the initial assessment when compared to autopsy, with some improvement after follow-up (Rizzo et al., 2016). Several diagnostic criteria or guidelines have been introduced the last decades to improve and facilitate the diagnostic process in PD, with the most recent being the Movement Disorder Society (MDS) Clinical Diagnostic Criteria for Parkinson's disease (MDS-PD criteria) published in 2015 (Postuma et al., 2015). The MDS-PD criteria encompass the two previous main sets of diagnostic criteria (United Kingdom PD Society Brain Bank and Gelb's criteria), retaining motor parkinsonism as a core feature of the disease. At the same time, the MDS-PD criteria also introduce new aspects, such as an increasing recognition given to non-motor manifestations (Postuma et al., 2015, Marsili et al., 2018).

A diagnosis of PD requires that the patient has parkinsonism. Parkinsonism is defined by bradykinesia, in combination with resting tremor, rigidity, or both. Bradykinesia describes slowing of movement, while rigidity is resistance to passive movement in a relaxed limb or neck. Resting tremor occurs when a part of the body is completely at rest and is typically suppressed during movement initiation. In later stages of the disease, PD patients may develop postural instability leading to troubles with balance and falls. After parkinsonism has been established, absolute exclusion criteria, red flags and supportive features are used to determine whether PD is the cause of parkinsonism. Examples of supportive features are a clear beneficial response to dopaminergic therapy and the presence of levodopa-induced dyskinesia. Exclusion criteria and red flags include clinical findings and signs clearly pointing to other causes of parkinsonism such as frontotemporal dementia, multisystem atrophy, corticobasal degeneration and drug-induced parkinsonism (Postuma et al., 2015).

For many patients, non-motor symptoms dominate the clinical picture and have a significant negative impact on quality of life (Chaudhuri et al., 2011). PD patients may suffer from a variety of non-motor symptoms including cognitive impairment, rapid eye movement sleep behavior disorder (RBD), bladder dysfunction, constipation, depression, anxiety, olfactory dysfunction and orthostatic hypotension. Non-motor features may precede motor symptoms and a clinical diagnosis of PD by several years. Prodromal disease may be defined as the stage when early symptoms and signs of PD neurodegeneration are present, but insufficient to set a classic clinical diagnosis of PD (Berg et al., 2015). Prodromal non-motor symptoms are not specific to PD, but individuals that present with a combination of these symptoms are at a greater risk of developing PD. RBD in particular has a high predictive value of a subsequent diagnosis of PD (Galbiati et al., 2019, Mahlknecht et al., 2015).

Imaging of the brain is not a decisive part of the diagnostic assessment, but may aid in differential diagnosis of PD. Magnetic resonance imaging is used to identify or rule out other causes or forms of parkinsonism (Armstrong and Okun, 2020). Functional imaging with single-photon emission computed tomography (SPECT) or positron emission tomography (PET) can reveal nigrostriatal cell loss, displayed by reduced or asymmetric uptake of striatal dopaminergic biomarkers (Balestrino and Schapira, 2020). Dopaminergic functional imaging may aid in the differential diagnosis between degenerative and nondegenerative parkinsonism.

1.1.3 Treatment

Treatment in PD is symptomatic, focused on alleviating motor and non-motor symptoms. There are currently no available disease-modifying therapies that halt or slow the rate of neurodegeneration. Dopamine replacement therapy is the standard treatment of motor symptoms of PD. Dopaminergic agents such as levodopa, dopamine agonists and monoamine oxidase-B inhibitors aim at reconstituting the dopaminergic signaling in striatum. Non-motor symptoms are generally refractory to dopaminergic medication and require therapeutic approaches targeting neurotransmitter systems other than dopamine, such as serotonin and acetylcholine (Armstrong and Okun, 2020).

Dopamine-based therapies typically provide good control of the initial motor symptoms. However, as the disease progresses, individuals tend to lose the long-duration response to dopamine, and also develop a diminished short-duration response (Armstrong and Okun, 2020). This leads to worsening of symptoms and an increase in disability when the medication wears off. Patients may also experience motor complications in the form of dyskinesia, dystonia or fluctuations. When motor complications are poorly managed by classical pharmacological therapies, patients may benefit from advanced treatments such as direct administration of levodopa-carbidopa gel into the duodenum by a pump through a gastrostomy catheter or deep brain stimulation (Balestrino and Schapira, 2020). Pharmacological treatment of PD should be complemented by non-pharmacological approaches. Rehabilitative therapy and physical activity may favorably affect speech, swallowing, gait and other aspects of PD (Armstrong and Okun, 2020, Gronek et al., 2021).

1.1.4 Etiology

The etiology of PD is still largely unknown. Most PD cases are idiopathic, meaning that no known cause of the disease has been identified. The major known risk factor in PD is aging. PD is uncommon in individuals younger than 50 years of age, but both the incidence and prevalence rise sharply after the age of 60 years. The prevalence is generally estimated at 1 % in individuals over 60 years of age (de Lau and Breteler, 2006). The prevalence peaks between 85 years and 89 years where it is estimated at 1.7% in men and 1.2 % in women (GBD 2016 Parkinson's Disease Collaborators, 2018). As these figures show, PD is more common in men and the male-to-female ratio is reported at 1.4:1.0 (GBD 2016 Parkinson's Disease Collaborators, 2018). Aging populations increase the global burden of PD and other neurodegenerative disorders. In 2016, 6.1 million individuals were estimated to have PD globally, which was more than a doubling compared to 1990. The increase in prevalence is however not solely explained by more people in higher age groups, but has probably additional contributing factors such as longer disease duration, greater awareness of diagnosis and possibly increased exposure to environmental factors related to the growing industrialization of the world (GBD 2016 Parkinson's Disease Collaborators, 2018).

A large number of environmental exposures have been investigated in epidemiological studies of PD and some do have substantial evidence of an association (Bellou et al., 2016). The interpretation of these findings is however complicated by study biases and causal inference has proven difficult. Among the identified risk factors are traumatic brain injury and exposure to pesticides, while smoking, coffee consumption, ibuprofen use and vigorous exercise show an inverse association with PD (Chen and Ritz, 2018). The relationship between pesticides and PD was discovered in the 1980s when it was reported that a group of drug users that had been exposed to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a substance structurally similar to the herbicide paraquat, developed parkinsonism indistinguishable from PD (Langston et al., 1983). MPTP and other exogenous toxic agents that damage the nigrostriatal dopaminergic pathway have been used to develop animal models of PD (Tieu, 2011).

The link between PD and environmental toxins incited the view of PD as a non-genetic disorder (Billingsley et al., 2018). This was supported by initial twin studies that lacked convincing evidence of any heritability in PD (Duvoisin et al., 1981, Ward et al., 1983). However, in the late 1990s a mutation in the *SNCA* gene was identified as causative in families with autosomal dominant PD (Polymeropoulos et al., 1997). This key discovery was followed by intensive genetic research demonstrating that genetic variants play a substantial role in the development of PD.

1.2 The genetic landscape of Parkinson’s disease

1.2.1 Molecular methods applied in Parkinson’s disease genetic research

Genetic discoveries in PD have followed the technical advances in molecular biology. The first DNA-based method was linkage analysis, a family-based method that estimates the co-segregation of genetic markers and a defined disease in pedigrees, with the aim of mapping the disease to a genomic region (Henriksen et al., 2017). The genomic region linked with disease is large, typically containing multiple coding genes and needs to be sequenced and analyzed for potential disease-causing mutations (Pihlstrom et al., 2017). Linkage analyses have been successful at identifying mutations that follow a classical Mendelian inheritance pattern and have identified several autosomal-dominant and autosomal-recessive genes that cause PD.

The majority of PD cases do however occur sporadically, resulting from a complex interplay between aging, environmental- and genetic factors (Figure 1). Complex sporadic disorders generally involve genetic variants with smaller impact on risk which require large-scale case-control association studies for identification. Candidate gene studies test whether frequencies of genetic variants of individuals that have a specific disease differ significantly from the control population. The candidate gene approach has been applied to a large number of genes selected based on the existing genetic, biological or clinical knowledge. Although candidate gene studies have contributed to the discovery of some well-established PD risk genes, most findings have proven difficult to replicate. This high rate of false-positive findings may have several reasons, including bias due to undetected population admixture and limited knowledge of the underlying genetic landscape (Lill, 2016, Henriksen et al., 2017). Many of these limitations have been overcome by the application of genome-wide genotyping arrays used in genome-wide association studies (GWASs), resulting in the successful identification of multiple risk signals in PD and other complex disorders.

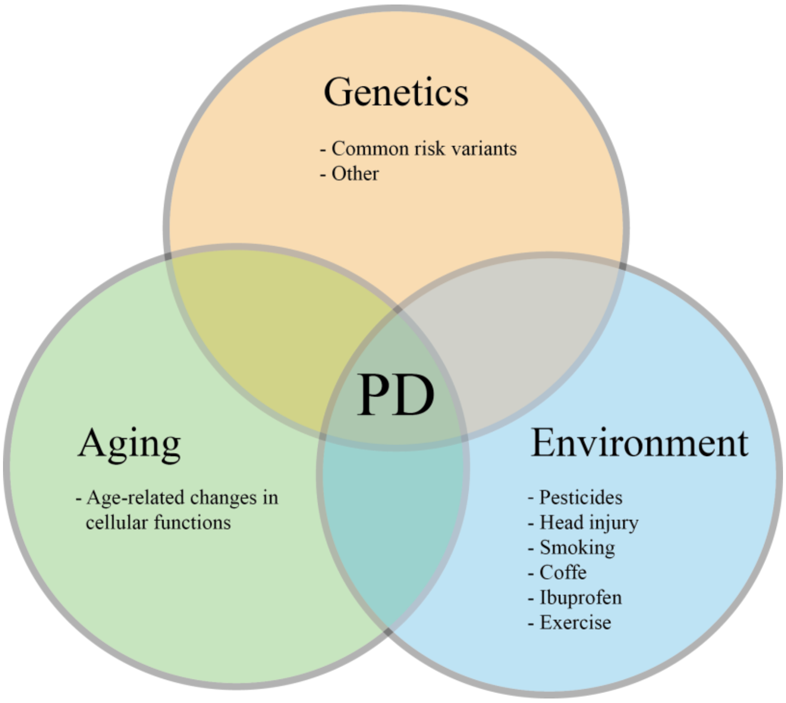


Figure 1. *Interplay of genetic-, environmental- and age-related factors underlying the pathogenesis of PD.*

In GWASs, up to several millions of variants across the genome are genotyped and tested for association with a disease, utilizing an unbiased hypothesis-free approach. This design relies on and exploits linkage disequilibrium (LD), which is the correlation that exists between genetic variants in the genome. LD makes it possible to reduce the number of markers that needs to be assayed since a few hundred thousand tagging variants can capture a sufficient proportion of the common variation in the human genome (Spain and Barrett, 2015). GWASs are mainly used to study common variants in common diseases. Common variants are typically defined as having an allele frequency above 1 %. The relatively inexpensive genotyping arrays enables inclusion of very large sample-sizes, up to more than a million individuals, enabling the discovery of risk loci with weak effect sizes.

1.2.2 Monogenic causes of Parkinson’s disease

In most populations less than 5 % of PD patients have a monogenic form of the disease, meaning that it can be contributed to a rare and highly penetrant pathogenic variant (Reed et al., 2019). The first Mendelian PD mutation was discovered in 1997 when the A53T mutation in the *SNCA* gene was identified as causative in families with autosomal dominant PD (Polymeropoulos et al., 1997). In addition to marking the starting point of PD genetics, this key discovery is a prime example of how genetic findings may lead to new biological insight. Shortly after the identification of a causative mutation in the *SNCA* gene, alpha-synuclein, which is the protein encoded by the *SNCA* gene, was identified as the main component in Lewy bodies and Lewy neurites (Spillantini et al., 1997). Additional mutations in the *SNCA* gene, and also duplications and triplications of *SNCA* have been identified as causative in PD. (Ibáñez et al., 2004, Farrer et al., 2004). It appears to be a clear *SNCA* genomic dosage-related phenotype so that patients with higher number of copies of the gene have more severe symptoms (Deng et al., 2018).

Since the discovery of the *SNCA* locus, at least 20 genes have been reported as causative for PD, including both autosomal dominant and autosomal recessive genes (Blauwendraat et al., 2020a, Deng et al., 2018). Genes reported as causative for PD are listed in Table 1 and have notably variable degree of confidence regarding the pathogenic relevance. The most commonly affected autosomal recessive gene in PD is Parkin, followed by *PINK1* and *DJ-1* (Kitada et al., 1998, Valente et al., 2004, Bonifati et al., 2003). Pathogenic mutations in recessive genes are very rare in the general PD population but occur at higher rates in patients with early onset PD, defined as AAO before 40 years of age (Kilarski et al., 2012). Mutations in *LRRK2* are the most common cause of autosomal dominant PD. Several pathogenic mutations have been identified in *LRRK2*, of which the G2019S mutation is the most common and well studied (Kachergus et al., 2005). The worldwide frequency of *LRRK2* G2019S is 1 % in sporadic PD and 4 % in familial PD but varies widely between different populations (Healy et al., 2008). *LRRK2* G2019S has an incomplete penetrance, meaning that there are carriers of G2019S that do not develop PD although reaching a high age. Actually, most causative mutations in PD, and possibly all, have an incomplete penetrance. The term monogenic PD may thus be regarded as an oversimplification, since additional genetic and environmental factors are likely to affect presentation of disease in carriers of these pathogenic mutations (Blauwendraat et al., 2020a).

Gene	**Mutation**	**Inheritance**	**Confidence as actual PD gene**
SNCA	Missense or multiplication	Dominant	Very high
PRKN	Missense or loss of function	Recessive	Very high
UCHL1	Missense	Dominant	Low
PARK7 (DJ-1)	Missense	Recessive	Very high
LRRK2	Missense	Dominant	Very high
PINK1	Missense or loss of function	Recessive	Very high
POLG	Missense or loss of function	Dominant	High
HTRA2	Missense	Dominant	Low
ATP13A2	Missense or loss of function	Recessive	Very high
FBXO7	Missense	Recessive	Very high
GIGYF2	Missense	Dominant	Low
PLA2G6	Missense or loss of function	Recessive	Very high
EIF4G1	Missense	Dominant	Low
VPS35	Missense	Dominant	Very high
DNAJC6	Missense or loss of function	Recessive	High
SYNJ1	Missense or loss of function	Recessive	High
DNAJC13	Missense	Dominant	Low
TMEM230	Missense	Dominant	Low
VPS13C	Missense or loss of function	Recessive	High
LRP10	Missense or loss of function	Dominant	Low

Table 1. *Mutations reported to cause PD. This is a modified version of a table published by Blauwendraat et al. (Blauwendraat et al., 2020a). Confidence as actual PD gene is based upon the number of reported families, functional evidence and number of reports that could not replicate the finding that this gene is a PD gene.*

1.2.3 The genetics of sporadic Parkinson’s disease

Investigations into the genetic basis of sporadic PD have been driven by GWASs. The first GWAS of PD was published in 2005, however early studies had too small sample sizes to convincingly identify risk loci (Maraganore et al., 2005, Fung et al., 2006). In 2009, two collaborating studies in Caucasian and Japanese subjects were the first studies to report risk loci of genome-wide significance in PD (Satake et al., 2009, Simon-Sanchez et al., 2009). In the first study, which included 5074 PD patients and 8551 controls of

European ancestry, *SNCA* and *MAPT* were confirmed as PD risk loci (Simon-Sanchez et al., 2009). The association at *SNCA* was replicated by the GWAS of Japanese subjects, where also *BST1*, *PARK16* and *LRRK2* were reported as GWAS loci (Satake et al., 2009). Several GWASs of increasing sample sizes have followed, identifying additional risk loci such as *HLA* and *GAK* (Hamza et al., 2010, Edwards et al., 2010, Saad et al., 2011, Spencer et al., 2011). In 2011, the International Parkinson's Disease Genomics Consortium performed the first meta-analysis of PD GWASs (Nalls et al., 2011). They confirmed six previously identified loci (*MAPT*, *SNCA*, *HLA*, *BST1*, *GAK* and *LRRK2*) and reported five novel loci (*ACMSD*, *STK39*, *MCCC1*/*LAMP3*, *SYT11*, and *CCDC62* /*HIP1R*). Subsequent meta-analyses have further expanded the list of PD risk loci (International Parkinson's Disease Genomics Consortium and Wellcome Trust Case Control Consortium, 2011, Lill et al., 2012, Pankratz et al., 2012, Nalls et al., 2014, Chang et al., 2017). The most recent PD GWAS meta-analysis identified 90 independent genome-wide risk signals (Nalls et al., 2019). This study included 37'688 cases, 18'618 proxy cases and 1.4 million controls, which is a striking increase in participants compared to the few thousand individuals assessed in the first GWASs.

Individually GWAS loci confer a relatively small amount of genetic risk with an odds ratio (OR) typically in range of 1,1–1,5. The genetic risk may thus instead be studied collectively as polygenetic risk scores (PRS) based on aggregation of the allelic status and effect size of multiple risk loci (Blauwendraat et al., 2020a). The PRS based on the 90 association signals identified by the most recent PD GWAS meta-analysis shows that individuals in the top decile of genetic risk are 6-fold more likely to have PD compared to those in the lowest decile of genetic risk (Nalls et al., 2019). In addition to affecting the risk of getting the disease, common genetic variants may also influence clinical progression and disease features of PD (Iwaki et al., 2019). PRS based on cumulative genetic PD risk has been shown to be a predictor of AAO in PD and a few PD risk loci have also been found to be individually associated with AAO (Blauwendraat et al., 2019, Nalls et al., 2015a, Escott-Price et al., 2015, Pihlstrom and Toft, 2015). However, the genetic determinants of PD AAO remain largely unknown.

1.2.4 Pleomorphic risk loci

Some of the PD risk loci identified by GWASs also harbor rare mutations known to cause monogenic PD, demonstrating that various genetic mechanisms contributing to PD may coexist at the same locus (Bandres-Ciga et al., 2020a). *SNCA* and *LRRK2* are examples of pleomorphic risk loci that contain both rare variants with large effects and common variants with smaller effect sizes. This links monogenic PD to sporadic PD, blurring the line between these two entities of the disease. Furthermore, new discoveries adding to the growing list of genetic causes of PD have made it increasingly clear that genetic risk is spread across a variety of allele frequencies and effect sizes (Figure 2).

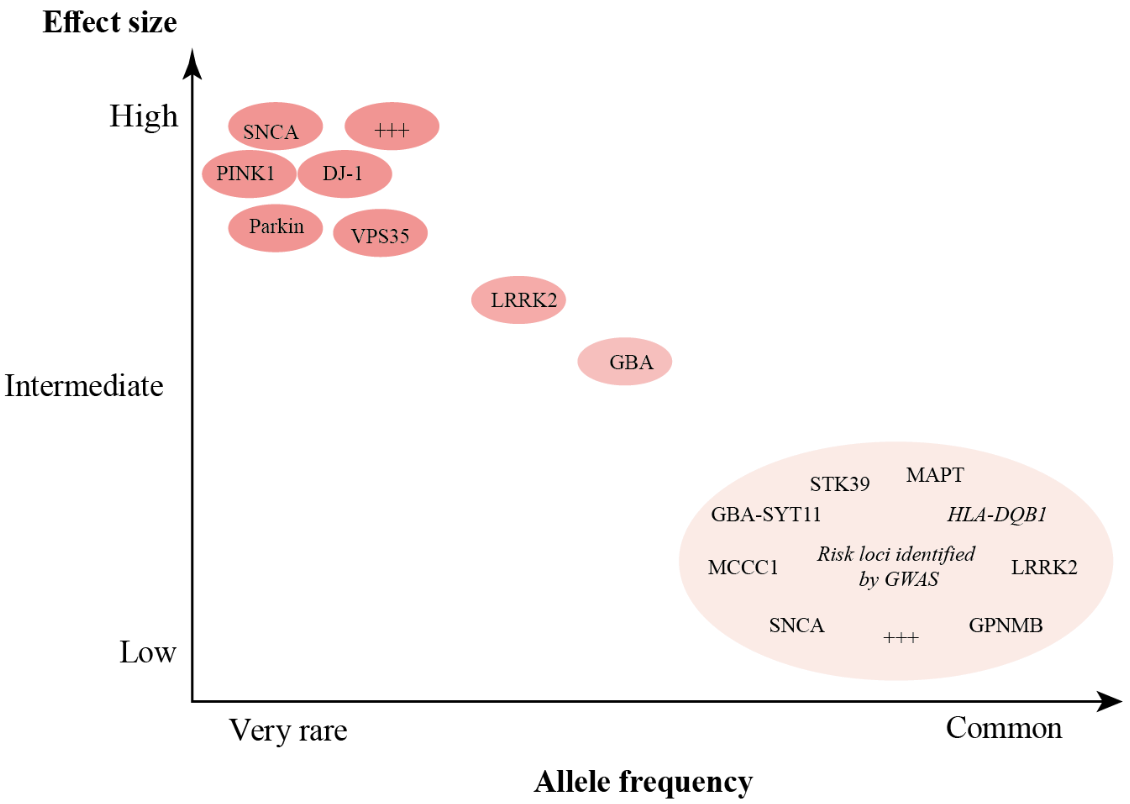


***Figure 2**. Continuous model of genetic risk in PD. Genetic variants associated with PD range from rare and highly penetrant disease-causing mutations to common risk variants with weak effect sizes discovered by large GWASs. The symbol +++ indicates additional known genes or risk loci.*

In line with such a continuous spectrum of genetic risk is the identification of risk variants at the *GBA* locus that are intermediate between the highly penetrant monogenic mutations and common GWAS risk variants. The *GBA* gene is also another example of a pleomorphic risk locus since some variants cause a rare Mendelian disorder in the homozygous state and are risk factors for PD in the heterozygous state (Sidransky, 2004, Blauwendraat et al., 2020a). *GBA* has been established as one of the most important risk genes in PD and events leading to this finding follow an original path differing from most other genetic discoveries.

1.2.5 GBA mutations

The link between PD and *GBA* was not identified by genetic family-based analyses or large-scale GWASs, but was instead first uncovered in the clinics of patients with Gaucher's disease (GD) (Sidransky and Lopez, 2012). GD is a rare lysosomal storage disorder caused by homozygous and compound heterozygous mutations in *GBA*. Several hundred mutations and re-arrangements in *GBA* have been reported (Grabowski, 2008, Hruska et al., 2008). *GBA* encodes the lysosomal enzyme glucocerebrosidase (GCase), which catalyzes the breakdown of the sphingolipid glucosylceramide to ceramide and glucose. In GD, *GBA* mutations lead to a pronounced decrease in activity of GCase, resulting in lysosomal accumulation of the undigested substrate glucosylceramide in tissue macrophages (Stirnemann et al., 2017).

GD is classified into three broad phenotypes based on the degree of neurological involvement. Type I non-neuronopathic GD is by far the most common form of the disease with clinical manifestations that include organomegaly, thrombocytopenia, anemia and bone pain. Type 2, acute-neuronopathic GD, is characterized by severe and rapid neurological decline resulting in early death. Patients with chronic-neuronopathic type 3 GD may have primarily visceral manifestations as described in type 1 GD in combination with oculomotor neurological involvement, or they can develop more severe neurological symptoms (Stirnemann et al., 2017).

The observation that some GD patients developed parkinsonian symptoms, and also the presentation of parkinsonism in relatives who were carriers of *GBA* mutations, led to investigations into the role of *GBA* mutations in PD (Tayebi et al., 2003, Neudorfer et al., 1996, Goker-Alpan et al., 2004). In 2009, a large multicenter study confirmed the association between heterozygous *GBA* mutations and PD with an estimated OR of 5.4 (Sidransky et al., 2009). This finding has been supported by further studies performed in different populations worldwide (Zhang et al., 2018). While initial GWASs failed to identify *GBA* as a susceptibility-gene for PD, a meta-analysis of several GWASs later found an association between a coding variant in *GBA* and PD (Pankratz et al., 2012). Interestingly, a strong association has also been reported between *GBA* mutations and dementia with Lewy bodies, where carriers have an 8-fold increase in the risk of developing the disease (Nalls et al., 2013).

There is a large variation in the distribution and frequency of *GBA* mutations between different populations. In the Ashkenazi Jewish population, which has an especially high frequency of *GBA* mutations, up to 30% of PD patients have been reported as carriers (Aharon-Peretz et al., 2004). Reported frequencies of heterozygous *GBA* mutations in European non-Ashkenazi Jewish PD patients vary between 2-10% (Migdalska-Richards and Schapira, 2016). N370S and L444P are the two most common *GBA* mutations worldwide (Sidransky et al., 2009). The L444P mutation is associated with severe phenotypes of GD with neurological involvement (type 2 and 3) and has a higher risk of developing PD in the heterozygous state compared to the milder N370S mutation which causes type 1 GD (Grabowski, 2008, Gan-Or et al., 2015a). E326K and T369M are low-frequency coding *GBA* variants that are often distinguished from other *GBA* mutations since they do not cause GD in homozygous carriers. Their effect on PD risk has been disputed, however emerging evidence points to a role for E326K, and possibly also T369M, as risk factors for PD (Duran et al., 2013, Mallett et al., 2016).

GBA-associated PD may be clinically indistinguishable from idiopathic PD when assessed during a routine examination. However, clinical studies show that *GBA* mutations are associated with distinct clinical characteristics, especially regarding the distribution and severity of non-motor symptoms. *GBA* mutations confer a higher risk of dementia during the course of PD and are associated with a more rapid cognitive decline (Cilia et al., 2016, Davis et al., 2016, Brockmann et al., 2015). Furthermore, patients with *GBA* mutations have been reported to have a higher prevalence of hyposmia, RBD, neuropsychiatric symptoms and autonomic dysfunction (Gan-Or et al., 2015b, Thaler et al., 2018, Cilia et al., 2016, Brockmann et al., 2011, Jesus et al., 2016). *GBA* mutations are associated with a younger AAO in PD and a more rapid progression of motor symptoms (Cilia et al., 2016, Jesus et al., 2016, Winder-Rhodes et al., 2013, Brockmann et al., 2015).

The underlying molecular mechanisms of how *GBA* mutations lead to PD are not fully understood. Interestingly, experimental evidence supports a bidirectional relationship between alpha-synuclein metabolism and GCase activity. Several studies in cell cultures and animal models show that inhibition of GCase induces alpha-synuclein accumulation with consequential neurotoxic effects (Mazzulli et al., 2011, Rockenstein et al., 2016, Schondorf et al., 2014). Alpha-synuclein, on the other hand, has been reported to disrupt the intracellular trafficking and lysosomal activity of GCase (Mazzulli et al., 2011). The link between *GBA* and PD has opened up for novel strategies in the pursuit of new therapeutic interventions in PD.

1.3 From genetic association to molecular function

GWASs have been successful at identifying thousands of statistically associated genetic loci with a wide variety of complex diseases and phenotypes (Buniello et al., 2019). A large proportion of these risk loci have been replicated, suggesting that they are true associations. There is however a huge gap between the number of robust risk loci and our understanding of the underlying molecular mechanisms.

GWAS results are typically reported as a list of loci labelled by the most strongly associated variant which is often referred to as the lead, index or top variant. The lead variant has however in most instances no biological function and is instead in LD with the actual causal variant (Schaub et al., 2012). Identification of the causal variant is challenging since GWAS signals typically comprise multiple genetic variants, sometimes hundreds, in high LD. The list of putative causal variants may be refined by statistical fine-mapping efforts which employ dense genotyping arrays to perform more complete analyses of the identified risk regions. Such fine-scale genotyping does however require large sample sizes to get sufficient statistical power, which together with the cost of the locus specific dense genotyping arrays require substantial resources (Spain and Barrett, 2015). Furthermore, sequence information alone is insufficient to prioritize between genetic variants in perfect LD.

Identification of the causal gene(s) at a GWAS locus is another important part of elucidating the underlying molecular mechanisms. GWAS loci have generally been named after the nearest gene(s) to the lead variant. Proximity is however not a good or sufficient measure of likely causality since genetic variants may affect distant genes. The majority of disease-associated variants identified by GWASs lie in the non-coding part of the genome, which complicates their functional characterization and the process of assigning a target gene (Maurano et al., 2012). Improved insight into the function and organization of the non-coding genome is crucial to the interpretation of GWAS findings.

1.3.1 The non-coding genome and its role in gene regulation

The predominant part of the genome is non-coding, with less than 2% constituting protein-coding sequence (ENCODE Project Consortium, 2012). The non-coding genome was once considered to be redundant and non-functional “junk” DNA. This view has however radically changed following the advancement of new technologies that have enabled extensive mapping of the non-coding genome, uncovering that it plays a crucial role in regulation of gene expression.

The complex interplay between trans-acting regulatory proteins and non-coding cis-regulatory elements, such as enhancers and promoters, control in which cells, at what time points and at what level genes are expressed. Promoters are located in immediate proximity to the transcription start site and this is where the transcriptional machinery binds to initiate transcription (Lenhard et al., 2012). Enhancers are located more distally from the transcription start site and provide additional input which is essential to ensure precise control of the transcriptional pattern. The classical definition of enhancers is that they are cis-regulatory elements that increase the transcription of genes and function independently of the orientation and position relative to the transcription start site (Banerji et al., 1981). Transcription factors are regulatory proteins that recognize and bind to short specific DNA sequences, referred to as motifs, at cis-regulatory elements. Over 1600 different transcription factors have so far been identified and to a variable degree characterized (Lambert et al., 2018). Importantly, cis-regulatory elements are typically bound by a combination of transcription factors, with the timing and spatial arrangement of transcription factor binding defining its regulatory function.

Initiation of transcription is controlled by the binding of transcription factors to enhancers and promoters which induce, via interaction with cofactors, the assembly of the RNA polymerase II machinery at the promoter region. To accomplish this, enhancers are brought into close proximity of target promoters through looping of the intervening DNA (Sanyal et al., 2012). Thus, instead of employing a linear representation of the genome, gene regulation needs to be viewed in a three-dimensional context.

1.3.2 Chromatin architecture and epigenetics

The three-dimensional organization of the genome is achieved through chromatin folding. Chromatin is the complex formed between DNA and proteins that enable tight packing of the genetic information. This is required to fit the entire genome into the microscopic nucleus. The nucleosome is the basic repeating core building block of chromatin, composed of a section of DNA wrapped around a core unit of histone proteins. Nucleosomes fold into shorter and thicker fibers, which form loops and are further compressed and tightly coiled into the chromatid of a chromosome. The occupancy of nucleosomes across the genome is dynamic, creating an accessibility continuum that ranges from closed chromatin to accessible chromatin (Klemm et al., 2019). Such remodeling of chromatin is necessary to make the tightly packed DNA accessible to regulatory proteins.

Chromatin accessibility, together with covalent modifications of histones and DNA, and the higher-order chromatin architecture, are important modes of epigenetic regulation. Epigenetics has been defined as "the study of mitotically (and potentially meiotically) heritable alterations in gene expression not caused by changes in DNA sequence" (Waterland, 2006). An even broader understanding of the term is often in use which does not require the epigenetic alteration to be heritable. The complete collection of epigenetic changes along the genome, referred to as the epigenome, varies between different cell types and represents a second dimension to the genomic sequence (Rivera and Ren, 2013). This enables cell type-specific gene expression patterns that are necessary to shape cell identity and produce the large variety of different human cell types.

The dynamic transition between the different states of chromatin is enabled by covalent modifications of histone proteins and DNA. These transient modifications may alter the structure and function of chromatin, enacting a pivotal role in the regulation of transcription, DNA repair and replication (Tessarz and Kouzarides, 2014). Methylation at the carbon-five position on cytosine residues is a widely studied nucleotide modification. More than ten post-translational modifications of histone tails have so far been identified, of which methylation and acetylation are best characterized (Chiarella et al., 2020). Cell type-specific transcriptional regulation is further impacted by the three-dimensional chromosomal structure. Chromatin looping forms the necessary contacts between distal enhancers and their target promoters. Studies of the three-dimensional genome organization show a complex and flexible interaction network where cis-regulatory elements typically have several interaction partners (Sanyal et al., 2012).

1.3.3 Mapping the epigenome

Assaying of epigenetic marks may be used to annotate the non-coding genome through the identification of putative regulatory elements. Technological advances, such as high-throughput sequencing (HTS), have enabled assaying of epigenetic features genome-wide. This has led to a shift from the characterization of single enhancers or promoters to modeling of the full regulatory genome in selected cell types.

DNA methylation is an epigenetic mark that has been extensively assayed. It is functionally linked to gene repression and is thus frequently described as a “silencing” epigenetic mark. Comprehensive high-throughput methods enable the construction of detailed whole-genome maps of DNA methylation. The ability to study complete methylomes shows that not only methylation near transcription start sites, but also in gene bodies and regulatory regions such as enhancers may be functionally important (Jones, 2012).

Profiling of chromatin accessibility can be used to identify a repertoire of putative regulatory regions across the genome. Genomic regions where transcription factors are bound lack nucleosomes and are thus preferentially digested by enzymes that modify DNA or are more easily fragmented by sonication (Nord and West, 2020). A principal method for measuring chromatin accessibility is DNase I hypersensitive sites sequencing

(DNase-seq) where the endonuclease DNase I is used to digest nucleosome-depleted DNA, followed by identification of these DNA fragments by HTS (Boyle et al., 2008). Assay for Transposase Accessible Chromatin followed by sequencing (ATAC-seq) is a more recent approach for chromatin accessibility profiling where hyperactive Tn5 transposase is utilized to simultaneously cut and ligate adapters for HTS at regions of open chromatin (Buenrostro et al., 2015).

Open chromatin is however non-specific when it comes to the function of the identified putative regulatory regions. Instead, or as a complement, analysis of histone modifications may be used to predict the function of regulatory elements. Patterns of specific post-translational histone modifications correlate with the functional state of the associated chromatin and constitute what is known as the “histone code” (Jenuwein and Allis, 2001, Strahl and Allis, 2000). For example, enhancers are marked by monomethylation of histone H3 at lysine 4 (H3K4me1), while promoters are marked by trimethylation at the same site (H3K4me3) (Heintzman et al., 2007). Another useful histone mark is the acetylation of histone H3 at lysine 27 (H3K27ac), which separates enhancers that are active from those that are inactive/poised (Creyghton et al., 2010).

Chromatin immunoprecipitation sequencing (ChIP-seq) is the most common method used to map histone modifications and is also used to identify the direct binding of transcription factors and other regulatory proteins to DNA. In ChIP-seq experiments, DNA binding proteins are first cross-linked to DNA, and then the DNA is sheared by sonication or enzymatic digestion. Then, specific antibodies directed against the target protein are used to precipitate the co-associated DNA fragments. HTS allows for genome-wide mapping of the assayed protein-DNA interactions (Nord and West, 2020, Johnson et al., 2007). ChIP-seq is considered as the gold standard for identification of transcription factor binding since it assays direct *in vivo* binding of the protein in the tested cell type.

To gain further insight into the regulatory genome, it is necessary to explore contacts made between cis-regulatory elements. Chromosome conformation capture (3C) and its high-throughput derivatives such as 4C, 5C and Hi-C, are used to study chromosomal contacts at different scale, ranging from single enhancer-promoter contacts to genome-wide contact maps (Sanyal et al., 2012, Dekker et al., 2002, Lieberman-Aiden et al., 2009). These methods are based on the concept that genomic regions that are closely positioned are more likely to be cross-linked and ligate to each other (Vermunt et al., 2019). Identification of enhancer-promoter interactions may determine which genes that are potentially regulated by a specific enhancer, and could also identify previously unknown enhancers that possibly contribute to regulation of a particular gene (Snetkova and Skok, 2018).

Epigenetic assays may be complemented by studies coupling gene expression levels with genetic variation. Genetic variants that are associated with altered expression level of a particular gene in a particular tissue are known as expression quantitative trait loci (eQTLs). RNA-sequencing in densely genotyped individuals allows for genome-wide mapping of eQTLs (Pickrell et al., 2010). Such eQTL datasets may aid in surveys of gene regulatory mechanisms, as well as in annotation of genetic variants.

The cell type and tissue specificity of the epigenome presents a paramount challenge to mapping of regulatory elements. This is because the different epigenetic marks need to be assayed in numerous cell types, which is extremely labor-intensive. To address this challenge, large international efforts such as NIH Roadmap Epigenomics Project and Encyclopedia of DNA Elements (ENCODE) Project, have been launched to systematically map epigenetic marks and develop new epigenomic technologies (Bernstein et al., 2010, Satterlee et al., 2019, ENCODE Project Consortium, 2012). The Genotype-Tissue Expression (GTEx) Consortium has analyzed eQTLs in more than 40 human tissues in hundreds of individuals (Battle et al., 2017). Such large-scale projects have generated human reference epigenomes and QTL maps from hundreds of cell types that have been made publicly available, providing a valuable resource to the scientific community.

1.3.4 Integration of GWAS findings with regulatory annotations

Disease-associated variants derived from GWAS studies in a range of complex disorders are more frequently located in regulatory annotations such as different histone marks, measures of chromatin accessibility and eQTLs (Trynka et al., 2013, Maurano et al., 2012, Nicolae et al., 2010). This suggests that alterations to regulatory elements, and the ensuing changes to gene expression, are important mechanisms of action by which genetic variants influence disease risk. Furthermore, the enrichments are typically more prominent in tissues or cell types of biological relevance to the disease (Kundaje et al., 2015, Trynka et al., 2013). For example, genetic variants associated with several cardiac traits (PR heart repolarization interval, blood pressure and aortic root size), are enriched in heart tissue enhancers (Kundaje et al., 2015). Enrichment analyses may however also nominate less obvious cell type-disease connections with a potential pathogenic role to be further explored.

The observation that GWAS signals tend to localize to non-coding regulatory regions indicates that alterations to transcription factor binding may be a mechanism of action. Variation in transcription factor-DNA binding is indeed believed to play an important role in mediating phenotypic diversity and has been linked to disease in several studies (Karczewski et al., 2013, Cowper-Sal lari et al., 2012). Transcription factor binding sites may be disrupted or created by genetic variants located in the transcription factor recognition motif, which may lead to changes in gene expression levels and ultimately in phenotype (Deplancke et al., 2016). This may be exemplified by the discovery identifying how a genetic variant in the *FTO* (fat mass and obesity) gene lead to disease by altering a conserved motif comprising the binding site for ARID5B (Claussnitzer et al., 2015). The “FTO story” is also a striking example of how GWAS signals may be integrated with functional genomic annotations in an effort to uncover the underlying molecular mechanisms. GWASs have consistently identified intronic *FTO* variants to be associated with elevated body mass index, leading to mechanistic studies exploring the role of *FTO* in obesity (Dina et al., 2007, Frayling et al., 2007). However, with the use of epigenomic- and gene expression data, later studies have shown that the associated variants are located in a regulatory element that controls the expression of the two distal genes *IRX3* and *IRX5* (Claussnitzer et al., 2015, Smemo et al., 2014). One risk allele disrupts the binding site of ARID5B, which results in over-expression of *IRX3* and *IRX5*. This ultimately leads to decreased mitochondrial energy generation and increased triglyceride accumulation in primary human adipocytes (Claussnitzer et al., 2015).

Functional characterization of GWAS loci is however the exception rather than the rule. A contributing factor to this may be the complexity of how genetic variants influence transcription factor binding and gene expression. Only a minority of variable transcription factor-DNA binding events involves nucleotide changes in the respective transcription factor recognition motif (Reddy et al., 2012). Genetic variants may affect binding of the studied transcription factor by altering proximal motifs, and even more distally located variants may have an effect through alterations of chromatin state or confirmation (Deplancke et al., 2016). In addition to effects on gene transcription, genetic variants may also alter gene expression levels through post-translational processing such as mRNA splicing and stability (Gallagher and Chen-Plotkin, 2018).

Annotation of the non-coding genome and enhanced insight into gene regulatory mechanisms will be essential to the process of bridging the gap between genetic discoveries and our understanding of underlying pathobiological processes in PD. Integration of GWAS results with cell type-specific functional genomic annotations may uncover pathogenic molecular mechanisms which in turn may lead to new therapeutic interventions.

2. Aims of the study

The overall aim of this study was to expand our understanding of the genetic contribution to sporadic PD. Enhanced knowledge of the genetic architecture of PD might lead to valuable insight into pathological mechanisms that are needed to develop new therapeutic strategies. More specifically, the aim of the presented work was to clarify the involvement of selected genetic variants to the risk of developing PD and in modifying AAO, as well as to untangle an important GWAS signal in PD. Furthermore, we aspired to identify trait-relevant biological mechanisms from the existing collection of GWAS signals in PD.

Aim of Paper 1: To study the complete *GBA* gene in PD patients and to investigate whether coding *GBA* variants may be driving reported GWAS signals.

Aim of Paper 2: A variant in the *DNM3* gene has been reported as a genetic modifier of AAO in *LRRK2*-associated PD. We sought to explore whether genetic variation in *DNM3* has an effect on AAO in idiopathic PD.

Aim of Paper 3: To identify transcription factor networks contributing to PD risk by integrating PD GWAS signals with open chromatin sites in brain coupled with transcription factor motif analysis.

3. Summary of results

Paper 1:

The GBA variant E326K is associated with Parkinson's disease and explains a genome-wide association signal

Coding mutations in the *GBA* gene have been identified as important genetic risk factors for PD. In addition, GWASs have identified associations with PD at the *SYT11-GBA* region on chromosome 1q22, but the relationship to coding *GBA* variants has been unclear. In Paper 1, we analyzed sequencing data covering all coding exons of the *GBA* gene in 366 Norwegian PD patients. We identified six rare mutations (1.6%) and two low-frequency coding variants in *GBA*. The two low-frequency coding variants E326K and T369M were genotyped in 786 patients and 713 controls from Norway and Sweden. We found that E326K was significantly more frequent in patients compared to controls, while there was no clear association between T369M and disease. To investigate whether E326K or T369M may be driving the reported nascent GWAS signals, two independent association signals within the *SYT11-GBA* locus were genotyped in the same patients and controls. We replicated the association between the primary GWAS hit and disease status, while the secondary GWAS hit had similar frequency in patients and controls. Evaluation of LD between the four genotyped variants showed that E326K and the primary GWAS hit are in very high LD (r^2 0.95). In conclusion, our results confirm that the *GBA* variant E326K is a susceptibility allele for PD and suggest that E326K may fully account for the primary association signal observed at chromosome 1q22 in previous GWASs of PD.

Paper 2:

No evidence for DNM3 as genetic modifier of age at onset in idiopathic Parkinson's disease

In this study, we analyzed the effect of *DNM3* variants on AAO in idiopathic PD. The *DNM3* variant rs2421947 had been identified as a modifier of AAO of PD in *LRRK2* G2019S carriers, with GG homozygotes reported to have a median AAO 12.5 years younger than CC homozygotes (Trinh et al., 2016). Since genetic variation at the *LRRK2* locus is also part of the genetic background of idiopathic PD, we wanted to test whether the association with *DNM3* reported in *LRRK2* mutation carriers was transferable to the much wider group of PD patients with no known highly penetrant disease causing mutation. We studied rs2421947 in a total of 5918 PD patients from seven different datasets. There was no significant association between rs2421947 and AAO in any of the individual studies. Meta-analysis of the seven studies was also non-significant. The analysis was extended to include all common variants within the *DNM3* gene and the flanking genomic region, of which none showed a significant association with AAO of PD. In conclusion, we found no evidence of a modifying effect of *DNM3* variants on AAO in idiopathic PD.

Paper 3:

Integrative analysis identifies bHLH transcription factors as contributors to Parkinson's disease risk mechanisms

While PD GWASs have identified multiple genetic association signals, the translation into underlying biological mechanisms has lagged behind. Emerging genomic functional annotations may be integrated with GWAS results to identify relevant cell types and molecular mechanisms important to PD pathogenesis. In the third paper presented in this thesis, we integrated association signals from the most recent PD GWAS with publicly available ATAC-seq data coupled with transcription factor motif analysis in an effort to identify transcriptional networks contributing to PD risk. We found that PD risk variants significantly overlap open chromatin sites in neurons of the superior temporal cortex, indicating that these cell types mediate genetic risk for PD. Neurons from other cortical regions approached the significance threshold, suggesting that a broader range of cortical regions may be implicated in PD risk. *In silico* motif analysis performed in neurons of the superior temporal cortex showed that PD risk variants concentrate in sites of open chromatin targeted by members of the basic helix-loop-helix (bHLH) transcription factor family, pointing to an involvement of these transcriptional networks in PD risk mechanisms.

4. Methodological considerations

4.1 Study population

In Paper 1, genetic variants at the *GBA* locus were studied in Norwegian and Swedish participants included from Oslo University Hospital, the ParkWest study in western Norway and the NYPUM study at Umeå University Hospital. A total of 786 PD patients and 713 controls were included in the analyses. All PD patients were examined by a neurologist and diagnosed according to either the revised United Kingdom PD Society Brain Bank criteria (Oslo and Umeå) or the Gelb criteria (ParkWest). The use of diagnostic criteria is important to ensure a clear definition of the included phenotype by increasing the accuracy of the diagnosis when no objective test exists. However, even when standard diagnostic criteria are applied and the clinical diagnosis is performed by experts in neurology, the diagnostic accuracy of PD is not absolute (Rizzo et al., 2016). Misclassification of disease may reduce power of genetic association studies and could have an influence on our study.

While the low-frequency *GBA* variants (E326K and T369M) and GWAS signals were genotyped in all patients and controls, sequencing was only performed in patients from Oslo University Hospital. The Department of Neurology at Oslo University Hospital is a tertiary care center for movement disorders where patients are referred for second opinion and advanced treatment. A large proportion of the sequenced patients have thus been treated with deep brain stimulation. Cognitive impairment is an exclusion criterion when evaluating PD patients for deep brain stimulation. We may have selected against carriers of *GBA* mutations since this group of PD patients have been reported to have an accelerated cognitive decline. This may be a contributing factor to the low *GBA* mutation carrier frequency of 1.6% that we find in these patients. The ParkWest study and NYPUM study are population-based cohort studies of incident PD patients and may thus be more representative of the PD population in general.

Control subjects should be representative of the population from which cases are obtained. In our study, control subjects were selected among spouses of patients, outpatients in primary care and healthy volunteers. All were without neurological disease and known parkinsonism among first degree relatives. In case-control studies, some of the controls may be expected to develop the tested disease later in life, and this could reduce the power of the study. This does however have larger implications in studies of more frequent phenotypes. PD is relatively rare, also in higher age groups, so only a minor fraction of the controls included in our study may be expected to develop PD. Inclusion of controls in higher age groups may reduce the chance of misclassification. Controls included in our study have a mean age at inclusion in the mid-60s from each of the study sites.

In Paper 2, we studied individual-level genotypes from seven different datasets including a total of 5918 PD patients. Genetic studies of PD from Oslo University Hospital and Mayo Clinic Jacksonville are in-house datasets, while the five remaining datasets were accessed from the Database of Genotypes and Phenotypes (dbGaP) or the Parkinson's Progression Markers Initiative (PPMI) (Tryka et al., 2014, Nalls et al., 2016). Patients included from Oslo University Hospital in Paper 2 largely overlap patients included from Oslo in Paper 1. The included datasets were selected based on having individual genotype information in PD patients with a reported AAO. All datasets have genome-wide genotypes, except patients from Mayo Clinic Jacksonville where only genotypes for rs2421947 were available. The datasets consisted mainly of participants of Caucasian non-Hispanic ethnicity and we filtered out the few patients that had another ethnicity. Since this was a study of idiopathic PD, carriers of *LRRK2* G2019S and other mutations causing monogenic forms of PD were excluded from analysis. Demographic characteristics of included datasets are presented in Paper 2 (Table 1).

AAO was mainly defined by patient reports of the initial manifestation of parkinsonian symptoms, with the exception of patients included from Mayo Clinic Jacksonville where age at diagnosis of PD was used as onset age. Self-reported AAO is a subjective measure that may be prone to recall bias. However, the correlation between reported age at symptom onset and age at diagnosis of PD has been found to be high (Blauwendraat et al., 2019). In Paper 3, we analyzed genome-wide significant risk signals accessed from a meta-analysis of 17 datasets from European ancestry PD GWASs (Nalls et al., 2019). At the time of our analysis, this was the largest genetic study of PD that had been performed.

4.2 Sequencing

The development and continuous evolvement of sequencing technologies have had a tremendous impact on biological research. The field of DNA sequencing was initiated by Sanger sequencing in 1977, which became the dominating sequencing technology for several decades to come (Sanger et al., 1977). Sanger sequencing is based on the use of di-deoxynucleotides (ddNTPs) in addition to the normal nucleotides found in DNA. ddNTPs are modified nucleotides that, when incorporated into the growing DNA strand, prevent the addition of further nucleotides. Dye-labeled ddNTPs are used to generate DNA fragments that terminate at different points. The DNA fragments may then be separated on the basis of size by gel electrophoresis. The DNA sequence can then be detected one nucleotide at a time based on the color of the dye and shown in a chromatogram. Multiple improvements have been made to Sanger sequencing leading to the development of increasingly automated DNA sequencing machines (Heather and Chain, 2016).

In the early 2000s, a new wave of sequence-technologies emerged. These technologies are normally referred to as high-throughput or next generation sequencing (HTS or NGS). Sanger sequencing gives high-quality sequence of short stretches of DNA, but the low throughput makes it unsuited for large-scale sequencing projects. A major technical advantage in HTS over Sanger sequencing is parallelization of the reaction, which allows for a greatly increased sequencing throughput from multiple samples. HTS has revolutionized the field of genomics and to a large degree replaced Sanger sequencing, enabling large-scale sequencing at much reduced costs. However, Sanger sequencing is still widely used in smaller sequencing-projects and for validation of HTS results.

In Paper 1, we used data from targeted pooled HTS of PD patients to identify coding *GBA* variants, which were then validated by Sanger sequencing. Targeted capture of all exons of 71 genes relevant to PD, including *GBA*, was combined with deep sequencing of DNA pools in an experiment performed in our laboratory by Lasse Pihlstrøm and colleagues. A subset of the genes sequenced in this experiment was analyzed in a study evaluating the performance of the targeted pooled HTS design (Pihlstrom et al., 2014). An advantage of pooling DNA are the reduced costs related to targeted enrichment and sequencing, as well as a reduction in the total amount of manual workload (Pihlstrom et al., 2014, Anand et al., 2016). Sequencing of pooled samples thus represents a cost- and time- effective strategy that may open up for genetic studies in large cohorts that would otherwise have been too resource-demanding. The pooled sequencing design does however also present some challenges related to accurate variant calling and allele frequency estimation (Anand et al., 2016). Lasse Pihlstrøm and colleagues tested the experiment's ability to detect rare variants where only one allele is present in a pool, finding a sensitivity of 97% (Pihlstrom et al., 2014). For the two variants that sequencing missed, the sequencing depth was below 80x. In our analysis of the *GBA* gene we therefore excluded pools with a sequencing depth below 80x at the relevant position. The challenge of correctly distinguishing true nonreference alleles present at low frequencies from the background of sequencing error may also lead to false positives (Anand et al., 2016). The sequenced pools that we analyzed were relatively small, each containing DNA from 10 individuals. Smaller pools favor detection of rare variants, while the accuracy of frequency estimations of common variants benefits from larger pools where inaccurate input of DNA from individuals even out. Evaluation of the targeted pooled HTS experiment does show that the number of nonreference alleles are called incorrectly in some cases (Pihlstrom et al., 2014).

Due to these challenges, positive findings from pooled sequencing should in many instances be followed up by additional methods. In our study of the *GBA* gene, we validated variants detected by pooled sequencing with Sanger sequencing off all individuals in the pool of detection. *GBA* is a large gene containing 11 exons and 7.6 kb of sequence (Horowitz et al., 1989). Consequently, complete Sanger sequencing of the entire gene is very laborious, and many studies have limited the analysis to genotyping the most common mutations or sequencing of selected exons. In our study design, the use of pooled sequencing data advantageously restricts Sanger sequencing to specific exons in selected pools where variants are detected. Sequencing of *GBA* is complicated by the presence of a highly homologous pseudogene (*GBAP1*) in close proximity to the *GBA* gene, resulting in complex gene-pseudogene rearrangements (Horowitz et al., 1989, Hruska et al., 2008). To avoid amplification of the pseudogene, we performed Sanger sequencing with primer sequences designed to DNA regions exclusive to the *GBA* gene (Neumann et al., 2009).

Since variants were validated by Sanger sequencing, false positives or incorrect allele count is not likely. However, we cannot exclude the possibility that some existing variants were not detected in the pooled-sequencing experiment and thus missed in our study. Since the pooled-sequencing experiment has been reported to have a high sensitivity, the number of missed variants is not expected to be high. It should although be noted that the estimated sensitivity is based on a limited number of investigated variants and would have been more robust if more variants had been assessed. Also, HTS analysis of *GBA* is challenging due to the presence of *GBA-GBAP1* complex rearrangements leading to possible misdetection of recombinant mutations (Zampieri et al., 2017). N370S and the recombinant mutation L444P had previously in our laboratory been genotyped in Norwegian patients largely overlapping the patients analyzed with pooled HTS, identifying the same N370S carrier and no L444P carriers as in our study. Furthermore, we did look through the Sanger sequencing reads without finding any additional mutations. However, this only covers a small fraction of the total sequence.

4.3 Genotyping

While sequencing reads every nucleotide in the covered genomic region, it is often the situation that we know exactly which variants we want to test. The variants of interest can then instead be genotyped with polymerase chain reaction (PCR)-based genotyping assays. This may be a good choice of method when the number of variants to be genotyped is limited. In Paper 1, genotyping of *GBA* variants and GWAS signals was performed by either KASP or TaqMan genotyping assays on a Viia7 instrument (Life Technologies, Foster City, CA, USA). Both assays are based on PCR-reactions where amplification of the region containing the genetic variant is detected by a fluorescence signal. Allele discrimination is achieved by allele-specific probes or primers containing distinct fluorescent dyes that bind to the target variant. TaqMan and KASP make use of a phenomenon called fluorescence resonant energy transfer (FRET) which occurs when the emission of a fluorescent dye is effectively captured and reduced by the presence of a nearby quencher dye. In KASP assays the fluorescent dye is attached directly to the tail of a set of allele-specific primers and is no longer quenched when amplification creates the complement strand that the primer tail binds to, resulting in a detectable signal. In TaqMan assays, the fluorescent dye is instead linked to an allele-specific probe which is degraded during PCR by 5’-nuclease activity of the Taq polymerase as it extends the

DNA from the PCR primers. This separates the fluorophore from the quencher and fluorescence is emitted.

Results from the genotyping experiment is depicted in an allelic discrimination plot and genotype calls are assigned to each sample according to its position on the plot (Figure 3). Positive controls may aid in cluster calling for the analysis algorithms. For the two *GBA* variants E326K and T369M, positive controls were available due to previous sequencing. We did not have any samples with known genotype for the primary and secondary GWAS signals, so the initial genotyping experiments were run without positive controls for these two assays. However, positive controls were included when genotypes containing the minor allele were identified. All genotyping experiments were run with negative controls.

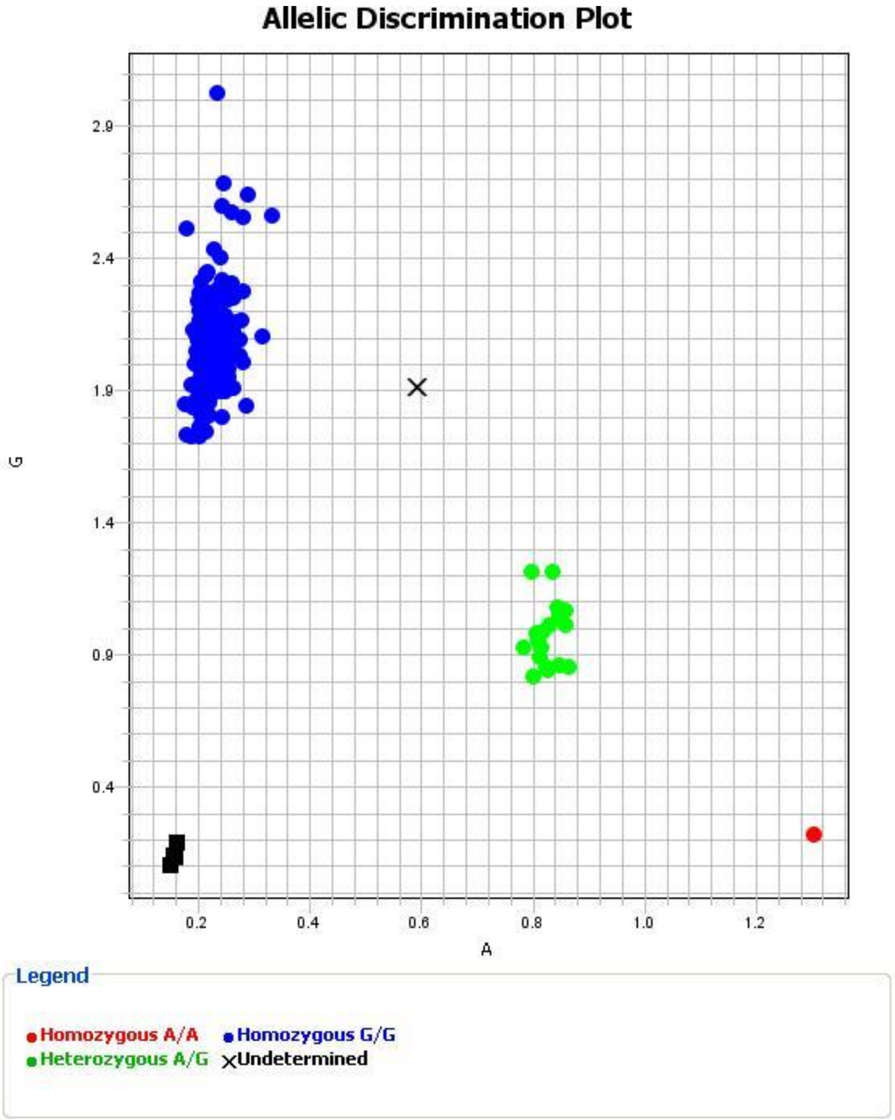


Figure 3. *Allelic discrimination plot from KASP genotyping. X-axis and Y-axis represent the fluorescence signal of the dye attached to the allele-specific-primers. Samples of the same genotype will have similar levels of fluorescence and will therefor cluster together on the plot. Black squares represent negative controls.*

Genotype call rate is used as a quality control measure of genotyping assays. The call rate was above 0.98 for each individual variant in our analyses, which is considered satisfactory. Hardy-Weinberg equilibrium (HWE) is another quality parameter used in analysis of genotype data. The HWE is a principal of population genetics stating that genotype frequencies in a population remain constant from one generation to the next in the absence of disturbing factors. Significant deviations from HWE predictions may indicate genotyping errors. In patients, however, deviation from HWE may also be a reflection of an association between a genotype and disease (Namipashaki et al., 2015). We tested for HWE in controls for all genotyped variants, observing no significant departure.

When the number of variants to be genotyped is large, such as in fine-mapping of specific genetic loci and in genome-wide genotyping, array-based genotyping technologies is the method of choice. The basic principle of genotyping microarrays is hybridization of immobilized complementary DNA probes to fragmented nucleotide sequences containing the variant site and subsequent detection of the hybridization events. This is a high throughput method capable of detecting hundreds of thousands of variants on the surface of oligonucleotide chips. Genotyping microarrays have been used to detect common genetic variants at genome-wide scale and are the basis of GWASs.

In Paper 2, we used genome-wide genotype data from in-house Oslo samples and additional GWAS datasets. Oslo samples were genotyped using the Illumina Infinium OmniExpress v.1.1 array. An overview of the genotyping arrays used by all included datasets is provided in Paper 2 (Table 1). In this study, imputation and pre-imputation quality control procedures were performed by Lasse Pihlstrøm, while Victoria Berge-Seidl conducted the genetic analyses. Quality control procedures is an important step in the analysis of genome-wide genotyping data to prevent errors that may bias the outcome of association tests. Genotyping and genotype-calling may be subjected to technical errors, resulting in variants and samples with low quality. Another potential source of biases that needs to be addressed is cryptic structures in the studied population, meaning similarities between individuals that are independent of the studied phenotype (Coleman et al., 2016). Quality control steps performed in our study include filtering of variants based on genotype-rate and deviations from HWE. Individuals with high genotype missingness, excess heterozygosity (may indicate sample contamination or inbreeding), evidence of cryptic relatedness, and those identified as ancestry outliers or having inconsistencies in assigned and genetic sex (may indicate sample mix-up) were removed (Marees et al., 2018). There is a risk of recruiting related individuals unknown to the investigator and the same individual may have been included in different datasets. This may influence association statistics. We assessed cryptic relatedness across studies to identify duplicates and related participants, which were then removed.

Imputation of genotypes is the prediction of genetic variants that were not included in the genotyping array and thus not directly assayed. The missing genotypes are statistically inferred from complete haplotype information in a reference panel. Imputation is performed to increase the statistical power and signal resolution, and is also useful to combine data from GWASs using different genotyping arrays (Hoffmann and Witte, 2015). In our study, all datasets were imputed using the Michigan Imputation Server with reference data from the Haplotype Reference Consortium (Das et al., 2016, McCarthy et al., 2016).

4.4 Prediction of transcription factor binding

In paper 3, we integrated genetic risk signals with predicted transcription factor binding sites in an effort to explore the involvement of transcriptional networks in PD. We made use of, and benefitted from, the growing amount of publicly available genomic datasets. When analyzing transcription factor binding, it is highly important to take cell type- and tissue-specificity into account. ChIP-seq allows for genome-wide detection of *in vivo* transcription factor binding and has been used to assay hundreds of transcription factors in multiple cell types. However, due to high experimental efforts and costs, only a fraction of transcription factor-cell type combinations has so far been assayed. Furthermore, transcription factor ChIP-seq data from neuronal cell types is particularly scarce.

Computational models of transcription factor binding specificities, such as position weight matrices (PWMs), may be used to scan the genome to identify putative transcription factor binding sites. PWMs describe the probability of a given nucleotide’s occurrence at each position in the binding motif of a transcription factor derived from observed transcription factor-DNA interactions (Inukai et al., 2017). These interactions have been obtained from *in vitro* assays (SELEX or protein binding microarrays) or from ChIP-based experiments (Fornes et al., 2020). Binding motifs represented as PWMs for a large number of transcription factors are collected in databases such as JASPAR, HOCOMOCO and CisBP (Fornes et al., 2020, Kulakovskiy et al., 2018, Weirauch et al., 2014). PWMs are limited by the assumption that positions within the motif are independent, which is not always true (Boeva, 2016). More complicated models have been developed to account for inter-dependencies between base pairs and other complexities, resulting in better performance for some transcription factors. However, in most cases the improvements are minor or not detectable and PWMs remain the most widely used computational model for analysis of transcription factor binding (Lambert et al., 2018).

Importantly, transcription factors only occupy a small proportion of the genomic sequences matching to their consensus binding sites. Transcription factor-DNA recognition is impacted by additional features such as sequence context, accessibility of chromatin and interactions among transcription factors (Wang et al., 2012, Inukai et al., 2017). Integrating cell type-specific experimental data, such as DNase-seq, has been shown to enhance transcription factor-DNA binding predictions (Pique-Regi et al., 2011, Sherwood et al., 2014). When setting out to predict transcription factor binding sites we therefore chose to combine transcription factor motif analysis with cell type-specific epigenomic annotations characterizing open or active genetic regions. A flow-chart displaying the analytical steps of this study is shown in Paper 3 (Fig 1).

4.4.1 Genomic annotations

In our study, we analyzed maps of open chromatin in neurons and non-neurons across 14 distinct brain regions of five individuals. The data was downloaded from the online database Brain Open Chromatin Atlas (BOCA) and has been described in a paper by Fullard and colleagues (Fullard et al., 2018). In generation of this dataset, fluorescence-activated nuclear sorting (FANS) was combined with ATAC-seq to create cell type-specific maps of open chromatin, distinguishing neuronal cells from non-neurons.

ATAC-seq has emerged as a powerful approach for genome-wide profiling of chromatin accessibility. One major advantage of ATAC-seq over other methods profiling chromatin accessibility is that it has a rapid and simple protocol containing few experimental steps. Another major benefit is the high sensitivity, enabling analysis of as few as 500-5000 cells (Buenrostro et al., 2013). This makes ATAC-seq especially suitable for situations where the starting number of input cells is low, such as cell populations sorted by fluorescence-activated cell sorting or its derivative FANS (Shashikant and Ettensohn, 2019).

ATAC-seq data needs to undergo thorough and sequential analysis where a variety of analytical tools are used. The initial steps typically include quality control and trimming of raw reads, mapping reads to the reference genome, followed by post-alignment processing and quality control (Yan et al., 2020). Then, peak calling is performed to identify areas in the genome that have been enriched with aligned reads and thus signify open and accessible chromatin. Fullard et al. used the peak calling algorithm model-based Analysis of ChIP-seq (MACS, v2.1) to generate ATAC-seq peaks representing open chromatin regions (OCRs). We used the chromosomal coordinates of these ATAC-seq peaks in our analyses.

PD is a neurodegenerative disorder with intra-neuronal protein inclusions reported in multiple regions of the brain. This guided our choice of tissue and cell type, leading us to focus the analysis on transcription factor binding in brain neurons. Targeted isolation of neurons from the non-neuronal cell population is a major strength of the dataset we analyzed. Another strength is the comprehensive coverage of the brain, which allowed for comparison of open chromatin in neurons across different cortical and subcortical regions. At the time of our analysis, this was to our knowledge the largest dataset of open chromatin in human brain with cell type-specificity. However, this dataset does not include substantia nigra, which naturally would have been an interesting region to study in the context of PD. We were not able to find other available data from ATAC-seq or comparable assays detecting accessible chromatin in human substantia nigra neurons.

Cellular resolution of genomic annotations is important since cellular heterogeneity may be masking signals. The brain is a tissue of immense complexity, which contains a heterogeneous mixture of cell types exhibiting different regulatory features (Darmanis et al., 2015, Girdhar et al., 2018, Fullard et al., 2018, Rizzardi et al., 2019). In analysis of bulk brain tissue, cell type-specific regulatory elements in neurons or glial subtypes may be diluted due to measurement of an average signal across a heterogeneous population of cells (Reynolds et al., 2019). Although Fullard et al. increased the cellular resolution by separating neurons from non-neurons, these components still represent broad categories of cell types. The non-neuronal cell population consists of astrocytes, oligodendrocytes, microglia, and epithelial cells. The lack of distinction between different glial subtypes contributed to our choice of not including this cellular fraction in the analysis. However, also the neuronal cellular component displays some degree of heterogeneity. There is a diverse set of neuronal cell types in the human brain with distinct patterns of connectivity, synaptic properties and expression profiles (Ecker et al., 2017, Lake et al., 2016).

4.4.2 De novo motif discovery

Transcription factors targeting binding motifs that are enriched in a set of regulatory regions in a cell may be regarded as candidate transcriptional regulators of that cell. *De novo* motif discovery methods aim at identifying over-represented motifs in a given set of sequences. This is however a difficult computational task and motif-finding algorithms have suffered from a high rate of false-positives (Lihu and Holban, 2015). Thus, to identify likely functional transcription factors in our cell type of interest, we employed two different motif discovery tools and performed the analyses in parallel. Analyses were performed with the softwares HOMER and MEME-ChIP that are both widely used and can handle large-scale datasets (Heinz et al., 2010, Ma et al., 2014). HOMER identifies motifs that are enriched in the target sequences relative to GC matched background sequences. In specifying the length of motifs to be found, we used the default setting of 8, 10 and 12 base pairs long motifs. MEME-ChIP is an ensemble tool that incorporates two different algorithms for motif discovery, multiple EM for motif elicitation (MEME) and discriminative regular expression motif elicitation (DREME). MEME is able to find relatively long motifs, but is highly labor intensive and can only analyze a very small fraction of the sequences in our dataset. *De novo* motifs identified by the MEME algorithm were therefore regarded as less relevant in our analysis of OCRs. DREME discovers short motifs up to 8 base pairs, is more computationally efficient and able to analyze all sequences in our dataset. HOMER identified 22 enriched motifs that were all included in further analyses. MEME-ChIP identified a much larger number of enriched motifs and further analyses were limited to the 25 most significant motifs, which were all identified by the DREME algorithm. Both HOMER and MEME-ChIP match *de novo* motifs to databases containing known motifs to identify candidate transcription factors (Figure 4).

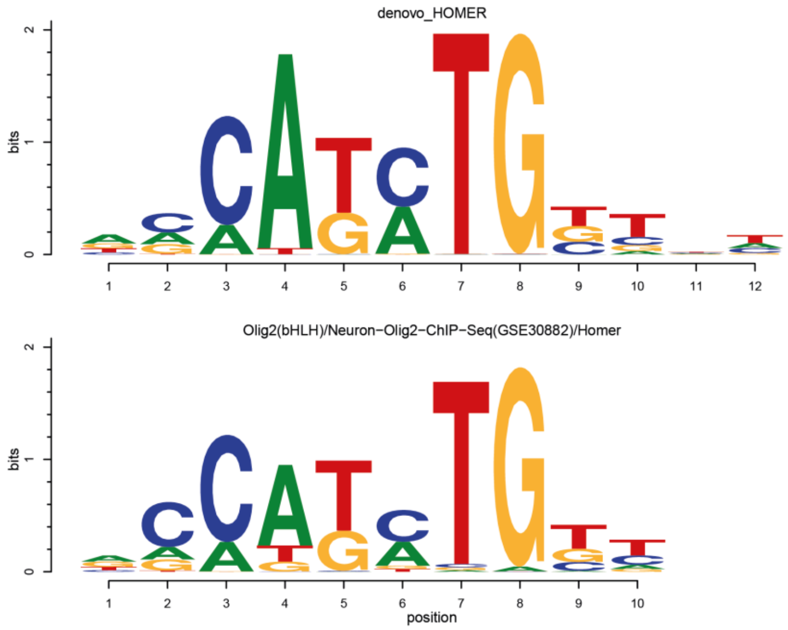


Figure 4. *Comparison of a de novo motif identified by HOMER to the highly similar known motif of the transcription factor Olig2. The motifs are shown as sequence logos which are graphical representations of position weight matrices.*

Motif sites were identified to create motif-containing OCR subsets that were tested for enrichment of PD risk variants. As part of the MEME-ChIP tool set, Find Individual Motif Occurrences (FIMO) was used to scan the sequence for motif sites. However, FIMO found no significant matches for the shortest motifs of 6 base pairs when scanning the large input sequence. This is because a high number of perfect matches to short motifs will occur by chance in large sequence sets. Matches were found for the 7 and 8 base pairs long *de novo* motifs and OCR subsets created. Based on the assumption that the longer known motifs matched to the short *de novo* motifs have a higher information content resulting in more accurate motif occurrences, motif-containing OCR subsets were also made that contained the best matched known motifs.

We consider it a strength of our study that two different motif discovery tools were included in the analyses. The analyses were performed in parallel and both showed an enrichment of PD risk variants in OCRs targeted by bHLH transcription factors, thus increasing the robustness of this finding. Two different methods of *de novo* motif discovery identified similar *de novo* motifs linked to bHLH transcription factors. Also, different methods were used to identify motif sites in the sequences. Enrichment analysis showed a significant overlap between PD risk variants and OCRs harboring the *de novo* motif identified by HOMER matched to bHLH transcription factors. However, no enrichment was found in OCRs containing the *de novo* motif identified by MEME-ChIP matched to bHLH transcription factors. We believe this may be due to the difference in length between the *de novo* motif discovered by HOMER (12 base pairs) and MEME-ChIP (8 base pairs), with likely less accurate motif occurrences identified for the shorter motif. Instead, a significant enrichment of PD risk variants was found in the OCRs harboring the known motif of the bHLH transcription factor NEUROD1 found by MEME-ChIP to be the best match to the *de novo* motif.

We were not able to pinpoint one specific transcription factor, but instead identified a family of transcription factors targeting the enriched OCR subset. Finding the likely binding candidates for *de novo* motifs is complicated by factors such as the incompleteness of the transcription factor motif catalogue, complex binding patterns and widespread sharing of similar motifs by multiple transcription factors (Deplancke et al., 2016, Lambert et al., 2018). A *de novo* motif may have no good matches, a single good match that is highly likely to be the targeting transcription factor, or there may be many transcription factors with high similarity score to the *de novo* motif. The latter is the case for the *de novo* motif targeted by bHLH transcription factors in our study.

We analyzed PD risk variants in OCRs containing the given motif, and not only in the short recognition motif itself (Figure 5). This complies with the understanding that only a small fraction of variability in transcription factor-DNA binding events appears to be caused by variants within the transcription factor recognition motif. By analyzing the open chromatin region surrounding the transcription factor motif, we were able to include proximal variants that may have an effect on binding through mechanisms such as cooperative or collaborative transcription factor-DNA binding (Deplancke et al., 2016). Transcription factor-DNA binding may however also be affected by more distally located variants that we do not capture in our analysis.

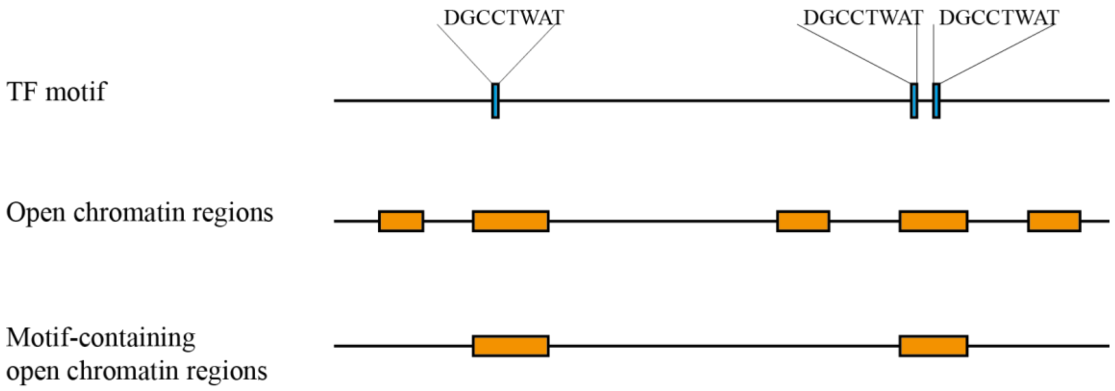


Figure 5. *PD risk variants were analyzed in motif-containing open chromatin regions. TF; transcription factor.*

4.5 Statistical methods

4.5.1 Statistical hypothesis testing

In statistical hypothesis testing, a p-value is used as a parameter of significance to determine the certainty of an observation. The p-value is the probability of rejecting the null hypothesis when it is true. Incorrect rejection of a true null hypothesis is termed Type I error. Another possible type of error is the failure to reject a null hypothesis when it is false, which is referred to as Type II error. A null hypothesis is a general statement of default position that in the context of a genetic association study could be that there is no association between the tested genetic marker and trait of interest. If the p-value is lower than the predefined significance level, then the null hypothesis is rejected. It is the standard in research to accept a 5% chance of obtaining a Type I error when conducting a statistical test, which corresponds to a significance level of 0.05. Statistical power is equal to one minus the probability of making a Type II error. This is the probability of rejecting a null hypothesis while the alternative hypothesis is true. In planning of studies, it is important to ensure that the power is sufficient to obtain meaningful results. A statistical power of 80% is a widely used threshold of adequate power to avoid false negative results and to determine cost-effective sample sizes (Hong and Park, 2012).

In Paper 2, power calculations were performed with the function pwr.f2.test in the R package pwr. The estimated effect size was given as a proportion of variance explained that was based on findings from previous studies of genetic determinants of AAO in PD.

We found that we had a high power (99%) in the primary analysis of the single *DNM3* variant and a moderate power (89%) in analysis of multiple variants in the *DNM3* locus. It may be argued that estimated effect sizes in power calculations should be lower than previous discoveries where the magnitude of effect is often inflated due to a phenomenon called “winner’s curse” (Button et al., 2013). Lowering the estimated effect size would consequently decrease the calculated power. We did not perform any power analyses in Paper 1. This would however have been beneficial, especially in interpretation of the negative findings.

4.5.2 Association analysis

Association analyses in Paper 1 and Paper 2 were performed with the statistical software PLINK 1.9 (Chang et al., 2015). In Paper 1, we performed a genetic association case-control analysis, testing the correlation between disease status and selected genetic variants. We used the basic chi-square allelic test to assess allele frequency differences between cases and controls (Clarke et al., 2011). The strength of the association was measured by the OR, comparing the odds of disease with the minor allele *a* to the odds of disease with the major allele *A*.

In Paper 2, we studied AAO as a quantitative trait and used linear regression to test for association with *DNM3* variants. The effect measure was given as a Beta estimate, which is the degree of change in the outcome variable for every unit of change in the predictor variable. As an alternative binary analysis, AAO was dichotomized by the median onset calculated across all datasets and logistic regression was used as the statistical test for association. Regression analysis allows for inclusion of additional covariates to correct for potential confounding factors. Confounding is the distortion of an association between the tested variable (independent variable) and an outcome (dependent variable) that occurs when the study groups differ in regard to other factors that influence the outcome. In genetic association studies, population stratification should be an important consideration since it may confound the association between a genetic variant and the trait of interest (Hellwege et al., 2017). A widely used method to address population stratification in genetic association studies is principal component analysis (Price et al., 2006). With this method, genome-wide level genotype data is used to estimate principal components representing features of genomic ancestry that capture population stratification (Zhao et al., 2018). Principal components may be used as covariables in association analyses to account for population stratification. In Paper 2, we utilized genome-wide datasets that allowed for calculation of principal components that we included as covariables in the regression analyses.

In candidate gene studies however, accounting for population stratification is more challenging due to the lack of genome-wide coverage of genetic factors from which ancestry may be inferred. Thus, to minimize the risk of population stratification, care should be taken in selection of the study population to prevent or limit the admixture of different ancestries. In Paper 1, analysis of genotyped *GBA* variants and GWAS signals was restricted to the Scandinavian population. Although the Scandinavian population is considered to be relatively homogenous compared to other study populations, genetic heterogeneity does exist between and within Scandinavian countries (Tian et al., 2008). While the number of cases and controls included from Oslo and western Norway was quite even, this was not the case for the Swedish participants. We cannot exclude the possibility that population stratification may have an effect on the case-control association analysis, however this is unlikely to affect LD calculations. LD between the four genotyped variants in Paper I was calculated and visualized with the Haploview software (Barrett et al., 2005). Measures of LD was provided both as the square of the correlation coefficient (r^2) and the normalized coefficient of linkage disequilibrium (D'). $r^2 = 1$ means that the loci are in perfect LD which happens when the loci have not been separated by recombination and also have the same allele frequencies.

Another approach to testing statistical hypotheses is permutation. In a permutation test, the observed test statistic is compared to the distribution of values you get when the observed data is resampled a number of times, called the null distribution. In Paper 3, we used permutation-based approaches when testing if there was a significant enrichment of PD risk variants in functional annotations.

4.5.3 Enrichment analysis

Functional enrichment analysis, also called colocalization analysis, may be used to test whether there is a statistically significant overlap between disease-associated variants and sets of annotated genomic regions, pinpointing genomic features (enhancers, promoters, exons, transcription factor binding sites) and cell types likely relevant to disease pathogenesis. The biological motivation behind enrichment analysis comes from the understanding that physical overlap or proximity of functional annotations implies some biological constraints or mechanistic relationship (Dozmorov, 2017).

Available enrichment tests use different concepts and null models to test the significance of overlap, and importantly the choice of null model is known to affect the subsequent conclusion (Simovski et al., 2018, Kanduri et al., 2019). A null model should appropriately preserve the distributional properties and dependency structure of the tested data (Kanduri et al., 2019). Genomic features do not occur uniformly across the genomic sequence, but instead clump in certain parts of the genome (e.g. close to gene rich regions). Failing to account for the non-uniform distribution of genetic variants and other genomic or epigenomic features may result in a higher rate of false-positives (Kanduri et al., 2019, Trynka et al., 2015). This challenge may be addressed by assessing the consistency of findings by tests employing different null models and parameter choices (Simovski et al., 2018, De et al., 2014).

We used the two methods GoShifter and GREGOR to analyze the overlap between PD GWAS signals and selected genomic annotations. We also tested for enrichment of GWAS signals from two non-brain related traits that were included as negative controls. GoShifter stringently controls for local genomic structure by locally shifting sites of the tested features within each risk locus to generate a null distribution of annotations overlapping associated variants by chance (Trynka et al., 2015). The second method applied, GREGOR, uses a snp-matching-based method to test for enrichment (Schmidt et al., 2015). The number of trait-associated signals where an index variant or one of its LD proxies overlaps a regulatory annotation is calculated, then the probability of the observed overlap of risk variants is estimated relative to expectation using a set of matched control variants. Control variants match the index variants for number of variants in LD, minor allele frequency and distance to nearest gene. Especially matching on the number of variants in LD has been found to be a critical step to avoid inflation of observed enrichment values (Trynka et al., 2015).

In both tests, variants in high LD with the index variant are included in the analysis, which is important since the index variant is often not the causal variant. We used a LD threshold of $r^2 > 0.8$ which is a frequently used cutoff when attempting to capture the causal variants within an association signal. However, we cannot exclude the possibility that the index variant has a lower degree of linkage to a causal variant.

In our study we tested genome-wide significant PD signals, although there is evidence that genetic variants below the level of genome-wide significance also contribute to the genetic heritability (Escott-Price et al., 2015). New methods have emerged, such as stratified LD score regression, which assesses whether the overall heritability of a trait is enriched within specified annotations (Finucane et al., 2015). Stratified LD score regression and similar methods use information from all common variants and thus require genome-wide association summary statistics. GWAS summary statistics are however often not made publicly available, which was the case for recent large-scale PD GWASs at the time of our analysis. There is a move towards increased sharing of data that may change this. Parts of the summary statistics from the most recent PD GWAS meta-analysis is currently publicly available, while it is possible to apply for access to the remaining data under an agreement that protects the privacy of participants (Nalls et al., 2019).

Statistically significant colocalization between two functional genomic annotations may be driven by colocalization with another genomic annotation that was not included in the analysis. This limits the inference of causality and has to be taken into account when interpreting results from enrichment analysis. We cannot exclude the possibility that an observed enrichment reported in our study may be due to unaccounted colocalization with other annotations. Such an effect of potential confounding features has been addressed by stratified LD score regression where a baseline model consisting of a range of main annotations that are not specific to any cell type is included in the analysis (Finucane et al., 2015).

4.5.4 Meta-analysis

In Paper 2, we performed a meta-analysis of the seven included studies using the GWAMA (Genome-Wide Association Meta-Analysis) software (Magi and Morris, 2010). Meta-analysis is a statistical method for combining results from different studies by weighting the data according to the amount of information in each study (Lee, 2015).

Combining results from different studies may increase the statistical power and provide a more precise estimate of the effect size. We employed the inverse variance method where the weight given to each study is the inverse of the variance of the effect estimate (one over the square of its standard error). Thus, larger studies are given more weight than smaller studies that have larger standard errors.

Assessment of the inter-study heterogeneity is an important step of meta-analysis. We used Cochran’s Q test and Higgins’s I statistic to test for inter-study heterogeneity, finding that the heterogeneity was low for both the primary tested variant (rs2421947) as well as for the vast majority of the tested common variants in the *DNM3* locus. While Cochran's Q test is used to determine whether significant heterogeneity in effect sizes between the primary studies exists, Higgins’s I statistic quantifies the effect of heterogeneity (Lee, 2015, Lunetta, 2013). We visualized the meta-analysis results with a forest plot, which graphically displays the results from each of the included studies along with the overall result from the meta-analysis.

4.5.5 Multiple testing correction

When performing multiple tests, the likelihood of making a Type I error increases. Genetic association studies usually test multiple genetic markers and controlling for multiple testing is important to prevent a high rate of false positive results. A widely applied approach to control for multiple comparisons is Bonferroni correction. This method generates a strict significance cutoff by dividing the original significance level by the number of performed tests. The Bonferroni correction assumes that the individual tests are independent of each other, which is often not the case. The inter-dependence between genetic variants needs to be taken into account when determining the significance level. In GWASs, based on an estimated testing burden of one million common independent variants genome-wide, a significance p-value threshold of 5×10^{-8} ($0.05/10^{-6}$) has become the standard (Pe'er et al., 2008).

In Paper 1, we tested for association between four variants in the *GBA* locus and PD, reporting p-values < 0.05 as significant. According to principles of Bayesian statistics, the interpretation of test statistics is largely dependent on assumptions about the prior probabilities of a research hypothesis being true. Our choice of significance level was based on a prior probability of true association since the two GWAS signals and two *GBA* variants that we tested have previous reports of an association with PD. In addition, the existence of high LD between some of the variants was also taken into consideration.

In Paper 2, we analyzed all common variation within *DNM3* as well as 100kb upstream and downstream of the gene. To estimate the degree of multiple testing, we identified the number of independent variants using a cutoff of LD > r^2=0.5 that we adjusted for by Bonferroni correction. This method takes the LD structure into account, however the threshold of LD is arbitrary. It could be argued that the threshold should be either higher or lower, which would consequentially alter the number of tests to correct for. This underscores the importance of defining the significance level prior to data analysis. In Paper 3, we adjust for the number of tested annotations by Bonferroni correction. There is a high degree of overlap between the different annotations, making it a very stringent estimate of significance level that may potentially lead to false negative results. Although most studies apply this strict approach, methods that calculate the number of independent annotations may be implemented to improve multiple testing correction (Iotchkova et al., 2019).

5. Ethical considerations

Genetic studies of Parkinson's disease at Oslo University Hospital was approved by the Regional Committees for Medical and Health Research Ethics-South East Norway (REC South East Norway). This approval covers the study performed in Paper 1 and Paper 2. Sample and data collection at other study sites were approved by local ethics committees. All participants gave written, informed consent. The consent form used at Oslo University Hospital informs participants that the genetic analyses are performed for research purposes only, and that they will not be individually informed regarding the results of these analyses. The voluntariness of participation and the right to withdraw from the study at any time point is clearly stated. Patient identity and personal data are stored on secure servers dedicated to research databases at Oslo University Hospital. Anonymized IDs were used in all sample handling, experiments and analyses. The study performed in Paper 3 does not require approval from an ethics committee since we analyze existing publicly available data.

Analysis of shared genomic and epigenomic data is central to some of the studies included in this thesis. ATAC-seq data analyzed in Paper 3 was publicly available in an open database. The individual-level genotype data analyzed in Paper 2 was made available through controlled-access sharing via dbGaP and PPMI. It is widely accepted that data sharing promotes scientific progress, maximizing the utility of generated datasets and reducing the burden to participants. Accessible data makes it possible for researchers to pool data to increase the power of studies and link different types of data, potentially leading to enhanced understanding of disease biology. For genetic data, there is a risk of reidentification that raises ethical and legal issues related to the privacy of research participants. The risk varies dependent on factors such as the dimensionality of the data, how frequent the disease is and the appending metadata. Thus, ethical data sharing requires consideration of both the value of the scientific data, and the privacy costs of participants (Byrd et al., 2020).

6. General discussion

Investigations into the genetic basis of PD could be essential to the development of effective therapeutic strategies (Blauwendraat et al., 2020a). Identification of genetic factors that influence disease risk, onset and progression may provide insight into molecular mechanisms initiating and driving PD. Paper 1 and Paper 2 join in the research of identifying and refining genetic risk. While in Paper 3, the aim was to use genetics to enhance our understanding of pathobiology. A major goal in disease-genetics is to translate genetic knowledge into drug targets for assessment in therapeutic trials. Furthermore, genetics may also play an important role in trial recruitment and later on clinical practice, since it is likely that future therapies will target genetic subgroups of PD (Figure 6).

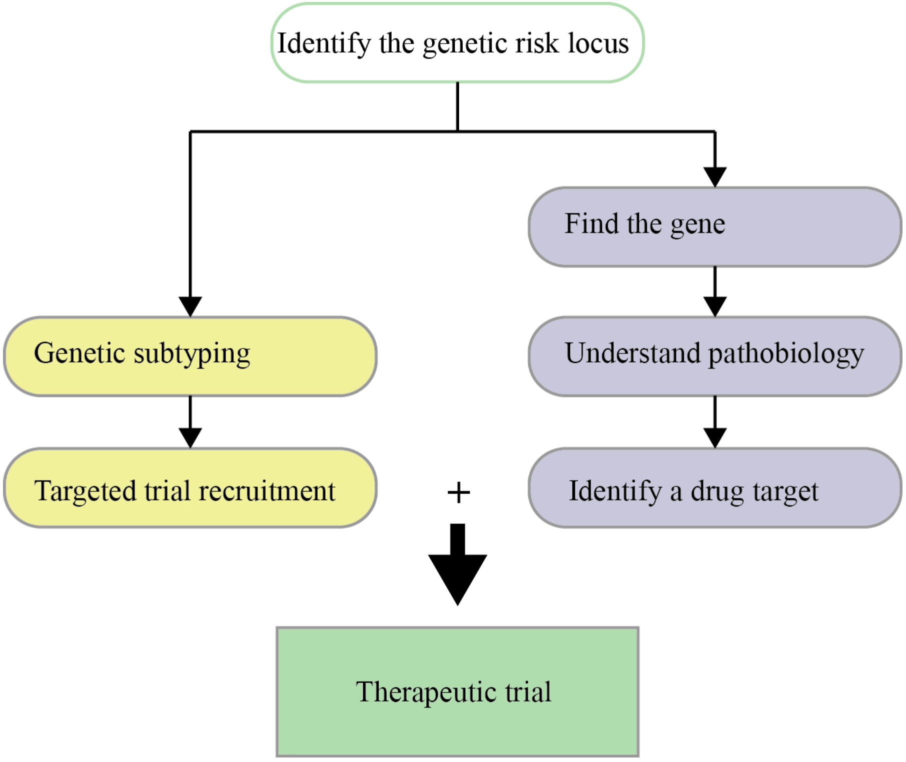


***Figure 6**. Workflow for drug discovery driven by genetics. The figure is inspired by (Blauwendraat et al., 2020a) and (Singleton and Hardy, 2016).*

6.1 The GBA locus and Parkinson's disease

6.1.1 Low-frequency coding GBA variants in Parkinson's disease

In Paper 1, we studied variants in the *GBA* gene and how they relate to an important PD GWAS signal. We found that the *GBA* variant E326K is significantly more frequent in PD patients compared to controls. This association has been reported by previous studies, although there have been some conflicting results (Sidransky et al., 2009, Lesage et al., 2011, Duran et al., 2013, Ran et al., 2016). In order to evaluate the effect of E326K on PD risk, a meta-analysis has been performed that found a significant association between E326K and PD risk in the Caucasians, Asians and the total population (Huang et al., 2018). They reported an OR of 1.82 in Caucasians, which is comparable to the OR we found at 1.62. Another meta-analysis studying a larger number of different *GBA* variants, found that E326K increases the risk of PD in the non-Ashkenazi Jewish population with an OR of 1.98 (Zhang et al., 2018). While it has become increasingly clear that E326K is a risk factor for PD, the pathological significance of T369M is considered less certain. However, in later years, meta-analyses and a large-scale case-control study have reported a significant association between T369M and PD risk (Zhang et al., 2018, Mallett et al., 2016, Blauwendraat et al., 2018). Studies analyzing both E326K and T369M find that the OR of T369M is a little lower compared to that of E326K (Blauwendraat et al., 2018, Zhang et al., 2018). In our analysis of T369M, we did not find a significant association with PD risk. This may be due to our study being underpowered to identify associations with weak effect sizes.

The discovery that E326K, and potentially also T369M, increase the risk of PD has been somewhat surprising since these two low-frequency *GBA* variants do not cause GD in the homozygous state. Both E326K and T369M are however associated with reduced GCase activity in the heterozygous state (Alcalay et al., 2015). Although the reduction of enzyme activity is not large enough to cause GD, it may still contribute to PD risk in combination with other genetic and biochemical alterations.

6.1.2 Coding variation in GBA underlie a Parkinson's disease GWAS signal

In our study of the *GBA* gene, an objective was to assess to what degree coding *GBA* variants are linked to the *GBA-SYT11* association signals reported for PD. An early PD GWAS reported an association signal with the top hit variant located within an intron of the *SYT11* gene on chromosome 1q22 (Nalls et al., 2011). Later, a meta-analysis of several GWASs included some *GBA* variants on the genotyping array used in replication analysis and found a significant association between E326K and PD (Pankratz et al., 2012). *GBA* is located about 650 kb from *SYT11*, within the same block of LD referred to as the *GBA-SYT11* locus. The largest and most recent meta-analysis of PD GWASs at the time of our investigation, reported two independent associations within the *GBA-SYT11* locus (Nalls et al., 2014). Although these association signals were located in an intergenic region hundreds of kilobases away from *SYT11*, they still kept the gene in the naming of the locus (*GBA-SYT11*). Furthermore, the relationship between the reported signals and coding *GBA* variants was not explored. We hypothesized that coding *GBA* variants underlie one or both of the association signals at the *GBA-SYT11* locus, which we refer to as the primary and secondary association signal. And interestingly, we did find that E326K is in very high LD (r^2 0.95) with the primary association signal at the *GBA-SYT11* locus. This result emphasizes E326K as the causal allele and consequently *GBA* as the causal gene behind this GWAS signal. We did however not find any evidence of the secondary association signal being related to any of the tested *GBA* variants.

Following our study, the relationship between *GBA* variants and the *GBA-SYT11* GWAS signals has been analyzed in a larger number of PD patients and controls (Blauwendraat et al., 2018). In line with our findings, their results are consistent with E326K being the effector allele underlying the primary association signal at the *GBA-SYT11* locus. They also explored the secondary association signal at the *GBA-SYT11* locus, finding that it remained significant after adjusting for E326K, T369M and N370S. This means that none of the tested *GBA* variants explains the secondary association signal at this locus and that it has yet to be untangled.

Identification of the casual genes behind GWAS signals is important to guide functional studies into disease related molecular mechanisms. *SYT11* has been referred to as a PD-related gene due to the initial association with an intronic variant in this gene and the naming of this GWAS locus. The protein encoded by *SYT11*, synaptotagmin 11, is involved in regulation of endocytosis and vesicle recycling processes in neurons (Wang et al., 2016). Our results may be considered to weaken the genetic evidence linking *SYT11* to PD, but do not exclude the possibility that *SYT11* plays a role in PD pathogenesis. Interestingly, a recent study performing targeted sequencing of PD loci genes identified an association between the burden of rare variants in *SYT11* and PD. The association was mainly driven by a rare variant independent of *GBA* variants, suggesting that a genetic link between *SYT11* and PD risk may exist (Rudakou et al., 2021).

6.2 DNM3 and age at onset of Parkinson's disease

In Paper 2, we investigated whether genetic variability reported to modify AAO of *LRRK2*-associated PD, has an effect on AAO in idiopathic PD. We analyzed *DNM3* rs2421947 and all other common variation in the *DNM3* locus in 5918 patients with idiopathic PD, finding no evidence of an association with AAO of disease.

Our study was the first to specifically assess the effect of *DNM3* variants on AAO of idiopathic PD based on the association reported in *LRRK2* G2019S carriers by Trinh et al. (Trinh et al., 2016). *LRRK2* is a gene with pleomorphic effects in PD. G2019S and other *LRRK2* mutations cause an autosomal dominant form of PD, while GWASs have shown consistent evidence that also common variants at this locus modulate PD risk. Since *LRRK2* is part of the genetic background for idiopathic PD, genetic modifiers of *LRRK2*-associated PD could also exert an effect in this much larger group of PD patients. However, as our results show, a potential modifying effect of *DNM3* on AAO cannot be generalized to idiopathic PD. The effect of *DNM3* rs2421947 on AAO in idiopathic PD has been assessed by an additional study, and consistent with our results, they found no significant association (Brown et al., 2021).

The largest and most recent GWAS of PD AAO to date identified two genome-wide significantly associated loci (*SNCA* and *TMEM175*), as well as some sub-significant loci (*GBA*, *SCARB2*, *BAG3* and *MCCC1*), of which all have previously been reported to influence PD risk (Blauwendraat et al., 2019). While known PD risk loci have been shown to be associated with AAO, the discovery of novel genetic modifiers of AAO in PD has proven difficult. Genetic modifiers of AAO could be restricted to sub-groups of patients that are carriers of specific mutations or susceptibility variants. *DNM3* may be a specific modifier of *LRRK2*-parkinsonism, however this finding has not yet been replicated. No significant association was found between *DNM3* rs2421947 and AAO of PD in *LRRK2* G2019S carriers in the Spanish population, or in Chinese individuals carrying Asian *LRRK2* risk alleles (Fernandez-Santiago et al., 2018, Foo et al., 2019, Yang et al., 2019). Furthermore, a recent analysis of *DNM3* rs2421947 in a multi-ethnic cohort of *LRRK2* G2019S carriers did not replicate the association between *DNM3* and PD AAO (Brown et al., 2021). The effect was analyzed in a new cohort of *LRRK2* G2019S heterozygotes and these data were meta-analyzed with previously published data. There was considerable inter-study heterogeneity that according to the authors could indicate ethnic or population-specific effects of *DNM3* (Brown et al., 2021). Still, the lack of replication warrants careful interpretation of the reported association between *DNM3* and AAO in *LRRK2* G2019S carriers.

6.3 Missing heritability

While GWASs have certainly expanded our understanding of the genetic basis of PD, many more risk variants remain to be discovered. The heritable component of PD due to common genetic variability is estimated to be around 22% and identified GWAS association signals to date represent only a proportion (16-36%) of this heritability (Nalls et al., 2019). Several explanations have been presented for the “missing heritability”, of which one is the insufficient power of current GWASs. Increasing the sample size of GWASs is expected to lead to identification of new susceptibility loci that are less common and with smaller effect sizes. Moreover, since the majority of genetic studies have been done in individuals with European ancestry, performing GWASs in more diverse populations may lead to the discovery of additional population specific risk factors (Blauwendraat et al., 2020a).

Another explanation to the “missing heritability” is the contribution of rare variants that are not well detected by current GWAS methodologies. Targeted resequencing of known GWAS loci may identify rare risk variants that are either independent of, or underlie the GWAS signal (Singleton and Hardy, 2016). Additional rare risk loci may be discovered by whole exome sequencing and whole genome sequencing, of which the availability is increasing due to falling costs. Other potential contributors to PD heritability are epistatic interactions between genetic variants, meaning that they act in a non-additive fashion, as well as gene-environment interactions (Bandres-Ciga et al., 2020a).

A common and reasonable criticism towards further large-scale genetic studies of PD is that resources instead should be used to shed light on pathobiological mechanisms underlying the already identified risk signals. An argument in support of continued genetic work is that expanding the list of genetic risk loci will improve our chances at biological insight through integration with large-scale functional data (Singleton and Hardy, 2016)

6.4 Gaining biological insight from common risk variants

PD risk loci have been tested for association with tissue, cell types and biological pathways (Nalls et al., 2019, Chang et al., 2017, Bandres-Ciga et al., 2020b). Efforts at linking PD genetic risk to specific transcription factor networks have however been scarce. In Paper 3, we coupled neuronal ATAC-seq data with transcription factor motif analysis in an effort to identify transcriptional networks contributing to PD risk mechanisms.

6.4.1 Cell types implicated in Parkinson’s disease

In our study, we found that PD risk signals significantly overlap with sites of open chromatin in neurons of the superior temporal cortex. Open chromatin sites in neurons of several other cortical regions approached the significance level, suggesting that a broader range of cortical regions are implicated in PD risk. The relevance of brain neurons in PD is highlighted by the most recent PD GWAS where nominated PD GWAS genes were integrated with expression data from 53 tissues. They found a significant enrichment for expression in 13 tissues, of which all were brain derived. To further explore the enrichment in brain tissues, PD GWAS genes were tested in a large number of brain cell types from mouse with results showing enrichment for expression only in neuronal cell types (Nalls et al., 2019). Furthermore, a recent single-nuclei transcriptomic atlas from cortex and substantia nigra revealed significant associations between PD genetic risk and neurons within both these brain regions (Agarwal et al., 2020). Interestingly, the same study also found a significant association between PD genetic risk and oligodendrocytes, a link that has previously been proposed based on analysis of mouse transcriptomic data (Bryois et al., 2020). Other non-neuronal cell types have been reported to be associated with PD risk, including immune cells, mesendoderm, liver- and fat cells (Coetzee et al., 2016, Gagliano et al., 2016). Further studies are needed to confirm, as well as understand, the potential role of glial cells and non-brain-related cell types in PD pathogenesis.

6.4.2 A potential role for bHLH transcription factors in Parkinson's disease risk mechanisms

In our analysis of putative transcription factor binding sites, we found that PD risk signals concentrate at sites of open chromatin targeted by members of the bHLH transcription factor family. bHLH transcription factors play essential developmental roles as regulators of neural cell fate specification and differentiation (Dennis et al., 2019). One could argue that transcription factors that are expressed and function in the developing nervous system are more likely to be involved in neurodevelopmental diseases rather that neurodegenerative disorders with an adult onset. However, a developmental component to PD pathogenesis is plausible and may be due to protection against future neurodegenerative effects. Genetic and epigenetic factors are suggested to affect PD risk by influencing the number of nigral dopaminergic neurons the individual is born with (von Linstow et al., 2020). Genetic risk variants with functional consequences during neurodevelopment may represent the first hit in a multi-hit hypothesis where a second hit, or potentially additional hits, are required for disease to develop (Sulzer, 2007). Interestingly, the bHLH transcription factor Srebf1 has been identified as a key regulator of midbrain dopaminergic neurogenesis (Toledo et al., 2020). There are also some bHLH transcription factors that function in adult neurons, such as TCF4 (Jung et al., 2018). Common variants at the TCF4 locus are associated with schizophrenia risk (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

6.4.3 Identification of transcription factors implicated in disease

Cell type-specific transcription factor binding annotations have been shown to outperform cell type-specific chromatin marks in calculations of heritability enrichment in a range of complex traits, confirming the importance of transcription factor binding in disease (van de Geijn et al., 2020). Computational methods are being developed in an attempt to improve the accuracy of transcription factor binding prediction based on chromatin accessibility and transcription factor binding motifs (Li et al., 2019a, Keilwagen et al., 2019). Transcription factor binding sites may also be inferred from investigation of chromatin accessibility patterns by computational footprinting. Protein bound DNA is more resistant to cleavage by enzymes, leaving short protected stretches of DNA called footprints that may be detected as sudden drops of coverage within peak regions of high coverage. Footprinting algorithms have been widely used in DNase-seq based studies, but is less explored and has a reduced performance in analysis of ATAC-seq (Karabacak Calviello et al., 2019). New footprinting methods tailored to the ATAC-seq protocol are emerging (Li et al., 2019b).

Importantly, transcription factor binding events may be nonfunctional interactions, meaning that there is no corresponding change in gene expression. Transcription factor activities have been predicted based on the expression of their target genes, however identification of such target genes has proven challenging (Garcia-Alonso et al., 2019). Interestingly, emerging cell type-specific HiC data may be coupled with transcription factor binding annotations, providing a physical basis for interactions between transcription factor binding sites at regulatory elements and predicted target genes. Target gene prioritization performed *a priori* of eQTL testing has been suggested as a strategy to increase detection sensitivity, revealing more causal associations between variants affecting transcription factor binding and gene expression (Mitchelmore et al., 2020).

Transcriptional data from individuals with the disease of interest and non-diseased controls may be used to identify differentially expressed genes between cases and controls. In a study of psychiatric diseases and Alzheimer’s disease, post-mortem prefrontal cortex gene expression profiles were analyzed in cases and controls (Pearl et al., 2019). Transcription factors with predicted target genes that were over-represented among the differentially expressed genes were considered key regulators of the disease. The transcription factor POU3F2 was identified as a key regulator in both schizophrenia and bipolar disorder. Luciferase reporter assays were performed to study the regulatory impact of a genetic variant associated with schizophrenia risk that was predicted to have a functional effect on POU3F2 binding. They also sought to validate and further study the trans-acting effect of POU3F2 by overexpressing it in primary human neural stem cells. The follow-up of predicted mechanistic hypotheses with *in vitro* functional assays, such as reporter assays or genome editing methods, is an important next step to validate findings and further explore pathological processes.

6.5 The role of genetics in therapeutic development

Genetic-, neuropathological- and clinical studies have shown that PD is a highly heterogeneous disorder. There may be subtypes of PD with distinct, or at least not fully overlapping, pathophysiology that possibly respond differently to therapeutic approaches (Singleton and Hardy, 2016). Thus, instead of treating PD as one single condition, precision medicine may be the approach that leads to long sought after therapeutic advances. Precision medicine aims at tailoring treatment to the individual characteristics of the patient, based on genotype or other biomarkers that ideally reflect disease-associated biological processes. PD subtyping may be used in clinical trials to create more homogenous groups within which to study treatment effects. In the clinical setting, a future goal is the use of PD subtypes to assist in counseling regarding prognosis and choice of treatment (Marras et al., 2020).

Building on the last decades' genetic discoveries in PD, therapies targeting genetic forms of the disease are now being tested in clinical trials. There is an active therapeutic development directed against both *GBA*- and *LRRK2*-related PD. Clinical trials that target the *GBA* pathway benefit from the overlap between PD and GD, since therapeutic strategies may be applicable to both diseases. Successful therapeutic approaches in GD involve restoration of GCase activity by enzyme replacement therapy and substrate reduction therapy with inhibitors of the glucosylceramide synthase enzyme. However, while hematological and visceral symptoms of GD are effectively treated, there is no improvement of the neurological manifestations of the disease since approved drugs do not penetrate the blood-brain barrier. Novel glucosylceramide synthase inhibitors have been developed that show good brain penetrance, improved alpha-synuclein processing and behavioral outcomes in synucleinopathy mouse models (Sardi et al., 2017). Based on these results, a double-blinded, placebo controlled phase 2 trial has been initiated to assess the safety and efficacy of the glucosylceramide synthase inhibitor Venglustat in PD patients carrying a *GBA* mutation (Peterschmitt et al., 2019).

Another therapeutic approach under investigation is the use of small molecule chaperones, such as ambroxol, that has been shown to increase brain GCase activity in mouse models with *GBA* mutations (Migdalska-Richards et al., 2016). Ambroxol has been tested in a non-controlled trial including PD patients with and without *GBA* mutations, finding that the drug penetrated CSF, was safe and well tolerated (Mullin et al., 2020). The effects of ambroxol therapy on the progression of PD will need to be assessed in placebo-controlled clinical trials. Additional treatment strategies are tested in *GBA*-positive patients, including other small molecule chaperones and gene therapy (Schneider and Alcalay, 2020). In *LRRK2*-associated PD, observations show that pathogenic *LRRK2* mutations increase the kinase activity. This proposed gain-of-function effect has emerged as a therapeutic target and several companies are pursuing *LRRK2* inhibitors (Sardi and Simuni, 2019). Potential peripheral side effects are however a concern, and the safety and tolerability of the drugs need further clarification.

Although precision medicine clinical trials may represent great progress towards effective therapies, there are several challenges. One major challenge is the identification of reliable outcome measures to detect disease modifying effects (Sardi and Simuni, 2019). Another obstacle to overcome is how to successfully recruit large enough numbers of mutation carriers, since only a small fraction of PD patients so far has been genotyped (Schneider and Alcalay, 2020). Interestingly, genetically targeted therapies may potentially be tested and effective in larger PD populations since some pathogenic mechanisms, such as impaired GCase activity, appear to be implicated also in non-mutation carriers with idiopathic PD (Gegg et al., 2012).

7.Future perspectives

Genetic studies of PD have revealed a complex genetic architecture with risk variants across a spectrum of frequency and penetrance. Alterations in the *GBA* gene may be considered the most important risk factor for PD, however the majority of *GBA* mutation carriers will not develop PD. An important question is thus, why some *GBA* mutation carriers develop PD while others do not? More specifically, what additional genetic and environmental factors influence *GBA*-associated risk of PD? A recent GWAS found that PD in *GBA* carriers is influenced by variants at loci that are known to be associated more generally with PD risk (Blauwendraat et al., 2020b). Common genetic factors were found to only explain some of the partial penetrance of *GBA* variants in PD, thus other factors such as rare variants probably contribute. Genetic variants that regulate the expression of *GBA* or influence other genes in the same pathway should be considered likely candidates to have an impact on *GBA* penetrance. Hence, inclusion of e-QTLs, Hi-C and other epigenomic data in the analysis may serve as a strategy to overcome power issues met in association testing in *GBA* mutation carriers or other subgroups of PD patients where the sample size is likely to be limited (Schierding et al., 2020). Integration of functional genomic data as biologic filters at early stages of genetic association analysis is an interesting approach that may be used to prioritize variants for testing and potentially reveal some of the missing heritability, such as rare variants and epistatic interactions (Castel et al., 2018, Manduchi et al., 2018).

Most genetic studies of PD have been performed in cross sectional patient collections where the clinical details are few. A better understanding of genetic factors influencing the progression of PD requires studies of more deeply phenotyped patient cohorts. Such detailed longitudinal clinical studies are extremely resource demanding, and this has obstructed the collection of large sample sizes. Overcoming this challenge depends upon ambitious long term international multi-center collaborations such as the PPMI (Marek et al., 2018). Large-scale collaborative efforts will also be key to enhance the collection of functional genomic data. Improved interpretation of disease-associated variants in PD and other brain disorders will probably depend on a growing amount of brain-relevant functional genomic annotations, a resource that until recently has been scarce. Increasing cellular resolution, as well as assaying of several molecular phenotypes within the same cell type, will improve the quality of annotations (Reynolds et al., 2019). Collection of such high quality data from multiple brain regions and cell types will be an expensive and laborious endeavor. It may however be feasible through collaborative initiatives that are brain- and/or disease-focused, of which some are ongoing (Wang et al., 2018, Fromer et al., 2016)

A major challenge to therapeutic development in PD lies in identification of individuals that will go on to be affected with the disease. It is likely that patients at early stages of the disease may be more responsive to neuroprotective treatments, compared to later stages when severe degradation of the nigrostriatal dopaminergic system has already occurred (Singleton and Hardy, 2016). Risk models based on genetics alone do not predict PD well enough to have clinical utility in the near future (Blauwendraat et al., 2020a). Instead, risk models based on the combination of genetic risk scores with other types of data, such as selected prodromal features or family history of PD, have shown improved predictive power (Nalls et al., 2015b). Multimodal predictive models may aid in selection of patients with prodromal PD for clinical trials, and potentially for future therapeutic options. Overall, genetics is expected to play a key role in achieving the ultimate goal of treating the right person, with the right therapy at the right time.

8. References

AGARWAL, D., SANDOR, C., VOLPATO, V., CAFFREY, T. M., MONZÓN-SANDOVAL, J., BOWDEN, R., ALEGRE-ABARRATEGUI, J., WADE-MARTINS, R. & WEBBER, C. 2020. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat Commun,* 11**,** 4183.

AHARON-PERETZ, J., ROSENBAUM, H. & GERSHONI-BARUCH, R. 2004. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. *N Engl J Med,* 351**,** 1972-7.

ALCALAY, R. N., LEVY, O. A., WATERS, C. C., FAHN, S., FORD, B., KUO, S. H., MAZZONI, P., PAUCIULO, M. W., NICHOLS, W. C., GAN-OR, Z., ROULEAU, G. A., CHUNG, W. K., WOLF, P., OLIVA, P., KEUTZER, J., MARDER, K. & ZHANG, X. 2015. Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. *Brain,* 138**,** 2648-58.

ANAND, S., MANGANO, E., BARIZZONE, N., BORDONI, R., SOROSINA, M., CLARELLI, F., CORRADO, L., MARTINELLI BONESCHI, F., D'ALFONSO, S. & DE BELLIS, G. 2016. Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. *Sci Rep,* 6**,** 33735.

ARMSTRONG, M. J. & OKUN, M. S. 2020. Diagnosis and Treatment of Parkinson Disease: A Review. *Jama,* 323**,** 548-560.

BALESTRINO, R. & SCHAPIRA, A. H. V. 2020. Parkinson disease. *Eur J Neurol,* 27**,** 27-42.

BANDRES-CIGA, S., DIEZ-FAIREN, M., KIM, J. J. & SINGLETON, A. B. 2020a. Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. *Neurobiol Dis,* 137**,** 104782.

BANDRES-CIGA, S., SAEZ-ATIENZAR, S., KIM, J. J., MAKARIOUS, M. B., FAGHRI, F., DIEZ-FAIREN, M., IWAKI, H., LEONARD, H., BOTIA, J., RYTEN, M., HERNANDEZ, D., GIBBS, J. R., DING, J., GAN-OR, Z., NOYCE, A., PIHLSTROM, L., TORKAMANI, A., SOLTIS, A. R., DALGARD, C. L., SCHOLZ, S. W., TRAYNOR, B. J., EHRLICH, D., SCHERZER, C. R., BOOKMAN, M., COOKSON, M., BLAUWENDRAAT, C., NALLS, M. A. & SINGLETON, A. B. 2020b. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. *Acta Neuropathol,* 140**,** 341-358.

BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell,* 27**,** 299-308.

BARRETT, J. C., FRY, B., MALLER, J. & DALY, M. J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics,* 21**,** 263-5.

BATTLE, A., BROWN, C. D., ENGELHARDT, B. E. & MONTGOMERY, S. B. 2017. Genetic effects on gene expression across human tissues. *Nature,* 550**,** 204-213.

BELLOU, V., BELBASIS, L., TZOULAKI, I., EVANGELOU, E. & IOANNIDIS, J. P. 2016. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. *Parkinsonism Relat Disord,* 23**,** 1-9.

BERG, D., POSTUMA, R. B., ADLER, C. H., BLOEM, B. R., CHAN, P., DUBOIS, B., GASSER, T., GOETZ, C. G., HALLIDAY, G., JOSEPH, L., LANG, A. E., LIEPELT-SCARFONE, I., LITVAN, I., MAREK, K., OBESO, J., OERTEL, W., OLANOW, C. W., POEWE, W., STERN, M. & DEUSCHL, G. 2015. MDS research criteria for prodromal Parkinson's disease. *Mov Disord,* 30**,** 1600-11.

BERNSTEIN, B. E., STAMATOYANNOPOULOS, J. A., COSTELLO, J. F., REN, B., MILOSAVLJEVIC, A., MEISSNER, A., KELLIS, M., MARRA, M. A., BEAUDET, A. L., ECKER, J. R., FARNHAM, P. J., HIRST, M., LANDER, E. S., MIKKELSEN, T. S. & THOMSON, J. A. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol,* 28**,** 1045-8.

BILLINGSLEY, K. J., BANDRES-CIGA, S., SAEZ-ATIENZAR, S. & SINGLETON, A. B. 2018. Genetic risk factors in Parkinson's disease. *Cell Tissue Res,* 373**,** 9-20.

BLAUWENDRAAT, C., BRAS, J. M., NALLS, M. A., LEWIS, P. A., HERNANDEZ, D. G. & SINGLETON, A. B. 2018. Coding variation in GBA explains the majority of the SYT11-GBA Parkinson's disease GWAS locus. *Mov Disord,* 33**,** 1821-1823.

BLAUWENDRAAT, C., HEILBRON, K., VALLERGA, C. L., BANDRES-CIGA, S., VON COELLN, R., PIHLSTROM, L., SIMON-SANCHEZ, J., SCHULTE, C., SHARMA, M., KROHN, L., SIITONEN, A., IWAKI, H., LEONARD, H., NOYCE, A. J., TAN, M., GIBBS, J. R., HERNANDEZ, D. G., SCHOLZ, S. W., JANKOVIC, J., SHULMAN, L. M., LESAGE, S., CORVOL, J. C., BRICE, A., VAN HILTEN, J. J., MARINUS, J., EEROLA-RAUTIO, J., TIENARI, P., MAJAMAA, K., TOFT, M., GROSSET, D. G., GASSER, T., HEUTINK, P., SHULMAN, J. M., WOOD, N., HARDY, J., MORRIS, H. R., HINDS, D. A., GRATTEN, J., VISSCHER, P. M., GAN-OR, Z., NALLS, M. A. & SINGLETON, A. B. 2019. Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and alpha-synuclein mechanisms. *Mov Disord,* 34**,** 866-875.

BLAUWENDRAAT, C., NALLS, M. A. & SINGLETON, A. B. 2020a. The genetic architecture of Parkinson's disease. *Lancet Neurol,* 19**,** 170-178.

BLAUWENDRAAT, C., REED, X., KROHN, L., HEILBRON, K., BANDRES-CIGA, S., TAN, M., GIBBS, J. R., HERNANDEZ, D. G., KUMARAN, R., LANGSTON, R., BONET-PONCE, L., ALCALAY, R. N., HASSIN-BAER, S., GREENBAUM, L., IWAKI, H., LEONARD, H. L., GRENN, F. P., RUSKEY, J. A., SABIR, M., AHMED, S., MAKARIOUS, M. B., PIHLSTRØM, L., TOFT, M., VAN HILTEN, J. J., MARINUS, J., SCHULTE, C., BROCKMANN, K., SHARMA, M., SIITONEN, A., MAJAMAA, K., EEROLA-RAUTIO, J., TIENARI, P. J., PANTELYAT, A., HILLIS, A. E., DAWSON, T. M., ROSENTHAL, L. S., ALBERT, M. S., RESNICK, S. M., FERRUCCI, L., MORRIS, C. M., PLETNIKOVA, O., TRONCOSO, J., GROSSET, D., LESAGE, S., CORVOL, J. C., BRICE, A., NOYCE, A. J., MASLIAH, E., WOOD, N., HARDY, J., SHULMAN, L. M., JANKOVIC, J., SHULMAN, J. M., HEUTINK, P., GASSER, T., CANNON, P., SCHOLZ, S. W., MORRIS, H., COOKSON, M. R., NALLS, M. A., GAN-OR, Z. & SINGLETON, A. B. 2020b. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain,* 143**,** 234-248.

BOEVA, V. 2016. Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Front Genet,* 7**,** 24.

BONIFATI, V., RIZZU, P., VAN BAREN, M. J., SCHAAP, O., BREEDVELD, G. J., KRIEGER, E., DEKKER, M. C., SQUITIERI, F., IBANEZ, P., JOOSSE, M., VAN DONGEN, J. W., VANACORE, N., VAN SWIETEN, J. C., BRICE, A., MECO, G., VAN DUIJN, C. M., OOSTRA, B. A. & HEUTINK, P. 2003. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science,* 299**,** 256-9.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell,* 132**,** 311-22.

BROCKMANN, K., SRULIJES, K., HAUSER, A. K., SCHULTE, C., CSOTI, I., GASSER, T. & BERG, D. 2011. GBA-associated PD presents with nonmotor characteristics. *Neurology,* 77**,** 276-80.

BROCKMANN, K., SRULIJES, K., PFLEDERER, S., HAUSER, A. K., SCHULTE, C., MAETZLER, W., GASSER, T. & BERG, D. 2015. GBA-associated Parkinson's disease: reduced survival and more rapid progression in a prospective longitudinal study. *Mov Disord,* 30**,** 407-11.

BROWN, E. E., BLAUWENDRAAT, C., TRINH, J., RIZIG, M., NALLS, M. A., LEVEILLE, E., RUSKEY, J. A., JONVIK, H., TAN, M. M. X., BANDRES-CIGA, S., HASSIN-BAER, S., BROCKMANN, K., INFANTE, J., TOLOSA, E., EZQUERRA, M., BEN ROMDHAN, S., BENMAHDJOUB, M., AREZKI, M., MHIRI, C., HARDY, J., SINGLETON, A. B., ALCALAY, R. N., GASSER, T., GROSSET, D. G., WILLIAMS, N. M., PITTMAN, A., GAN-OR, Z., FERNANDEZ-SANTIAGO, R., BRICE, A., LESAGE, S., FARRER, M., WOOD, N. & MORRIS, H. R. 2021. Analysis of DNM3 and VAMP4 as genetic modifiers of LRRK2 Parkinson's disease. *Neurobiol Aging,* 97**,** 148.e17-148.e24.

BRYOIS, J., SKENE, N. G., HANSEN, T. F., KOGELMAN, L. J. A., WATSON, H. J., LIU, Z., BRUEGGEMAN, L., BREEN, G., BULIK, C. M., ARENAS, E., HJERLING-LEFFLER, J. & SULLIVAN, P. F. 2020. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat Genet,* 52**,** 482-493.

BRAAK, H., DEL TREDICI, K., RUB, U., DE VOS, R. A., JANSEN STEUR, E. N. & BRAAK, E. 2003. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging,* 24**,** 197-211.

BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods,* 10**,** 1213-8.

BUENROSTRO, J. D., WU, B., CHANG, H. Y. & GREENLEAF, W. J. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol,* 109**,** 21.29.1-9.

BUNIELLO, A., MACARTHUR, J. A. L., CEREZO, M., HARRIS, L. W., HAYHURST, J., MALANGONE, C., MCMAHON, A., MORALES, J., MOUNTJOY, E., SOLLIS, E., SUVEGES, D., VROUSGOU, O., WHETZEL, P. L., AMODE, R., GUILLEN, J. A., RIAT, H. S., TREVANION, S. J., HALL, P., JUNKINS, H., FLICEK, P., BURDETT, T., HINDORFF, L. A., CUNNINGHAM, F. & PARKINSON, H. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res,* 47**,** D1005-d1012.

BUTTON, K. S., IOANNIDIS, J. P., MOKRYSZ, C., NOSEK, B. A., FLINT, J., ROBINSON, E. S. & MUNAFÒ, M. R. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci,* 14**,** 365-76.

BYRD, J. B., GREENE, A. C., PRASAD, D. V., JIANG, X. & GREENE, C. S. 2020. Responsible, practical genomic data sharing that accelerates research. *Nat Rev Genet,* 21**,** 615-629.

CASTEL, S. E., CERVERA, A., MOHAMMADI, P., AGUET, F., REVERTER, F., WOLMAN, A., GUIGO, R., IOSSIFOV, I., VASILEVA, A. & LAPPALAINEN, T. 2018. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat Genet,* 50**,** 1327-1334.

CHANG, C. C., CHOW, C. C., TELLIER, L. C., VATTIKUTI, S., PURCELL, S. M. & LEE, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience,* 4**,** 7.

CHANG, D., NALLS, M. A., HALLGRIMSDOTTIR, I. B., HUNKAPILLER, J., VAN DER BRUG, M., CAI, F., KERCHNER, G. A., AYALON, G., BINGOL, B., SHENG, M., HINDS, D., BEHRENS, T. W., SINGLETON, A. B., BHANGALE, T. R. & GRAHAM, R. R. 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet,* 49**,** 1511-1516.

CHAUDHURI, K. R., ODIN, P., ANTONINI, A. & MARTINEZ-MARTIN, P. 2011. Parkinson's disease: the non-motor issues. *Parkinsonism Relat Disord,* 17**,** 717-23.

CHEN, H. & RITZ, B. 2018. The Search for Environmental Causes of Parkinson's Disease: Moving Forward. *J Parkinsons Dis,* 8**,** S9-s17.

CHIARELLA, A. M., LU, D. & HATHAWAY, N. A. 2020. Epigenetic Control of a Local Chromatin Landscape. *Int J Mol Sci,* 21.

CILIA, R., TUNESI, S., MAROTTA, G., CEREDA, E., SIRI, C., TESEI, S., ZECCHINELLI, A. L., CANESI, M., MARIANI, C. B., MEUCCI, N., SACILOTTO, G., ZINI, M., BARICHELLA, M., MAGNANI, C., DUGA, S., ASSELTA, R., SOLDA, G., SERESINI, A., SEIA, M., PEZZOLI, G. & GOLDWURM, S. 2016. Survival and dementia in GBA-associated Parkinson's disease: The mutation matters. *Ann Neurol,* 80**,** 662-673.

CLARKE, G. M., ANDERSON, C. A., PETTERSSON, F. H., CARDON, L. R., MORRIS, A. P. & ZONDERVAN, K. T. 2011. Basic statistical analysis in genetic case-control studies. *Nat Protoc,* 6**,** 121-33.

CLAUSSNITZER, M., DANKEL, S. N., KIM, K. H., QUON, G., MEULEMAN, W., HAUGEN, C., GLUNK, V., SOUSA, I. S., BEAUDRY, J. L., PUVIINDRAN, V., ABDENNUR, N. A., LIU, J., SVENSSON, P. A., HSU, Y. H., DRUCKER, D. J., MELLGREN, G., HUI, C. C., HAUNER, H. & KELLIS, M. 2015. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med,* 373**,** 895-907.

COETZEE, S. G., PIERCE, S., BRUNDIN, P., BRUNDIN, L., HAZELETT, D. J. & COETZEE, G. A. 2016. Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology. *Sci Rep,* 6**,** 30509.

COLEMAN, J. R., EUESDEN, J., PATEL, H., FOLARIN, A. A., NEWHOUSE, S. & BREEN, G. 2016. Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Brief Funct Genomics,* 15**,** 298-304.

COWPER-SAL LARI, R., ZHANG, X., WRIGHT, J. B., BAILEY, S. D., COLE, M. D., EECKHOUTE, J., MOORE, J. H. & LUPIEN, M. 2012. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet,* 44**,** 1191-8.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

DARMANIS, S., SLOAN, S. A., ZHANG, Y., ENGE, M., CANEDA, C., SHUER, L. M., HAYDEN GEPHART, M. G., BARRES, B. A. & QUAKE, S. R. 2015. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A,* 112**,** 7285-90.

DAS, S., FORER, L., SCHONHERR, S., SIDORE, C., LOCKE, A. E., KWONG, A., VRIEZE, S. I., CHEW, E. Y., LEVY, S., MCGUE, M., SCHLESSINGER, D., STAMBOLIAN, D., LOH, P. R., IACONO, W. G., SWAROOP, A., SCOTT, L. J., CUCCA, F., KRONENBERG, F., BOEHNKE, M., ABECASIS, G. R. & FUCHSBERGER, C. 2016. Next-generation genotype imputation service and methods. *Nat Genet,* 48**,** 1284-1287.

DAVIS, M. Y., JOHNSON, C. O., LEVERENZ, J. B., WEINTRAUB, D., TROJANOWSKI, J. Q., CHEN-PLOTKIN, A., VAN DEERLIN, V. M., QUINN, J. F., CHUNG, K. A., PETERSON-HILLER, A. L., ROSENTHAL, L. S., DAWSON, T. M., ALBERT, M. S., GOLDMAN, J. G., STEBBINS, G. T., BERNARD, B., WSZOLEK, Z. K., ROSS, O. A., DICKSON, D. W., EIDELBERG, D., MATTIS, P. J., NIETHAMMER, M., YEAROUT, D., HU, S. C., CHOLERTON, B. A., SMITH, M., MATA, I. F., MONTINE, T. J., EDWARDS, K. L. & ZABETIAN, C. P. 2016. Association of GBA Mutations and the E326K Polymorphism With Motor and Cognitive Progression in Parkinson Disease. *JAMA Neurol,* 73**,** 1217-1224.

DE LAU, L. M. & BRETELER, M. M. 2006. Epidemiology of Parkinson's disease. *Lancet Neurol,* 5**,** 525-35.

DE, S., PEDERSEN, B. S. & KECHRIS, K. 2014. The dilemma of choosing the ideal permutation strategy while estimating statistical significance of genome-wide enrichment. *Brief Bioinform,* 15**,** 919-28.

DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. *Science,* 295**,** 1306-11.

DENG, H., WANG, P. & JANKOVIC, J. 2018. The genetics of Parkinson disease. *Ageing Res Rev,* 42**,** 72-85.

DENNIS, D. J., HAN, S. & SCHUURMANS, C. 2019. bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res,* 1705**,** 48-65.

DEPLANCKE, B., ALPERN, D. & GARDEUX, V. 2016. The Genetics of Transcription Factor DNA Binding Variation. *Cell,* 166**,** 538-554.

DINA, C., MEYRE, D., GALLINA, S., DURAND, E., KÖRNER, A., JACOBSON, P., CARLSSON, L. M., KIESS, W., VATIN, V., LECOEUR, C., DELPLANQUE, J., VAILLANT, E., PATTOU, F., RUIZ, J., WEILL, J., LEVY-MARCHAL, C., HORBER, F., POTOCZNA, N., HERCBERG, S., LE STUNFF, C., BOUGNÈRES, P., KOVACS, P., MARRE, M., BALKAU, B., CAUCHI, S., CHÈVRE, J. C. & FROGUEL, P. 2007. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet,* 39**,** 724-6.

DOZMOROV, M. G. 2017. Epigenomic annotation-based interpretation of genomic data: from enrichment analysis to machine learning. *Bioinformatics,* 33**,** 3323-3330.

DURAN, R., MENCACCI, N. E., ANGELI, A. V., SHOAI, M., DEAS, E., HOULDEN, H., MEHTA, A., HUGHES, D., COX, T. M., DEEGAN, P., SCHAPIRA, A. H., LEES, A. J., LIMOUSIN, P., JARMAN, P. R., BHATIA, K. P., WOOD, N. W., HARDY, J. & FOLTYNIE, T. 2013. The glucocerebrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease. *Mov Disord,* 28**,** 232-6.

DUVOISIN, R. C., ELDRIDGE, R., WILLIAMS, A., NUTT, J. & CALNE, D. 1981. Twin study of Parkinson disease. *Neurology,* 31**,** 77-80.

ECKER, J. R., GESCHWIND, D. H., KRIEGSTEIN, A. R., NGAI, J., OSTEN, P., POLIOUDAKIS, D., REGEV, A., SESTAN, N., WICKERSHAM, I. R. & ZENG, H.

2017. The BRAIN Initiative Cell Census Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. *Neuron,* 96**,** 542-557.

EDWARDS, T. L., SCOTT, W. K., ALMONTE, C., BURT, A., POWELL, E. H., BEECHAM, G. W., WANG, L., ZUCHNER, S., KONIDARI, I., WANG, G., SINGER, C., NAHAB, F., SCOTT, B., STAJICH, J. M., PERICAK-VANCE, M., HAINES, J., VANCE, J. M. & MARTIN, E. R. 2010. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Ann Hum Genet,* 74**,** 97-109.

ENCODE PROJECT CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

ESCOTT-PRICE, V., NALLS, M. A., MORRIS, H. R., LUBBE, S., BRICE, A., GASSER, T., HEUTINK, P., WOOD, N. W., HARDY, J., SINGLETON, A. B. & WILLIAMS, N. M. 2015. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann Neurol,* 77**,** 582-91.

FARRER, M., KACHERGUS, J., FORNO, L., LINCOLN, S., WANG, D. S., HULIHAN, M., MARAGANORE, D., GWINN-HARDY, K., WSZOLEK, Z., DICKSON, D. & LANGSTON, J. W. 2004. Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. *Ann Neurol,* 55**,** 174-9.

FERNANDEZ-SANTIAGO, R., GARRIDO, A., INFANTE, J., GONZALEZ-ARAMBURU, I., SIERRA, M., FERNANDEZ, M., VALLDEORIOLA, F., MUNOZ, E., COMPTA, Y., MARTI, M. J., RIOS, J., TOLOSA, E. & EZQUERRA, M. 2018. alpha-synuclein (SNCA) but not dynamin 3 (DNM3) influences age at onset of leucine-rich repeat kinase 2 (LRRK2) Parkinson's disease in Spain. *Mov Disord,* 33**,** 637-641.

FINUCANE, H. K., BULIK-SULLIVAN, B., GUSEV, A., TRYNKA, G., RESHEF, Y., LOH, P. R., ANTTILA, V., XU, H., ZANG, C., FARH, K., RIPKE, S., DAY, F. R., PURCELL, S., STAHL, E., LINDSTROM, S., PERRY, J. R., OKADA, Y., RAYCHAUDHURI, S., DALY, M. J., PATTERSON, N., NEALE, B. M. & PRICE, A. L. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet,* 47**,** 1228-35.

FOO, J. N., TAN, L. C., AU, W. L., PRAKASH, K. M., LIU, J. & TAN, E. K. 2019. No association of DNM3 with age of onset in Asian Parkinson's disease. *Eur J Neurol,* 26**,** 827-829.

FORNES, O., CASTRO-MONDRAGON, J. A., KHAN, A., VAN DER LEE, R., ZHANG, X., RICHMOND, P. A., MODI, B. P., CORREARD, S., GHEORGHE, M., BARANAŠIĆ, D., SANTANA-GARCIA, W., TAN, G., CHÈNEBY, J., BALLESTER, B., PARCY, F., SANDELIN, A., LENHARD, B., WASSERMAN, W. W. & MATHELIER, A. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res,* 48**,** D87-d92.

FRAYLING, T. M., TIMPSON, N. J., WEEDON, M. N., ZEGGINI, E., FREATHY, R. M., LINDGREN, C. M., PERRY, J. R., ELLIOTT, K. S., LANGO, H., RAYNER, N. W., SHIELDS, B., HARRIES, L. W., BARRETT, J. C., ELLARD, S., GROVES, C. J., KNIGHT, B., PATCH, A. M., NESS, A. R., EBRAHIM, S., LAWLOR, D. A., RING, S. M., BEN-SHLOMO, Y., JARVELIN, M. R., SOVIO, U., BENNETT, A. J., MELZER, D., FERRUCCI, L., LOOS, R. J., BARROSO, I., WAREHAM, N. J., KARPE, F., OWEN, K. R., CARDON, L. R., WALKER, M., HITMAN, G. A., PALMER, C. N., DONEY, A. S., MORRIS, A. D., SMITH, G. D., HATTERSLEY, A. T. & MCCARTHY, M. I. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science,* 316**,** 889-94.

FROMER, M., ROUSSOS, P., SIEBERTS, S. K., JOHNSON, J. S., KAVANAGH, D. H., PERUMAL, T. M., RUDERFER, D. M., OH, E. C., TOPOL, A., SHAH, H. R., KLEI, L. L., KRAMER, R., PINTO, D., GÜMÜŞ, Z. H., CICEK, A. E., DANG, K. K., BROWNE, A., LU, C., XIE, L., READHEAD, B., STAHL, E. A., XIAO, J., PARVIZI, M., HAMAMSY, T., FULLARD, J. F., WANG, Y. C., MAHAJAN, M. C., DERRY, J. M., DUDLEY, J. T., HEMBY, S. E., LOGSDON, B. A., TALBOT, K., RAJ, T., BENNETT, D. A., DE JAGER, P. L., ZHU, J., ZHANG, B., SULLIVAN, P. F., CHESS, A., PURCELL, S. M., SHINOBU, L. A., MANGRAVITE, L. M., TOYOSHIBA, H., GUR, R. E., HAHN, C. G., LEWIS, D. A., HAROUTUNIAN, V., PETERS, M. A., LIPSKA, B. K., BUXBAUM, J. D., SCHADT, E. E., HIRAI, K., ROEDER, K., BRENNAND, K. J., KATSANIS, N., DOMENICI, E., DEVLIN, B. & SKLAR, P. 2016. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci,* 19**,** 1442-1453.

FULLARD, J. F., HAUBERG, M. E., BENDL, J., EGERVARI, G., CIRNARU, M. D., REACH, S. M., MOTL, J., EHRLICH, M. E., HURD, Y. L. & ROUSSOS, P. 2018. An atlas of chromatin accessibility in the adult human brain. *Genome Res,* 28**,** 1243-1252.

FUNG, H. C., SCHOLZ, S., MATARIN, M., SIMON-SANCHEZ, J., HERNANDEZ, D., BRITTON, A., GIBBS, J. R., LANGEFELD, C., STIEGERT, M. L., SCHYMICK, J., OKUN, M. S., MANDEL, R. J., FERNANDEZ, H. H., FOOTE, K. D., RODRIGUEZ, R. L., PECKHAM, E., DE VRIEZE, F. W., GWINN-HARDY, K., HARDY, J. A. & SINGLETON, A. 2006. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol,* 5**,** 911-6.

GAGLIANO, S. A., POUGET, J. G., HARDY, J., KNIGHT, J., BARNES, M. R., RYTEN, M. & WEALE, M. E. 2016. Genomics implicates adaptive and innate immunity in Alzheimer's and Parkinson's diseases. *Ann Clin Transl Neurol,* 3**,** 924-933.

GALBIATI, A., VERGA, L., GIORA, E., ZUCCONI, M. & FERINI-STRAMBI, L. 2019. The risk of neurodegeneration in REM sleep behavior disorder: A systematic review and meta-analysis of longitudinal studies. *Sleep Med Rev,* 43**,** 37-46.

GALLAGHER, M. D. & CHEN-PLOTKIN, A. S. 2018. The Post-GWAS Era: From Association to Function. *Am J Hum Genet,* 102**,** 717-730.

GAN-OR, Z., AMSHALOM, I., KILARSKI, L. L., BAR-SHIRA, A., GANA-WEISZ, M., MIRELMAN, A., MARDER, K., BRESSMAN, S., GILADI, N. & ORR-URTREGER, A. 2015a. Differential effects of severe vs mild GBA mutations on Parkinson disease. *Neurology,* 84**,** 880-7.

GAN-OR, Z., MIRELMAN, A., POSTUMA, R. B., ARNULF, I., BAR-SHIRA, A., DAUVILLIERS, Y., DESAUTELS, A., GAGNON, J. F., LEBLOND, C. S., FRAUSCHER, B., ALCALAY, R. N., SAUNDERS-PULLMAN, R., BRESSMAN, S. B., MARDER, K., MONACA, C., HOGL, B., ORR-URTREGER, A., DION, P. A., MONTPLAISIR, J. Y., GILADI, N. & ROULEAU, G. A. 2015b. GBA mutations are associated with Rapid Eye Movement Sleep Behavior Disorder. *Ann Clin Transl Neurol,* 2**,** 941-5.

GARCIA-ALONSO, L., HOLLAND, C. H., IBRAHIM, M. M., TUREI, D. & SAEZ-RODRIGUEZ, J. 2019. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res,* 29**,** 1363-1375.

GBD 2016 PARKINSON'S DISEASE COLLABORATORS 2018. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol,* 17**,** 939-953.

GEGG, M. E., BURKE, D., HEALES, S. J., COOPER, J. M., HARDY, J., WOOD, N. W. & SCHAPIRA, A. H. 2012. Glucocerebrosidase deficiency in substantia nigra of parkinson disease brains. *Ann Neurol,* 72**,** 455-63.

GIGUERE, N., BURKE NANNI, S. & TRUDEAU, L. E. 2018. On Cell Loss and Selective Vulnerability of Neuronal Populations in Parkinson's Disease. *Front Neurol,* 9**,** 455.

GIRDHAR, K., HOFFMAN, G. E., JIANG, Y., BROWN, L., KUNDAKOVIC, M., HAUBERG, M. E., FRANCOEUR, N. J., WANG, Y. C., SHAH, H., KAVANAGH, D. H., ZHAROVSKY, E., JACOBOV, R., WISEMAN, J. R., PARK, R., JOHNSON, J. S., KASSIM, B. S., SLOOFMAN, L., MATTEI, E., WENG, Z., SIEBERTS, S. K., PETERS, M. A., HARRIS, B. T., LIPSKA, B. K., SKLAR, P., ROUSSOS, P. & AKBARIAN, S. 2018. Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nat Neurosci,* 21**,** 1126-1136.

GOEDERT, M., SPILLANTINI, M. G., DEL TREDICI, K. & BRAAK, H. 2013. 100 years of Lewy pathology. *Nat Rev Neurol,* 9**,** 13-24.

GOKER-ALPAN, O., SCHIFFMANN, R., LAMARCA, M. E., NUSSBAUM, R. L., MCINERNEY-LEO, A. & SIDRANSKY, E. 2004. Parkinsonism among Gaucher disease carriers. *J Med Genet,* 41**,** 937-40.

GRABOWSKI, G. A. 2008. Phenotype, diagnosis, and treatment of Gaucher's disease. *Lancet,* 372**,** 1263-71.

GRONEK, P., HAAS, A. N., CZARNY, W., PODSTAWSKI, R., DELABARY, M. D. S., CLARK, C. C., BORACZYŃSKI, M., TARNAS, M., WYCICHOWSKA, P., PAWLACZYK, M. & GRONEK, J. 2021. The Mechanism of Physical Activity-induced Amelioration of Parkinson's Disease: A Narrative Review. *Aging Dis,* 12**,** 192-202.

HAMZA, T. H., ZABETIAN, C. P., TENESA, A., LAEDERACH, A., MONTIMURRO, J., YEAROUT, D., KAY, D. M., DOHENY, K. F., PASCHALL, J., PUGH, E., KUSEL, V. I., COLLURA, R., ROBERTS, J., GRIFFITH, A., SAMII, A., SCOTT, W. K., NUTT, J., FACTOR, S. A. & PAYAMI, H. 2010. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat Genet,* 42**,** 781-5.

HASSLER, R. 1938. Zur Pathologie der Paralysis agitans und des postenzephalitischen Parkinsonismus. *J. Psychol. Neurol.,* 48**,** 387-455.

HAWKES, C. H., DEL TREDICI, K. & BRAAK, H. 2007. Parkinson's disease: a dual-hit hypothesis. *Neuropathol Appl Neurobiol,* 33**,** 599-614.

HEALY, D. G., FALCHI, M., O'SULLIVAN, S. S., BONIFATI, V., DURR, A., BRESSMAN, S., BRICE, A., AASLY, J., ZABETIAN, C. P., GOLDWURM, S., FERREIRA, J. J., TOLOSA, E., KAY, D. M., KLEIN, C., WILLIAMS, D. R., MARRAS, C., LANG, A. E., WSZOLEK, Z. K., BERCIANO, J., SCHAPIRA, A. H., LYNCH, T., BHATIA, K. P., GASSER, T., LEES, A. J. & WOOD, N. W. 2008. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol,* 7**,** 583-90.

HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8.

HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C., CHING, K. A., WANG, W., WENG, Z., GREEN, R. D., CRAWFORD, G. E. & REN, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet,* 39**,** 311-8.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HELLWEGE, J. N., KEATON, J. M., GIRI, A., GAO, X., VELEZ EDWARDS, D. R. & EDWARDS, T. L. 2017. Population Stratification in Genetic Association Studies. *Curr Protoc Hum Genet,* 95**,** 1.22.1-1.22.23.

HENRIKSEN, M. G., NORDGAARD, J. & JANSSON, L. B. 2017. Genetics of Schizophrenia: Overview of Methods, Findings and Limitations. *Front Hum Neurosci,* 11**,** 322.

HOFFMANN, T. J. & WITTE, J. S. 2015. Strategies for Imputing and Analyzing Rare Variants in Association Studies. *Trends Genet,* 31**,** 556-563.

HONG, E. P. & PARK, J. W. 2012. Sample size and statistical power calculation in genetic association studies. *Genomics Inform,* 10**,** 117-22.

HOROWITZ, M., WILDER, S., HOROWITZ, Z., REINER, O., GELBART, T. & BEUTLER, E. 1989. The human glucocerebrosidase gene and pseudogene: structure and evolution. *Genomics,* 4**,** 87-96.

HRUSKA, K. S., LAMARCA, M. E., SCOTT, C. R. & SIDRANSKY, E. 2008. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum Mutat,* 29**,** 567-83.

HUANG, Y., DENG, L., ZHONG, Y. & YI, M. 2018. The Association between E326K of GBA and the Risk of Parkinson's Disease. *Parkinsons Dis,* 2018**,** 1048084.

IBÁÑEZ, P., BONNET, A. M., DÉBARGES, B., LOHMANN, E., TISON, F., POLLAK, P., AGID, Y., DÜRR, A. & BRICE, A. 2004. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet,* 364**,** 1169-71.

INTERNATIONAL PARKINSON'S DISEASE GENOMICS CONSORTIUM AND WELLCOME TRUST CASE CONTROL CONSORTIUM 2011. A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet,* 7**,** e1002142.

INUKAI, S., KOCK, K. H. & BULYK, M. L. 2017. Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev,* 43**,** 110-119.

IOTCHKOVA, V., RITCHIE, G. R. S., GEIHS, M., MORGANELLA, S., MIN, J. L., WALTER, K., TIMPSON, N. J., DUNHAM, I., BIRNEY, E. & SORANZO, N. 2019. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet,* 51**,** 343-353.

IWAKI, H., BLAUWENDRAAT, C., LEONARD, H. L., LIU, G., MAPLE-GRØDEM, J., CORVOL, J. C., PIHLSTRØM, L., VAN NIMWEGEN, M., HUTTEN, S. J., NGUYEN, K. H., RICK, J., EBERLY, S., FAGHRI, F., AUINGER, P., SCOTT, K. M., WIJEYEKOON, R., VAN DEERLIN, V. M., HERNANDEZ, D. G., DAY-WILLIAMS, A. G., BRICE, A., ALVES, G., NOYCE, A. J., TYSNES, O. B., EVANS, J. R., BREEN, D. P., ESTRADA, K., WEGEL, C. E., DANJOU, F., SIMON, D. K., RAVINA, B., TOFT, M., HEUTINK, P., BLOEM, B. R., WEINTRAUB, D., BARKER, R. A., WILLIAMS-GRAY, C. H., VAN DE WARRENBURG, B. P., VAN HILTEN, J. J., SCHERZER, C. R., SINGLETON, A. B. & NALLS, M. A. 2019. Genetic risk of Parkinson disease and progression:: An analysis of 13 longitudinal cohorts. *Neurol Genet,* 5**,** e348.

JENUWEIN, T. & ALLIS, C. D. 2001. Translating the histone code. *Science,* 293**,** 1074-80.

JESUS, S., HUERTAS, I., BERNAL-BERNAL, I., BONILLA-TORIBIO, M., CACERES-REDONDO, M. T., VARGAS-GONZALEZ, L., GOMEZ-LLAMAS, M., CARRILLO, F., CALDERON, E., CARBALLO, M., GOMEZ-GARRE, P. & MIR, P. 2016. GBA Variants Influence Motor and Non-Motor Features of Parkinson's Disease. *PLoS One,* 11**,** e0167749.

JOHNSON, D. S., MORTAZAVI, A., MYERS, R. M. & WOLD, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science,* 316**,** 1497-502.

JONES, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet,* 13**,** 484-92.

JUNG, M., HABERLE, B. M., TSCHAIKOWSKY, T., WITTMANN, M. T., BALTA, E. A., STADLER, V. C., ZWEIER, C., DORFLER, A., GLOECKNER, C. J. & LIE, D. C. 2018. Analysis of the expression pattern of the schizophrenia-risk and intellectual disability gene TCF4 in the developing and adult brain suggests a role in development and plasticity of cortical and hippocampal neurons. *Mol Autism,* 9**,** 20.

KACHERGUS, J., MATA, I. F., HULIHAN, M., TAYLOR, J. P., LINCOLN, S., AASLY, J., GIBSON, J. M., ROSS, O. A., LYNCH, T., WILEY, J., PAYAMI, H., NUTT, J., MARAGANORE, D. M., CZYZEWSKI, K., STYCZYNSKA, M., WSZOLEK, Z. K., FARRER, M. J. & TOFT, M. 2005. Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations. *Am J Hum Genet,* 76**,** 672-80.

KANDURI, C., BOCK, C., GUNDERSEN, S., HOVIG, E. & SANDVE, G. K. 2019. Colocalization analyses of genomic elements: approaches, recommendations and challenges. *Bioinformatics,* 35**,** 1615-1624.

KARABACAK CALVIELLO, A., HIRSEKORN, A., WURMUS, R., YUSUF, D. & OHLER, U. 2019. Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol,* 20**,** 42.

KARCZEWSKI, K. J., DUDLEY, J. T., KUKURBA, K. R., CHEN, R., BUTTE, A. J., MONTGOMERY, S. B. & SNYDER, M. 2013. Systematic functional regulatory assessment of disease-associated variants. *Proc Natl Acad Sci U S A,* 110**,** 9607-12.

KEILWAGEN, J., POSCH, S. & GRAU, J. 2019. Accurate prediction of cell type-specific transcription factor binding. *Genome Biol,* 20**,** 9.

KILARSKI, L. L., PEARSON, J. P., NEWSWAY, V., MAJOUNIE, E., KNIPE, M. D., MISBAHUDDIN, A., CHINNERY, P. F., BURN, D. J., CLARKE, C. E., MARION, M. H., LEWTHWAITE, A. J., NICHOLL, D. J., WOOD, N. W., MORRISON, K. E., WILLIAMS-GRAY, C. H., EVANS, J. R., SAWCER, S. J., BARKER, R. A., WICKREMARATCHI, M. M., BEN-SHLOMO, Y., WILLIAMS, N. M. & MORRIS, H. R. 2012. Systematic review and UK-based study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. *Mov Disord,* 27**,** 1522-9.

KITADA, T., ASAKAWA, S., HATTORI, N., MATSUMINE, H., YAMAMURA, Y., MINOSHIMA, S., YOKOCHI, M., MIZUNO, Y. & SHIMIZU, N. 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature,* 392**,** 605-8.

KLEMM, S. L., SHIPONY, Z. & GREENLEAF, W. J. 2019. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet,* 20**,** 207-220.

KORDOWER, J. H., OLANOW, C. W., DODIYA, H. B., CHU, Y., BEACH, T. G., ADLER, C. H., HALLIDAY, G. M. & BARTUS, R. T. 2013. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. *Brain,* 136**,** 2419-31.

KULAKOVSKIY, I. V., VORONTSOV, I. E., YEVSHIN, I. S., SHARIPOV, R. N., FEDOROVA, A. D., RUMYNSKIY, E. I., MEDVEDEVA, Y. A., MAGANA-MORA, A., BAJIC, V. B., PAPATSENKO, D. A., KOLPAKOV, F. A. & MAKEEV, V. J. 2018. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res,* 46**,** D252-d259.

KUNDAJE, A., MEULEMAN, W., ERNST, J., BILENKY, M., YEN, A., HERAVI-MOUSSAVI, A., KHERADPOUR, P., ZHANG, Z., WANG, J., ZILLER, M. J., AMIN, V., WHITAKER, J. W., SCHULTZ, M. D., WARD, L. D., SARKAR, A., QUON, G., SANDSTROM, R. S., EATON, M. L., WU, Y. C., PFENNING, A. R., WANG, X., CLAUSSNITZER, M., LIU, Y., COARFA, C., HARRIS, R. A., SHORESH, N., EPSTEIN, C. B., GJONESKA, E., LEUNG, D., XIE, W., HAWKINS, R. D., LISTER, R., HONG, C., GASCARD, P., MUNGALL, A. J., MOORE, R., CHUAH, E., TAM, A., CANFIELD, T. K., HANSEN, R. S., KAUL, R., SABO, P. J., BANSAL, M. S., CARLES, A., DIXON, J. R., FARH, K. H., FEIZI, S., KARLIC, R., KIM, A. R., KULKARNI, A., LI, D., LOWDON, R., ELLIOTT, G., MERCER, T. R., NEPH, S. J., ONUCHIC, V., POLAK, P., RAJAGOPAL, N., RAY, P., SALLARI, R. C., SIEBENTHALL, K. T., SINNOTT-ARMSTRONG, N. A., STEVENS, M., THURMAN, R. E., WU, J., ZHANG, B., ZHOU, X., BEAUDET, A. E., BOYER, L. A., DE JAGER, P. L., FARNHAM, P. J., FISHER, S. J., HAUSSLER, D., JONES, S. J., LI, W., MARRA, M. A., MCMANUS, M. T., SUNYAEV, S., THOMSON, J. A., TLSTY, T. D., TSAI, L. H., WANG, W., WATERLAND, R. A., ZHANG, M. Q., CHADWICK, L. H., BERNSTEIN, B. E., COSTELLO, J. F., ECKER, J. R., HIRST, M., MEISSNER, A., MILOSAVLJEVIC, A., REN, B., STAMATOYANNOPOULOS, J. A., WANG, T. & KELLIS, M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature,* 518**,** 317-30.

LAKE, B. B., AI, R., KAESER, G. E., SALATHIA, N. S., YUNG, Y. C., LIU, R., WILDBERG, A., GAO, D., FUNG, H. L., CHEN, S., VIJAYARAGHAVAN, R., WONG, J., CHEN, A., SHENG, X., KAPER, F., SHEN, R., RONAGHI, M., FAN, J. B., WANG, W., CHUN, J. & ZHANG, K. 2016. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science,* 352**,** 1586-90.

LAMBERT, S. A., JOLMA, A., CAMPITELLI, L. F., DAS, P. K., YIN, Y., ALBU, M., CHEN, X., TAIPALE, J., HUGHES, T. R. & WEIRAUCH, M. T. 2018. The Human Transcription Factors. *Cell,* 172**,** 650-665.

LANGSTON, J. W., BALLARD, P., TETRUD, J. W. & IRWIN, I. 1983. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science,* 219**,** 979-80.

LEE, Y. H. 2015. Meta-analysis of genetic association studies. *Ann Lab Med,* 35**,** 283-7.

LENHARD, B., SANDELIN, A. & CARNINCI, P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet,* 13**,** 233-45.

LESAGE, S., ANHEIM, M., CONDROYER, C., POLLAK, P., DURIF, F., DUPUITS, C., VIALLET, F., LOHMANN, E., CORVOL, J. C., HONORE, A., RIVAUD, S., VIDAILHET, M., DURR, A. & BRICE, A. 2011. Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease. *Hum Mol Genet,* 20**,** 202-10.

LEWY, F. H. 1912. Paralysis agitans. I. Patologische Anatomie. *In:* LEWANDOWSKY, M. A., G. (ed.) *Handbuch der Neurologie Vol. 3.* Berlin: Springer-Verlag.

LI, H., QUANG, D. & GUAN, Y. 2019a. Anchor: trans-cell type prediction of transcription factor binding sites. *Genome Res,* 29**,** 281-292.

LI, Z., SCHULZ, M. H., LOOK, T., BEGEMANN, M., ZENKE, M. & COSTA, I. G. 2019b. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol,* 20**,** 45.

LIEBERMAN-AIDEN, E., VAN BERKUM, N. L., WILLIAMS, L., IMAKAEV, M., RAGOCZY, T., TELLING, A., AMIT, I., LAJOIE, B. R., SABO, P. J., DORSCHNER, M. O., SANDSTROM, R., BERNSTEIN, B., BENDER, M. A., GROUDINE, M., GNIRKE, A., STAMATOYANNOPOULOS, J., MIRNY, L. A., LANDER, E. S. & DEKKER, J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science,* 326**,** 289-93.

LIHU, A. & HOLBAN, S. 2015. A review of ensemble methods for de novo motif discovery in ChIP-Seq data. *Brief Bioinform,* 16**,** 964-73.

LILL, C. M. 2016. Genetics of Parkinson's disease. *Mol Cell Probes,* 30**,** 386-396.

LILL, C. M., ROEHR, J. T., MCQUEEN, M. B., KAVVOURA, F. K., BAGADE, S., SCHJEIDE, B. M., SCHJEIDE, L. M., MEISSNER, E., ZAUFT, U., ALLEN, N. C., LIU, T., SCHILLING, M., ANDERSON, K. J., BEECHAM, G., BERG, D., BIERNACKA, J. M., BRICE, A., DESTEFANO, A. L., DO, C. B., ERIKSSON, N., FACTOR, S. A., FARRER, M. J., FOROUD, T., GASSER, T., HAMZA, T., HARDY, J. A., HEUTINK, P., HILL-BURNS, E. M., KLEIN, C., LATOURELLE, J. C., MARAGANORE, D. M., MARTIN, E. R., MARTINEZ, M., MYERS, R. H., NALLS, M. A., PANKRATZ, N., PAYAMI, H., SATAKE, W., SCOTT, W. K., SHARMA, M., SINGLETON, A. B., STEFANSSON, K., TODA, T., TUNG, J. Y., VANCE, J., WOOD, N. W., ZABETIAN, C. P., YOUNG, P., TANZI, R. E., KHOURY, M. J., ZIPP, F., LEHRACH, H., IOANNIDIS, J. P. & BERTRAM, L. 2012. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet,* 8**,** e1002548.

LUNETTA, K. L. 2013. Methods for meta-analysis of genetic data. *Curr Protoc Hum Genet,* Chapter 1**,** Unit1.24.

MA, W., NOBLE, W. S. & BAILEY, T. L. 2014. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat Protoc,* 9**,** 1428-50.

MAGI, R. & MORRIS, A. P. 2010. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics,* 11**,** 288.

MAHLKNECHT, P., SEPPI, K. & POEWE, W. 2015. The Concept of Prodromal Parkinson's Disease. *J Parkinsons Dis,* 5**,** 681-97.

MALLETT, V., ROSS, J. P., ALCALAY, R. N., AMBALAVANAN, A., SIDRANSKY, E., DION, P. A., ROULEAU, G. A. & GAN-OR, Z. 2016. GBA p.T369M substitution in Parkinson disease: Polymorphism or association? A meta-analysis. *Neurol Genet,* 2**,** e104.

MANDUCHI, E., CHESI, A., HALL, M. A., GRANT, S. F. A. & MOORE, J. H. 2018. Leveraging putative enhancer-promoter interactions to investigate two-way epistasis in Type 2 Diabetes GWAS. *Pac Symp Biocomput,* 23**,** 548-558.

MARAGANORE, D. M., DE ANDRADE, M., LESNICK, T. G., STRAIN, K. J., FARRER, M. J., ROCCA, W. A., PANT, P. V., FRAZER, K. A., COX, D. R. & BALLINGER, D. G. 2005. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet,* 77**,** 685-93.

MAREES, A. T., DE KLUIVER, H., STRINGER, S., VORSPAN, F., CURIS, E., MARIE-CLAIRE, C. & DERKS, E. M. 2018. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res,* 27**,** e1608.

MAREK, K., CHOWDHURY, S., SIDEROWF, A., LASCH, S., COFFEY, C. S., CASPELL-GARCIA, C., SIMUNI, T., JENNINGS, D., TANNER, C. M., TROJANOWSKI, J. Q., SHAW, L. M., SEIBYL, J., SCHUFF, N., SINGLETON, A., KIEBURTZ, K., TOGA, A. W., MOLLENHAUER, B., GALASKO, D., CHAHINE, L. M., WEINTRAUB, D., FOROUD, T., TOSUN-TURGUT, D., POSTON, K., ARNEDO, V., FRASIER, M. & SHERER, T. 2018. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann Clin Transl Neurol,* 5**,** 1460-1477.

MARRAS, C., CHAUDHURI, K. R., TITOVA, N. & MESTRE, T. A. 2020. Therapy of Parkinson's Disease Subtypes. *Neurotherapeutics,* 17**,** 1366-1377.

MARSILI, L., RIZZO, G. & COLOSIMO, C. 2018. Diagnostic Criteria for Parkinson's Disease: From James Parkinson to the Concept of Prodromal Disease. *Front Neurol,* 9**,** 156.

MAURANO, M. T., HUMBERT, R., RYNES, E., THURMAN, R. E., HAUGEN, E., WANG, H., REYNOLDS, A. P., SANDSTROM, R., QU, H., BRODY, J., SHAFER, A., NERI, F., LEE, K., KUTYAVIN, T., STEHLING-SUN, S., JOHNSON, A. K., CANFIELD, T. K., GISTE, E., DIEGEL, M., BATES, D., HANSEN, R. S., NEPH, S., SABO, P. J., HEIMFELD, S., RAUBITSCHEK, A., ZIEGLER, S., COTSAPAS, C., SOTOODEHNIA, N., GLASS, I., SUNYAEV, S. R., KAUL, R. & STAMATOYANNOPOULOS, J. A. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science,* 337**,** 1190-5.

MAZZULLI, J. R., XU, Y. H., SUN, Y., KNIGHT, A. L., MCLEAN, P. J., CALDWELL, G. A., SIDRANSKY, E., GRABOWSKI, G. A. & KRAINC, D. 2011. Gaucher disease glucocerebrosidase and alpha-synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell,* 146**,** 37-52.

MCCARTHY, S., DAS, S., KRETZSCHMAR, W., DELANEAU, O., WOOD, A. R., TEUMER, A., KANG, H. M., FUCHSBERGER, C., DANECEK, P., SHARP, K., LUO, Y., SIDORE, C., KWONG, A., TIMPSON, N., KOSKINEN, S., VRIEZE, S., SCOTT, L. J., ZHANG, H., MAHAJAN, A., VELDINK, J., PETERS, U., PATO, C., VAN DUIJN, C. M., GILLIES, C. E., GANDIN, I., MEZZAVILLA, M., GILLY, A., COCCA, M., TRAGLIA, M., ANGIUS, A., BARRETT, J. C., BOOMSMA, D., BRANHAM, K., BREEN, G., BRUMMETT, C. M., BUSONERO, F., CAMPBELL, H., CHAN, A., CHEN, S., CHEW, E., COLLINS, F. S., CORBIN, L. J., SMITH, G. D., DEDOUSSIS, G., DORR, M., FARMAKI, A. E., FERRUCCI, L., FORER, L., FRASER, R. M., GABRIEL, S., LEVY, S., GROOP, L., HARRISON, T., HATTERSLEY, A., HOLMEN, O. L., HVEEM, K., KRETZLER, M., LEE, J. C., MCGUE, M., MEITINGER, T., MELZER, D., MIN, J. L., MOHLKE, K. L., VINCENT, J. B., NAUCK, M., NICKERSON, D., PALOTIE, A., PATO, M., PIRASTU, N., MCINNIS, M., RICHARDS, J. B., SALA, C., SALOMAA, V., SCHLESSINGER, D., SCHOENHERR, S., SLAGBOOM, P. E., SMALL, K., SPECTOR, T., STAMBOLIAN, D., TUKE, M., TUOMILEHTO, J., VAN DEN BERG, L. H., VAN RHEENEN, W., VOLKER, U., WIJMENGA, C., TONIOLO, D., ZEGGINI, E., GASPARINI, P., SAMPSON, M. G., WILSON, J. F., FRAYLING, T., DE BAKKER, P. I., SWERTZ, M. A., MCCARROLL, S., KOOPERBERG, C., DEKKER, A., ALTSHULER, D., WILLER, C., IACONO, W., RIPATTI, S., et al. 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet,* 48**,** 1279-83.

MIGDALSKA-RICHARDS, A., DALY, L., BEZARD, E. & SCHAPIRA, A. H. 2016. Ambroxol effects in glucocerebrosidase and alpha-synuclein transgenic mice. *Ann Neurol,* 80**,** 766-775.

MIGDALSKA-RICHARDS, A. & SCHAPIRA, A. H. 2016. The relationship between glucocerebrosidase mutations and Parkinson disease. *J Neurochem,* 139 Suppl 1**,** 77-90.

MITCHELMORE, J., GRINBERG, N. F., WALLACE, C. & SPIVAKOV, M. 2020. Functional effects of variation in transcription factor binding highlight long-range gene regulation by epromoters. *Nucleic Acids Res,* 48**,** 2866-2879.

MULLIN, S., SMITH, L., LEE, K., D'SOUZA, G., WOODGATE, P., ELFLEIN, J., HÄLLQVIST, J., TOFFOLI, M., STREETER, A., HOSKING, J., HEYWOOD, W. E., KHENGAR, R., CAMPBELL, P., HEHIR, J., CABLE, S., MILLS, K., ZETTERBERG, H., LIMOUSIN, P., LIBRI, V., FOLTYNIE, T. & SCHAPIRA, A. H. V. 2020. Ambroxol for the Treatment of Patients With Parkinson Disease With and Without Glucocerebrosidase Gene Mutations: A Nonrandomized, Noncontrolled Trial. *JAMA Neurol,* 77**,** 427-434.

NALLS, M. A., BLAUWENDRAAT, C., VALLERGA, C. L., HEILBRON, K., BANDRES-CIGA, S., CHANG, D., TAN, M., KIA, D. A., NOYCE, A. J., XUE, A., BRAS, J., YOUNG, E., VON COELLN, R., SIMÓN-SÁNCHEZ, J., SCHULTE, C., SHARMA, M., KROHN, L., PIHLSTRØM, L., SIITONEN, A., IWAKI, H., LEONARD, H., FAGHRI, F., GIBBS, J. R., HERNANDEZ, D. G., SCHOLZ, S. W., BOTIA, J. A., MARTINEZ, M., CORVOL, J. C., LESAGE, S., JANKOVIC, J., SHULMAN, L. M., SUTHERLAND, M., TIENARI, P., MAJAMAA, K., TOFT, M., ANDREASSEN, O. A., BANGALE, T., BRICE, A., YANG, J., GAN-OR, Z., GASSER, T., HEUTINK, P., SHULMAN, J. M., WOOD, N. W., HINDS, D. A., HARDY, J. A., MORRIS, H. R., GRATTEN, J., VISSCHER, P. M., GRAHAM, R. R. & SINGLETON, A. B. 2019. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol,* 18**,** 1091-1102.

NALLS, M. A., DURAN, R., LOPEZ, G., KURZAWA-AKANBI, M., MCKEITH, I. G., CHINNERY, P. F., MORRIS, C. M., THEUNS, J., CROSIERS, D., CRAS, P., ENGELBORGHS, S., DE DEYN, P. P., VAN BROECKHOVEN, C., MANN, D. M., SNOWDEN, J., PICKERING-BROWN, S., HALLIWELL, N., DAVIDSON, Y., GIBBONS, L., HARRIS, J., SHEERIN, U. M., BRAS, J., HARDY, J., CLARK, L., MARDER, K., HONIG, L. S., BERG, D., MAETZLER, W., BROCKMANN, K., GASSER, T., NOVELLINO, F., QUATTRONE, A., ANNESI, G., DE MARCO, E. V., ROGAEVA, E., MASELLIS, M., BLACK, S. E., BILBAO, J. M., FOROUD, T., GHETTI, B., NICHOLS, W. C., PANKRATZ, N., HALLIDAY, G., LESAGE, S., KLEBE, S., DURR, A., DUYCKAERTS, C., BRICE, A., GIASSON, B. I., TROJANOWSKI, J. Q., HURTIG, H. I., TAYEBI, N., LANDAZABAL, C., KNIGHT, M. A., KELLER, M., SINGLETON, A. B., WOLFSBERG, T. G. & SIDRANSKY, E. 2013. A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. *JAMA Neurol,* 70**,** 727-35.

NALLS, M. A., ESCOTT-PRICE, V., WILLIAMS, N. M., LUBBE, S., KELLER, M. F., MORRIS, H. R. & SINGLETON, A. B. 2015a. Genetic risk and age in Parkinson's disease: Continuum not stratum. *Mov Disord,* 30**,** 850-4.

NALLS, M. A., KELLER, M. F., HERNANDEZ, D. G., CHEN, L., STONE, D. J. & SINGLETON, A. B. 2016. Baseline genetic associations in the Parkinson's Progression Markers Initiative (PPMI). *Mov Disord,* 31**,** 79-85.

NALLS, M. A., MCLEAN, C. Y., RICK, J., EBERLY, S., HUTTEN, S. J., GWINN, K., SUTHERLAND, M., MARTINEZ, M., HEUTINK, P., WILLIAMS, N. M., HARDY, J., GASSER, T., BRICE, A., PRICE, T. R., NICOLAS, A., KELLER, M. F., MOLONY, C., GIBBS, J. R., CHEN-PLOTKIN, A., SUH, E., LETSON, C., FIANDACA, M. S., MAPSTONE, M., FEDEROFF, H. J., NOYCE, A. J., MORRIS, H., VAN DEERLIN, V. M., WEINTRAUB, D., ZABETIAN, C., HERNANDEZ, D. G., LESAGE, S., MULLINS, M., CONLEY, E. D., NORTHOVER, C. A., FRASIER, M., MAREK, K., DAY-WILLIAMS, A. G., STONE, D. J., IOANNIDIS, J. P. & SINGLETON, A. B. 2015b. Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. *Lancet Neurol,* 14**,** 1002-9.

NALLS, M. A., PANKRATZ, N., LILL, C. M., DO, C. B., HERNANDEZ, D. G., SAAD, M., DESTEFANO, A. L., KARA, E., BRAS, J., SHARMA, M., SCHULTE, C., KELLER, M. F., AREPALLI, S., LETSON, C., EDSALL, C., STEFANSSON, H., LIU, X., PLINER, H., LEE, J. H., CHENG, R., IKRAM, M. A., IOANNIDIS, J. P., HADJIGEORGIOU, G. M., BIS, J. C., MARTINEZ, M., PERLMUTTER, J. S., GOATE, A., MARDER, K., FISKE, B., SUTHERLAND, M., XIROMERISIOU, G., MYERS, R. H., CLARK, L. N., STEFANSSON, K., HARDY, J. A., HEUTINK, P., CHEN, H., WOOD, N. W., HOULDEN, H., PAYAMI, H., BRICE, A., SCOTT, W. K., GASSER, T., BERTRAM, L., ERIKSSON, N., FOROUD, T. & SINGLETON, A. B. 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet,* 46**,** 989-93.

NALLS, M. A., PLAGNOL, V., HERNANDEZ, D. G., SHARMA, M., SHEERIN, U. M., SAAD, M., SIMON-SANCHEZ, J., SCHULTE, C., LESAGE, S., SVEINBJORNSDOTTIR, S., STEFANSSON, K., MARTINEZ, M., HARDY, J., HEUTINK, P., BRICE, A., GASSER, T., SINGLETON, A. B. & WOOD, N. W. 2011. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet,* 377**,** 641-9.

NAMIPASHAKI, A., RAZAGHI-MOGHADAM, Z. & ANSARI-POUR, N. 2015. The Essentiality of Reporting Hardy-Weinberg Equilibrium Calculations in Population-Based Genetic Association Studies. *Cell J,* 17**,** 187-92.

NEUDORFER, O., GILADI, N., ELSTEIN, D., ABRAHAMOV, A., TUREZKITE, T., AGHAI, E., RECHES, A., BEMBI, B. & ZIMRAN, A. 1996. Occurrence of Parkinson's syndrome in type I Gaucher disease. *Qjm,* 89**,** 691-4.

NEUMANN, J., BRAS, J., DEAS, E., O'SULLIVAN, S. S., PARKKINEN, L., LACHMANN, R. H., LI, A., HOLTON, J., GUERREIRO, R., PAUDEL, R., SEGARANE, B., SINGLETON, A., LEES, A., HARDY, J., HOULDEN, H., REVESZ, T. & WOOD, N. W. 2009. Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. *Brain,* 132**,** 1783-94.

NICOLAE, D. L., GAMAZON, E., ZHANG, W., DUAN, S., DOLAN, M. E. & COX, N. J. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet,* 6**,** e1000888.

NORD, A. S. & WEST, A. E. 2020. Neurobiological functions of transcriptional enhancers. *Nat Neurosci,* 23**,** 5-14.

PANKRATZ, N., BEECHAM, G. W., DESTEFANO, A. L., DAWSON, T. M., DOHENY, K. F., FACTOR, S. A., HAMZA, T. H., HUNG, A. Y., HYMAN, B. T., IVINSON, A. J., KRAINC, D., LATOURELLE, J. C., CLARK, L. N., MARDER, K., MARTIN, E. R., MAYEUX, R., ROSS, O. A., SCHERZER, C. R., SIMON, D. K., TANNER, C., VANCE, J. M., WSZOLEK, Z. K., ZABETIAN, C. P., MYERS, R. H., PAYAMI, H., SCOTT, W. K., FOROUD, T. & CONSORTIUM, P. G. 2012. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann Neurol,* 71**,** 370-84.

PARKINSON, J. 1817. *An essay on the shaking palsy,* London, Sherwood, Neely and Jones.

PE'ER, I., YELENSKY, R., ALTSHULER, D. & DALY, M. J. 2008. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol,* 32**,** 381-5.

PEARL, J. R., COLANTUONI, C., BERGEY, D. E., FUNK, C. C., SHANNON, P., BASU, B., CASELLA, A. M., OSHONE, R. T., HOOD, L., PRICE, N. D. & AMENT, S. A. 2019. Genome-Scale Transcriptional Regulatory Network Models of Psychiatric and Neurodegenerative Disorders. *Cell Syst,* 8**,** 122-135.e7.

PETERSCHMITT, M. J., GASSER, T., ISAACSON, S., KULISEVSKY, J., MIR, P., SIMUNI, T., WILLS, A.-M., GUEDES, L. C., SVENNINGSSON, P. & WATERS, C. 2019. Safety, tolerability and pharmacokinetics of oral venglustat in Parkinson disease patients with a GBA mutation. *Molecular Genetics and Metabolism,* 126**,** S117.

PICKRELL, J. K., MARIONI, J. C., PAI, A. A., DEGNER, J. F., ENGELHARDT, B. E., NKADORI, E., VEYRIERAS, J. B., STEPHENS, M., GILAD, Y. & PRITCHARD, J. K. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature,* 464**,** 768-72.

PIHLSTROM, L., RENGMARK, A., BJORNARA, K. A. & TOFT, M. 2014. Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease. *Ann Hum Genet,* 78**,** 243-52.

PIHLSTROM, L. & TOFT, M. 2015. Cumulative genetic risk and age at onset in Parkinson's disease. *Mov Disord,* 30**,** 1712-3.

PIHLSTROM, L., WIETHOFF, S. & HOULDEN, H. 2017. Genetics of neurodegenerative diseases: an overview. *Handb Clin Neurol,* 145**,** 309-323.

PIQUE-REGI, R., DEGNER, J. F., PAI, A. A., GAFFNEY, D. J., GILAD, Y. & PRITCHARD, J. K. 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res,* 21**,** 447-55.

POLYMEROPOULOS, M. H., LAVEDAN, C., LEROY, E., IDE, S. E., DEHEJIA, A., DUTRA, A., PIKE, B., ROOT, H., RUBENSTEIN, J., BOYER, R., STENROOS, E. S., CHANDRASEKHARAPPA, S., ATHANASSIADOU, A., PAPAPETROPOULOS, T., JOHNSON, W. G., LAZZARINI, A. M., DUVOISIN, R. C., DI IORIO, G., GOLBE, L. I. & NUSSBAUM, R. L. 1997. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science,* 276**,** 2045-7.

POSTUMA, R. B., BERG, D., STERN, M., POEWE, W., OLANOW, C. W., OERTEL, W., OBESO, J., MAREK, K., LITVAN, I., LANG, A. E., HALLIDAY, G., GOETZ, C. G., GASSER, T., DUBOIS, B., CHAN, P., BLOEM, B. R., ADLER, C. H. & DEUSCHL, G. 2015. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord,* 30**,** 1591-601.

PRICE, A. L., PATTERSON, N. J., PLENGE, R. M., WEINBLATT, M. E., SHADICK, N. A. & REICH, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet,* 38**,** 904-9.

RAN, C., BRODIN, L., FORSGREN, L., WESTERLUND, M., RAMEZANI, M., GELLHAAR, S., XIANG, F., FARDELL, C., NISSBRANDT, H., SÖDERKVIST, P., PUSCHMANN, A., YGLAND, E., OLSON, L., WILLOWS, T., JOHANSSON, A., SYDOW, O., WIRDEFELDT, K., GALTER, D., SVENNINGSSON, P. & BELIN, A. C. 2016. Strong association between glucocerebrosidase mutations and Parkinson's disease in Sweden. *Neurobiol Aging,* 45**,** 212.e5-212.e11.

REDDY, T. E., GERTZ, J., PAULI, F., KUCERA, K. S., VARLEY, K. E., NEWBERRY, K. M., MARINOV, G. K., MORTAZAVI, A., WILLIAMS, B. A., SONG, L., CRAWFORD, G. E., WOLD, B., WILLARD, H. F. & MYERS, R. M. 2012. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res,* 22, 860-9.

REED, X., BANDRES-CIGA, S., BLAUWENDRAAT, C. & COOKSON, M. R. 2019. The role of monogenic genes in idiopathic Parkinson's disease. *Neurobiol Dis,* 124, 230-239.

REYNOLDS, R. H., HARDY, J., RYTEN, M. & GAGLIANO TALIUN, S. A. 2019. Informing disease modelling with brain-relevant functional genomic annotations. *Brain,* 142, 3694-3712.

RIVERA, C. M. & REN, B. 2013. Mapping human epigenomes. *Cell,* 155, 39-55.

RIZZARDI, L. F., HICKEY, P. F., RODRIGUEZ DIBLASI, V., TRYGGVADOTTIR, R., CALLAHAN, C. M., IDRIZI, A., HANSEN, K. D. & FEINBERG, A. P. 2019. Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. *Nat Neurosci,* 22, 307-316.

RIZZO, G., COPETTI, M., ARCUTI, S., MARTINO, D., FONTANA, A. & LOGROSCINO, G. 2016. Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis. *Neurology,* 86, 566-76.

ROCKENSTEIN, E., CLARKE, J., VIEL, C., PANARELLO, N., TRELEAVEN, C. M., KIM, C., SPENCER, B., ADAME, A., PARK, H., DODGE, J. C., CHENG, S. H., SHIHABUDDIN, L. S., MASLIAH, E. & SARDI, S. P. 2016. Glucocerebrosidase modulates cognitive and motor activities in murine models of Parkinson's disease. *Hum Mol Genet,* 25, 2645-2660.

RUDAKOU, U., YU, E., KROHN, L., RUSKEY, J. A., ASAYESH, F., DAUVILLIERS, Y., SPIEGELMAN, D., GREENBAUM, L., FAHN, S., WATERS, C. H., DUPRÉ, N., ROULEAU, G. A., HASSIN-BAER, S., FON, E. A., ALCALAY, R. N. & GAN-OR, Z. 2021. Targeted sequencing of Parkinson's disease loci genes highlights SYT11, FGF20 and other associations. *Brain,* 144, 462-472.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74, 5463-7.

SANYAL, A., LAJOIE, B. R., JAIN, G. & DEKKER, J. 2012. The long-range interaction landscape of gene promoters. *Nature,* 489, 109-13.

SARDI, S. P. & SIMUNI, T. 2019. New Era in disease modification in Parkinson's disease: Review of genetically targeted therapeutics. *Parkinsonism Relat Disord,* 59, 32-38.

SARDI, S. P., VIEL, C., CLARKE, J., TRELEAVEN, C. M., RICHARDS, A. M., PARK, H., OLSZEWSKI, M. A., DODGE, J. C., MARSHALL, J., MAKINO, E., WANG, B., SIDMAN, R. L., CHENG, S. H. & SHIHABUDDIN, L. S. 2017. Glucosylceramide synthase inhibition alleviates aberrations in synucleinopathy models. *Proc Natl Acad Sci U S A,* 114, 2699-2704.

SATAKE, W., NAKABAYASHI, Y., MIZUTA, I., HIROTA, Y., ITO, C., KUBO, M., KAWAGUCHI, T., TSUNODA, T., WATANABE, M., TAKEDA, A., TOMIYAMA, H., NAKASHIMA, K., HASEGAWA, K., OBATA, F., YOSHIKAWA, T., KAWAKAMI, H., SAKODA, S., YAMAMOTO, M., HATTORI, N., MURATA, M., NAKAMURA, Y. & TODA, T. 2009. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet,* 41, 1303-7.

SATTERLEE, J. S., CHADWICK, L. H., TYSON, F. L., MCALLISTER, K., BEAVER, J., BIRNBAUM, L., VOLKOW, N. D., WILDER, E. L., ANDERSON, J. M. & ROY, A. L. 2019. The NIH Common Fund/Roadmap Epigenomics Program: Successes of a comprehensive consortium. *Sci Adv,* 5, eaaw6507.

SCHAUB, M. A., BOYLE, A. P., KUNDAJE, A., BATZOGLOU, S. & SNYDER, M. 2012. Linking disease associations with regulatory information in the human genome. *Genome Res,* 22**,** 1748-59.

SCHIERDING, W., FARROW, S., FADASON, T., GRAHAM, O. E. E., PITCHER, T. L., QUBISI, S., DAVIDSON, A. J., PERRY, J. K., ANDERSON, T. J., KENNEDY, M. A., COOPER, A. & O'SULLIVAN, J. M. 2020. Common Variants Coregulate Expression of GBA and Modifier Genes to Delay Parkinson's Disease Onset. *Mov Disord,* 35**,** 1346-1356.

SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS CONSORTIUM 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature,* 511**,** 421-7.

SCHMIDT, E. M., ZHANG, J., ZHOU, W., CHEN, J., MOHLKE, K. L., CHEN, Y. E. & WILLER, C. J. 2015. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics,* 31**,** 2601-6.

SCHNEIDER, S. A. & ALCALAY, R. N. 2020. Precision medicine in Parkinson's disease: emerging treatments for genetic Parkinson's disease. *J Neurol,* 267**,** 860-869.

SCHONDORF, D. C., AURELI, M., MCALLISTER, F. E., HINDLEY, C. J., MAYER, F., SCHMID, B., SARDI, S. P., VALSECCHI, M., HOFFMANN, S., SCHWARZ, L. K., HEDRICH, U., BERG, D., SHIHABUDDIN, L. S., HU, J., PRUSZAK, J., GYGI, S. P., SONNINO, S., GASSER, T. & DELEIDI, M. 2014. iPSC-derived neurons from GBA1-associated Parkinson's disease patients show autophagic defects and impaired calcium homeostasis. *Nat Commun,* 5**,** 4028.

SHASHIKANT, T. & ETTENSOHN, C. A. 2019. Genome-wide analysis of chromatin accessibility using ATAC-seq. *Methods Cell Biol,* 151**,** 219-235.

SHERWOOD, R. I., HASHIMOTO, T., O'DONNELL, C. W., LEWIS, S., BARKAL, A. A., VAN HOFF, J. P., KARUN, V., JAAKKOLA, T. & GIFFORD, D. K. 2014. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol,* 32**,** 171-178.

SIDRANSKY, E. 2004. Gaucher disease: complexity in a "simple" disorder. *Mol Genet Metab,* 83**,** 6-15.

SIDRANSKY, E. & LOPEZ, G. 2012. The link between the GBA gene and parkinsonism. *Lancet Neurol,* 11**,** 986-98.

SIDRANSKY, E., NALLS, M. A., AASLY, J. O., AHARON-PERETZ, J., ANNESI, G., BARBOSA, E. R., BAR-SHIRA, A., BERG, D., BRAS, J., BRICE, A., CHEN, C. M., CLARK, L. N., CONDROYER, C., DE MARCO, E. V., DURR, A., EBLAN, M. J., FAHN, S., FARRER, M. J., FUNG, H. C., GAN-OR, Z., GASSER, T., GERSHONI-BARUCH, R., GILADI, N., GRIFFITH, A., GUREVICH, T., JANUARIO, C., KROPP, P., LANG, A. E., LEE-CHEN, G. J., LESAGE, S., MARDER, K., MATA, I. F., MIRELMAN, A., MITSUI, J., MIZUTA, I., NICOLETTI, G., OLIVEIRA, C., OTTMAN, R., ORR-URTREGER, A., PEREIRA, L. V., QUATTRONE, A., ROGAEVA, E., ROLFS, A., ROSENBAUM, H., ROZENBERG, R., SAMII, A., SAMADDAR, T., SCHULTE, C., SHARMA, M., SINGLETON, A., SPITZ, M., TAN, E. K., TAYEBI, N., TODA, T., TROIANO, A. R., TSUJI, S., WITTSTOCK, M., WOLFSBERG, T. G., WU, Y. R., ZABETIAN, C. P., ZHAO, Y. & ZIEGLER, S. G. 2009. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med,* 361**,** 1651-61.

SIMON-SANCHEZ, J., SCHULTE, C., BRAS, J. M., SHARMA, M., GIBBS, J. R., BERG, D., PAISAN-RUIZ, C., LICHTNER, P., SCHOLZ, S. W., HERNANDEZ, D. G., KRUGER, R., FEDEROFF, M., KLEIN, C., GOATE, A., PERLMUTTER, J., BONIN, M., NALLS,

M. A., ILLIG, T., GIEGER, C., HOULDEN, H., STEFFENS, M., OKUN, M. S., RACETTE, B. A., COOKSON, M. R., FOOTE, K. D., FERNANDEZ, H. H., TRAYNOR, B. J., SCHREIBER, S., AREPALLI, S., ZONOZI, R., GWINN, K., VAN DER BRUG, M., LOPEZ, G., CHANOCK, S. J., SCHATZKIN, A., PARK, Y., HOLLENBECK, A., GAO, J., HUANG, X., WOOD, N. W., LORENZ, D., DEUSCHL, G., CHEN, H., RIESS, O., HARDY, J. A., SINGLETON, A. B. & GASSER, T. 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet,* 41**,** 1308-12.

SIMOVSKI, B., KANDURI, C., GUNDERSEN, S., TITOV, D., DOMANSKA, D., BOCK, C., BOSSINI-CASTILLO, L., CHIKINA, M., FAVOROV, A., LAYER, R. M., MIRONOV, A. A., QUINLAN, A. R., SHEFFIELD, N. C., TRYNKA, G. & SANDVE, G. K. 2018. Coloc-stats: a unified web interface to perform colocalization analysis of genomic features. *Nucleic Acids Res,* 46**,** W186-w193.

SINGLETON, A. & HARDY, J. 2016. The Evolution of Genetics: Alzheimer's and Parkinson's Diseases. *Neuron,* 90**,** 1154-1163.

SMEMO, S., TENA, J. J., KIM, K. H., GAMAZON, E. R., SAKABE, N. J., GÓMEZ-MARÍN, C., ANEAS, I., CREDIDIO, F. L., SOBREIRA, D. R., WASSERMAN, N. F., LEE, J. H., PUVIINDRAN, V., TAM, D., SHEN, M., SON, J. E., VAKILI, N. A., SUNG, H. K., NARANJO, S., ACEMEL, R. D., MANZANARES, M., NAGY, A., COX, N. J., HUI, C. C., GOMEZ-SKARMETA, J. L. & NÓBREGA, M. A. 2014. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature,* 507**,** 371-5.

SNETKOVA, V. & SKOK, J. A. 2018. Enhancer talk. *Epigenomics,* 10**,** 483-498.

SPAIN, S. L. & BARRETT, J. C. 2015. Strategies for fine-mapping complex traits. *Hum Mol Genet,* 24**,** R111-9.

SPENCER, C. C., PLAGNOL, V., STRANGE, A., GARDNER, M., PAISAN-RUIZ, C., BAND, G., BARKER, R. A., BELLENGUEZ, C., BHATIA, K., BLACKBURN, H., BLACKWELL, J. M., BRAMON, E., BROWN, M. A., BROWN, M. A., BURN, D., CASAS, J. P., CHINNERY, P. F., CLARKE, C. E., CORVIN, A., CRADDOCK, N., DELOUKAS, P., EDKINS, S., EVANS, J., FREEMAN, C., GRAY, E., HARDY, J., HUDSON, G., HUNT, S., JANKOWSKI, J., LANGFORD, C., LEES, A. J., MARKUS, H. S., MATHEW, C. G., MCCARTHY, M. I., MORRISON, K. E., PALMER, C. N., PEARSON, J. P., PELTONEN, L., PIRINEN, M., PLOMIN, R., POTTER, S., RAUTANEN, A., SAWCER, S. J., SU, Z., TREMBATH, R. C., VISWANATHAN, A. C., WILLIAMS, N. W., MORRIS, H. R., DONNELLY, P. & WOOD, N. W. 2011. Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Hum Mol Genet,* 20**,** 345-53.

SPILLANTINI, M. G., SCHMIDT, M. L., LEE, V. M., TROJANOWSKI, J. Q., JAKES, R. & GOEDERT, M. 1997. Alpha-synuclein in Lewy bodies. *Nature,* 388**,** 839-40.

STEINER, J. A., QUANSAH, E. & BRUNDIN, P. 2018. The concept of alpha-synuclein as a prion-like protein: ten years after. *Cell Tissue Res,* 373**,** 161-173.

STIRNEMANN, J., BELMATOUG, N., CAMOU, F., SERRATRICE, C., FROISSART, R., CAILLAUD, C., LEVADE, T., ASTUDILLO, L., SERRATRICE, J., BRASSIER, A., ROSE, C., BILLETTE DE VILLEMEUR, T. & BERGER, M. G. 2017. A Review of Gaucher Disease Pathophysiology, Clinical Presentation and Treatments. *Int J Mol Sci,* 18**,** 441.

STRAHL, B. D. & ALLIS, C. D. 2000. The language of covalent histone modifications. *Nature,* 403**,** 41-5.

SULZER, D. 2007. Multiple hit hypotheses for dopamine neuron loss in Parkinson's disease. *Trends Neurosci,* 30**,** 244-50.

SAAD, M., LESAGE, S., SAINT-PIERRE, A., CORVOL, J. C., ZELENIKA, D., LAMBERT, J. C., VIDAILHET, M., MELLICK, G. D., LOHMANN, E., DURIF, F., POLLAK, P., DAMIER, P., TISON, F., SILBURN, P. A., TZOURIO, C., FORLANI, S., LORIOT, M. A., GIROUD, M., HELMER, C., PORTET, F., AMOUYEL, P., LATHROP, M., ELBAZ, A., DURR, A., MARTINEZ, M. & BRICE, A. 2011. Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Hum Mol Genet,* 20**,** 615-27.

TAYEBI, N., WALKER, J., STUBBLEFIELD, B., ORVISKY, E., LAMARCA, M. E., WONG, K., ROSENBAUM, H., SCHIFFMANN, R., BEMBI, B. & SIDRANSKY, E. 2003. Gaucher disease with parkinsonian manifestations: does glucocerebrosidase deficiency contribute to a vulnerability to parkinsonism? *Mol Genet Metab,* 79**,** 104-9.

TESSARZ, P. & KOUZARIDES, T. 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol,* 15**,** 703-8.

THALER, A., BREGMAN, N., GUREVICH, T., SHINER, T., DROR, Y., ZMIRA, O., GAN-OR, Z., BAR-SHIRA, A., GANA-WEISZ, M., ORR-URTREGER, A., GILADI, N. & MIRELMAN, A. 2018. Parkinson's disease phenotype is influenced by the severity of the mutations in the GBA gene. *Parkinsonism Relat Disord,* 55**,** 45-49.

TIAN, C., GREGERSEN, P. K. & SELDIN, M. F. 2008. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet,* 17**,** R143-50.

TIEU, K. 2011. A guide to neurotoxic animal models of Parkinson's disease. *Cold Spring Harb Perspect Med,* 1**,** a009316.

TOLEDO, E. M., YANG, S., GYLLBORG, D., VAN WIJK, K. E., SINHA, I., VARAS-GODOY, M., GRIGSBY, C. L., LONNERBERG, P., ISLAM, S., STEFFENSEN, K. R., LINNARSSON, S. & ARENAS, E. 2020. Srebf1 Controls Midbrain Dopaminergic Neurogenesis. *Cell Rep,* 31**,** 107601.

TRETIAKOFF, C. 1919. *Contribution à l'étude de l'anatomie pathologique du locus niger de Soemmering avec quelques déductions relatives à la pathogénie des troubles du tonus musculaire et de la maladie de Parkinson.* University of Paris

TRINH, J., GUSTAVSSON, E. K., VILARINO-GUELL, C., BORTNICK, S., LATOURELLE, J., MCKENZIE, M. B., TU, C. S., NOSOVA, E., KHINDA, J., MILNERWOOD, A., LESAGE, S., BRICE, A., TAZIR, M., AASLY, J. O., PARKKINEN, L., HAYTURAL, H., FOROUD, T., MYERS, R. H., SASSI, S. B., HENTATI, E., NABLI, F., FARHAT, E., AMOURI, R., HENTATI, F. & FARRER, M. J. 2016. DNM3 and genetic modifiers of age of onset in LRRK2 Gly2019Ser parkinsonism: a genome-wide linkage and association study. *Lancet Neurol,* 15**,** 1248-1256.

TRYKA, K. A., HAO, L., STURCKE, A., JIN, Y., WANG, Z. Y., ZIYABARI, L., LEE, M., POPOVA, N., SHAROPOVA, N., KIMURA, M. & FEOLO, M. 2014. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res,* 42**,** D975-9.

TRYNKA, G., SANDOR, C., HAN, B., XU, H., STRANGER, B. E., LIU, X. S. & RAYCHAUDHURI, S. 2013. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet,* 45**,** 124-30.

TRYNKA, G., WESTRA, H. J., SLOWIKOWSKI, K., HU, X., XU, H., STRANGER, B. E., KLEIN, R. J., HAN, B. & RAYCHAUDHURI, S. 2015. Disentangling the Effects of

Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. *Am J Hum Genet,* 97**,** 139-52.

VALENTE, E. M., ABOU-SLEIMAN, P. M., CAPUTO, V., MUQIT, M. M., HARVEY, K., GISPERT, S., ALI, Z., DEL TURCO, D., BENTIVOGLIO, A. R., HEALY, D. G., ALBANESE, A., NUSSBAUM, R., GONZÁLEZ-MALDONADO, R., DELLER, T., SALVI, S., CORTELLI, P., GILKS, W. P., LATCHMAN, D. S., HARVEY, R. J., DALLAPICCOLA, B., AUBURGER, G. & WOOD, N. W. 2004. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science,* 304**,** 1158-60.

VAN DE GEIJN, B., FINUCANE, H., GAZAL, S., HORMOZDIARI, F., AMARIUTA, T., LIU, X., GUSEV, A., LOH, P. R., RESHEF, Y., KICHAEV, G., RAYCHAUDURI, S. & PRICE, A. L. 2020. Annotations capturing cell type-specific TF binding explain a large fraction of disease heritability. *Hum Mol Genet,* 29**,** 1057-1067.

VERMUNT, M. W., ZHANG, D. & BLOBEL, G. A. 2019. The interdependence of gene-regulatory elements and the 3D genome. *J Cell Biol,* 218**,** 12-26.

VON LINSTOW, C. U., DELANO-TAYLOR, M., KORDOWER, J. H. & BRUNDIN, P. 2020. Does Developmental Variability in the Number of Midbrain Dopamine Neurons Affect Individual Risk for Sporadic Parkinson's Disease? *J Parkinsons Dis,* 10**,** 405-411.

WANG, C., WANG, Y., HU, M., CHAI, Z., WU, Q., HUANG, R., HAN, W., ZHANG, C. X. & ZHOU, Z. 2016. Synaptotagmin-11 inhibits clathrin-mediated and bulk endocytosis. *EMBO Rep,* 17**,** 47-63.

WANG, D., LIU, S., WARRELL, J., WON, H., SHI, X., NAVARRO, F. C. P., CLARKE, D., GU, M., EMANI, P., YANG, Y. T., XU, M., GANDAL, M. J., LOU, S., ZHANG, J., PARK, J. J., YAN, C., RHIE, S. K., MANAKONGTREECHEEP, K., ZHOU, H., NATHAN, A., PETERS, M., MATTEI, E., FITZGERALD, D., BRUNETTI, T., MOORE, J., JIANG, Y., GIRDHAR, K., HOFFMAN, G. E., KALAYCI, S., GUMUS, Z. H., CRAWFORD, G. E., ROUSSOS, P., AKBARIAN, S., JAFFE, A. E., WHITE, K. P., WENG, Z., SESTAN, N., GESCHWIND, D. H., KNOWLES, J. A. & GERSTEIN, M. B. 2018. Comprehensive functional genomic resource and integrative model for the human brain. *Science,* 362.

WANG, J., ZHUANG, J., IYER, S., LIN, X., WHITFIELD, T. W., GREVEN, M. C., PIERCE, B. G., DONG, X., KUNDAJE, A., CHENG, Y., RANDO, O. J., BIRNEY, E., MYERS, R. M., NOBLE, W. S., SNYDER, M. & WENG, Z. 2012. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res,* 22**,** 1798-812.

WARD, C. D., DUVOISIN, R. C., INCE, S. E., NUTT, J. D., ELDRIDGE, R. & CALNE, D. B. 1983. Parkinson's disease in 65 pairs of twins and in a set of quadruplets. *Neurology,* 33**,** 815-24.

WATERLAND, R. A. 2006. Epigenetic mechanisms and gastrointestinal development. *J Pediatr,* 149**,** S137-42.

WEIRAUCH, M. T., YANG, A., ALBU, M., COTE, A. G., MONTENEGRO-MONTERO, A., DREWE, P., NAJAFABADI, H. S., LAMBERT, S. A., MANN, I., COOK, K., ZHENG, H., GOITY, A., VAN BAKEL, H., LOZANO, J. C., GALLI, M., LEWSEY, M. G., HUANG, E., MUKHERJEE, T., CHEN, X., REECE-HOYES, J. S., GOVINDARAJAN, S., SHAULSKY, G., WALHOUT, A. J. M., BOUGET, F. Y., RATSCH, G., LARRONDO, L. F., ECKER, J. R. & HUGHES, T. R. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell,* 158**,** 1431-1443.

WINDER-RHODES, S. E., EVANS, J. R., BAN, M., MASON, S. L., WILLIAMS-GRAY, C. H., FOLTYNIE, T., DURAN, R., MENCACCI, N. E., SAWCER, S. J. & BARKER, R. A. 2013. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. *Brain,* 136**,** 392-9.

YAN, F., POWELL, D. R., CURTIS, D. J. & WONG, N. C. 2020. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol,* 21**,** 22.

YANG, Z. H., LI, Y. S., SHI, M. M., YANG, J., LIU, Y. T., MAO, C. Y., FAN, Y., HU, X. C., SHI, C. H. & XU, Y. M. 2019. SNCA but not DNM3 and GAK modifies age at onset of LRRK2-related Parkinson's disease in Chinese population. *J Neurol,* 266**,** 1796-1800.

ZAMPIERI, S., CATTAROSSI, S., BEMBI, B. & DARDIS, A. 2017. GBA Analysis in Next-Generation Era: Pitfalls, Challenges, and Possible Solutions. *J Mol Diagn,* 19**,** 733-741.

ZHANG, Y., SHU, L., SUN, Q., ZHOU, X., PAN, H., GUO, J. & TANG, B. 2018. Integrated Genetic Analysis of Racial Differences of Common GBA Variants in Parkinson's Disease: A Meta-Analysis. *Front Mol Neurosci,* 11**,** 43.

ZHAO, H., MITRA, N., KANETSKY, P. A., NATHANSON, K. L. & REBBECK, T. R. 2018. A practical approach to adjusting for population stratification in genome-wide association studies: principal components and propensity scores (PCAPS). *Stat Appl Genet Mol Biol,* 17.

Paper I

Neuroscience Letters 658 (2017) 48–52

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Research paper

The *GBA* variant E326K is associated with Parkinson's disease and explains a genome-wide association signal

Victoria Berge-Seidl[a,b], Lasse Pihlstrøm[b], Jodi Maple-Grødem[c,d], Lars Forsgren[e], Jan Linder[e], Jan Petter Larsen[f], Ole-Bjørn Tysnes[g], Mathias Toft[a,*]

[a] Department of Neurology, Oslo University Hospital, Oslo, Norway
[b] Faculty of Medicine, University of Oslo, Oslo, Norway
[c] The Norwegian Centre for Movement Disorders, Stavanger University Hospital, Stavanger, Norway
[d] The Centre for Organelle Research, University of Stavanger, Stavanger, Norway
[e] Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå, Sweden
[f] Network for Medical Sciences, University of Stavanger, Stavanger, Norway
[g] Department of Neurology, Haukeland University Hospital, Bergen, Norway

ARTICLE INFO

Keywords:
Parkinson's disease
Glucocerebrosidase
E326K
T369M
Synaptotagmin 11
GWAS

ABSTRACT

Objective: Coding variants in the *GBA* gene have been identified as the numerically most important genetic risk factors for Parkinson's disease (PD). In addition, genome-wide association studies (GWAS) have identified associations with PD in the *SYT11-GBA* region on chromosome 1q22, but the relationship to *GBA* coding variants have remained unclear. The aim of this study was to sequence the complete *GBA* gene in a clinical cohort and to investigate whether coding variants within the *GBA* gene may be driving reported association signals.
Methods: We analyzed high-throughput sequencing data of all coding exons of *GBA* in 366 patients with PD. The identified low-frequency coding variants were genotyped in three Scandinavian case-controls series (786 patients and 713 controls). Previously reported risk variants from two independent association signals within the *SYT11-GBA* locus on chromosome 1 were also genotyped in the same samples. We performed association analyses and evaluated linkage disequilibrium (LD) between the variants.
Results: We identified six rare mutations (1.6%) and two low-frequency coding variants in *GBA*. E326K (rs2230288) was significantly more frequent in PD patients compared to controls (OR 1.65, $p = 0.03$). There was no clear association of T369M (rs75548401) with disease (OR 1.43, $p = 0.24$). Genotyping the two GWAS hits rs35749011 and rs114138760 in the same sample set, we replicated the association between rs35749011 and disease status (OR 1.67, $p = 0.03$), while rs114138760 was found to have similar allele frequencies in patients and controls. Analyses revealed that E326K and rs35749011 are in very high LD (r^2 0.95).
Conclusions: Our results confirm that the *GBA* variant E326K is a susceptibility allele for PD. The results suggest that E326K may fully account for the primary association signal observed at chromosome 1q22 in previous GWAS of PD.

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder and is generally estimated to affect 1% of people over 60 years of age, with increasing prevalence in higher age groups. PD is mainly a sporadic disease, but family and candidate gene studies have identified a number of genes related to PD pathogenesis [1]. There is particular interest in the *GBA* gene and its relationship to risk for PD. Homozygous *GBA* mutations cause the autosomal recessive lysosomal storage disorder Gaucher disease. However, heterozygous *GBA* mutations have been identified as the numerically most important genetic risk factors for PD, and 5–10% of PD patients have been reported to carry *GBA* mutations [2]. The two *GBA* coding variants E326K and T369M do not cause Gaucher disease in the homozygous state and were initially considered to be benign polymorphisms. There is now increasing evidence in support of the variant E326K as a risk factor for PD, while the association between T369M and PD has been less clear.

Genome wide association studies (GWAS) have linked a number of

[*] Corresponding author at: Department of Neurology, Oslo University Hospital, P.O. Box 4950 Nydalen, N-0424 Oslo, Norway.
E-mail addresses: berge.victoria@gmail.com (V. Berge-Seidl), lasse.pihlstrom@gmail.com (L. Pihlstrøm), jodi.maple.grodem@sus.no (J. Maple-Grødem), lars.forsgren@umu.se (L. Forsgren), jan.linder@umu.se (J. Linder), nevrologlarsen@gmail.com (J.P. Larsen), ole-bjorn.tysnes@helse-bergen.no (O.-B. Tysnes), mathias.toft@gmail.com (M. Toft).

http://dx.doi.org/10.1016/j.neulet.2017.08.040
Received 7 June 2017; Received in revised form 31 July 2017; Accepted 15 August 2017
Available online 19 August 2017
0304-3940/ © 2017 Elsevier B.V. All rights reserved.

risk loci to PD susceptibility [3]. Association signals emerging from GWAS typically involve dozens of gene variants in high linkage disequilibrium (LD) encompassing several genes. This complicates the identification of the functionally relevant variants within risk loci.

In PD, an early GWAS reported an intronic disease-associated polymorphism within the *SYT11* gene on chromosome 1q22 [4]. Later, a *meta*-analysis of several GWAS found an association between a coding variant in the *GBA* gene, E326K, and PD [5]. *GBA* is located about 650 kb from *SYT11*, within the same block of LD referred to as the *GBA-SYT11* locus. The largest and most recent *meta*-analysis of PD GWAS reported two independent associations within the *GBA-SYT11* locus, but the relationship between the reported signals and *GBA* coding variants was not examined in detail [3].

The aim of this study was to investigate the frequency of *GBA* mutations in our population and if the two *GBA* variants E326K and T369M are associated with PD in a Scandinavian case-control series. We also wanted to assess to what degree coding *GBA* variants are linked to the *GBA-SYT11* association signals reported for PD.

2. Methods

2.1. Patients and controls

We included samples from three Scandinavian biobanks in our study. From Oslo University Hospital 486 patients (mean age at onset 56 years; SD 11 years) and 473 controls (mean age at inclusion 62 years; SD 11 years) were included. 173 patients (mean age at onset 66 years; SD 9 years) and 187 controls (mean age at inclusion 66 years; SD 9 years) originated from the ParkWest study. 127 patients (mean age at onset 68 years; SD 10 years) and 53 controls (mean age at inclusion 65 years; SD 7 years) were from the NYPUM study at Umeå University Hospital. All PD patients were examined by a neurologist and diagnosed according to the revised UKPDSBB criteria (Oslo and Umeå) or Gelb criteria (ParkWest). The majority of patients were screened for the *LRRK2* G2019S mutation, in addition a large subset of patients was also sequenced for genes causing Mendelian forms of PD (*SNCA*, *PRKN*, *PINK1*, *DJ-1*, *LRRK2*, and *VPS35*). Patients with pathogenic mutations in these genes were excluded from the study. Control subjects consist of spouses of patients, outpatients in primary care and healthy volunteers, all without neurological disease and known parkinsonism among first degree relatives. The study was approved by the Regional Committee for Medical Research Ethics (Oslo, Norway). Sample and data collection at each study site was approved by local ethics committees. All participants gave written, informed consent.

2.2. Identification of GBA coding variants

To identify all coding variants in the *GBA* gene we analyzed sequencing data from 366 patients from the Oslo patient series. All coding exons of the *GBA* gene were part of a gene panel examined by targeted deep sequencing of DNA pools as described previously [6]. Putative variants were identified by bioinformatic analyses and individually validated by Sanger sequencing. Pools with a read depth below 80 x at the relevant position were excluded from analysis of that specific variant. The *GBA* gene was amplified in distinct fragments. To avoid amplification of the pseudogene, we used primer sequences designed to DNA regions exclusive to the *GBA* gene. PCR products were sequenced with a selection of previously described sequencing primers (all primer sequences are available upon request). The conventional nomenclature for *GBA* alleles was used, excluding the 39-residue signal peptide. *In silico* prediction of deleteriousness of the identified variants was performed by the use of Combined Annotation Dependent Depletion (C-ADD) v1.3, a method integrating and combining multiple genome annotations [7].

2.3. Genotyping and statistical analyses

Two identified *GBA* variants, E326K (rs2230288) and T369M (rs75548401), were genotyped in all 786 cases and 713 controls. We also genotyped the primary risk SNP (rs35749011) and a second independent risk SNP (rs114138760) located within the *GBA-SYT11* locus identified by a recent *meta*-analysis of PD GWAS [3]. Genotyping was performed by KASP and TaqMan SNP genotyping assays on a Viia7 instrument (Life Technologies, Foster City, CA, USA). The genotype call rate was above 98% for each individual variant. Statistical analyses were performed in PLINK (https://www.cog-genomics.org/plink/1.9/). We tested for Hardy-Weinberg equilibrium in controls, observing no significant departure. We assessed the association between each single variant and disease status with Chi-square test and calculated odds ratio (OR). LD between *GBA* coding variants and GWAS risk SNPs were analyzed by using Haploview 4.2 software (https://www.broadinstitute.org/haploview/haploview).

3. Results

We identified two low-frequency coding variants in *GBA* (E326K and T369M) in the sequenced samples. Five additional coding variants and one potential splicing variant were identified by sequencing, each variant only occurring once. Only three of these variants have been described in Gaucher disease patients (N370S, R463C, IVS3 + 1G > A). The remaining three variants are to our knowledge novel and thus of unknown significance (V457A, G377D, W357R). The novel variants all have a CADD score above the suggested cutoff on deleteriousness. Information on the *GBA* variants identified by sequencing is summarized in Table 1.

Subsequent genotyping of the two low-frequency variants in all samples revealed that E326K (rs2230288) was significantly more frequent in PD patients compared to controls (OR 1.65, p = 0.03). There

Table 1
GBA variants identified by sequencing.

dbSNP ID	Position	Allele name	AA change	Function	n	MAF PD (n total alleles = 732)	CADD/PHRED
rs80356771	1:155204987	R463C	p.Arg502Cys	missense	1	0,14%	29,6
–	1:155205004	V457A (a)	p.Val496Ala	missense	1	0,14%	22,8
–	1:155205613	G377D (a)	p.Gly416Asp	missense	1	0,14%	29,0
rs76763715	1:155205634	N370S	p.Asn409Ser	missense	1	0,15% (b)	22,7
rs75548401	1:155206037	T369M	p.Thr408Met	missense	13	1,78%	22,2
–	1:155206074	W357R (a)	pTrp396Arg	missense	1	0,14%	26,5
rs2230288	1:155206167	E326K	p.Glu365Lys	missense	20	3,03% (b)	17,3
–	1:155209676	IVS3 + 1G > A	–	splicing	1	0,14%	23,3

Position refers to chromosomal position of the variant in build 37 of the Genome Reference Consortium human genome. Reported allele names follow the common nomenclature and refer to the processed protein, excluding the 39-residue signal peptide. Phred-like scaled C-scores are calculated using CADD. A Phred-value between 10 and 20 has been suggested as a possible cutoff for deleteriousness (http://cadd.gs.washington.edu/info). (a) Previously undescribed mutation. (b) Due to exclusion of pools with a read depth < 80 x the total number of alleles examined was 678 for rs76763715 and 660 for rs2230288. MAF: minor allele frequency. OR: odds ratio. AA: Amino acid. PD: Parkinsońs disease.

Table 2
Frequencies of *SYT11-GBA* locus genotypes in PD patients and controls.

dbSNP ID	Position	Alleles	MAF PD (%)	MAF Controls (%)	MAF Databases (%)	OR (CI)	p-value
rs114138760	1:154898185	C/G	0.6	0.6	1.2	0.91 (0.36 – 2.29)	0.83
rs35749011	1:155135036	A/G	3.4	2.1	2.2	1.67 (1.06 – 2.64)	0.03
rs75548401	1:155206037	A/G	1.8	1.3	1.0	1.43 (0.79 – 2.60)	0.24
rs2230288	1:155206167	T/C	3.5	2.1	1.2	1.65 (1.05 – 2.60)	0.03

Position refers to chromosomal position of the variant in build 37 of the Genome Reference Consortium human genome. Database frequencies are taken from the ExAC European (Non Finnish) population for coding variants and the 1000genomes European population for non-coding variants. MAF: minor allele frequency. OR: odds ratio. CI: confidence interval. PD: Parkinsońs disease

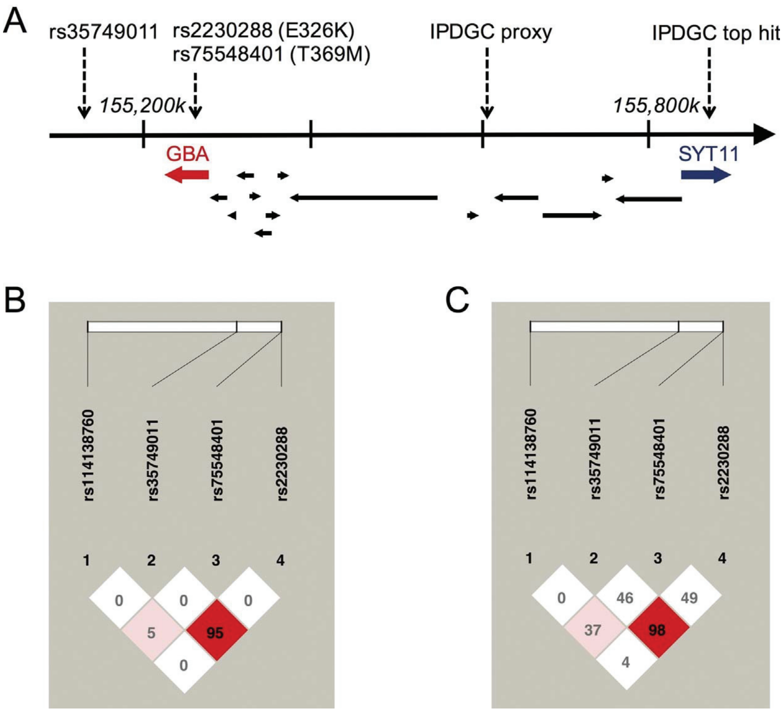


Fig. 1. *SYT11-GBA* locus and locations of genotyped variants with their pairwise linkage disequilibrium patterns.
A) Location of the *SYT11* and *GBA* genes, as well as top hit SNPs from previous GWAS and low-frequency *GBA* variants on chromosome 1. IPDGC refers to publication 4 in the reference list. rs114138760 is located further upstream of the *GBA* gene and is not shown on the figure. B) Pairwise r^2 between the four genotyped variants, and C) Pairwise D' between the four genotyped variants.

was no clear association of T369M (rs75548401) with disease (OR 1.43, p = 0.24). When genotyping the two *meta*-GWAS hits rs35749011 and rs114138760 in the same sample set we observed a significant association of rs35749011 in PD patients (OR 1.67, p = 0.03), while rs114138760 was found to have similar allele frequencies in patients and controls (OR 0.91, p = 0.83) (Table 2).

The location of the *SYT11* and *GBA* genes, as well as top hit SNPs from previous GWAS are shown in Fig. 1 a. Analyses of the pairwise LD between the four genotyped variants revealed that E326K and rs35749011 are in very high LD with a r^2 of 0.95 (D́ = 0.98) (Fig. 1 b and c). Therefore, it is likely that E326K in *GBA* explains the association observed at rs35749011 in previous studies. The LD between T369M and rs114138760 was low, indicating that the secondary association signal reported by Nalls et al. is independent of this coding variant.

4. Discussion

Our results confirm that the *GBA* variant E326K is a susceptibility allele for PD. The frequency of E326K and T369M seem to be higher in our Scandinavian case-control series compared to other European populations. Our study was nevertheless underpowered to identify the previously reported association between PD and T369M. However, we note that the odds ratio was similar to that reported by a recent *meta*-analysis of T369M [8].

GBA mutations may cause a deficiency of the enzyme glucocerebrosidase (GCase) leading to an accumulation of glucocerebroside within lysosomes. Although E326K and T369M do not cause Gaucher disease in the homozygous state, they have been shown to modify GCase activity. Studies expressing *GBA* constructs with E326K suggest that this polymorphism reduces enzyme activity [9]. An association between T369M and reduced enzyme activity has also been reported in carriers of this variant [10]. Such a modification of GCase activity may contribute to PD risk in concert with other risk variants/small biochemical alterations.

We found a low frequency of *GBA* mutations in our study, as only 6 of 366 (1.6%) carry known or novel rare mutations. The patients sequenced in our study are included from a tertiary care hospital, and a large proportion of these patients have been treated with deep brain stimulation (DBS). Cognitive impairment is an exclusion criterion when evaluating PD patients for DBS. We may have selected against carriers of *GBA* mutations since this group of PD patients have been reported to have an accelerated cognitive decline [11,12]. The mutation frequency in a previous Norwegian study is low, indicating that *GBA* mutations may be rare in hospital-based studies from this population [13].

In this study *GBA* mutations were identified by analyses of data from a pooled sequencing experiment. We have previously reported a high sensitivity of this approach [6]. Furthermore, a high number of exons were Sanger sequenced to validate both rare mutations and low frequency variants, without identifying any additional mutations. We thus find it unlikely that the low frequency of *GBA* mutations should be caused by low sensitivity of our sequencing method.

Mutations in *GBA* play an important role in PD, as *GBA* mutation carriers have an increased disease risk, earlier age at onset, and faster progression. In addition to cognitive decline, various other nonmotor symptoms including REM sleep behavior disorder, hyposmia, and autonomic dysfunction seem to be more frequent [14]. Interestingly, it has recently been demonstrated that also the E326K variant predicts a more rapid progression of cognitive dysfunction and motor symptoms in patients with PD [15]. Thus, *GBA* variants influence the heterogeneity in symptom progression observed in PD. This observation may have important clinical implications, especially if *GBA*-specific treatment will become available.

Our results suggest that the low-frequency *GBA* variant E326K may fully account for the primary association signal observed at the chromosome 1 *SYT11-GBA* locus in previous GWAS of PD. This is in accordance with a previous report by Pankratz et al. where E326K reaches genome-wide significance [5]. Recent GWAS have not clearly reported the relationship between identified association signals and *GBA* variants, which could inform functional studies. *SYT11* has therefore been considered a potential PD-related gene, since a GWAS reported an intronic disease-associated polymorphism within this gene [4]. Further genetic evidence linking *SYT11* to PD has however been scarce. The largest and most recent *meta*-analysis of PD GWAS to date located the association signal in an intergenic region hundreds of kilobases away from *SYT11*, but still kept the gene in the naming of the locus. In an attempt to functionally characterize this locus, several studies of synaptotagmin 11 (SYT11) and its role in PD pathogenesis have recently been performed [16]. We report very high LD between E326K and the primary association signal, emphasizing *GBA* as the causal gene at the chromosome 1 *SYT11-GBA* locus. On the other hand, we found no evidence that the secondary signal at this locus was related to the coding *GBA* variant T369M. In the *meta*-analysis by Pankratz et al. the *GBA* mutation N370S is detected as a second independent signal at the *SYT11-GBA* locus [5]. We are not able to study this due to the very low frequency of N370S in our population.

Identifying the functionally relevant variants within disease risk loci identified by GWAS is important to understand the disease mechanisms involved in disease pathogenesis of PD. Most genetic risk variants fall outside coding regions and do not alter the amino acid sequence of proteins. Until recently, the functional characterization of risk-associated loci has been hindered by the limited annotation of the human genome outside coding sequences. However, approaches to successfully characterize the functional nature of these loci are emerging. Future studies will hopefully lead to the identification of specific genes and pathways that could serve as actionable therapeutic targets.

Conflict of interest

The authors report no conflicts of interest concerning the research related to the manuscript.

Author contributions

The study was designed by VBS, with support and advice by LP and MT. LP, JPL, OBT, LF, JL and MT designed clinical studies and collected data. VBS and JMG carried out the genetic analyses. VBS performed statistical analyses and analyzed the data. VBS and MT drafted the manuscript. All the co-authors critically revised the manuscript for intellectual content and approved the final version for publication.

Funding

The study was funded by a grant from the South-Eastern Norway Regional Health Authority and the Norwegian Parkinson Research Fund. Mathias Toft is also supported by the Research Council of Norway. Lasse Pihlstrøm is supported by the Norwegian Health Association. The ParkWest study is currently supported by the Western Norway Regional Health Authority (grant# 912049) and the Norwegian Parkinson Research Foundation. The funding organization had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors thank the participating patients and control subjects. The authors acknowledge the contributions to data collection (Camilla Sætehaug), technical assistance (Sandra Pilar Henriksen) and supervision of personnel (Aina Haugen Rengmark) at Oslo University Hospital. All other personnel involved in the NYPUM and ParkWest studies are thanked for their contributions.

References

[1] M.K. Lin, M.J. Farrer, Genetics and genomics of Parkinson's disease, Genome Med. 6 (2014) 48.

[2] E. Sidransky, M.A. Nalls, J.O. Aasly, J. Aharon-Peretz, G. Annesi, E.R. Barbosa, A. Bar-Shira, D. Berg, J. Bras, A. Brice, C.M. Chen, L.N. Clark, C. Condroyer, E.V. De Marco, A. Durr, M.J. Eblan, S. Fahn, M.J. Farrer, H.C. Fung, Z. Gan-Or, T. Gasser, R. Gershoni-Baruch, N. Giladi, A. Griffith, T. Gurevich, C. Januario, P. Kropp, A.E. Lang, G.J. Lee-Chen, S. Lesage, K. Marder, I.F. Mata, A. Mirelman, J. Mitsui, I. Mizuta, G. Nicoletti, C. Oliveira, R. Ottman, A. Orr-Urtreger, L.V. Pereira, A. Quattrone, E. Rogaeva, A. Rolfs, H. Rosenbaum, R. Rozenberg, A. Samii, T. Samaddar, C. Schulte, M. Sharma, A. Singleton, M. Spitz, E.K. Tan, N. Tayebi, T. Toda, A.R. Troiano, S. Tsuji, M. Wittstock, T.G. Wolfsberg, Y.R. Wu, C.P. Zabetian, Y. Zhao, S.G. Ziegler, Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease, N. Engl. J. Med. 361 (2009) 1651–1661.

[3] M.A. Nalls, N. Pankratz, C.M. Lill, C.B. Do, D.G. Hernandez, M. Saad, A.L. DeStefano, E. Kara, J. Bras, M. Sharma, C. Schulte, M.F. Keller, S. Arepalli, C. Letson, C. Edsall, H. Stefansson, X. Liu, H. Pliner, J.H. Lee, R. Cheng, M.A. Ikram, J.P. Ioannidis, G.M. Hadjigeorgiou, J.C. Bis, M. Martinez, J.S. Perlmutter, A. Goate, K. Marder, B. Fiske, M. Sutherland, G. Xiromerisiou, R.H. Myers, L.N. Clark, K. Stefansson, J.A. Hardy, P. Heutink, H. Chen, N.W. Wood, H. Houlden, H. Payami, A. Brice, W.K. Scott, T. Gasser, L. Bertram, N. Eriksson, T. Foroud, A.B. Singleton, Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease, Nat. Genet. 46 (2014) 989–993.

[4] M.A. Nalls, V. Plagnol, D.G. Hernandez, M. Sharma, U.M. Sheerin, M. Saad, J. Simon-Sanchez, C. Schulte, S. Lesage, S. Sveinbjornsdottir, K. Stefansson, M. Martinez, J. Hardy, P. Heutink, A. Brice, T. Gasser, A.B. Singleton, N.W. Wood, Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies, Lancet 377 (2011) 641–649.

[5] N. Pankratz, G.W. Beecham, A.L. DeStefano, T.M. Dawson, K.F. Doheny, S.A. Factor, T.H. Hamza, A.Y. Hung, B.T. Hyman, A.J. Ivinson, D. Krainc, J.C. Latourelle, L.N. Clark, K. Marder, E.R. Martin, R. Mayeux, O.A. Ross, C.R. Scherzer, D.K. Simon, C. Tanner, J.M. Vance, Z.K. Wszolek, C.P. Zabetian, R.H. Myers, H. Payami, W.K. Scott, T. Foroud, P.G. Consortium, Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2, Ann. Neurol. 71 (2012) 370–384.

[6] L. Pihlstrom, A. Rengmark, K.A. Bjornara, M. Toft, Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease, Ann. Hum. Genet. 78 (2014) 243–252.

[7] M. Kircher, D.M. Witten, P. Jain, B.J. O'Roak, G.M. Cooper, A general framework for estimating the relative pathogenicity of human genetic variants, Nat. Genet. 46 (2014) 310–315.

[8] V. Mallett, J.P. Ross, R.N. Alcalay, A. Ambalavanan, E. Sidransky, P.A. Dion, G.A. Rouleau, Z. Gan-Or, GBA p.T369M substitution in Parkinson disease: polymorphism or association? A meta-analysis, Neurology. Genetics 2 (2016) e104.

[9] E. Malini, S. Grossi, M. Deganuto, C. Rosano, R. Parini, S. Dominisini, R. Cariati, S. Zampieri, B. Bembi, M. Filocamo, A. Dardis, Functional analysis of 11 novel GBA alleles, Eur. J. Hum. Gene.: EJHG 22 (2014) 511–516.

[10] R.N. Alcalay, O.A. Levy, C.C. Waters, S. Fahn, B. Ford, S.H. Kuo, P. Mazzoni, M.W. Pauciulo, W.C. Nichols, Z. Gan-Or, G.A. Rouleau, W.K. Chung, P. Wolf, P. Oliva, J. Keutzer, K. Marder, X. Zhang, Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations, Brain: J. Neurol. 138 (2015) 2648–2658.

[11] T. Oeda, A. Umemura, Y. Mori, S. Tomita, M. Kohsaka, K. Park, K. Inoue, H. Fujimura, H. Hasegawa, H. Sugiyama, H. Sawada, Impact of glucocerebrosidase mutations on motor and nonmotor complications in Parkinson's disease, Neurobiol. Aging 36 (2015) 3306–3313.

[12] K. Brockmann, K. Srulijes, S. Pflederer, A.K. Hauser, C. Schulte, W. Maetzler, T. Gasser, D. Berg, GBA-associated Parkinson's disease: reduced survival and more rapid progression in a prospective longitudinal study, Mov. Disord.: Off. J. Mov. Disord. Soc. (2014), http://dx.doi.org/10.1002/mds.26071.

[13] M. Toft, L. Pielsticker, O.A. Ross, J.O. Aasly, M.J. Farrer, Glucocerebrosidase gene mutations and Parkinson disease in the Norwegian population, Neurology 66 (2006) 415–417.

[14] S. Jesus, I. Huertas, I. Bernal-Bernal, M. Bonilla-Toribio, M.T. Caceres-Redondo, L. Vargas-Gonzalez, M. Gomez-Llamas, F. Carrillo, E. Calderon, M. Carballo, P. Gomez-Garre, P. Mir, GBA variants influence motor and non-Motor features of parkinson's disease, PLoS One 11 (2016) e0167749.

[15] M.Y. Davis, C.O. Johnson, J.B. Leverenz, D. Weintraub, J.Q. Trojanowski, A. Chen-Plotkin, V.M. Van Deerlin, J.F. Quinn, K.A. Chung, A.L. Peterson-Hiller, L.S. Rosenthal, T.M. Dawson, M.S. Albert, J.G. Goldman, G.T. Stebbins, B. Bernard, Z.K. Wszolek, O.A. Ross, D.W. Dickson, D. Eidelberg, P.J. Mattis, M. Niethammer, D. Yearout, S.C. Hu, B.A. Cholerton, M. Smith, I.F. Mata, T.J. Montine, K.L. Edwards, C.P. Zabetian, Association of GBA mutations and the E326K polymorphism with motor and cognitive progression in parkinson disease, JAMA Neurol. 73 (2016) 1217–1224.

[16] C.F. Bento, A. Ashkenazi, M. Jimenez-Sanchez, D.C. Rubinsztein, The Parkinson's disease-associated genes ATP13A2 and SYT11 regulate autophagy via a common pathway, Nat. Commun. 7 (2016) 11803.

Paper II

Neurobiology of Aging 74 (2019) 236.e1–236.e5

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

No evidence for *DNM3* as genetic modifier of age at onset in idiopathic Parkinson's disease

Victoria Berge-Seidl [a,b,*], Lasse Pihlstrøm [a], Zbigniew K. Wszolek [c], Owen A. Ross [d], Mathias Toft [a,b]

[a] *Department of Neurology, Oslo University Hospital, Oslo, Norway*
[b] *Faculty of Medicine, University of Oslo, Oslo, Norway*
[c] *Department of Neurology, Mayo Clinic, Jacksonville, FL, USA*
[d] *Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA*

ARTICLE INFO

Article history:
Received 9 November 2017
Received in revised form 14 September 2018
Accepted 15 September 2018
Available online 22 September 2018

Keywords:
Parkinson's disease
Age at onset
DNM3

ABSTRACT

Parkinson's disease (PD) is a disorder with highly variable clinical phenotype. The identification of genetic variants modifying age at onset and other traits is of great interest because it may provide insight into disease mechanisms and potential therapeutic targets. A variant in the *DNM3* gene (rs2421947) has been reported as a genetic modifier of age at onset in *LRRK2*-associated PD. To test the possible effect of genetic variation in *DNM3* on age at onset in idiopathic PD, we examined rs2421947 in a total of 5918 patients with PD from seven data sets. We also assessed the potential effect of all common variants in the *DNM3* locus. There was no significant association between rs2421947 and age at onset in any of the individual studies. Meta-analysis of the seven studies was nonsignificant and the between-study heterogeneity was minimal. No other common variants within the *DNM3* locus affected age at onset. In conclusion, we find no evidence of an association between *DNM3* variants and age at onset in idiopathic PD.

1. Introduction

Parkinson's disease (PD) is a common neurodegenerative disorder with a complex etiology. A small proportion of patients with PD have a monogenic form of the disease with highly penetrant mutations following an autosomal dominant or recessive inheritance pattern. However, most PD cases are idiopathic, presumably caused by the combined action of multiple genetic variants in interplay with epigenetic, environmental, and stochastic factors (Lill, 2016). To date, genome-wide association studies (GWASs) have linked more than 40 risk loci to PD susceptibility (Chang et al., 2017).

In addition to affecting the risk of disease development, genetic variants may also affect the clinical phenotype once the disease has manifested. The clinical heterogeneity of PD is characterized by a marked variation in the pattern and progression of motor, cognitive, and other nonmotor symptoms. Both rare monogenic mutations and common genetic variants have been shown to contribute to the clinical heterogeneity of PD (Pihlstrom et al., 2016; Puschmann, 2013; Winder-Rhodes et al., 2013).

There is a broad range of age at onset in PD, varying between debut in early adulthood to patients reaching the 8th and 9th decade of life before the onset of motor symptoms. Several studies have investigated the effect of PD risk loci on the onset age. Cumulative genetic risk scores calculated across PD risk loci have been shown to have a small, but consistent, effect on age at onset (Escott-Price et al., 2015; Lill et al., 2015; Nalls et al., 2015; Pihlstrom and Toft, 2015). In addition, risk loci having the greatest effect in PD GWAS meta-analysis (*GBA*, *SNCA*, *MAPT,* and *TMEM175)* (Nalls et al., 2014) are reported to individually be associated with age at onset (Brockmann et al., 2013; Davis et al., 2016; Lill et al., 2015; Nalls et al., 2015).

Linking common variation to age at onset represents an interesting step toward a better understanding of how genetics affects the PD phenotype. In a recent genome-wide study of genetic modifiers of age at onset in leucine-rich repeat kinase 2 (LRRK2) p.G2019S carriers, a *DNM3* haplotype tagged by rs2421947 was identified (Trinh et al., 2016). This *LRRK2* mutation is the most frequent genetic cause of PD in many populations, estimated to have a frequency of 1% in white North

[*] Corresponding author at: Department of Neurology, Oslo University Hospital, P.O. Box 4950 Nydalen, N-0424 Oslo, Norway. Tel.: +47 23079022.
E-mail address: victoria.berge@medisin.uio.no (V. Berge-Seidl).

0197-4580/$ – see front matter © 2018 Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.neurobiolaging.2018.09.022

American and as high as 39% in North African Arab patients (Healy et al., 2008).

LRRK2 mutations cause an autosomal dominant form of PD often segregating in families, while GWASs provide consistent evidence that common variation at this locus also modulates disease risk. Because *LRRK2* is part of the genetic background for idiopathic PD, variants that modulate age at onset in *LRRK2* parkinsonism may also exert an effect in a much wider group of patients. Herein we report analyses of data from seven studies of PD from Europe and North America to determine associations between the *DNM3* rs2421947 variant and age at onset of idiopathic PD. To study the potential effect of other *DNM3* variants, we also performed a complete assessment of common variation in the gene locus.

2. Methods

2.1. Study populations

We analyzed individual-level genotypes from seven different data sets. Samples originated from genetic studies of PD from Oslo University Hospital and Mayo Clinic Jacksonville. The remaining five data sets were publicly available and selected due to available individual genotype information in PD patients with a reported age at onset. The following four data sets were accessed from dbGaP: 1) CIDR: Genome Wide Association Study in Familial Parkinson Disease (Accession number: phs000126.v1.p1), 2) Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson's Solutions) Collaboration (Accession number: phs000048.v1.p1), 3) National Institute of Neurological Disorders and Stroke (NINDS) Genome-Wide Genotyping in Parkinson's Disease (Accession number: phs000089.v3.p2), 4) Genome-Wide Association Study of Parkinson Disease: Genes and Environment performed by the NeuroGenetics Research Consortium (NGRC) (Accession number: phs000196.v2.p1). The last data set is made available by the Parkinson's Progression Markers Initiative (http://www.ppmi-info.org).

All patients have been examined by a neurologist. The Oslo and Mayo Clinic patients were diagnosed according to the revised UKPDSBB criteria. Inclusion criteria for the other studies have previously been described (Hamza et al., 2010; Maraganore et al., 2005; Nalls et al., 2016; Pankratz et al., 2009; Simon-Sanchez et al., 2009). Age at onset is either reported as age at symptom onset or age at diagnosis. In the current analysis, patients with PD reporting other than Caucasian non-Hispanic ethnicity have been excluded along with LRRK2 p.G2019S carriers that were identified by imputation or had previously been genotyped. A large subset of the Oslo and Mayo Clinic patients was sequenced for genes causing Mendelian forms of PD, and mutation carriers were excluded from the analysis. In addition, no known carriers of Mendelian PD mutations in the publicly available data sets were included in the analysis. Demographic characteristics are summarized in Table 1. All participants gave written informed consent. Sample and data collection at each study site was approved by local ethics committees. The study was approved by the Regional Committee for Medical Research Ethics (Oslo, Norway).

2.2. Genotyping and quality control

Mayo Clinic patients were genotyped for rs2421947 using a TaqMan assay. A subset of genotypes was validated by Sanger sequencing with complete concordance. For the other studies genome-wide genotypes were available. Oslo samples were genotyped using the Illumina Infinium OmniExpress v.1.1 array. Pre-imputation quality filtering included filtering out variants with genotype rate <0.95 or Hardy-Weinberg equilibrium $p < 10^{-6}$ and removal of individuals with call rate <0.95, excess heterozygosity >4 standard deviations from mean, and evidence of cryptic relatedness or sex-check failure. Population outliers were excluded from analysis after inspection of principal component analysis plot (>2.5 standard deviation from mean). Details of genotyping methods and data quality assessments for the publicly available GWASs and the Parkinson's Progression Markers Initiative study are described in previous publications (Hamza et al., 2010; Maraganore et al., 2005; Nalls et al., 2016; Pankratz et al., 2009; Simon-Sanchez et al., 2009).

Common, pruned, genotyped variants (minor allele frequency >0.05 and $r^2 < 0.5$) were used to calculate principal components for each of the six genome-wide data sets. For all genome-wide data sets, imputation was performed using the Michigan imputation server (Das et al., 2016) with reference data from the Haplotype Reference Consortium (McCarthy et al., 2016) setting a quality cutoff of $r^2 > 0.3$ for variants included in the analysis. A set of common, pruned variants from each imputed data set was merged to assess cryptic relatedness across studies. Duplicates and related samples were removed. The final sample sets included in analysis after quality control comprise a total of 5918 patients with PD (Table 1).

2.3. Statistical analyses

First we tested all seven studies individually for association between the *DNM3* rs2421947 variant and age at onset under an additive linear regression model. In an alternative binary analysis, age at onset was dichotomized by the median onset calculated across all seven data sets (61 years of age) and logistic regression was used to test for association with the *DNM3* rs2421947 variant within each data set. In both regression analyses, sex and the first five principal components were used as covariates in the genome-wide data sets, while sex was the single covariate in analysis of the Mayo Clinic study. Association analyses were performed in the PLINK 1.9 software (https://www.cog-genomics.org/plink/1.9/) (Chang et al., 2015). Inverse-variance, fixed-effects meta-analysis of the seven studies was conducted using the Genome-wide Association Meta-Analysis software (Magi and Morris, 2010). Between-study heterogeneity was assessed using Cochran's Q test and Higgins's I^2 statistic. Owing to the increased burden of recessive

Table 1
Demographic characteristics of study samples

Study	Population	Patients with PD	% Male	Age at onset ±SD	Genotyping method
Oslo	Norway	472	64	55.9 ± 11.2	Illumina Infinium OmniExpress v.1.1.
Mayo	USA	987	64	65.5 ± 11.8	Taqman assay
CIDR	North America, Europe and Australia	823	59	62.0 ± 10.7	Illumina HumanCNV370 BeadChip
LEAPS	USA	439	62	60.9 ± 11.1	Perlegen DNA chip (85k SNP markers)
NINDS	USA	912	60	58.4 ± 13.2	Illumina HumanHap550 BeadChip
NGRC	USA	1971	68	58.4 ± 11.9	Illumina HumanOmni1_Quad
PPMI	USA and Europe	314	68	59.7 ± 9.8	Illumina NeuroX

Key: CIDR, Center for Inherited Disease Research; LEAPS, Linked Efforts to Accelerate Parkinson's Solutions; NINDS, National Institute of Neurological Disorders and Stroke; NGRC, NeuroGenetics Research Consortium; PPMI, Parkinson's Progression Markers Initiative; PD, Parkinson's disease; SD, standard deviation.

disease-causing mutations in early-onset PD (Puschmann, 2013), we repeated the linear regression analysis excluding all patients with an age at onset <40 years of age.

Next, common variation within *DNM3* as well as 100kb upstream and downstream of the gene was analyzed by linear regression in the six genome-wide data sets. Variants with a minor allele frequency below 0.01 in each data set were excluded from analysis. Association analysis, both within the individual data sets and meta-analysis, were performed as described for *DNM3* rs2421947. To estimate the degree of multiple testing, we generated a combined, pruned data set using a cutoff of linkage disequilibrium $> r^2 = 0.5$, leaving 226 independent variants. Adjusting for 226 independent tests by Bonferroni correction, a *p*-value < 0.0002 was considered significant. Power calculations were performed with the function pwr.f2.test in the R package pwr (version 1.2–1; https://cran.r-project.org/web/packages/pwr/index.html).

3. Results

We found no significant association between rs2421947 and age at onset in any of the individual studies, both when analyzed as a quantitative trait and in the alternative binary analysis. The frequency of the alternative allele G of rs2421947 was similar in all the seven studies, varying between 54% and 56%. Meta-analysis of the seven studies analyzed as a quantitative trait was nonsignificant ($p = 0.55$), and the between-study heterogeneity was minimal ($I^2 = 0\%$, $p = 0.77$). Results from linear regression analysis of the individual studies and meta-analysis are shown in Fig. 1. We obtained similar results when we excluded patients with an age at onset <40 years from the analysis. Meta-analysis of age at onset as a binary trait was also nonsignificant (OR = 1.03, 95% CI = 0.96–1.11, $p = 0.44$) and between-study heterogeneity was minimal ($I^2 = 0\%$, $p = 0.85$).

Next, we analyzed all common variants within the *DNM3* gene and the flanking genomic region. Variants covered in at least four of the six analyzed genome-wide data sets were included in the meta-analysis, constituting 1932 variants. None of the meta-analyzed variants had a significant association with age at onset when corrected for multiple testing. Results from the extended *DNM3* analysis are provided in Supplementary Table 1. Tests for statistical heterogeneity indicated that heterogeneity was low ($I^2 < 50\%$) for the vast majority (97%) of the meta-analyzed variants. Studies of the combined impact of PD risk loci on age at onset report a phenotypic variance explained by the calculated genetic risk score of 0.6% and 0.7%, with the signal mostly being driven by two individually age at onset-associated variants (Lill et al., 2015; Nalls et al., 2015). Assuming a variance explained of ~0.5% for the tested variant, we have a power of over 99% for the primary analysis of rs2421947 (N = 5918) and a power of 89.5% to achieve a *p*-value of 0.0002 (N = 4931) in the extended *DNM3* analysis.

The effect of rs2421947 on age at onset was also assessed in the LRRK2 p.G2019S carriers (N = 39) that had been excluded from the aforementioned analyses, although this test was limited by a small sample size. Association with age at onset analyzed as a quantitative trait, with sex and data set as covariables, was nonsignificant for rs2421947, with the trend for direction of effect reversed relative to the original report by Trinh et al., 2016 (effect allele = G, beta = 1.79, 95% CI =−3.31–6.89, $p = 0.50$). Trinh et al. report a correlation between rs2421947 genotype and *DNM3* mRNA levels in striatal brain tissue (Trinh et al., 2016). We explored the Genotype-Tissue Expression (GTEx) Portal (version 7; https://www.gtexportal.org/home/) and found that significant expression quantitative trait loci (eQTLs) for *DNM3* are reported in cerebellar hemisphere and cerebellum, but not in any of the other brain regions examined by the GTEx project. Interestingly, rs2421947 and variants in high linkage disequilibrium ($r^2 \geq 0.6$) are not reported as significant eQTLs for *DNM3* in any brain tissue.

4. Discussion

In this study, we found no evidence for a modifying effect of rs2421947 or other common *DNM3* variants on age at onset in idiopathic PD. A *DNM3* haplotype tagged by rs2421947 was identified by Trinh et al. as a modifier of age at onset in LRRK2 p.G2019S carriers. They reported that the median age at onset of *DNM3* GG homozygotes was 12.5 years younger than that of CC homozygotes. This is a large difference in the onset age compared with the effect of other variants associated with age at onset in PD, and could be meaningful in the clinical setting. As shown by the 95% confidence interval of our primary analysis, it is highly unlikely that we did not detect an effect altering the onset of PD more than a few months per G allele.

We performed a meta-analysis including a total of 5918 patients with PD, and the high number of analyzed individuals is a strength of our study. However, by including cases from different study sites with variations in study design, heterogeneity may be introduced in meta-analysis of the genetic data. We assessed this and found low heterogeneity. We accounted for population substructure within the individual studies by including five eigenvectors in the regression model. The meta-analyzed studies use mostly self-

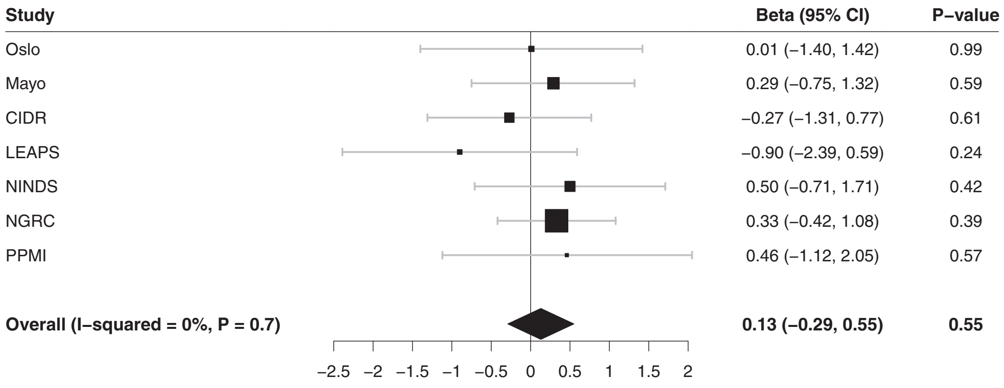


Fig. 1. Study-specific and meta-analysis results for the *DNM3* variant rs2421947. Forest plot showing the effect of rs2421947 on age at onset in idiopathic Parkinson's disease in individual studies and meta-analysis. The effect size of the G allele is given as a beta estimate with a 95% confidence interval (CI). The size of the squares indicates the size of the data sets. Abbreviations: CIDR, Center for Inherited Disease Research; LEAPS, Linked Efforts to Accelerate Parkinson's Solutions; NINDS, National Institute of Neurological Disorders and Stroke; NGRC, NeuroGenetics Research Consortium; PPMI, Parkinson's Progression Markers Initiative.

reported symptom onset. Age at onset is subjective and may be prone to recall bias. Nevertheless, the reliability of self- and family-reported age at onset compared with medical records is high and all three methods have been regarded as valid (Reider et al., 2003).

A recent study of LRRK2 p.G2019S carriers in the Spanish population did not find any association between *DNM3* rs2421947 and age at onset of PD (Fernandez-Santiago et al., 2018). Linkage patterns vary among populations, and the possibility that disease-relevant variation could be tagged by different genetic markers in Europeans/North Americans as compared with the Arab-Berber population studied by Trinh et al. prompted us to extend our analysis to all imputed common variants across the *DNM3* locus. The genomic region flanking the *DNM3* gene is included in our analysis to cover potential regulatory variants, although this increases the multiple testing burden. On the other hand, regulatory variants affecting *DNM3* expression may reside in an even more distal part of the genome not covered in our analysis. eQTL data could be used to identify potential regulatory variants and reduce the number of tests to adjust for. However, the currently available databases are incomplete because eQTLs, in addition to depending on tissue and cell type, also may vary between different physiological conditions (Albert and Kruglyak, 2015).

Despite recent efforts to elucidate the genetic architecture behind age at onset and other clinical characteristics of PD, the vast majority of genetic variation affecting PD phenotypes remains unexplained. Many studies have limited their analysis to known risk loci of PD, while attempts at identifying novel genetic modifiers of age at onset have proven challenging. Genetic modifiers of age at onset may be limited to subgroups of patients carrying specific mutations or susceptibility variants. Variations in the *MAPT* gene have been found to be associated with age at onset in patients with PD carrying a *LRRK2* mutation (Gan-Or et al., 2012; Golub et al., 2009). A recent GWAS of age at onset analyzed PD patients with and without a family history of the disease separately. No significant association was found in those without a family history of PD, while two signals were detected in individuals reporting to have a first- or second-degree relative with PD. Both these signals mapped to gene regions that are not known to affect PD risk (Hill-Burns et al., 2016).

Discovering genetic modifiers of phenotype in PD is important because it may provide insight into disease mechanisms and help in identifying potential therapeutic targets. *DNM3* has previously not been identified by GWASs of disease risk or onset age (Chang et al., 2017; Hill-Burns et al., 2016; Latourelle et al., 2009; Nalls et al., 2014). We found no association between common *DNM3* variants and age at onset in idiopathic PD, but the possible contribution of rare variants within this genetic locus cannot be excluded. Variability within the *DNM3* locus may be a specific modifier of *LRRK2* parkinsonism, although this has yet to be replicated in independent studies. *DNM3* encodes the protein dynamin-3 which is highly expressed in neurons (Raimondi et al., 2011). The LRRK2 protein has been shown to interact with dynamin-3 and other members of the dynamin GTPase superfamily that regulate membrane dynamics important for endocytosis and mitochondrial morphology (Stafa et al., 2014). Trinh et al. report a correlation between rs2421947 genotype and *DNM3* mRNA levels in striatal tissue, but such an association is not observed in the GTEx portal. There are several methodological differences between these analyses and additional evidence is needed before conclusions regarding rs2421947 as an eQTL for *DNM3* in the brain can be drawn.

Further insight into disease mechanisms may be gained by examining complex genetic interactions. A novel epistatic interaction between two genetic variants was reported by a recent study incorporating genetic, molecular, and clinical data into models to predict motor progression in PD (Latourelle et al., 2017). Larger patient cohorts with comprehensive characterization of disease phenotype will benefit future studies of how genetics affect clinical heterogeneity.

Disclosure statement

The authors report no conflicts of interest concerning the research related to the article.

Acknowledgements

The authors thank all study participants, investigators, and funding partners for their contribution to the data analyzed in the current publication. The funding organizations had no role in study design, data collection and analysis, decision to publish, or preparation of the article. VBS and MT are funded by a grant from the South-Eastern Norway Regional Health Authority. MT is also supported by the Research Council of Norway. LP is supported by the Norwegian Health Association. Mayo Clinic is a Morris K. Udall Parkinson's Disease Research Center of Excellence (NINDS P50 #NS072187 to OR and ZKW); ZKW is also funded by Mayo Clinic Center for Individualized Medicine, Mayo Clinic Neuroscience Focused Research Team (Cecilia and Dan Carmichael Family Foundation, and the James C. and Sarah K. Kennedy Fund for Neurodegenerative Disease Research at Mayo Clinic in Florida), the gift from Carl Edward Bolch, Jr, and Susan Bass Bolch, the Sol Goldman Charitable Trust, and Donald G. and Jodi P. Heeringa. OR is supported by NINDS R01 NS078086, Department of Defense award W81XWH-16-PRMRP-IIRA, Michael J. Fox Foundation. The analyses presented in the current publication are based on the use of study data downloaded from the dbGaP website, under dbGaP accession phs000126.v1.p1, phs000048.v1.p1, phs000089.v3.p2 and phs000196.v2.p1. Data were also obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners including AbbVie, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi Genzyme, Servier, Takeda, Teva and UCB.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.neurobiolaging.2018.09.022.

References

Albert, F.W., Kruglyak, L., 2015. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16, 197–212.

Brockmann, K., Schulte, C., Hauser, A.K., Lichtner, P., Huber, H., Maetzler, W., Berg, D., Gasser, T., 2013. SNCA: major genetic modifier of age at onset of Parkinson's disease. Mov Disord. 28, 1217–1221.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7.

Chang, D., Nalls, M.A., Hallgrimsdottir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., Kerchner, G.A., Ayalon, G., Bingol, B., Sheng, M., Hinds, D., Behrens, T.W., Singleton, A.B., Bhangale, T.R., Graham, R.R., 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat. Genet. 49, 1511–1516.

Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R., Fuchsberger, C., 2016. Next-generation genotype imputation service and methods. Nat. Genet. 48, 1284–1287.

Davis, A.A., Andruska, K.M., Benitez, B.A., Racette, B.A., Perlmutter, J.S., Cruchaga, C., 2016. Variants in GBA, SNCA, and MAPT influence Parkinson disease risk, age at onset, and progression. Neurobiol. Aging 37, 209.e201–209.e207.

Escott-Price, V., Nalls, M.A., Morris, H.R., Lubbe, S., Brice, A., Gasser, T., Heutink, P., Wood, N.W., Hardy, J., Singleton, A.B., Williams, N.M., 2015. Polygenic risk of Parkinson disease is correlated with disease age at onset. Ann. Neurol. 77, 582–591.

Fernandez-Santiago, R., Garrido, A., Infante, J., Gonzalez-Aramburu, I., Sierra, M., Fernandez, M., Valldeoriola, F., Munoz, E., Compta, Y., Marti, M.J., Rios, J., Tolosa, E., Ezquerra, M., 2018. alpha-synuclein (SNCA) but not dynamin 3 (DNM3) influences age at onset of leucine-rich repeat kinase 2 (LRRK2) Parkinson's disease in Spain. Mov Disord. 33, 637–641.

Gan-Or, Z., Bar-Shira, A., Mirelman, A., Gurevich, T., Giladi, N., Orr-Urtreger, A., 2012. The age at motor symptoms onset in LRRK2-associated Parkinson's disease is affected by a variation in the MAPT locus: a possible interaction. J. Mol. Neurosci. 46, 541–544.

Golub, Y., Berg, D., Calne, D.B., Pfeiffer, R.F., Uitti, R.J., Stoessl, A.J., Wszolek, Z.K., Farrer, M.J., Mueller, J.C., Gasser, T., Fuchs, J., 2009. Genetic factors influencing age at onset in LRRK2-linked Parkinson disease. Parkinsonism Relat. Disord. 15, 539–541.

Hamza, T.H., Zabetian, C.P., Tenesa, A., Laederach, A., Montimurro, J., Yearout, D., Kay, D.M., Doheny, K.F., Paschall, J., Pugh, E., Kusel, V.I., Collura, R., Roberts, J., Griffith, A., Samii, A., Scott, W.K., Nutt, J., Factor, S.A., Payami, H., 2010. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat. Genet. 42, 781–785.

Healy, D.G., Falchi, M., O'Sullivan, S.S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C.P., Goldwurm, S., Ferreira, J.J., Tolosa, E., Kay, D.M., Klein, C., Williams, D.R., Marras, C., Lang, A.E., Wszolek, Z.K., Berciano, J., Schapira, A.H., Lynch, T., Bhatia, K.P., Gasser, T., Lees, A.J., Wood, N.W., 2008. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol. 7, 583–590.

Hill-Burns, E.M., Ross, O.A., Wissemann, W.T., Soto-Ortolaza, A.I., Zareparsi, S., Siuda, J., Lynch, T., Wszolek, Z.K., Silburn, P.A., Mellick, G.D., Ritz, B., Scherzer, C.R., Zabetian, C.P., Factor, S.A., Breheny, P.J., Payami, H., 2016. Identification of genetic modifiers of age-at-onset for familial Parkinson's disease. Hum. Mol. Genet. 25, 3849–3862.

Latourelle, J.C., Beste, M.T., Hadzi, T.C., Miller, R.E., Oppenheim, J.N., Valko, M.P., Wuest, D.M., Church, B.W., Khalil, I.G., Hayete, B., Venuto, C.S., 2017. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. Lancet Neurol. 16, 908–916.

Latourelle, J.C., Pankratz, N., Dumitriu, A., Wilk, J.B., Goldwurm, S., Pezzoli, G., Mariani, C.B., DeStefano, A.L., Halter, C., Gusella, J.F., Nichols, W.C., Myers, R.H., Foroud, T., 2009. Genomewide association study for onset age in Parkinson disease. BMC Med. Genet. 10, 98.

Lill, C.M., 2016. Genetics of Parkinson's disease. Mol. Cell. Probes 30, 386–396.

Lill, C.M., Hansen, J., Olsen, J.H., Binder, H., Ritz, B., Bertram, L., 2015. Impact of Parkinson's disease risk loci on age at onset. Mov Disord. 30, 847–850.

Magi, R., Morris, A.P., 2010. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 11, 288.

Maraganore, D.M., de Andrade, M., Lesnick, T.G., Strain, K.J., Farrer, M.J., Rocca, W.A., Pant, P.V., Frazer, K.A., Cox, D.R., Ballinger, D.G., 2005. High-resolution whole-genome association study of Parkinson disease. Am. J. Hum. Genet. 77, 685–693.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L.J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C.M., Gillies, C.E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J.C., Boomsma, D., Branham, K., Breen, G., Brummett, C.M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F.S., Corbin, L.J., Smith, G.D., Dedoussis, G., Dorr, M., Farmaki, A.E., Ferrucci, L., Forer, L., Fraser, R.M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O.L., Hveem, K., Kretzler, M., Lee, J.C., McGue, M., Meitinger, T., Melzer, D., Min, J.L., Mohlke, K.L., Vincent, J.B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, J.B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P.E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L.H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M.G., Wilson, J.F., Frayling, T., de Bakker, P.I., Swertz, M.A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C.A., Myers, R.M., Boehnke, M., McCarthy, M.I., Durbin, R., 2016. A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. 48, 1279–1283.

Nalls, M.A., Escott-Price, V., Williams, N.M., Lubbe, S., Keller, M.F., Morris, H.R., Singleton, A.B., 2015. Genetic risk and age in Parkinson's disease: continuum not stratum. Mov Disord. 30, 850–854.

Nalls, M.A., Keller, M.F., Hernandez, D.G., Chen, L., Stone, D.J., Singleton, A.B., 2016. Baseline genetic associations in the Parkinson's progression markers initiative (PPMI). Mov Disord. 31, 79–85.

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M.F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J.H., Cheng, R., Ikram, M.A., Ioannidis, J.P., Hadjigeorgiou, G.M., Bis, J.C., Martinez, M., Perlmutter, J.S., Goate, A., Marder, K., Fiske, B., Sutherland, M., Xiromerisiou, G., Myers, R.H., Clark, L.N., Stefansson, K., Hardy, J.A., Heutink, P., Chen, H., Wood, N.W., Houlden, H., Payami, H., Brice, A., Scott, W.K., Gasser, T., Bertram, L., Eriksson, N., Foroud, T., Singleton, A.B., 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. 46, 989–993.

Pankratz, N., Wilk, J.B., Latourelle, J.C., DeStefano, A.L., Halter, C., Pugh, E.W., Doheny, K.F., Gusella, J.F., Nichols, W.C., Foroud, T., Myers, R.H., 2009. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum. Genet. 124, 593–605.

Pihlstrom, L., Morset, K.R., Grimstad, E., Vitelli, V., Toft, M., 2016. A cumulative genetic risk score predicts progression in Parkinson's disease. Mov Disord. 31, 487–490.

Pihlstrom, L., Toft, M., 2015. Cumulative genetic risk and age at onset in Parkinson's disease. Mov Disord. 30, 1712–1713.

Puschmann, A., 2013. Monogenic Parkinson's disease and parkinsonism: clinical phenotypes and frequencies of known mutations. Parkinsonism Relat. Disord. 19, 407–415.

Raimondi, A., Ferguson, S.M., Lou, X., Armbruster, M., Paradise, S., Giovedi, S., Messa, M., Kono, N., Takasaki, J., Cappello, V., O'Toole, E., Ryan, T.A., De Camilli, P., 2011. Overlapping role of dynamin isoforms in synaptic vesicle endocytosis. Neuron 70, 1100–1114.

Reider, C.R., Halter, C.A., Castelluccio, P.F., Oakes, D., Nichols, W.C., Foroud, T., 2003. Reliability of reported age at onset for Parkinson's disease. Mov Disord. 18, 275–279.

Simon-Sanchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G., Kruger, R., Federoff, M., Klein, C., Goate, A., Perlmutter, J., Bonin, M., Nalls, M.A., Illig, T., Gieger, C., Houlden, H., Steffens, M., Okun, M.S., Racette, B.A., Cookson, M.R., Foote, K.D., Fernandez, H.H., Traynor, B.J., Schreiber, S., Arepalli, S., Zonozi, R., Gwinn, K., van der Brug, M., Lopez, G., Chanock, S.J., Schatzkin, A., Park, Y., Hollenbeck, A., Gao, J., Huang, X., Wood, N.W., Lorenz, D., Deuschl, G., Chen, H., Riess, O., Hardy, J.A., Singleton, A.B., Gasser, T., 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet. 41, 1308–1312.

Stafa, K., Tsika, E., Moser, R., Musso, A., Glauser, L., Jones, A., Biskup, S., Xiong, Y., Bandopadhyay, R., Dawson, V.L., Dawson, T.M., Moore, D.J., 2014. Functional interaction of Parkinson's disease-associated LRRK2 with members of the dynamin GTPase superfamily. Hum. Mol. Genet. 23, 2055–2077.

Trinh, J., Gustavsson, E.K., Vilarino-Guell, C., Bortnick, S., Latourelle, J., McKenzie, M.B., Tu, C.S., Nosova, E., Khinda, J., Milnerwood, A., Lesage, S., Brice, A., Tazir, M., Aasly, J.O., Parkkinen, L., Haytural, H., Foroud, T., Myers, R.H., Sassi, S.B., Hentati, E., Nabli, F., Farhat, E., Amouri, R., Hentati, F., Farrer, M.J., 2016. DNM3 and genetic modifiers of age of onset in LRRK2 Gly2019Ser parkinsonism: a genome-wide linkage and association study. Lancet Neurol. 15, 1248–1256.

Winder-Rhodes, S.E., Evans, J.R., Ban, M., Mason, S.L., Williams-Gray, C.H., Foltynie, T., Duran, R., Mencacci, N.E., Sawcer, S.J., Barker, R.A., 2013. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. Brain 136 (Pt 2), 392–399.

Paper III

www.nature.com/scientificreports

scientific reports

OPEN

Integrative analysis identifies bHLH transcription factors as contributors to Parkinson's disease risk mechanisms

Victoria Berge-Seidl[1,2], Lasse Pihlstrøm[1] & Mathias Toft[1,2]✉

Genome-wide association studies (GWAS) have identified multiple genetic risk signals for Parkinson's disease (PD), however translation into underlying biological mechanisms remains scarce. Genomic functional annotations of neurons provide new resources that may be integrated into analyses of GWAS findings. Altered transcription factor binding plays an important role in human diseases. Insight into transcriptional networks involved in PD risk mechanisms may thus improve our understanding of pathogenesis. We analysed overlap between genome-wide association signals in PD and open chromatin in neurons across multiple brain regions, finding a significant enrichment in the superior temporal cortex. The involvement of transcriptional networks was explored in neurons of the superior temporal cortex based on the location of candidate transcription factor motifs identified by two de novo motif discovery methods. Analyses were performed in parallel, both finding that PD risk variants significantly overlap with open chromatin regions harboring motifs of basic Helix-Loop-Helix (bHLH) transcription factors. Our findings show that cortical neurons are likely mediators of genetic risk for PD. The concentration of PD risk variants at sites of open chromatin targeted by members of the bHLH transcription factor family points to an involvement of these transcriptional networks in PD risk mechanisms.

Parkinson's disease (PD) is a progressive neurodegenerative disorder affecting about 1% of the population above 60 years of age[1]. The cause of neuronal death is poorly understood, and this is obstructing the path toward more effective treatments. The largest-to-date genome-wide association study (GWAS) for PD identified 90 independent association signals, of which a large proportion were new compared to previous reports[2]. In spite of the last two decades' successful identification of genetic association signals in PD and other complex diseases, translation into underlying biological mechanisms has been scarce. GWAS signals typically involve multiple variants in high linkage disequilibrium (LD), making it difficult to pinpoint the actual causal variants. In addition, most risk variants are located in the noncoding part of the genome, where the functional impact may be challenging to predict[3,4]. There is however a growing amount of epigenomic and transcriptomic data that may be integrated with GWAS findings to discover disease-relevant regulatory networks.

In previous studies, PD risk variants have been integrated with gene expression data, epigenomic annotations and functionally related gene sets to identify cell types and pathways implicated in PD pathogenesis[5–7]. Studies coupling PD risk to transcription factor binding are however scarce and there is consequently limited knowledge concerning transcriptional networks central to PD pathogenesis. Altered transcription factor binding has been shown to play an important role in human diseases[8–10]. Transcription factors bind to short and specific DNA sequences, referred to as motifs, to alter gene expression. Genetic variants may alter the binding of a transcription factor through disruption of the transcription factor recognition motif. However, the majority of variability in transcription factor-DNA binding events appear to be caused by variants outside the transcription factor recognition motif[11]. A fine-mapping study of autoimmune diseases found that predicted causal variants tend to occur near binding sites for immune related transcription factors, but only a fraction alter recognizable transcription factor binding motifs[12].

Transcription factor binding patterns vary between cell types and may be directly assessed through chromatin immunoprecipitation sequencing (ChIP-seq)[13]. This requires one transcription factor to be tested at a time and only a fraction of transcription factor-cell type combinations has so far been assayed. Intersection of disease risk

[1]Department of Neurology, Oslo University Hospital, P.O. Box 4956 Nydalen, 0424 Oslo, Norway. [2]Faculty of Medicine, University of Oslo, Oslo, Norway. ✉email: mathias.toft@medisin.uio.no

variants for 213 phenotypes with an extensive catalogue of ChIP-seq derived transcription factor binding datasets identified more than 2000 significant transcription factor-disease relationships[14]. There were however sparse findings in regard to transcription factors associated with PD, which may be explained by the small number of transcription factors assayed in neuronal cell types.

Position weight matrices, which are widely used models to describe the DNA sequence binding preferences of transcription factors, may be used to scan the genome to predict transcription factor binding sites. Importantly, transcription factors only occupy a small proportion of the genomic sequences matching to their consensus binding sites. This is because transcription factor binding is influenced by additional features such as sequence context, accessibility of chromatin and interactions among transcription factors[15,16]. Integration of genome sequence information together with cell type specific experimental data has been shown to improve the accuracy of inference of transcription factor binding[17].

Through analysis of the overlap between Alzheimer's disease risk variants and open chromatin sites containing specific transcription factor motifs, Tansey et al. provided evidence suggestive of specific transcriptional networks being central to Alzheimer's disease risk mechanisms[18]. We use a similar approach integrating PD risk variants with open chromatin sites in brain, coupled with transcription factor motif analysis, to identify transcription factor networks contributing to PD risk.

Methods

Genomic annotations.

Assay for Transposase Accessible Chromatin followed by sequencing (ATAC-seq) is a fast and sensitive method used to map genome-wide accessibility of chromatin[19]. We downloaded maps of open chromatin in neurons and non-neurons across 14 distinct brain regions of five individuals from the online database Brain Open Chromatin Atlas (BOCA). A detailed description of data generation and quality control of this dataset has been published[20]. In brief, ATAC-seq was applied to neuronal and non-neuronal nuclei isolated from frozen brain tissue by fluorescence-activated nuclear sorting. Reads were mapped to the hg19 (GRCh37) reference genome using STAR aligner v2.5.0 and peaks representing open chromatin regions (OCRs) were called using model-based Analysis of ChIP-seq (MACS) v2.1[21,22].

The 14 analysed brain regions include different areas of neocortex, in addition to subcortical regions such as hippocampus, thalamus, amygdala, putamen and nucleus accumbens. Substantia nigra, which has a well established role in the pathogenesis of PD due to the loss of neurons in this region, was however not part of this dataset. There was to our knowledge no other data from ATAC-seq or similar assays analysing the accessibility of chromatin in human dopaminergic neurons of substantia nigra available at the time of our analysis. Pairwise intersections between genomic annotations were computed and visualized with the command line tool Intervene (version 0.6.4)[23]. Jaccard statistic was used as measure of similarity, where 0 means no overlap and 1 means full overlap.

Genetic association signals.

Genome-wide significant PD risk signals were accessed from a recent meta-analysis, which is the largest genetic study of PD to date[2]. This study, which involved the analysis of 37.7K cases, 18.6K UK Biobank proxy-cases and 1.4M controls, identified 90 independent association signals that we included in enrichment analyses. Published top-hits were accessed from Table S2 and we included the 90 association signals that were marked as having passed final filtering. We performed an additional analysis excluding the three PD risk signals located within the extended major histocompatibility complex (MHC) region (chr6: 26–34 Mb), due to the unusual LD and genetic architecture at this locus[24].

As negative controls, we selected GWASs from non-brain related disorders that had a number of independent association signals (p-value $< 5 \times 10^{-8}$) comparable to that of the included PD meta-analysis. A GWAS of inflammatory bowel disease (IBD) (study accession GCST004131) with 94 association signals and a GWAS of peak expiratory flow (PEF) (study accession GCST007430) with 91 association signals were accessed from the GWAS catalogue[25]. As for the PD association signals, additional analyses were performed excluding one IBD risk signal and two PEF risk signals located within the extended MHC region.

Testing for enrichment of PD risk variants in open chromatin regions.

Two methods were used to evaluate the statistical enrichment of PD risk variants and the two negative controls in OCRs defined by ATAC-seq in neurons across the 14 brain regions. We chose not to further analyse the non-neuronal cell population due to the cellular heterogeneity in this group, which contains different glial subtypes in addition to a small component of vascular cells and nucleated blood cells[26]. The workflow of our analysis is depicted in Fig. 1. First, enrichment was calculated with GoShifter, which includes genome-wide significant index variants and their LD proxies in the analysis[27]. We identified variants in LD with the index variants with the webserver Snipa (v3.3, http://www.snipa.org), using the European subset of 1000 Genomes Phase 3 v5 data and a LD threshold of $r^2 > 0.8$ (Supplementary Table S1)[28]. GoShifter calculates the proportion of risk loci where at least one linked variant overlaps the tested annotation. The observed overlap is then compared to a null distribution generated by randomly shuffling the annotations within each locus, thus preserving the local genomic structure. After each shuffle, the proportion of loci overlapping annotations is calculated. We carried out 10,000 permutations to draw the null distribution.

The second method applied, GREGOR, uses a snp-matching-based method to test for enrichment[29]. The number of trait-associated signals where an index variant or one of its LD proxies overlaps a regulatory annotation is calculated, then the probability of the observed overlap of risk variants is estimated relative to expectation using a set of matched control variants. Control variants match the index variants for number of variants in LD, minor allele frequency and distance to nearest gene. European 1000 Genomes Phase 1 data is implemented in

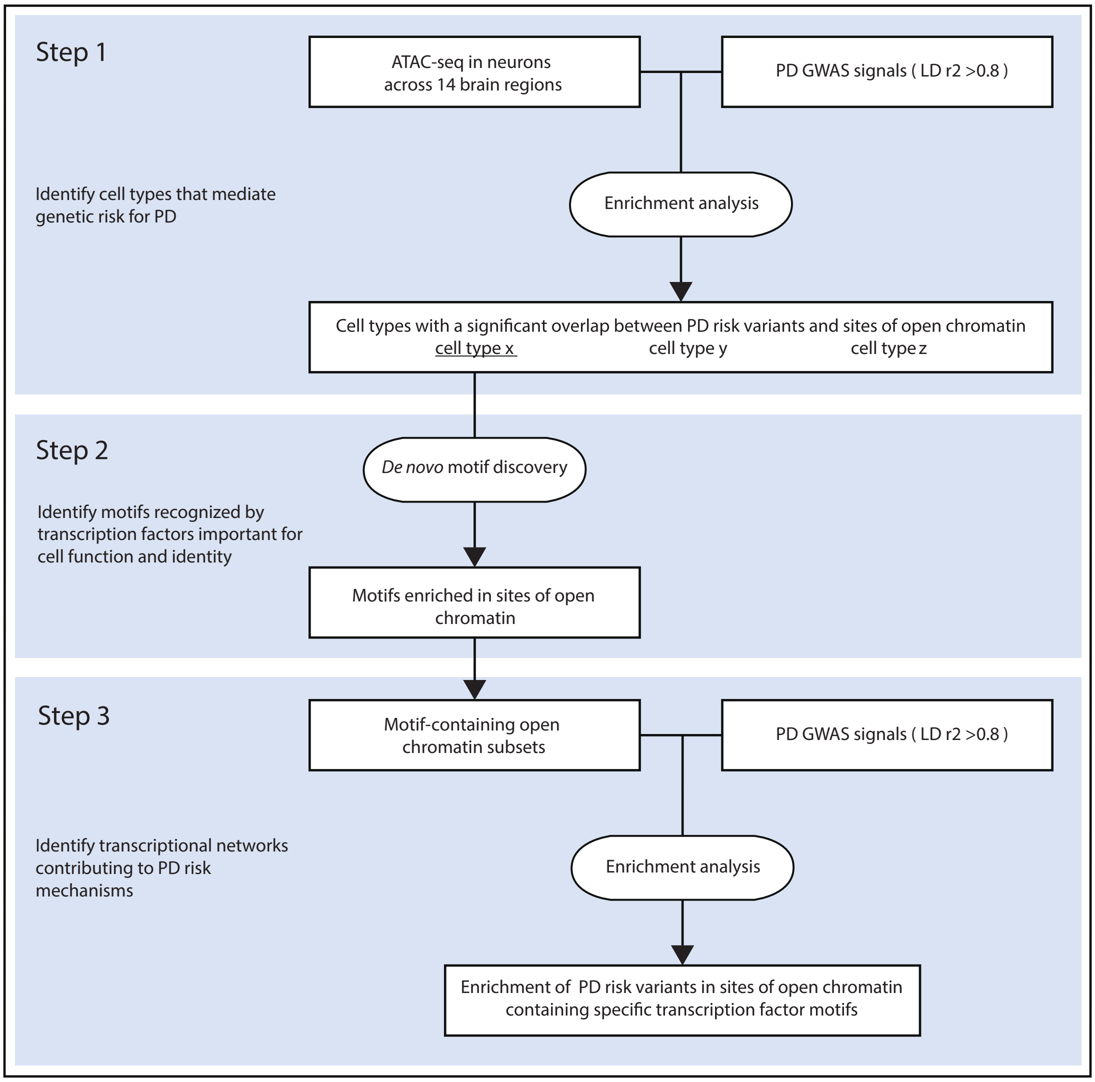

Step 1
Identify cell types that mediate genetic risk for PD
ATAC-seq in neurons across 14 brain regions
PD GWAS signals (LD r2 >0.8)
Enrichment analysis
Cell types with a significant overlap between PD risk variants and sites of open chromatin
cell type x
cell type y
cell type z
Step 2
Identify motifs recognized by transcription factors important for cell function and identity
De novo motif discovery
Motifs enriched in sites of open chromatin
Step 3
Identify transcriptional networks contributing to PD risk mechanisms
Motif-containing open chromatin subsets
PD GWAS signals (LD r2 >0.8)
Enrichment analysis
Enrichment of PD risk variants in sites of open chromatin containing specific transcription factor motifs

Figure 1. Workflow of the data analysis. PD, Parkinson's disease; GWAS, Genome-wide association study; LD, Linkage disequilibrium.

GREGOR and was used to identify LD proxies with the threshold set to $r^2 > 0.8$ and a LD window at 1 Mb. The minimum number of control variants for each index variant was set at 500.

We adjusted for multiple testing by Bonferroni correction, adjusting for 14 tests in the analysis of OCRs in neurons from different brain regions and 23 tests when analysing OCRs containing specific transcription factor motifs identified by de novo motif discovery with HOMER. In enrichment analysis of OCRs harboring de novo motifs and best matched known transcription factor motifs identified with MEME-ChIP, we adjusted for 13 and 18 tests. Brain annotations that pass the significance threshold with both GoShifter (adj. $p < 0.05$) and GREGOR (adj. $p < 0.05$) are reported as significantly enriched in the text.

De novo motif discovery and assignment to open chromatin regions.

Transcription factors targeting binding motifs that are enriched in a set of regulatory regions in a cell may be regarded as candidate transcriptional regulators of that cell. We performed de novo motif analysis with two different softwares, HOMER v 4.10.3 and MEME-ChIP v 5.1.1, to identify motifs significantly enriched in OCRs in superior temporal cortex neurons[30,31]. HOMER identifies motifs that are enriched in the target sequences relative to GC matched background sequences. In our analysis with HOMER we used the findMotifsGenome.pl script with –size given, -mask and otherwise default settings. De novo motifs are compared against a library of known motifs in the HOMER Motif Database and all motifs in Jaspar. The identified enriched de novo motifs were assigned to OCRs using the annotatePeaks.pl script with default parameters.

MEME-ChIP performs comprehensive motif analysis of large nucleotide datasets through the combination of several motif discovery and analysis tools. Although MEME-ChIP was designed for the analysis of peak regions identified by ChIP-seq, it may also be used to identify motifs associated with genetic elements obtained by other high-throughput assays such as ATAC-seq[31]. Bedtools v2.28.2 (getfasta sub-command) was used to extract superior temporal cortex ATAC-seq peak sequences in FASTA format from the hg19 reference genome obtained from the UCSC Genome Browser[32]. The ATAC-seq peak FASTA file was used as input to analysis with the command-line version of MEME-ChIP. MEME-ChIP executes two de novo motif discovery algorithms, multiple EM for motif elicitation (MEME) and discriminative regular expression motif elicitation (DREME). MEME can find relatively long motifs, while DREME discovers short motifs up to 8 bp and is more computationally efficient. In contrast to MEME, the DREME algorithm analyses all sequences. As a default, MEME-ChIP only performs motif discovery on the central 100 bp. In our analysis the –ccut parameter was set to 0 which indicates that the full length sequences should be analysed. We used the JASPAR 2018 Core vertebrates non-redundant database for motif comparison and otherwise default settings.

The discovered motifs are grouped by similarity to each other and compared to known motifs by the Tomtom algorithm. As part of the MEME-ChIP tool set FIMO uses the most significant motif in each cluster to scan the input sequence. We used the 25 de novo motifs (most significant motif in each cluster) with lowest E-value as a basis for further analysis. All of these motifs had been identified by DREME and were thus between 6 and 8 bp long. Due to the low information content in the short 6 bp motifs, FIMO found no matches passing the default p-value threshold of 1×10^{-4} when scanning the large input sequence. Bedtools v 2.28.2 was used to identify ATAC-seq peak subsets containing each of the de novo motifs[32].

The known motifs from the Jaspar database were generally longer than the de novo motifs they were matched to. Based on the assumption that the higher information content of the known motifs results in more accurate motif occurrences, we identified additional ATAC-seq peak subsets containing the best matched known motifs. Occurrences of the known motifs best matched (lowest Tomtom p-value) to the 25 most significant de novo motifs were identified with the command-line version of FIMO v 5.1.1. We used the default p-value threshold of 1×10^{-4} and the same Markov background model that was calculated from the input sequences by analysis with MEME-ChIP. As for the analysis of de novo motifs, Bedtools was used to identify ATAC-seq peak subsets containing each of the best matched known motifs.

Computations were performed on resources provided by UNINETT Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway. Figures comparing multiple motifs were created with the R/Bioconductor package MotifStack v1.18.0[33]. Motif matrices provided in the HOMER Motif Database, the Jaspar database and in the output from de novo motif discovery were used as input to MotifStack.

Results

Similarity measures of open chromatin between brain regions and cell types.

Pairwise intersections in terms of Jaccard statistics of ATAC-seq peaks representing OCRs in the different cell types show a separation between neurons and non-neurons, with the inter-region similarity being higher between the non-neurons (Supplementary Figure S1). Among the neurons, mediodorsal thalamus, putamen and nucleus accumbens differ the most from the other brain regions, while cortical regions cluster together. These results are in concordance with findings by Fullard et al., where differences were assessed between all individual samples using MDS clustering and pi1 estimates[20].

Enrichment of PD risk variants in neuronal open chromatin regions.

PD risk variants are significantly enriched in OCRs of neurons of the superior temporal cortex (GoShifter adj. $p = 0.028$, GREGOR adj. $p = 6.94 \times 10^{-05}$). There is a tendency that the lowest p-values, although not significant with both enrichment tests, are in cortical regions rather than subcortical regions (Table 1). This should be viewed in relation to the high inter-region similarity between OCRs in the different cortical regions (Supplementary Figure S1). There is no evidence of an enrichment of PEF risk variants or IBD risk variants in neurons from any of the tested brain regions (Supplementary Table S2). This indicates that the enrichment of risk variants in OCRs of neurons of the superior temporal cortex is specific to PD risk variants and not to disease-associated variants in general.

Enrichment of PD risk variants in open chromatin regions harboring specific transcription factor motifs identified by HOMER.

Candidate transcriptional regulators were assessed in OCRs in neurons of the superior temporal cortex, since this was the ATAC-seq peak set passing the significance threshold with both enrichment tests. We performed de novo motif discovery with the HOMER software and found that 22 motifs were enriched in open chromatin (Supplementary Table S3). ATAC-seq peaks were divided into 22 subsets containing each of the enriched motifs. We also created one subset with all the enriched motifs being absent (noMotif), which was intended as a negative control. HOMER compares the de novo motifs to a library of known motifs, presenting a list of the best matched known motifs based on a similarity score. The ATAC-seq peak subsets are named after the best matched known transcription factor.

When analysing HOMER motif-containing OCR subsets we found that PD risk variants were significantly enriched in OCRs harboring the de novo motif matched to the Olig2 motif (GoShifter adj. $p = 0.025$, GREGOR adj. $p = 1.39 \times 10^{-03}$) (Table 2). None of the other motif-containing OCR subsets were significantly enriched when both GREGOR and GoShifter were subjected to Bonferroni correction. There are however some OCR subsets that have an adjusted p-value < 0.05 with GREGOR and a nominally significant p-value with GoShifter (POL010.1_DCE_S_III, NRF1 and NFIA). There is a high degree of concordance between the highest ranked motif-containing OCR sets resulting from analysis with GoShifter and GREGOR. Also, none of the negative controls are enriched in the Olig2 OCR subset or any of the other motif-containing OCR subsets (Supplementary Table S4).

16 out of the 90 PD association signals have one or more variants in high LD located in OCRs containing the de novo motif best matched to oligodendrocyte transcription factor 2 (Olig2). Several other transcription

Cell type	GoShifter Adj. p-val (p-val)	GREGOR Adj. p-val (p-val)	No. ATAC-seq peaks
STC*	**0.028 (2.00×10^{-03})**	**6.94×10^{-05} (4.96×10^{-06})**	76145
VLPFC	0.162 (0.012)	**2.67×10^{-03} (1.90×10^{-04})**	86082
ITC	0.204 (0.015)	**4.30×10^{-04} (3.07×10^{-05})**	65346
PMC	0.206 (0.015)	**3.73×10^{-03} (2.66×10^{-04})**	84995
ACC	0.325 (0.023)	**7.20×10^{-04} (5.14×10^{-05})**	70654
OFC	0.403 (0.029)	**3.19×10^{-03} (2.28×10^{-04})**	81621
INS	0.468 (0.033)	**3.97×10^{-03} (2.84×10^{-04})**	68261
DLPFC	1 (0.075)	**0.021 (1.49×10^{-03})**	74825
PVC	1 (0.093)	**0.014 (1.02×10^{-03})**	51874
NAC	1 (0.105)	0.191 (0.014)	77290
MDT	1 (0.117)	**2.97×10^{-03} (2.12×10^{-04})**	69913
HIPP	1 (0.135)	**0.037 (2.66×10^{-03})**	80571
AMY	1 (0.151)	0.072 (5.16×10^{-03})	38564
PUT	1 (0.166)	0.145 (0.01)	100752

Table 1. Enrichment of PD risk variants within open chromatin regions in neurons from different brain regions. The cell type passing the significance threshold with both GoShifter and GREGOR is marked with a star and is reported in the text as significantly enriched with PD risk variants. We adjusted for multiple testing by Bonferroni correction, adjusting for 14 tests. Unadjusted p-values are provided in parenthesis. Adjusted p-val < 0.05 are written in bold. No. ATAC-seq peaks refers to the total number of peaks, representing open chromatin regions, in the analysed cell types. PD, Parkinson's disease; ACC, Anterior cingulate cortex; AMY, Amygdala; DLPFC, Dorsolateral prefrontal cortex; HIPP, Hippocampus; INS, Insula; ITC, Inferior temporal cortex; MDT, Mediodorsal thalamus; NAC, Nucleus Accumbens; OFC, Orbitofrontal cortex; PMC, Primary motor cortex; PUT, Putamen; PVC, Primary visual cortex; STC, Superior temporal cortex; VLPFC, Ventrolateral prefrontal cortex.

factors are also closely matched to this de novo motif since they share very similar binding motifs (Supplementary Figure S2). This provides additional candidates potentially targeting the enriched subset of open chromatin. All candidates do however belong to the basic Helix-Loop-Helix (bHLH) transcription factor family. The noMotif OCR subset does not show a significant enrichment of PD risk variants. This subset may however not be that well suited as a negative control since it is among the smallest OCR subsets, only constituting 1,6% of the total number of OCRs.

Enrichment of PD risk variants in open chromatin regions harboring specific transcription factor motifs identified by MEME-ChIP. Motif analysis with MEME-ChIP identified 88 de novo motifs (118 motifs clustered by similarity) to be enriched (E-value ≤ 0.05). Further analysis was limited to the 25 most significantly enriched motifs (Supplementary Table S5). Known motifs matched to the 25 most significant MEME-ChIP de novo motifs overlap several of the known motifs matched to HOMER de novo motifs (Supplementary Table S6). Transcription factors confidently matched to de novo motifs by both motif discovery tools have been described to function in neurons, such as MEF2C, SP2/SP1, NRF1 and NEUROD1/bHLH transcription factors[34–37]. We created ATAC-seq peak subsets containing each of the enriched de novo motifs. Seven de novo motifs that had no significant motif occurrences and six de novo motifs that had not been matched to any known motif (of which two had no significant motif occurrences) were excluded from further analysis. One additional subset was excluded since it was smaller than 1000 OCRs. This left 13 de novo motif-containing ATAC-seq peak subsets to be tested with enrichment analysis, of which none were significantly enriched with PD risk variants, nor with any of the negative controls (Supplementary Table S7).

Based on the assumption that known motifs matched to the short de novo motifs have a higher information content resulting in more accurate motif occurrences, ATAC-seq peaks were divided into subsets based on the location of the best matched known motifs. 19 out of the 25 de novo motifs with lowest E-value were matched to known motifs of which two were matched to the same known motif. Enrichment analysis of ATAC-seq peak subsets containing each of the 18 best matched known motifs show a significant enrichment of PD risk variants in the OCRs containing the neurogenic differentiation factor 1 (NEUROD1) motif (GoShifter adj. $p = 7.20 \times 10^{-03}$, GREGOR adj. $p = 7.63 \times 10^{-04}$) (Table 3). There is a high degree of concordance between the highest ranked motif-containing OCR sets resulting from analysis with GoShifter and GREGOR. Also, none of the negative controls are enriched in the NEUROD1 OCR subset or any of the other motif-containing OCR subsets (Supplementary Table S8). 13 out of the 90 PD association signals have one or more variants in high LD located in OCRs containing a NEUROD1 motif. NEUROD1 is a bHLH transcription factor and is interestingly among the ten known motifs best matched to the de novo motif located in the enriched ATAC-seq peak subset based on analysis with HOMER (Supplementary Figure S2).

Enrichment testing performed with exclusion of risk signals in the extended MHC region shows similar results in all analyses to those found when including this region. OCRs in superior temporal cortex neurons, OCR subsets containing motifs linked to Olig2 and OCRs containing the NEUROD1 motif were all significantly enriched with PD risk variants also when excluding the extended MHC region. No additional OCR sets were

Motif-containing OCR sets	GoShifter Adj. p-val (p-val)	GREGOR Adj. p-val (p-val)	No. ATAC-seq peaks
Olig2*	**0.025 (1.10×10^{-03})**	**1.39×10^{-03} (6.05×10^{-05})**	21924
POL010.1_DCE	0.064 (2.80×10^{-03})	**7.18×10^{-05} (3.12×10^{-06})**	37574
NRF1	0.407 (0.018)	**6.26×10^{-03} (2.72×10^{-04})**	7729
NFIA	0.580 (0.025)	**0.022 (9.51×10^{-04})**	37903
Sp2	1 (0.052)	0.147 (6.39×10^{-03})	7566
Egr2	1 (0.073)	0.103 (4.46×10^{-03})	25202
NFY	1 (0.078)	0.152 (6.63×10^{-03})	5806
PB0080.1_Tbp_1	1 (0.087)	0.774 (0.034)	5118
ETV2	1 (0.171)	0.113 (4.91×10^{-03})	10961
CTCF	1 (0.189)	1 (0.091)	6257
Mef2c	1 (0.205)	1 (0.050)	17566
PB0013.1_Eomes_1	1 (0.221)	1 (0.053)	29438
Atf1	1 (0.297)	0.331 (0.014)	5809
BORIS	1 (0.333)	1 (0.138)	6018
POL002.1_INR	1 (0.336)	1 (0.136)	34917
SPDEF	1 (0.390)	0.172 (0.007)	18884
MafF	1 (0.523)	1 (0.219)	34091
GFY	1 (0.683)	1 (0.543)	1196
Rfx5	1 (0.759)	1 (0.382)	7410
Fra1	1 (0.851)	1 (0.601)	9557
NFIL3	1 (0.901)	1 (0.723)	4431
Rfx1	1 (1)	1 (1)	3337
noMotif	1 (1)	0.542 (0.024)	1196

Table 2. Enrichment of PD risk variants within motif-containing open chromatin region sets identified with HOMER. The motif-containing OCR set passing the significance threshold with both GoShifter and GREGOR is marked with a star and is reported in the text as significantly enriched with PD risk variants. We adjusted for multiple testing by Bonferroni correction, adjusting for 23 tests. Unadjusted p-values are provided in parenthesis. Adjusted p-val < 0.05 are written in bold. No. ATAC-seq peaks refers to the total number of peaks, representing OCRs, in the analysed motif-containing OCR sets. The total number of ATAC-seq peaks in superior temporal cortex neurons is 76145. PD, Parkinson's disease; OCR, Open chromatin region.

significant with both GoShifter and GREGOR, and also no significant enrichments were found for any of the negative controls.

Analysis of open chromatin region subsets based on two different de novo motif discovery methods point to the same transcription factor family: bHLH transcription factors.

Analysis of OCR subsets based on de novo motif discovery with HOMER and MEME-ChIP both show a significant enrichment of PD risk variants in the subset targeted by bHLH transcription factors. The HOMER de novo motif matched to Olig2 and the MEME-ChIP de novo motif matched to NEUROD1 (with bHLH transcription factor motifs bhlha and TAL1::TCF3 as second and third best match) are highly similar. The similarity between these two de novo motifs, as well as between the de novo motifs and best matched known motifs, are illustrated in Fig. 2. bHLH transcription factors are known to bind to E-box motifs with the consensus sequence CANNTG, corresponding with the identified de novo motifs. In E-box motifs, the central two nucleotides and the surrounding nucleotides provide specificity of binding[34].

The PD association signals and corresponding proxy variants that overlap the NEUROD1 OCR subset and Olig2 OCR subset are listed in Supplementary Table S9. We find high concordance between PD risk variants overlapping the two enriched motif-containing OCR subsets. 12 out of the 13 association signals and 17 out of the 20 proxy variants that locate to the NEUROD1 OCR subset are also located in the Olig2 OCR subset (Supplementary Figure S3).

Discussion

Characterization of disease-related transcriptional networks is essential to improve our understanding of pathogenic processes and possible therapeutic targets. Identification of transcriptional networks that contribute to genetic risk mechanisms may be explored through integration of GWAS findings with epigenomic data and in silico motif analysis. This has been done in a recent study by Tansey et al., where results point to SPI1 and MEF2A/C transcriptional networks as central to Alzheimer's disease risk mechanisms[18]. In support of these findings, variants in the proximity of both SPI1 and MEF2C have earlier been identified as significant Alzheimer's disease risk loci[38,39]. Intriguingly, this suggests that a transcription factor may be implicated in genetic disease risk both by variants altering expression of the transcription factor itself, as well as through variants altering its binding affinity to regulatory DNA at other loci[18].

Motif-containing OCR sets	GoShifter Adj. p-val (p-val)	GREGOR Adj. p-val (p-val)	No. ATAC-seq peaks
NEUROD1*	**7.20×10^{-03} (4.00×10^{-04})**	**7.63×10^{-04} (4.24×10^{-05})**	12197
SP1	0.058 (3.20×10^{-03})	**5.79×10^{-06} (3.21×10^{-07})**	19373
ZNF263	0.472 (0.026)	**2.85×10^{-03} (1.59×10^{-04})**	27496
NHLH1	0.770 (0.043)	0.552 (0.031)	8288
TEAD2	1 (0.133)	0.097 (5.41×10^{-03})	7000
RBPJ	1 (0.137)	0.790 (0.044)	11637
KLF9	1 (0.206)	0.188 (0.010)	12364
NRF1	1 (0.254)	0.071 (3.92×10^{-03})	7740
SPIC	1 (0.261)	1 (0.075)	8737
SPIB	1 (0.291)	0.422 (0.023)	9300
ZIC1	1 (0.335)	0.221 (0.012)	6952
Stat5a::Stat5b	1 (0.457)	1 (0.140)	9783
ZNF384	1 (0.510)	1 (0.644)	14373
MEF2C	1 (0.557)	1 (0.279)	16923
FOSL2	1 (0.854)	1 (0.519)	9141
FOXP1	1 (0.888)	1 (0.919)	6847
TBP	1 (0.911)	1 (0.816)	5011
CREB1	1 (0.911)	1 (0.363)	2994

Table 3. Enrichment of PD risk variants within open chromatin region sets containing known motifs identified with MEME-ChIP. The motif-containing OCR set passing the significance threshold with both GoShifter and GREGOR is marked with a star and is reported in the text as significantly enriched with PD risk variants. We adjusted for multiple testing by Bonferroni correction, adjusting for 18 tests. Unadjusted p-values are provided in parenthesis. Adjusted p-val < 0.05 are written in bold. No. ATAC-seq peaks refers to the total number of peaks, representing OCRs, in the analysed motif-containing open chromatin sets. The total number of ATAC-seq peaks in superior temporal cortex neurons is 76145. PD, Parkinson's disease; OCR, Open chromatin region.

In our study, we integrated association signals from the most recent PD GWAS with publicly available ATAC-seq data coupled with transcription factor motif analysis in an effort to identify transcriptional networks contributing to PD risk. Enrichment analysis shows that PD risk variants are concentrated at sites of open chromatin in neurons of the superior temporal cortex indicating that these cell types mediate genetic risk for PD. The finding that neurons from additional cortical regions approach the significance threshold by being significant upon multiple testing with one enrichment test and nominally significant with the other enrichment test, suggests that a broader range of cortical regions are implicated in PD risk.

The involvement of transcriptional networks was explored in neurons of the superior temporal cortex based on the location of candidate motifs identified by de novo motif discovery. Enrichment analysis shows a significant overlap between PD risk variants and OCRs harboring motifs matched to transcription factors within a distinct family, suggesting that risk variants localize to specific transcription factor targeted OCRs. We find an enrichment of PD risk variants in OCRs targeted by bHLH transcription factors. There is a high degree of similarity between recognition motifs of members of the large bHLH transcription factor family, which provides several binding candidates. bHLH transcription factors are key determinants of neural cell fate specification and differentiation[34]. Many of the transcription factors that are candidates to target this subset of open chromatin are mainly expressed and function in the developing nervous system, and thus more likely to be involved in neurodevelopmental diseases. However, a developmental component to PD pathogenesis cannot be excluded, conceivably laying the groundworks for the brain's future vulnerability to or resilience against adult onset neurodegeneration[34]. Some bHLH transcription factors also function in adult neurons, such as transcription factor 4 (TCF4), which is the second best match to the de novo motif identified by HOMER[40]. Autosomal dominant mutations and deletions in TCF4 cause the neurodevelopmental disorder Pitt-Hopkins syndrome, while common variants at the TCF4 locus are associated with schizophrenia risk[41–44].

Epigenomic studies of the brain have predominantly been conducted in bulk tissue, which may perturb the detection of cell type specific regulatory elements due to measurement of an average signal across a heterogeneous population of cells. In contrast, Fullard et al. applied ATAC-seq to sorted nuclei[20]. This enables the distinction between OCRs in neurons vs non-neuronal cells, which we consider to be a major strength of this dataset.

We draw our conclusions based on results from two different enrichment tests. Due to the overlap between OCRs in the different cell types and motif-subsets, adjustment for multiple testing by Bonferroni correction may be considered to be a very strict significance threshold potentially leading to false negatives. This is mostly relevant to GoShifter, which is reported to have very conservative estimates[45]. It should however be noted that it is only the motif-containing OCR subset passing our set significance threshold which is also significant in analyses based on the alternative de novo motif discovery method. We consider it a strength of our study that we employ two different methods for de novo motif discovery. HOMER and MEME-ChIP are widely used tools for motif

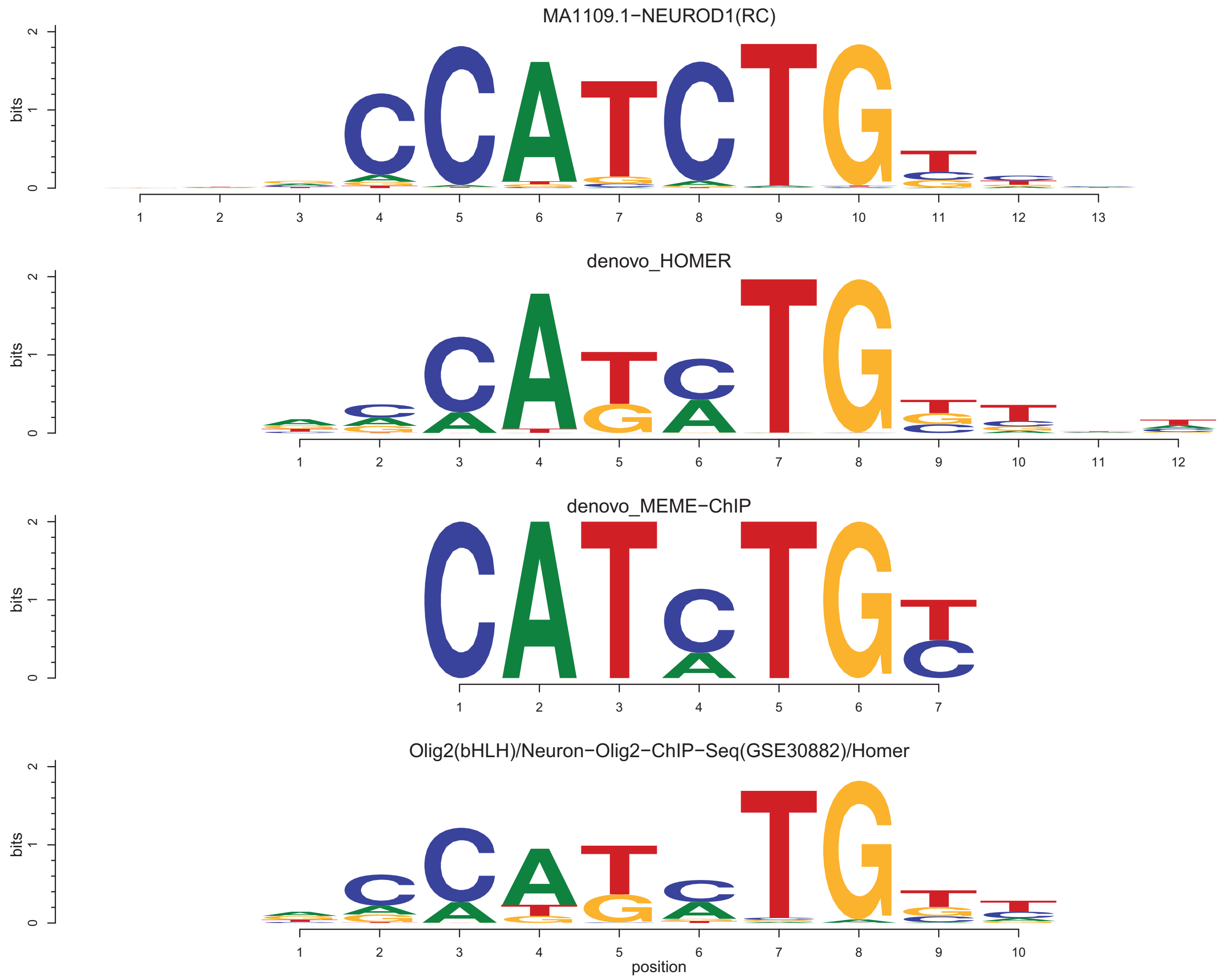


Figure 2. Comparison of de novo motifs matched to bHLH transcription factors. Denovo_HOMER is the de novo motif identified by HOMER, while denovo_MEME-ChIP refers to the de novo motif identified by MEME-ChIP. Olig2(bHLH)/Neuron-Olig2-ChIP-Seq (GSE30882)/Homer is the known motif best matched to denovo_HOMER and is part of the HOMER Motif Database. MA1109.1-NEUROD1 is the known motif best matched to denovo_MEME-ChIP and is part of the JASPAR 2018 Core vertebrates non-redundant database. bHLH, Basic Helix-Loop-Helix; RC, Reverse complement.

analysis of large DNA sequence data sets. The analyses are performed in parallel and both show an enrichment of PD risk variants in OCRs targeted by bHLH transcription factors, thus increasing the robustness of this finding.

We analyse two non-brain related disorders as negative controls and find no evidence of enrichment in cortical neurons, showing some degree of specificity of our findings to PD. In further interpretations of our results it is important to recognize that the detection of motifs and potential binding of bHLH transcription factors could be a marker of an active regulatory region also bound by other regulatory factors, of which one exerts the true causal effect on PD risk. We cannot exclude the possibility that an observed enrichment is due to unaccounted colocalization with other annotations. This limits the inference of causality and must be taken into account when interpreting results from enrichment analysis.

In our study, integration of GWAS signals with sites of open chromatin suggests that neurons in the superior temporal cortex and additional cortical regions mediate genetic risk for PD. Motif analysis performed in neurons of the superior temporal cortex shows that PD risk variants significantly overlap OCRs targeted by members of the bHLH transcription factor family, pointing to an involvement of these transcriptional networks in PD risk mechanisms. Additional investigations are needed to further explore the role of bHLH transcription factors in PD. Our study also demonstrates that ATAC-seq data coupled with motif analysis may be used in the assessment of hundreds of different transcription factors in a relevant cellular context, something that is not possible with existing transcription factor ChIP-seq data. Future studies addressing regulatory mechanisms in PD will benefit from improved computational approaches to predict transcription factor binding sites as a complement to ChIP-seq. Novel computational methods highlight the importance of both motif-based and chromatin accessibility features as pivotal to yield high performance predictions for most transcription factors[46,47]. Generation of epigenomic data with increased cellular resolution in brain related cell types would thus provide another valuable resource to study the involvement of transcription factors in neurodegenerative diseases.

Data availability

The datasets analysed during the current study are available from Brain Open Chromatin Atlas (BOCA) (https://bendlj01.u.hpc.mssm.edu/multireg/resources/boca_peaks.zip) and UCSC Genome Browser (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/latest/hg19.fa.masked.gz).

Received: 6 November 2019; Accepted: 26 January 2021
Published online: 10 February 2021

References

1. de Lau, L. M. & Breteler, M. M. Epidemiology of Parkinson's disease. *Lancet Neurol.* **5**, 525–535. https://doi.org/10.1016/s1474-4422(06)70471-9 (2006).
2. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102. https://doi.org/10.1016/s1474-4422(19)30320-5 (2019).
3. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195. https://doi.org/10.1126/science.1222794 (2012).
4. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA.* **106**, 9362–9367. https://doi.org/10.1073/pnas.0903103106 (2009).
5. Reynolds, R. H. *et al.* Moving beyond neurons: The role of cell type-specific gene regulation in Parkinson's disease heritability. *NPJ Parkinson's Dis.* **5**, 6. https://doi.org/10.1038/s41531-019-0076-6 (2019).
6. Coetzee, S. G. *et al.* Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology. *Sci. Rep.* **6**, 30509. https://doi.org/10.1038/srep30509 (2016).
7. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* **49**, 1511–1516. https://doi.org/10.1038/ng.3955 (2017).
8. Karczewski, K. J. *et al.* Systematic functional regulatory assessment of disease-associated variants. *Proc. Natl. Acad. Sci. USA.* **110**, 9607–9612. https://doi.org/10.1073/pnas.1219099110 (2013).
9. Cowper-Sal lari, R. *et al.* Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.* **44**, 1191–1198. https://doi.org/10.1038/ng.2416 (2012).
10. Claussnitzer, M. *et al.* FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907. https://doi.org/10.1056/NEJMoa1502214 (2015).
11. Deplancke, B., Alpern, D. & Gardeux, V. The genetics of transcription factor DNA binding variation. *Cell* **166**, 538–554. https://doi.org/10.1016/j.cell.2016.07.012 (2016).
12. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343. https://doi.org/10.1038/nature13835 (2015).
13. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein–DNA interactions. *Science* **316**, 1497–1502. https://doi.org/10.1126/science.1141319 (2007).
14. Harley, J. B. *et al.* Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707. https://doi.org/10.1038/s41588-018-0102-3 (2018).
15. Wang, J. *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812. https://doi.org/10.1101/gr.139105.112 (2012).
16. Inukai, S., Kock, K. H. & Bulyk, M. L. Transcription factor-DNA binding: Beyond binding site motifs. *Curr. Opin. Genet. Dev.* **43**, 110–119. https://doi.org/10.1016/j.gde.2017.02.007 (2017).
17. Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21**, 447–455. https://doi.org/10.1101/gr.112623.110 (2011).
18. Tansey, K. E., Cameron, D. & Hill, M. J. Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks. *Genome Med.* **10**, 14. https://doi.org/10.1186/s13073-018-0523-8 (2018).
19. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protocols Mol. Biol.* **109**, 21–29. https://doi.org/10.1002/0471142727.mb2129s109 (2015).
20. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28**, 1243–1252. https://doi.org/10.1101/gr.232488.117 (2018).
21. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. https://doi.org/10.1093/bioinformatics/bts635 (2013).
22. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. https://doi.org/10.1186/gb-2008-9-9-r137 (2008).
23. Khan, A. & Mathelier, A. Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinform.* **18**, 287. https://doi.org/10.1186/s12859-017-1708-7 (2017).
24. de Bakker, P. I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172. https://doi.org/10.1038/ng1885 (2006).
25. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005-d1012. https://doi.org/10.1093/nar/gky1120 (2019).
26. Tulloch, J. *et al.* Glia-specific APOE epigenetic changes in the Alzheimer's disease brain. *Brain Res.* **1698**, 179–186. https://doi.org/10.1016/j.brainres.2018.08.006 (2018).
27. Trynka, G. *et al.* Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. *Am. J. Hum. Genet.* **97**, 139–152. https://doi.org/10.1016/j.ajhg.2015.05.016 (2015).
28. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmüller, G. SNiPA: An interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334–1336. https://doi.org/10.1093/bioinformatics/btu779 (2015).
29. Schmidt, E. M. *et al.* GREGOR: Evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606. https://doi.org/10.1093/bioinformatics/btv201 (2015).
30. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004 (2010).
31. Ma, W., Noble, W. S. & Bailey, T. L. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat. Protoc.* **9**, 1428–1450. https://doi.org/10.1038/nprot.2014.083 (2014).
32. Quinlan, A. R. BEDTools: The Swiss-army tool for genome feature analysis. *Curr. Protocols Bioinform.* **47**, 11–12. https://doi.org/10.1002/0471250953.bi1112s47 (2014).
33. Ou, J., Wolfe, S. A., Brodsky, M. H. & Zhu, L. J. motifStack for the analysis of transcription factor binding site evolution. *Nat. Methods* **15**, 8–9. https://doi.org/10.1038/nmeth.4555 (2018).

34. Dennis, D. J., Han, S. & Schuurmans, C. bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res.* **1705**, 48–65. https://doi.org/10.1016/j.brainres.2018.03.013 (2019).
35. Ma, Q. & Telese, F. Genome-wide epigenetic analysis of MEF2A and MEF2C transcription factors in mouse cortical neurons. *Commun. Integr. Biol.* **8**, e1087624. https://doi.org/10.1080/19420889.2015.1087624 (2015).
36. Ryu, H. *et al.* Sp1 and Sp3 are oxidative stress-inducible, antideath transcription factors in cortical neurons. *J. Neurosci.* **23**, 3597–3606. https://doi.org/10.1523/jneurosci.23-09-03597.2003 (2003).
37. Dhar, S. S., Ongwijitwat, S. & Wong-Riley, M. T. Nuclear respiratory factor 1 regulates all ten nuclear-encoded subunits of cytochrome c oxidase in neurons. *J. Biol. Chem.* **283**, 3120–3129. https://doi.org/10.1074/jbc.M707587200 (2008).
38. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458. https://doi.org/10.1038/ng.2802 (2013).
39. Escott-Price, V. *et al.* Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. *PLoS ONE* **9**, e94661. https://doi.org/10.1371/journal.pone.0094661 (2014).
40. Jung, M. *et al.* Analysis of the expression pattern of the schizophrenia-risk and intellectual disability gene TCF4 in the developing and adult brain suggests a role in development and plasticity of cortical and hippocampal neurons. *Mol. Autism* **9**, 20. https://doi.org/10.1186/s13229-018-0200-1 (2018).
41. Consortium, S. W. G. o. t. P. G. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427. https://doi.org/10.1038/nature13595 (2014).
42. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747. https://doi.org/10.1038/nature08186 (2009).
43. Brockschmidt, A. *et al.* Severe mental retardation with breathing abnormalities (Pitt-Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. *Hum. Mol. Genet.* **16**, 1488–1494. https://doi.org/10.1093/hmg/ddm099 (2007).
44. Amiel, J. *et al.* Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* **80**, 988–993. https://doi.org/10.1086/515582 (2007).
45. Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* https://doi.org/10.1038/s41588-018-0322-6 (2019).
46. Li, H., Quang, D. & Guan, Y. Anchor: Trans-cell type prediction of transcription factor binding sites. *Genome Res.* **29**, 281–292. https://doi.org/10.1101/gr.237156.118 (2019).
47. Keilwagen, J., Posch, S. & Grau, J. Accurate prediction of cell type-specific transcription factor binding. *Genome Biol.* **20**, 9. https://doi.org/10.1186/s13059-018-1614-y (2019).

Acknowledgements

We thank the investigators and participants who have contributed to the generation of data made accessible in publicly available resources that was analysed in this study.

Author contributions

V.B.S., L.P. and M.T. designed the study. V.B.S. performed the analyses. V.B.S. drafted the manuscript. All authors were involved in project discussions and revision of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the South-Eastern Norway Regional Health Authority (Project no. 2016057 to MT), Research Council of Norway (Project no. 250597 to MT) and Norwegian Health Association (to LP).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-83087-2.

Correspondence and requests for materials should be addressed to M.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Integrative analysis identifies bHLH transcription factors as contributors to Parkinson's disease risk mechanisms

Victoria Berge-Seidl[1,2], Lasse Pihlstrøm[1], Mathias Toft[1,2*]

[1]Department of Neurology, Oslo University Hospital, Oslo, Norway

[2]Faculty of Medicine, University of Oslo, Oslo, Norway

* Corresponding author

This file includes:
Captions for Supplementary Table S1 – S9
Supplementary Table S2, S4, S7 and S8
Supplementary Figure S1 – S3

Supplementary Table S1. List of variants in high linkage disequilibrium with index variants representing independent genome-wide significant association signals in PD, IBD and PEF. For PD, all index variants passed analysis with the webserver Snipa with the exception of rs34637584 and rs76763715 due to not being in the reference set or population. The IBD index variants rs75900472 and rs144344067, and the PEF index variants rs9274247 and rs79412431 did not pass analysis for the same reason. QRSID is the query variant and defines the index variant, while RSID defines a proxy variant in high linkage disequilibrium ($r^2 > 0.8$) with the index variant. POS1 refers to the index variant position and POS2 is the proxy variant position. The degree of linkage disequilibrium between QRSID and RSID is provided as R-squared (R2) and DPRIME. PD, Parkinson's disease; IBD, Inflammatory bowel disease; PEF, Peak expiratory flow; MAF, Minor allele frequency.

Supplementary Table S2. Results from enrichment analysis of negative controls within open chromatin regions in brain neurons.

	PEF		**IBD**		
Cell type	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	No. ATAC-seq peaks
STC	1 (0.314)	0.93 (0.066)	1 (0.46)	1 (0.319)	76145
VLPFC	1 (0.424)	0.696 (0.05)	1 (0.627)	1 (0.481)	86082
ITC	1 (0.557)	1 (0.171)	1 (0.699)	1 (0.816)	65346
PMC	1 (0.482)	1 (0.074)	1 (0.949)	1 (0.9)	84995
ACC	1 (0.811)	1 (0.291)	1 (0.789)	1 (0.677)	70654
OFC	1 (0.666)	1 (0.184)	1 (0.866)	1 (0.649)	81621
INS	1 (0.747)	1 (0.306)	1 (0.366)	1 (0.406)	68261
DLPFC	1 (0.515)	1 (0.117)	1 (0.344)	1 (0.232)	74825
PVC	1 (0.215)	1 (0.099)	1 (0.722)	1 (0.637)	51874
NAC	1 (0.884)	1 (0.71)	1 (0.981)	1 (0.83)	77290
MDT	1 (0.193)	1 (0.169)	1 (0.828)	1 (0.679)	69913
HIPP	1 (0.616)	1 (0.282)	1 (0.797)	1 (0.705)	80571
AMY	1 (0.718)	1 (0.529)	1 (0.681)	1 (0.449)	38564
PUT	1 (0.874)	1 (0.419)	1 (0.959)	1 (0.577)	100752

There are no adj. p-val < 0.05 in any of the tested cell types. We adjusted for multiple testing by Bonferroni correction, adjusting for 14 tests. Unadjusted p-values are provided in parenthesis. No. ATAC-seq peaks refers to the total number of peaks, representing open chromatin regions, in the analysed cell types. PEF, Peak expiratory flow; IBD, Inflammatory bowel disease; ACC, Anterior cingulate cortex; AMY, Amygdala; DLPFC, Dorsolateral prefrontal cortex; HIPP, Hippocampus; INS, Insula; ITC, Inferior temporal cortex; MDT, Mediodorsal thalamus; NAC, Nucleus Accumbens; OFC, Orbitofrontal cortex; PMC, Primary motor cortex; PUT, Putamen; PVC, Primary visual cortex; STC, Superior temporal cortex; VLPFC, Ventrolateral prefrontal cortex.

Supplementary Table S3. Results from *de novo* motif analysis of open chromatin regions in superior temporal cortex neurons performed with HOMER

Supplementary Table S4. Results from enrichment analysis of negative controls within motif-containing open chromatin region sets identified with HOMER.

	PEF		**IBD**		
Motif-containing OCR sets	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	No. ATAC-seq peaks
Olig2*	1 (0.214)	1 (0.063)	1 (0.692)	1 (0.472)	21924
POL010.1_DCE	1 (0.293)	1 (0.089)	1 (0.571)	1 (0.507)	37574
NRF1	1 (0.436)	1 (0.663)	1 (0.820)	1 (0.563)	7729
NFIA	1 (0.510)	1 (0.178)	1 (0.371)	1 (0.309)	37903
Sp2	1 (0.861)	1 (0.972)	1 (0.918)	1 (0.592)	7566
Egr2	1 (0.223)	1 (0.148)	1 (0.926)	1 (0.682)	25202
NFY	1 (0.657)	1 (0.893)	1 (0.598)	1 (0.722)	5806
PB0080.1_Tbp_1	1 (0.302)	1 (0.109)	1 (1)	1 (1)	5118
ETV2	1 (0.252)	1 (0.235)	1 (0.734)	1 (0.369)	10961
CTCF	1 (0.817)	1 (0.338)	1 (0.258)	1 (0.161)	6257
Mef2c	1 (0.237)	0.239 (0.010)	1 (0.942)	1 (0.722)	17566
PB0013.1_Eomes_1	1 (0.756)	1 (0.448)	1 (0.330)	1 (0.221)	29438
Atf1	1 (0.653)	1 (0.645)	1 (0.341)	1 (0.180)	5809
BORIS	1 (0.874)	1 (0.444)	1 (0.371)	1 (0.623)	6018
POL002.1_INR	1 (0.408)	1 (0.161)	1 (0.884)	1 (0.959)	34917
SPDEF	1 (0.976)	1 (0.883)	1 (0.748)	1 (0.383)	18884
MafF	1 (0.967)	1 (0.840)	1 (0.674)	1 (0.495)	34091
GFY	1 (0.261)	1 (0.415)	1 (1)	1 (1)	1196
Rfx5	1 (0.300)	0.901 (0.039)	1 (0.725)	1 (0.522)	7410
Fra1	1 (0.592)	1 (0.119)	1 (0.665)	1 (0.296)	9557
NFIL3	1 (0.628)	1 (0.581)	1 (1)	1 (1)	4431
Rfx1	1 (0.841)	1 (0.518)	1 (1)	1 (1)	3337
noMotif	1 (1)	1 (1)	1 (1)	1 (1)	1196

There are no adj. p-val < 0.05 in any of the motif-containing OCR sets. We adjusted for multiple testing by Bonferroni correction, adjusting for 23 tests. Unadjusted p-values are provided in parenthesis. No. ATAC-seq peaks refers to the total number of peaks, representing OCRs, in the analysed motif-containing OCR sets. The total number of ATAC-seq peaks in superior temporal cortex neurons is 76145. PEF, Peak expiratory flow; IBD, Inflammatory bowel disease; OCR, Open chromatin region.

Supplementary Table S5. The 25 most significant results from *de novo* motif analysis of open chromatin regions in superior temporal cortex neurons performed with MEME-ChIP. The three most similar known motifs are listed. Only known motifs with a TOMTOM similarity E-value of less than 1.0 to the discovered motif are shown.

Supplementary Table S6. Known motifs matched to *de novo* motifs identified with HOMER and MEME-ChIP. Known motifs with a similarity score of 0.70 and higher to the 22 *de novo* motifs discovered with HOMER, known motifs with TOMTOM similarity E-value of less than 1.0 to the 25 most significant *de novo* motifs discovered with MEME-ChIP, and known motifs matched to both HOMER and MEME-ChIP *de novo* motifs are listed. Known motifs matched to *de novo* motifs identified with HOMER are either from Jaspar motif database (J), Homer motif database (H), or from both. All known motifs matched to *de novo* motifs identified with MEME-ChIP are from Jaspar motif database. A known motif may be matched to more than one *de novo* motif, but is only listed once.

Supplementary Table S7. Results from enrichment analysis of PD risk variants and negative controls in open chromatin region sets containing *de novo* motifs identified with MEME-ChIP.

	PD		**PEF**		**IBD**		
Motif – containing OCR sets	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	No. ATAC-seq peaks
SP1	0.403 (0.031)	**4.07×10^{-03} (3.13×10^{-04})**	1 (0.693)	1 (0.793)	1 (0.917)	1 (0.656)	11838
ZNF263	0.579 (0.045)	**1.69×10^{-03} (1.30×10^{-04})**	1 (0.838)	1 (0.724)	1 (0.986)	1 (0.868)	17280
RBPJ	1 (0.082)	1 (0.088)	1 (0.954)	1 (0.927)	1 (0.602)	1 (0.651)	8710
NEUROD1_2	1 (0.090)	0.097 (7.50×10^{-03})	1 (0.760)	1 (0.480)	1 (0.928)	1 (0.614)	9948
SPIB	1 (0.110)	0.673 (0.052)	1 (0.829)	1 (0.324)	1 (0.361)	1 (0.238)	12998
TBP	1 (0.263)	1 (0.333)	1 (0.511)	1 (0.121)	1 (0.347)	1 (0.222)	10679
KLF9	1 (0.425)	0.323 (0.025)	1 (0.901)	1 (0.856)	1 (0.976)	1 (0.945)	8071
NEUROD1_1	1 (0.548)	1 (0.469)	0.689 (0.053)	0.793 (0.061)	1 (1)	1 (1)	6772
NRF1	1 (0.589)	1 (0.103)	1 (1)	1 (1)	1 (0.873)	1 (0.656)	3538
NHLH1	1 (0.620)	1 (0.225)	1 (0.490)	1 (0.470)	1 (0.833)	1 (0.501)	9684
FOSL2	1 (0.640)	1 (0.170)	1 (0.568)	1 (0.087)	1 (0.650)	1 (0.321)	9021
MEF2C	1 (0.647)	1 (0.945)	1 (0.458)	1 (0.116)	1 (0.305)	1 (0.367)	11541
ZNF384	1 (0.726)	1 (0.195)	1 (0.802)	1 (0.316)	1 (0.220)	1 (0.501)	16464

The motif-containing OCR sets are named after the best matched known motif. We adjusted for multiple testing by Bonferroni correction, adjusting for 13 tests. Unadjusted p-values are provided in parenthesis. Adjusted p-val < 0.05 are written in bold. No. ATAC-seq peaks refers to the total number of peaks, representing OCRs, in the analysed motif-containing OCR sets. The total number of ATAC-seq peaks in superior temporal cortex neurons is 76145. PD, Parkinson’s disease; PEF, Peak expiratory flow; IBD, Inflammatory bowel disease; OCR, Open chromatin region.

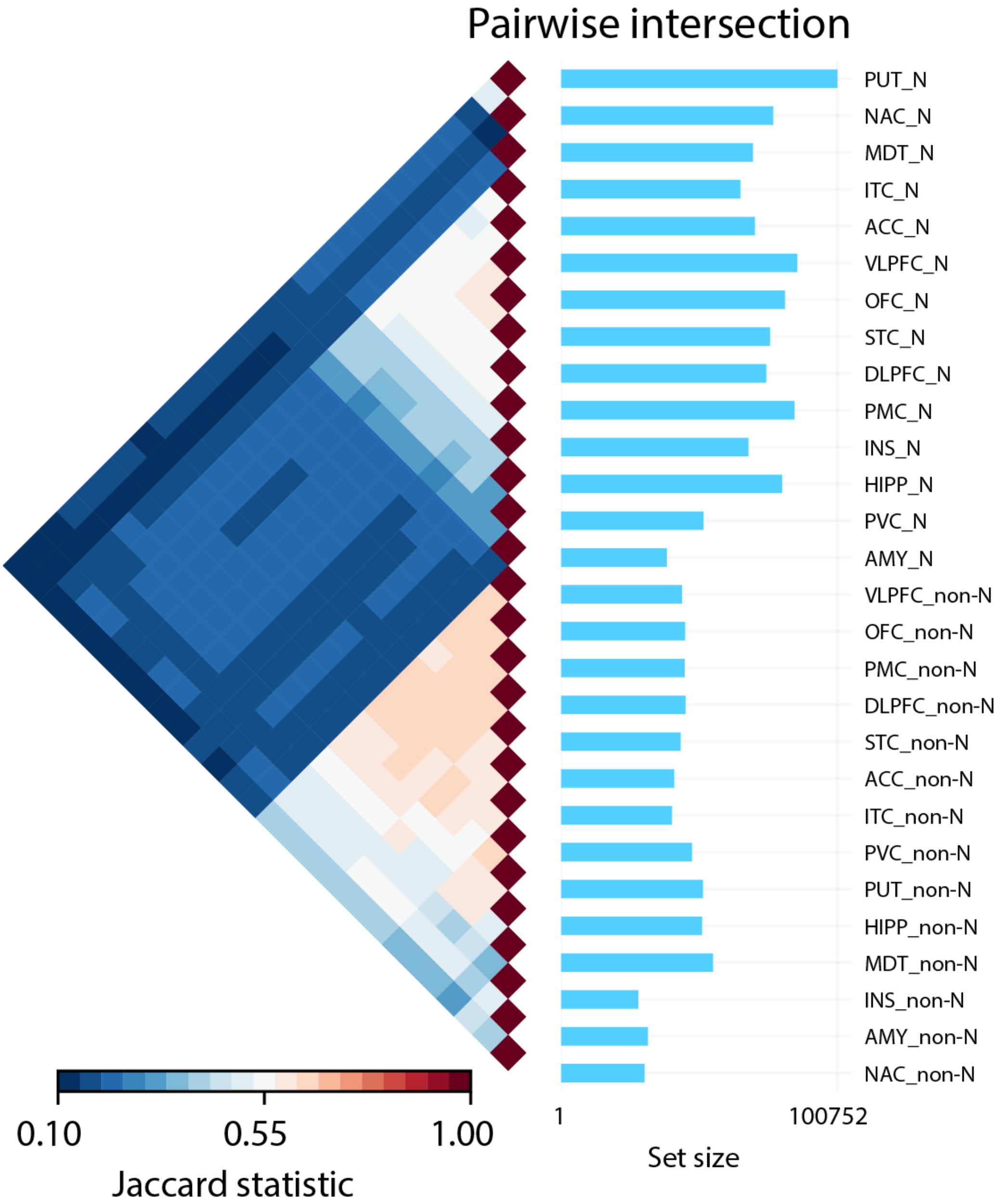
Supplementary Table S8. Results from enrichment analysis of negative controls within open chromatin region sets containing known motifs identified with MEME-ChIP.

	PEF		IBD		
Motif-containing OCR sets	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	GoShifter adj. p-val (p-val)	GREGOR adj. p-val (p-val)	No. ATAC-seq peaks
NEUROD1	1 (0.244)	1 (0.359)	1 (0.992)	1 (0.977)	12197
SP1	1 (0.691)	1 (0.867)	1 (0.900)	1 (0.553)	19373
ZNF263	1 (0.059)	0.256 (0.014)	1 (0.560)	1 (0.257)	27496
NHLH1	1 (0.925)	1 (0.904)	1 (0.862)	1 (0.671)	8288
TEAD2	1 (0.927)	1 (0.455)	1 (0.971)	1 (0.899)	7000
RBPJ	1 (0.846)	1 (0.578)	1 (0.351)	1 (0.296)	11637
KLF9	1 (0.164)	1 (0.241)	1 (0.962)	1 (0.869)	12364
NRF1	1 (0.742)	1 (0.852)	1 (0.688)	1 (0.142)	7740
SPIC	1 (1)	1 (1)	1 (0.234)	1 (0.126)	8737
SPIB	1 (1)	1 (1)	1 (0.603)	1 (0.433)	9300
ZIC1	1 (0.675)	1 (0.707)	1 (0.790)	1 (0.784)	6952
Stat5a::Stat5b	1 (0.540)	1 (0.457)	1 (0.561)	1 (0.648)	9783
ZNF384	1 (0.500)	1 (0.172)	1 (0.478)	1 (0.549)	14373
MEF2C	1 (0.432)	0.555 (0.031)	1 (0.974)	1 (0.901)	16923
FOSL2	1 (0.392)	1 (0.122)	1 (0.655)	1 (0.356)	9141
FOXP1	1 (0.907)	1 (0.801)	1 (0.681)	1 (0.485)	6847
TBP	1 (0.433)	1 (0.103)	1 (0.562)	1 (0.410)	5011
CREB1	1 (0.546)	1 (0.174)	1 (0.463)	1 (0.171)	2994

There are no adj. p-val < 0.05 in any of the motif-containing OCR sets. We adjusted for multiple testing by Bonferroni correction, adjusting for 18 tests. Unadjusted p-values are provided in parenthesis. No. ATAC-seq peaks refers to the total number of peaks, representing OCRs, in the analysed motif-containing OCR sets. The total number of ATAC-seq peaks in superior temporal cortex neurons is 76145. PEF, Peak expiratory flow; IBD, Inflammatory bowel disease; OCR, Open chromatin region.

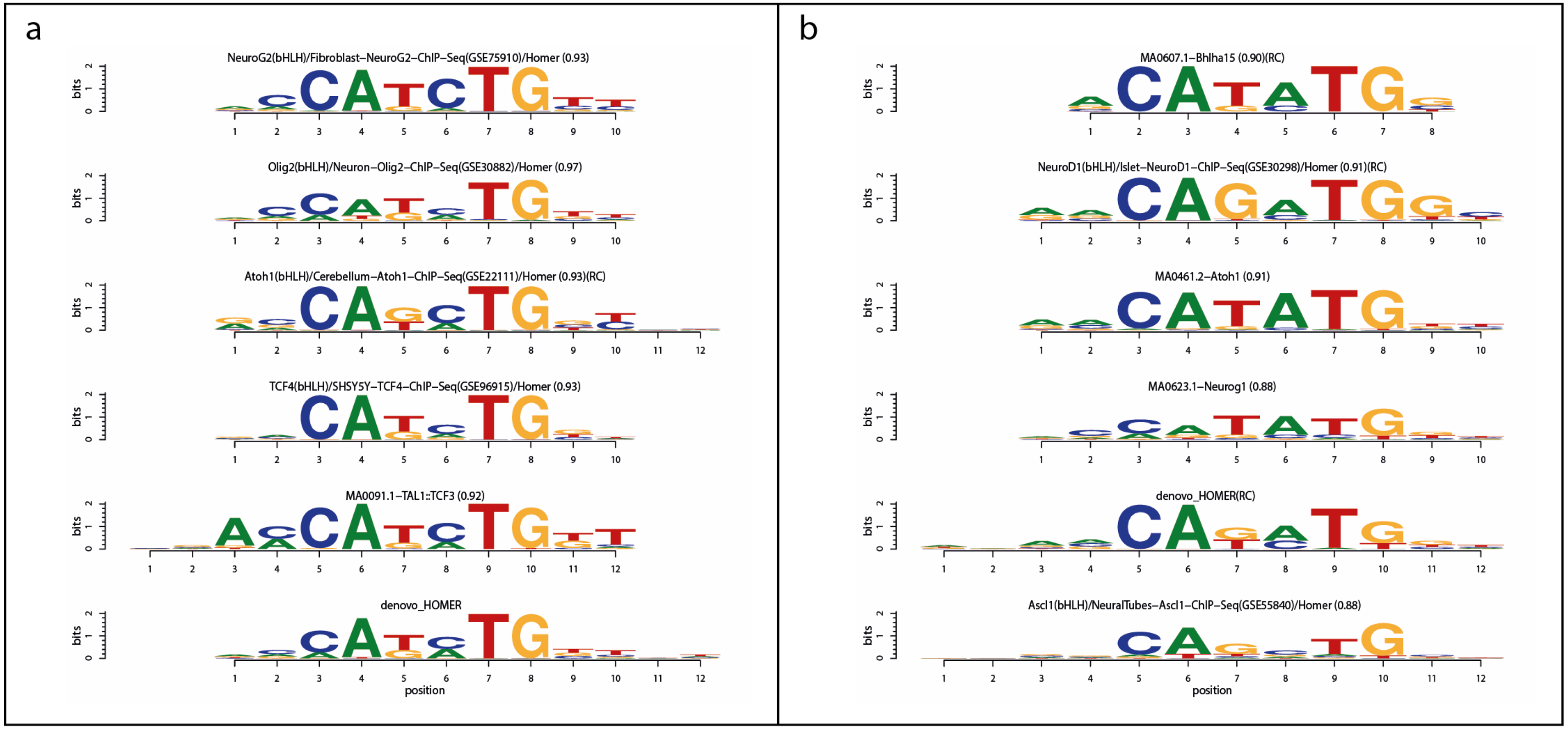
Supplementary Table S9. PD association signals and proxy variants that overlap open chromatin region subsets targeted by bHLH transcription factors. PD association signals (represented by the top-hit variant) and proxy variants that overlap the OCR subset containing the HOMER *de novo* motif best matched to Olig2 and the OCR subset containing the NEUROD1 motif identified by MEME-ChIP are listed. The overlap has been identified by analysis with GoShifter. PD association signals and proxy variants that overlap both the Olig2 OCR subset and the NEUROD1 OCR subset are written in bold. PD, Parkinson's disease; bHLH, Basic Helix-Loop-Helix; OCR, Open chromatin region.

Supplementary Figure S1. Heatmap of pairwise intersections of Jaccard statistic of open chromatin regions in brain.



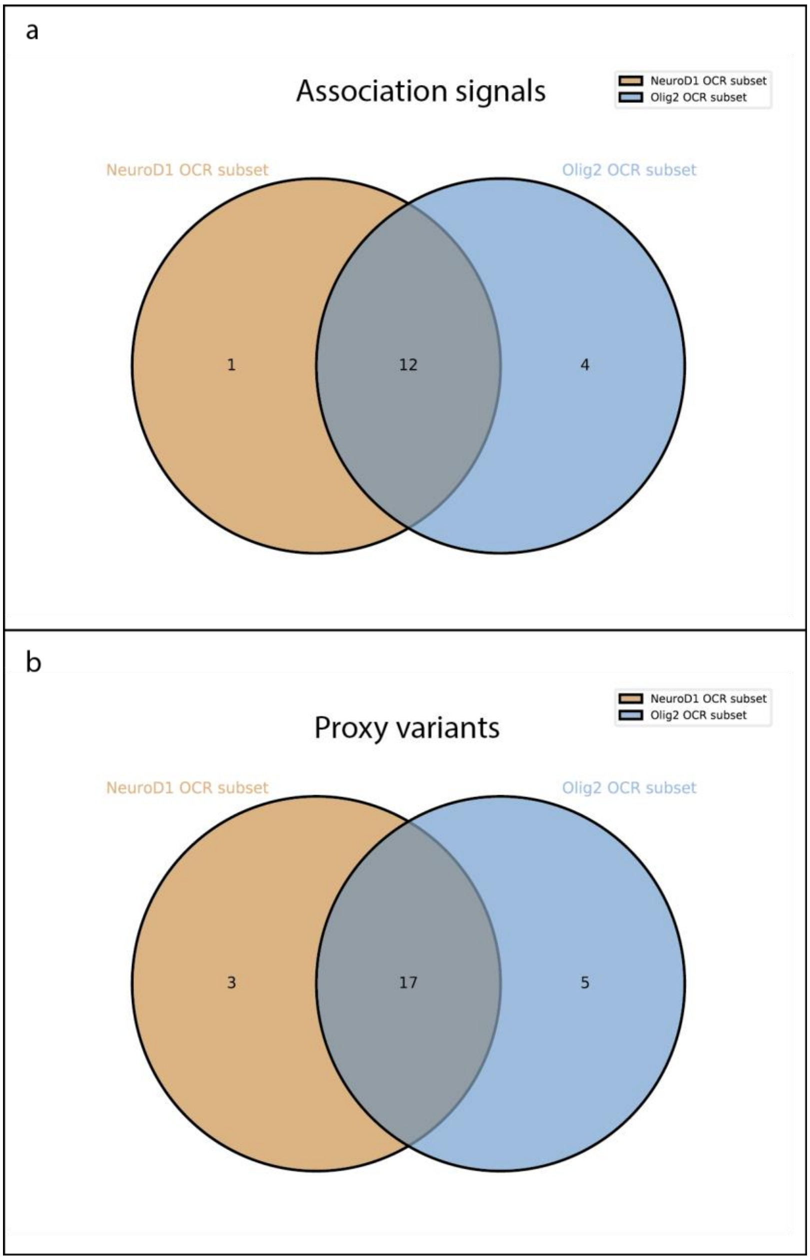
The set size refers to the number of open chromatin regions in each dataset. N, Neuronal; non-N, non-Neuronal; ACC, Anterior cingulate cortex; AMY, Amygdala; DLPFC, Dorsolateral prefrontal cortex; HIPP, Hippocampus; INS, Insula; ITC, Inferior temporal cortex; MDT, Mediodorsal thalamus; NAC, Nucleus Accumbens; OFC, Orbitofrontal cortex; PMC, Primary motor cortex; PUT, Putamen; PVC, Primary visual cortex; STC, Superior temporal cortex; VLPFC, Ventrolateral prefrontal cortex.

Supplementary Figure S2. Known motifs matched to the HOMER *de novo* motif located in the open chromatin region subset enriched with Parkinson’s disease risk variants



a) *De novo* motif compared to the five known motifs with highest similarity scores. b) *De novo* motif compared to known motifs with 6th – 10th highest similarity scores. The similarity score is provided in parenthesis after the motif name. RC, Reverse complement.

Supplementary Figure S3. Overlap between PD risk variants that locate to the two enriched motif-containing open chromatin region subsets



a) Venn diagram illustrating the overlap between PD association signals that locate to the NEUROD1 OCR subset and the Olig2 OCR subset, showing that 12 association signals locate to both subsets. b) Venn diagram illustrating the overlap between proxy variants that locate to the NEUROD1 OCR subset and the Olig2 OCR subset, showing that 17 proxy variants locate to both subsets. Venn diagrams have been created with the software Intervene. PD, Parkinson’s disease; OCR, Open chromatin region.