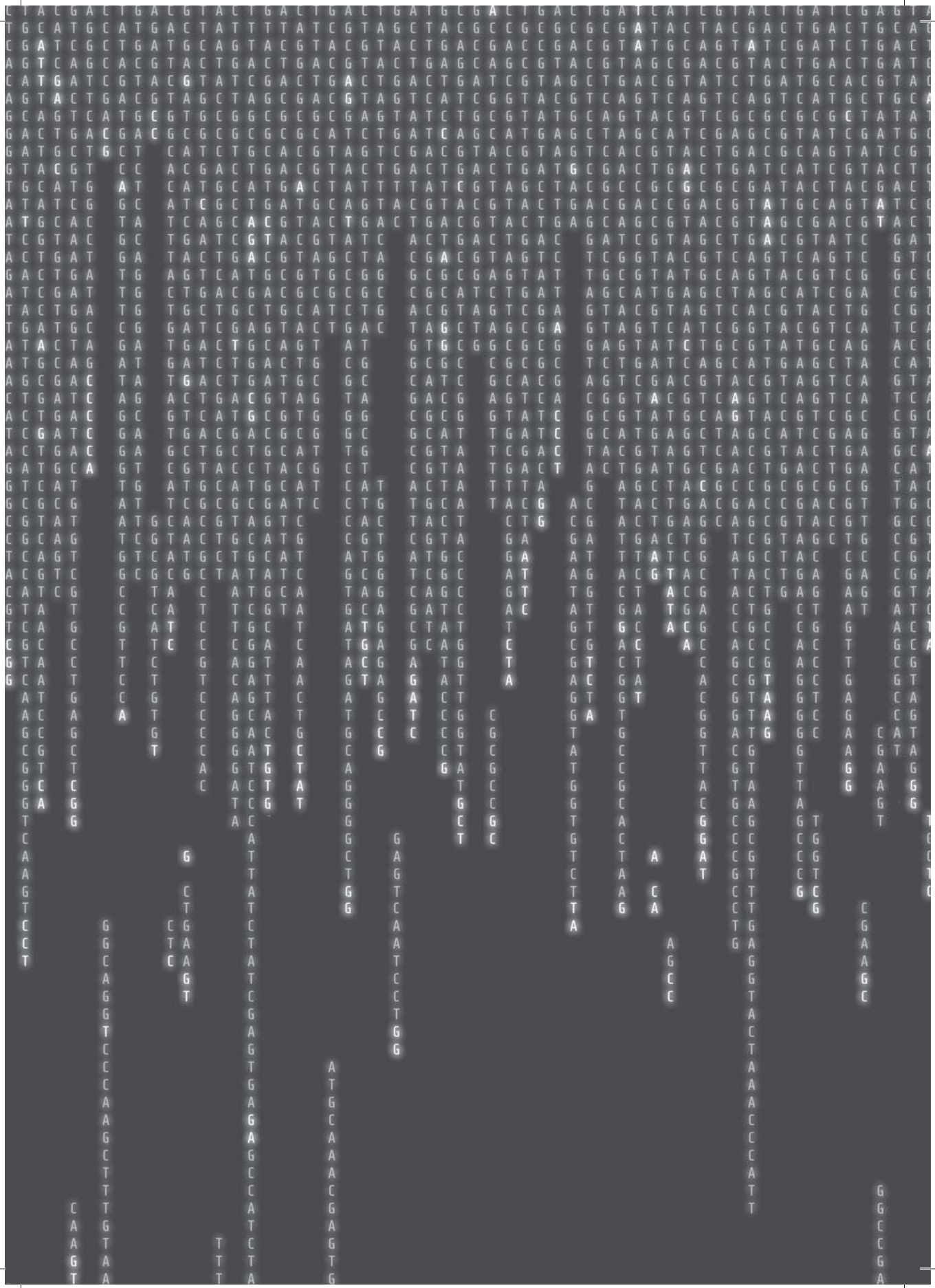
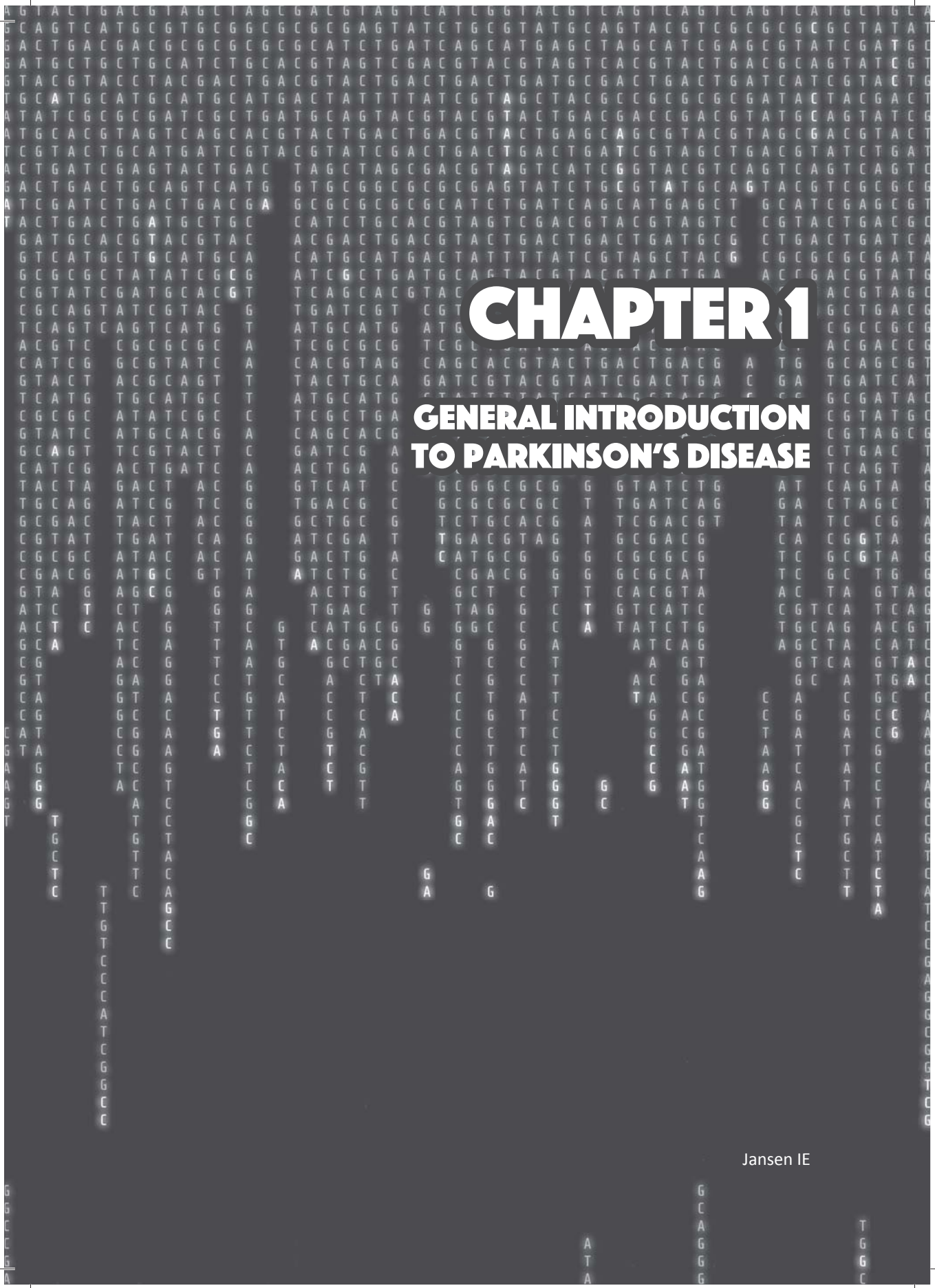
Cracking the code-ing sequence for Parkinson’s disease

Jansen, I.E.

2017

document version
Publisher's PDF, also known as Version of record





CHAPTER 1

GENERAL INTRODUCTION TO PARKINSON'S DISEASE

Jansen IE

Parkinson's disease (PD) is a progressive neurodegenerative disorder, which is typically recognized by the public as the movement disease of the elderly. Indeed the risk to develop PD increases with age, but PD is also diagnosed in younger individuals. The prevalence ranges from 0.04% for the age group between 40 and 49 years, to 1.9% for individuals over 80 years.[1] PD is a complex disorder, which is reflected in the variety of symptoms, pathologies and genetic causes. These heterogeneous features complicate the development of disease reversing treatment, which is the ultimate goal of the PD research field.

Heterogeneous features of Parkinson's disease

Clinical symptoms characterizing PD are heterogeneous and encompass both the classical motor symptoms and the more recently recognized non-motor symptoms, which frequently precede the motor features by a decade. The motor symptoms consist of bradykinesia (extreme slowness of movements and reflexes), rigidity (stiffness), resting tremor and postural disturbances.[2] Based on the motor symptoms, two clinical subtypes are suggested that separate a tremor-dominant type and a non-tremor-dominant type.[3] The first subtype comes with a slower progression rate and little functional impairment compared to the non-tremor-dominant type. The second type is mainly defined by rigidity and postural deteriorations. In addition to the motor symptoms, various non-motor symptoms hinder the quality of the patient's life.[4] The non-motor symptoms often present themselves before diagnosis, and are initially not suggestive for a diagnosis of PD. Prodromal features are constipation, sleep disorders, olfactory dysfunction and depression, which continue to evolve once the diagnosis has been determined.[5] Additional non-motor symptoms post diagnosis include fatigue, low blood pressure, urinary symptoms and mild cognitive impairment that potentially progresses into dementia.[3,6] The premotor phase ranges from years to decades.[7,8] Extensive research of the prodromal period has the potential to identify prodromal biomarkers that could enable the PD diagnosis in an earlier stage before the onset of motor symptoms, which might be beneficial for treatment of the disease.[5]

Besides the division of clinical subtypes based on motor symptoms, other subtypes are based on the age of onset. PD patients with a young onset (cutoff ranges from 40 to 55 years[9-11]) comprise between 5% and 10% of the PD population.[10,11] The main overall difference between young-onset Parkinson's disease (YOPD) and typical PD is the slower disease progression in the younger patients.[12,13] Furthermore, YOPD has been associated with increased levodopa-induced motor complications[14], earlier motor impairments (e.g. dyskinesia and dystonia)[14,15], increased risk of psychiatric symptoms[4,16-18] and less cognitive decline[15,19]. As distinct outcomes have been reported on differences in non-motor symptoms between YOPD and classical PD,[11,20] more extensive research is required.

Two main pathological hallmarks characterize PD. The loss of dopaminergic neurons in the Substantia Nigra pars compacta (SNpc) was the first macroscopic pathological feature observed for PD.[21,22] Neurodegeneration in this part of the midbrain results in basal ganglia impairments causing the motor symptoms in PD patients.[23,24] At the moment of clinical onset, more than 50% of the dopaminergic neurons in the SNpc are degenerated.[25,26] Although neuronal loss is most prominent in the SNpc, other brain regions are also affected such as the glossopharyngeal and vagal nerves, olfactory bulb, locus coeruleus, medulla, mesocortex and neocortex.[27] The neuronal degeneration in these regions is accompanied by the second microscopic pathological hallmark: Lewy bodies.

Lewy bodies (LB) are abnormal protein aggregates observed in the cytoplasm of neuronal cells in PD patients.[28] LB are a characteristic pathologic feature of PD, although a small proportion of post-mortem brain tissue of PD patients lack LB-pathology (e.g. mutation carriers of the parkinsonism gene *parkin*[29,30]), suggesting additional undiscovered forms of protein aggregates. The primary component of LB is α-synuclein, which is a small protein (140 amino acid residues weighting ~40 kDA) encoded by *SNCA* that is mainly expressed in brain and, in healthy state, has its role in synaptic function and localizes to the presynaptic terminals.[31] Increased levels of α-synuclein causes a migration of the protein to the cytoplasm and generates misfolded protein inclusions in the cell body (Lewy bodies) and in axons or dendrites (Lewy neurites).[32] Distinct aggregate types exist, ranging from thread-like α-synuclein structures through soluble monomer-composing oligomers.[33] Although the traditional theory viewed the large insoluble fibrillary protein aggregates as the toxic components inducing neurodegeneration, more recent research proposes that the smaller soluble oligomers are the type causing cytotoxicity.[34] It is currently unknown which specific type of α-synuclein aggregates are the toxic species.

α-synuclein aggregates are observed in the SNpc, but are also present in other brain regions lacking macroscopic changes.[35] Research on the distribution of LB pathology throughout the brain proposed 6 different stages of neurodegeneration following a specific anatomical route, currently well known as Braak staging.[27] The first inclusions appear in two specific cranial nerves (peripheral nervous system) or the olfactory bulb, followed by appearance in the medulla and pons in stage 2. At stage 3, α-synuclein depositions are for the first time observed in the SNpc and amygdala. The pathology progresses to the cortex in the subsequent stages, affecting the temporal cortex in the fourth and neocortex in the fifth and sixth stage. The pathological progression is accompanied by clinical progression matching the affected brain regions, explaining the clinical course of symptoms. Olfactory dysfunction or sleeping problems could be accounted for by the degeneration of regions in the first stages. The classical motor symptom manifestation is parallel in time with stage 3 midbrain α-synuclein inclusions. The affection of the cortex in the later stages presumably

contributes to cognitive impairment observed in the later phases of the clinical disease progression.[36,37]

Recent studies have investigated the pathological progression from neuron to neuron in distinct brain regions. The first hint originates from observations of research transplanting healthy fetal dopamine cells into the striatum of PD patients.[38,39] Pathological examination of these brains after one or two decades revealed that the grafted neurons also developed LB pathology, suggesting prion-like transmission of α-synuclein inclusions. This hypothesis led to multiple in-vivo models showing the transmission of injected α-synuclein fibrils within various brain regions to previously unaffected other brain regions in transgenic *SNCA* or wildtype mice.[40-42] Furthermore, a macaque model, a species closer related to human, showed similar findings by injecting α-synuclein inclusion from postmortem brain tissue of PD patients into the SNpc or striatum of macaque.[43] Progressive nigrostriatal neurodegeneration upon LB extracts inoculations was observed. Although these animal models support the hypothesis of a prion-like transmission of α-synuclein aggregates, novel techniques relying on direct observations of the total time-course of progression in PD patients (instead of snapshots in animal models) are required to allow definite conclusions about this topic.[44]

Diagnosis of PD is predominantly based on clinical symptoms, but only definite after pathological confirmation based on post-mortem brain tissue. The diagnostic criteria have recently been updated resulting in the Movement Disorder Society Clinical Diagnostic Criteria for Parkinson's Disease.[45] Although the main diagnostic criterion continues to be the motor symptoms, the non-motor symptoms are increasingly emphasized. However, for this thesis the earlier criteria were used for the diagnosis of PD patients. Until October 2015, the general consensus was to use the UK Parkinson's Disease Society Brain Bank criteria[2], of which the highest sensitivity rate is estimated to be at 90%.[46] The patient should present with bradykinesia and one other motor symptom, including mascular rigidity, resting tremor or postural instability. A response to levodopa, a drug that increases the dopamine level, strengthens the diagnosis.[47] Other differential diagnoses should be excluded considering diseases such as multiple system atrophy, progressive supranuclear palsy, corticobasal degeneration, essential tremor, Alzheimer's disease and vascular parkinsonsim. The latter three diseases are not caused by a loss of neurons in the SNpc.

Positron emission tomography (PET) and single photon emission computed tomography (SPECT) are imaging techniques that help differentiate between PD and diseases not related to SNpc degeneration, as they measure the density of presynaptic dopaminergic terminals within the striatum.[48] Longitudinal studies that monitored newly diagnosed PD patients without evidence of dopaminergic deficits (accessed by PET or SPECT) suggest PET and SPECT to be valuable diagnostic tools as the majority of these patients would not develop PD.[49] However, PET and SPECT only detect substantial

dopaminergic loss. It would be preferable to apply a diagnostic tool that could observe PD-related symptomatic or pathologic changes earlier in the disease process still leaving the possibility to slow or alter the progressive disease. As the olfactory bulb is one of the first affected brain regions, accompanied by a loss of smell, the University of Pennsylvania smell identification test is used to improve earlier diagnosis.[50]

Despite the absence of generally accepted pathological criteria,[35] the standard is to confirm clinical diagnosis pathologically in post-mortem brain tissue by the presence of the two discussed pathological hallmarks: neurodegeneration in the SNpc and Lewy body pathology in surviving dopamine neurons, and the absence of pathological evidence of other diseases producing similar clinical motor symptoms.[51] The acknowledgement of non-traditional features of PD that are not standard accounted for in diagnosis, such as genetically confirmed PD patients without Lewy Body pathology or the importance of non-motor symptoms, emphasizes the necessity for continuous improvement of the diagnostic criteria.[52] For future improvements, early diagnosis before the onset of motor symptoms in the prodromal phase would be most effective for drugs aiming to slow or alter underlying neurodegenerative processes.

Although the development of such therapies has a high priority in ongoing PD research, the current drugs only treat the symptoms, thereby improving the quality of the patient's life. Treatment that diminish motor symptoms increase brain dopamine levels or stimulate dopamine receptors, of which the most commonly administered drug is levodopa.[53] As levodopa is accompanied by side-effects causing motor complications, such as dyskinesia and motor fluctuations, the administration of levodopa is postponed until the patient experiences discomfort. Other motor symptom medications encompass dopamine agonists, monoamine oxidase type B inhibitors, and, less commonly, amantadine targeting tremors, rigidity and bradykinesia.[53,54] Besides the therapies addressing motor symptoms, a limited number of treatments for non-motor symptoms are also available aiming to diminish sleeping problems, psychiatric symptoms and autonomic malfunctioning.[53,55] As last, surgical treatment is another therapy option if the motor complications of levodopa become disabling in moderately to severely progressed PD (10 to 13 years post PD diagnosis).[56] Deep brain stimulation of the subthalamic nucleus or globus pallidus internus could benefit both motor symptoms and non-motor symptoms.[57]

As implied by the heterogeneity of the clinical symptoms and the pathological features: PD is a multifactorial heterogeneous disease. The different clinical PD subtypes are hypothesized to originate from distinct pathogenesis.[58] The underlying affected biological mechanisms contributing to PD pathogenesis range from synaptic function to mitochondrial maintenance.[59] Therefore, targeting specific molecular defects in biological pathways with therapeutic strategies in a patient-specific manner might be the most effective treatment. Fundamental research establishing the causal factors that increase the risk to develop PD is a valuable source to improve our understanding of underlying

dysfunctional biological mechanisms. The identification and characterization of genetic defects, as causal factors for PD, will continue to have an important influence on the development of PD-related therapies.

Genetic etiology of PD

Another heterogeneous feature of PD is its genetic etiology. The heritability of PD is estimated to be around 40%.[9,60,61] The discovery of PD genes followed a similar trend as observed for the general progression of genetic and genomics techniques (Box A). First, family-based linkage analyses and population-based genome-wide association studies (GWAS) dominated the genetic research field for decades, which was followed by the application of next generation sequencing (NGS) methods during the last decade. The first gene to be linked to PD was *SNCA,* encoding for the main protein component of Lewy bodies, of which rare and highly penetrant mutations were shown to aggregate with the disease within families showing autosomal dominant inheritance patterns.[62] Subsequent research showed both missense mutations and exonic multiplications to be the causative defects.[63] Further family-based linkage studies reported the involvement of additional genes in the development of monogenic forms of parkinsonism. Another autosomal dominant inherited parkinsonism gene is *LRRK2,*[64] of which the pathogenic mutations are predominantly located within the catalytic domains of the encoded protein.[65] *LRRK2* mutations are furthermore the most common genetic cause for idiopathic PD.[66] *Parkin,*[67] *PINK1*[68] and *DJ-1*[69] are genes associated to parkinsonism with an autosomal recessive inheritance pattern, which is mostly seen in the YOPD subtype.[10] Besides the familial PD cases, sporadic PD cases with an early onset of PD have also been observed with pathogenic mutations in *PINK1*[73,74] and *parkin*[75-77], making the latter gene the most common genetic factor (10-20%) for YOPD.[78] These 5 PD genes were defined as the genetic discoveries with conclusive evidence resulting from the linkage analysis era.

Three additional genes linked to monogenic forms of parkinsonism have been discovered with NGS, of which *VPS35* was the first one to be identified. Two independent studies reported *VPS35* mutations in Austrian and Swiss families diagnosed with autosomal dominant PD.[79,80] More recently, *DNAJC13*[81] and *CHCHD2*[82] have also been linked to dominantly inherited parkinsonism by implementing whole exome sequencing (WES). Although PD case-specific *CHCHD2* variants have also been observed in a second study,[83] independent identification of convincing pathogenic mutations in familial or sporadic PD cases will strengthen the evidence for both *DNAJC13* and *CHCHD2.* However, convincing replication of similar rare variants in these genes requires NGS screening of large cohorts with thousands of PD cases.

Box A. Genome-wide research techniques

The **linkage method** was developed during the eighties of the last century and was the first technique to explore the full genome in an unbiased fashion lacking assumptions about gene functions putatively involved in biological pathways related to the disease pathogenesis of interest.[70] Relatives share large pieces of DNA, therefore enabling the comparison of identical genomic regions between family members with a relative small number of genetic markers. By selecting those genomic regions that overlap in patients but not in healthy relatives it is possible to decrease the genomic region that might contain the rare and highly-penetrant causal variant.

The completion of the full human DNA sequence in 2001 was followed by the identification of millions of single nucleotide polymorphisms (SNPs), which are in general defined as variants that are common in the general population. These SNPs served as the foundation upon which **genome-wide association studies (GWAS)** would build on, resulting in the first published GWAS on a specific type of blindness in 2005.[71] SNPs that are relatively closely located to one another are likely to be simultaneously inherited as a block, and are called to be in linkage disequilibrium (LD).[72] In principle, one SNP for each LD block of the genome would be sufficient to test for the effect of genetic variation to a trait. Therefore, GWAS uses a selection of SNPs that capture most of the allelic variation in the surrounding region. GWAS has a distinct study design from linkage, as the study subjects encompass unrelated individuals, typically consisting of a case and control group for disease-related genetic research.

During the last decade, **next-generation sequencing (NGS)** has revolutionized the genomics field. This technique uses massively parallel sequencing, meaning that millions of DNA-molecules are simultaneously processed allowing a single genome to be sequenced in a time-frame of approximately one day. All nucleotides of the sequence are covered therefore enabling to analyze every single base, increasing the likelihood to find the true causal variant. Although the costs for genome sequencing have decreased substantially, it is still relatively expensive. An alternative and cheaper NGS technique is targeted sequencing where specific genomic regions, that are expected to contain disease-associated variants, are sequenced. One example of targeted sequencing is whole-exome sequencing.

Besides the rare variants with high effect sizes explaining the genetic variability in Mendelian types of PD, many GWAS[84-88] identified common risk variants contributing to multifactorial PD, which was expected to explain an additional part of the heritability estimates.[89] The most recent and largest meta-analysis of GWAS data, including 13,708 PD cases and 95,282 healthy controls of European ancestry, identified 24 loci that contribute to PD susceptibility confirming many previously identified PD genes.[90] Among these 24 risk loci are genes related to monogenic PD (*LRRK2* and *SNCA*), implying different variants within the same gene to have distinct magnitudes of effect. In addition, *GBA* was

confirmed as a genome-wide significant genetic risk factor. Initially *GBA* was suggested to contribute to PD risk based on clinical observations where some Gaucher disease patients, often caused by *GBA* mutations with autosomal recessive inheritance pattern, presented symptoms resembling Parkinsonism.[91,92] Subsequent large-scale research aiming to convincingly verify the importance of *GBA* for PD, reported a 5-fold increase to develop PD for heterozygous *GBA* carriers.[93] Although the 24 independent common genetic risk factors with moderate effect sizes are not of great value for individual diagnosis, they are important for highlighting impaired biological processes in PD pathogenesis. Future studies should decipher which genes underneath the GWAS peaks are the actual causal ones and through which molecular pathways they influence the development of PD.

Both the discovery of the Mendelian genes and the genetic risk loci expanded the knowledge about underlying biological mechanisms that experience defects contributing to PD susceptibility. First, the genes related to monogenic forms of PD have put emphasize on three biological mechanisms: synaptic function, lysosome-mediated autophagy and mitochondrial maintenance.[59] The hypothesis that synaptic transmission impairments are important in PD emerged from the discovery of mutations in *SNCA,* the main component of LB, and is a presynaptic protein promoting exocytosis.[94] The identification of *LRRK2,* a protein in the postsynaptic terminal suggested to have a function in endocytosis,[95] strengthens this theory. Furthermore, the YOPD-associated gene *parkin* inhibits the activity of excitatory synapses.[96] Pre- and postsynaptic processes might therefore be suitable targets for future treatment strategies. *SNCA* and *LRRK2* have additional roles in autophagy processes (intracellular degradation system), especially in chaperon-mediated autophagy.[97-99] This specific form of autophagy, selecting particular cargo, is a type of lysosomal degradation.[100] *VPS35* also has a role in lysosomal degradation through its function in mediating retrograde transport of cargo and is a component of the retromer and its defects could lead to trafficking issues of a lysosome protease involved in α-synuclein degradation.[101] *GBA* encodes for a lysosomal acid hydrolases and mutations in *GBA* are suggested to result in overall lysosomal dysfunction and autophagy impairments.[102,103] In addition, homozygous mutations within *GBA* cause Gaucher disease, which is a lysosomal storage disorder.[104] These studies imply that *LRRK2, SNCA, VPS35* and *GBA* all have a function in the lysosomal-mediated autophagy process, providing a putative target for research ultimately resulting in improved therapeutic interventions.

The processes to maintain mitochondrial health are additional molecular mechanisms that are implied to influence PD development by the discovery of the YOPD genes *parkin, PINK1* and *DJ-1. Parkin* and *PINK1* are suggested to be closely related as they both have roles in mitophagy,[105] which encompasses autophagy of deficient mitochondria. *Parkin* is in general present in the cytosol but localizes to the mitochondria when they depolarize to serve as a mediator for autophagy of the specific damaged mitochondria.[106] This recruitment of *parkin* is accomplished by *PINK1,* as it recognizes

dysfunctional mitochondria, accumulates on them and subsequently signals *parkin.*[107] *Parkin* ubiquinates mitochondrial substrates on the outer membrane, thereby informing the lysosome which mitochondrion should be degraded. The third YOPD gene *DJ-1* is not active in mitophagy, but seems to work in parallel to protect mitochondria against oxidative stress.[108] The recently discovered PD gene *CHCHD2* is suggested to have a functional role in mitochondrial respiration.[82] Therefore, a total of 4 genes point to the importance of healthy mitochondrial processes for adequate functioning of dopaminergic neurons.

The identification of the common genetic risk factors by GWAS has revealed an additional biological pathway to be involved in PD pathogenesis. Different loci (near *HLA-DRA, HLA-DRB1, HLA-DRB5* and *HLA-DQB1*) of the HLA region on chromosome 6 are associated to PD.[87,90,109,110] This region harbors many genes encoding molecules that are important for adaptive immune response. The discovery of these loci strengthened the involvement of the immune system in PD etiology, which was previously implied by observations of neuroinflammation and innate immunity in the cerebrospinal fluid and post-mortem brain tissues of PD patients.[111] These genetic findings furthermore imply that the observed neuroinflammation is a cause rather than a result of neuronal loss.

Although the identified genetic factors are not applicable for individual diagnosis or as direct targets for therapeutic interventions, they have led to an improved understanding of the underlying cellular processes contributing to PD pathogenesis. Extensive comprehension of these affected biological mechanisms is required to develop treatments that are not only aiming to relieve the symptoms, but would be able to alter the disease progression or ultimately cure the disease. The identification of Mendelian genes with strong effects and risk loci with moderate effects explain only up to 50% of the genetic variation,[9,60] implying the existence of uncovered genetic factors contributing to PD pathogenesis. Establishing these unrecognized genes could potentially further improve our knowledge on how to develop treatment. It is crucial to consider which variant type to focus on, as novel genes will presumable be revealed in areas that were previously untouched. Figure 1 is an overview of the established genes and loci for PD categorized according to the type of the associated variants. The mutations related to monogenic forms of PD are mostly present in the upper left corner, representing very rare variants with a high penetrance. The lower right corner of the figure contains the common risk variants with a low to moderate effect, representing the risk loci revealed by PD GWAS. It is implausible that the right upper corner holds genetic factors for PD, because variants common in the general population are unlikely to have strong pathogenic effects. The position in Figure 1 that is indicated by the green cloud could be a fruitful area to explore for novel PD genes. This variant category would encompass rare to very rare variant with moderate effect size. Whole-exome sequencing is a valuable genomic technique to examine these types of variants by the performance of population studies, comparing unrelated cases and controls.

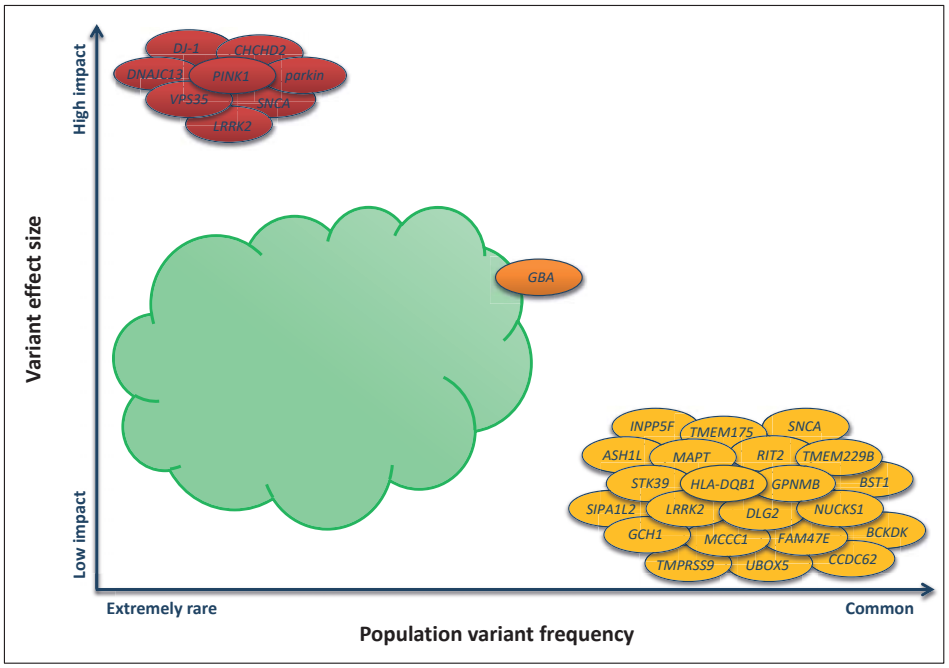


Figure 1. PD-associated genes and risk loci in perspective of variant type.

Whole exome sequencing

Subsequent to the popular family-based linkage analyses in the last two decades of the 20th century that revealed many novel genes causing Mendelian types of diseases, the genetic research community established methods to genetically study populations encompassing unrelated samples,[112] as described in box A. By exploring the genetic involvement of common risk variants in complex multifactorial disorders, instead of the rare variants with strong effects identified by the family based studies, it was anticipated that these GWAS would explain a large part of the heritability estimates.[89] The success rate of the GWAS was at the beginning mostly limited by the sample size. However with the reduction of genotype expenses,[113] the sample size per study increased, starting from 96 cases and 50 controls for the first GWAS study focusing on macular degeneration study,[71] to 36,989 schizophrenia patients and 113,075 control samples encompassing a meta-analysis GWAS in 2014.[114] These high numbers of samples had already revealed over 14,000 variants that are associated to more than 700 traits by the end of 2013.[115]

Despite the large amount of newly identified common genetic factors, the GWAS design was accompanied by new issues. The associated SNP often captures a large haplotype (LD block) of a few Mb encompassing multiple genes, therefore lacking the ability to pinpoint the true causal variant or gene. The biological relevance was unclear

for the vast majority of novel identified genetic factors.[116] Secondly, contrary to what was anticipated; the identified common variants were not explaining a large part of the heritability estimates, implying the existence of additional undiscovered genetic risk factors.[117] As an example, height heritability is estimated to lie between 80% and 90%.[118] A multi-institutional approach in 2014 studying 253,288 individuals identified over 423 loci associated to height but only explaining 20% of the heritability.[119]

The missing heritability could be due to the existence of genuine associated SNPs with too small individual variant effects to pass conservative significance test, as proposed by Yang et al.,[120] or the contribution of other types of genetic factors that were unable to detect by GWAS. The latter explanation resulted in the need for next-generation sequencing (NGS) methods, enabling the study of rare variants.[121] It was hypothesized that the identification of rare causal variants would increase the proportion of heritability that could be explained. Furthermore, by sequencing the full genome or essential regions of the genome, the causal variants would be more easily detected, enabling a more straightforward approach to identify the causal and biological relevant variant.

During the last decade, NGS has revolutionized the genomics research field. NGS overruled the conventional Sanger-based sequencing techniques, as it was faster and cheaper while maintaining a high sensitivity.[121] NGS processes millions of DNA-molecules simultaneously and is therefore called a massively parallel throughput method. The manifestation of NGS challenged the bioinformatics community to develop competent approaches to store and analyze the huge amount of data that were produced by the sequencing machines.[122] Raw sequence data has to go through multiple analyses steps encompassing read quality control, alignment of reads to reference genome, recalibration of aligned reads and variants calling, including small insertions/deletions (indels) and single nucleotide variants (SNVs). Especially when simultaneously calling variants for hundreds or thousands of samples, a computer cluster is required. Many different tools have been developed to deal with these analyses steps, aiming to analyze the data in the fastest way while maintaining the highest quality.[123,124]

Sequencing the full genome by NGS was the most attractive approach, as you sequence every nucleotide and thereby increase the likelihood to uncover the causal variant. However, the costs to perform whole-genome sequencing were excessively high at the beginning of the NGS era, especially for large-scale cohorts. Around 2008, the costs for a single genome were around 10 million dollar.[125] As a comparison, anno 2016 the sequencing costs are close to thousand dollars per genome. As an alternative, research groups focused on specific regions of the genome. One approach involved targeted sequencing, where custom-made designs are the source for library preparation kits.[126] For example, a study on ulcerative colitis identified novel rare variants in 3 genes by resequencing 55 genes originating from 21 associated GWAS LD-regions. [127] Another option for targeted sequencing is to focus on a specific region based on family-based

linkage analyses, as was the case for the discovery of *C9orf72* for FTD-ALS.[128,129] Another option is WES, which is standard commercially available by every sequencing technology company. As the costs per total genome were still relatively high (>$5,000) the moment this PhD thesis was started in 2011, it was decided to use WES to study all coding regions of the genome in a relatively unbiased fashion.

WES aims to target every exon of all protein coding genes.[130] Exons are the components of the gene that are transcribed into mRNA and eventually translated into amino acids which are the building blocks of all proteins in the human body. By sequencing all coding regions of the genome we anticipated to increase the chance to detect a variant that might be damaging for the function of the corresponding protein.[131,132] Variants in functional domains of the protein could for instance interrupt its physical interactions with other proteins. Or variants changing the amino acid, the nonsynonymous variants, could have an effect on the folding behavior of the protein. Various exonic variant categories exist which have different magnitudes of effect on the gene's function (Box B and Figure 2).

To predict the functional effect of exonic variants, the bioinformatics field has developed various algorithms which consider biological metrics including conservation scores (e.g. PhyloP[133]), transcription information (e.g. position with reference to exon-intron boundaries[134]), protein level scores (e.g. SIFT[135]) and regulatory information (e.g. regions of DNase I hypersensitivity[136]). The general consensus is to use annotation algorithms such as the Combined Annotation–Dependent Depletion (CADD)[137] framework or Functional Analysis through Hidden Markov Models (FATHMM-MKL)[138], which aim to incorporate all the aspects of previous prediction algorithms, while covering the total genome. The advantage of CADD and FTAHMM-MKL, over the older prediction algorithms (like SIFT and Polyphen), is that the predictions have been calculated for every possible nucleotide change in the genome, including exonic changes not affecting the corresponding amino acid or non-coding variants located in intronic and intergenic regions.

The first successful WES study was published in 2011, identifying *DHODH* as the causal gene for the rare Mendelian disorder Miller syndrome.[139] Four affected individuals from three different families harbored different rare mutations that were all predicted to be pathogenic within the same gene. Many outstanding exome studies followed, mostly focusing on Mendelian disorders by examining familial cases.[140] WES mainly focuses on rare variants, which are more straightforward to detect in families as they share large portions of their genome. Population-based cohorts are more complex to address with WES. The samples are unrelated, therefore decreasing the probability of having the exact same rare variant. However, these sporadic cases are more abundant than familial cases, and are a great source to examine. As the costs for WES continued to drop, the number of samples per study increased, facilitating the exploration of exomes in larger populations of unrelated samples. One commonly used approach for WES population studies is to

Box B. Exonic variant types

Loss-of-function variants (LoF)

LoF variants are anticipated to have a high impact by disrupting the protein-coding genes, therefore also called disruptive variants. This subset of variants is assumed to diminish the corresponding gene-expression, thereby decreasing the function of the gene. Three types of LoF variants exist: 1) stopgain/stoploss variants, 2) splicing variants and 3) frameshift indels. The first LoF type influences the position of the stopcodon, which is the marker for a DNA polymerase to terminate the transcription process. Splicing variants are close to the intron-exon boundaries, therefore presumably influencing the splice activity of the transcript. Frameshifts, which cause a different amino acid sequence from the position where the indel occurs, often result in a stopcodon relatively close to start of the frameshift. In general, LoF variants are rare in the human population as strong purifying selection against deleterious variants is at play.[141,142]

Nonsynonymous variants

These amino acid changing variants are furthermore called missense variants. Although such variants may harm the function of the gene, many are also benign. The benign **nonsynonymous** variants are often common in the general population as natural selection tolerates non-damaging variants.[142] Multiple efficient bioinformatics strategies have been developed to judge the functionality of variants, including the nonsynonymous ones. The predictions are based on conservation scores (e.g. PhyloP[133]), transcription information (e.g. position with reference to exon-intron boundaries[134]), protein level scores (e.g. SIFT[135]) and regulatory information (e.g. regions of DNase I hypersensitivity[136]). These algorithms are required to select the variants that have a high likelihood to be involved in the disease pathogenesis.

Non-frameshift indels

This category encompasses the small insertions and deletions that are within the reading frame. **Non-frameshift indels** affect three adjacently located nucleotides, therefore only changing the amino acids at that exact position.[143] Again, such variants have benign or harmful effects and bioinformatics functional prediction algorithms are applied to determine the deleteriousness.

Synonymous variants

Exonic **synonymous** variants, also called silent mutations, are traditionally expected to have low impact on the function of a gene, comparable to intronic, 3' and 5' UTR variants, as the amino acid is preserved. However, more recent studies show that such variants, changing the codon but not the amino acid, could have significant effects on substrate specificity or translation speed.[144,145] Some functional prediction algorithms (e.g. CADD and FTAHMM-MKL) calculate scores for every possible nucleotide change in the genome, including synonymous changes.

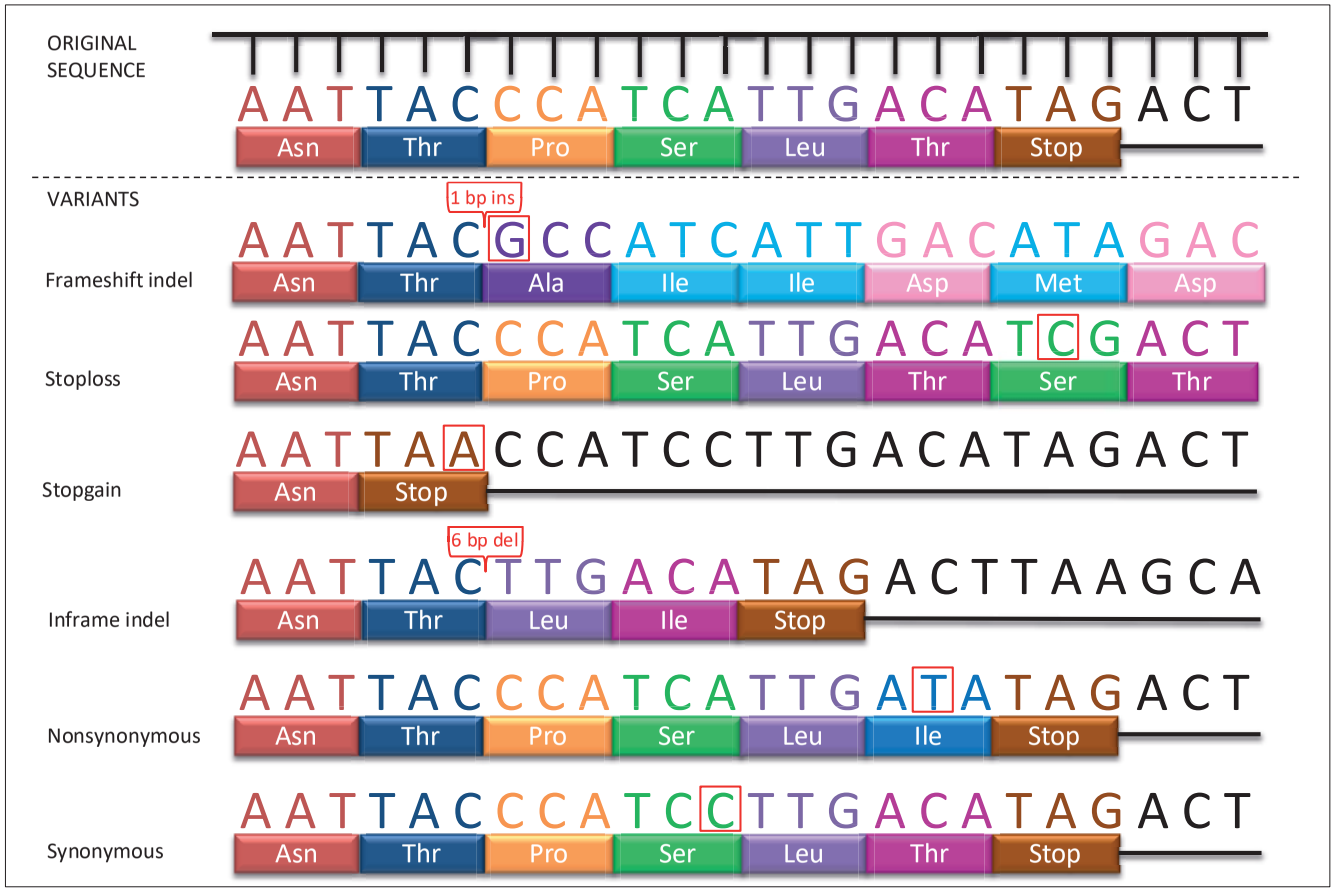


Figure 2. Variant types with distinct effects.

aggregate rare variants per gene and test for the combined association to a trait. Although population association studies seemed the next logical step, the accompanied issues were underestimated.[146]

As for traditional GWAS, very large sample sizes are required, particularly when focusing on complex diseases caused by rare variants with mild effects. Furthermore, rare variant aggregation tests are more sensitive to confounding factors such as population stratification, especially considering the difference in number of rare variants in different ethnic populations.[147] Another important aspect to consider when performing rare variant association test is which type of variants to include. Ultimately, one would prefer to test only those variants that have a deleterious effect on the function of a gene and exclude benign variants that dilute an association signal. By only studying variants below a certain rare minor allele frequency (MAF) cutoff (damaging variants are expected to have a lower MAF in the general population due to natural selection) and variants predicted to be damaging by functional prediction algorithms, the selection of functional variants is hypothesized to be maximized.[148,149] Another possibility to boost power is to aggregate variants over sets of genes belonging to similar biological pathways, increasing the effect size by the joint number of variants that are tested.[148]

Although many of the above-mentioned aspects should be considered when performing high-quality association tests with WES data in populations, few studies generated valuable results by proper study design. A WES study focusing on schizophrenia, including over 5,000 samples with balanced case-control distribution, identified a polygenic burden of rare LoF variants in several gene-sets.[150] Besides showing the contribution of rare risk factors in genes involved in calcium ion channels and signaling complexes of the postsynaptic density, they moreover showed that proper WES study design enables the identification of rare risk alleles in population-based research. A second population-based WES association study (2,869 amyotrophic lateral sclerosis patients and 6,405 controls) discovered a novel risk gene, *TBK1*, by running a standard aggregation association method that collapses rare variants for every gene of the genome.[151] Both these studies highlight the importance of careful consideration on how to choose sample size, type of variants and MAF cutoffs.

While WES is frequently preferred over whole genome sequencing (WGS) for economic reasons, WES is still substantially more expensive than genotype arrays. For research projects specifically interested in exonic regions in large sample sizes, but lacking the financial resources to accomplish WES, an affordable alternative are exome genotyping arrays.[152] These arrays cost <$50 per sample. Over 90% of the variants on this chip are located within exons and most of the variants are rare. Typically, the exome chips contain variants that have been observed in more than one sequencing study. Multiple publications have proven to be successful by the usage of the Illumina HumanExome Beadchip, discovering rare variant associations for novel genes influencing insulin

processes, [153] Type 2 Diabetes[154] and schizophrenia.[155] A specific exome array of Illumina has been created in the context of genetic factors contributing to neurodegenerative diseases.[156] In addition to the standard Illumina exome backbone of ~240,000 variants, this NeuroX exome array also includes ~25,000 variants associated to neurological diseases. This chip enables exome association studies that are enriched for variants important for neurodegenerative processes. Although exome chips are beneficial in terms of sample size they have the disadvantage that novel variants remain uncovered and is therefore limited to known genetic factors. When money is not a limiting factor or cohorts with smaller sample sizes are being studied, WES is the preferred option.

International Parkinson's Disease Genomics Consortium

Effective generation of large-scale WES datasets, which are required for adequate statistical testing of genetic associations, is depending on joined forces of multiple scientific entities (e.g. institutes, universities, funds, etc.). Consortia are furthermore of great value to share the high experimental expenses. For genetic research in PD, a worldwide collaboration was started in 2007 between genetic scientists from the Unites States, United Kingdom, The Netherlands, France and Germany, forming the International Parkinson's Disease Genomics Consortium (IPDGC; www.pdgenetics.org). The initial goal encompassed the performance of a GWAS, reaching a sample size allowing for sufficient power to identify novel PD risk factors. With more than 100 members and over 25 publications, the consortium has proven to be of great value to the PD research community by the identification of both Mendelian and risk genes important for PD.

The IPDGC evolved along with the temporal changes in genomic techniques, advancing to NGS strategies by starting a WES project in 2010 aiming to sequence 1,500 PD subjects. The laboratorial activities to generate the WES dataset covered 5 years, resulting in various data freezes thereby creating multiple versions of the WES dataset with increasing number of sample size. Besides the classical single-marker associations, the rare variant aggregation analysis was performed for every single gene. In addition to these relatively straightforward analyses procedures, members of the consortium are given the freedom to conceptualize hypotheses and perform genetic analyses to test them.

Scope of the thesis

The general aim of the thesis was to improve our understanding of PD etiology by the exploration of known PD genes and the discovery of novel genetic factors. Exonic regions of the genome were explored, either obtained by WES or the NeuroX array of which the data was generated by the IPDGC. The first two experimental chapters of this thesis survey the effect of rare variants in recognized PD genes, whereas the last three experimental chapters focus on rare variants in novel PD genes.

Regarding the known PD genes, **chapter 2** describes a study showing a straightforward approach that turns the genetic IPDGC datasets into helpful resources for replication of recently discovered genes in the PD research community. By performing standard variant identification, annotation and gene-based association tests, the WES and NeuroX datasets allow to relatively quickly examine novel PD genes discovered by other research groups. We describe an inspection on a potential novel PD gene (*CHCHD2*), which was initially identified in the Asian population and determined the genetic effects of variants in *CHCHD2* in the European population.

Assessment of the IPDGC WES and NeuroX datasets further permits to examine the role of rare variants within the genomic areas of previously identified PD risk loci. **Chapter 3** uses a strategy, which is based on functional similarities, to select the most probable causal genes underlying the associated PD GWAS peaks, and tests for the joint effect of coding variants on PD. Genes are not only tested individually for their burden of rare variants, but are additionally studied as a gene-set. Grouping all genes into a gene-set enhances the power of the design, if the majority is genuinely associated to PD. This study specifically highlights the analytical obstacles we need to overcome when performing rare variant aggregation studies, specifically of gene-sets.

Although unraveling the exact contribution of previously published PD genes by using novel approaches or datasets is meaningful, it is yet more exciting to search for novel and previously unrecognized genes that contribute to PD susceptibility. We decreased the searching area by focusing on genes that are implied to have a function in biological mechanisms involved in PD pathogenesis. **Chapter 4** uses the knowledge of the involvement of the lysosomal dysfunction in PD, which has been suggested through the affected lysosomal-mediated autophagy in PD patients and through the identification of *GBA* (initially only related to a specific lysosomal storage disorder (LSD)) as PD risk gene. By focusing on additional genes that have been genetically linked to LSD, chapter 4 aims to establish the effect of rare variants within such genes to PD. **Chapter 5** searches for novel genetic influences within molecular processes that are suggested to be involved in PD through genetics or transcriptomics. This study furthermore determines whether the observed gene-expression changes in brains of early stage PD progression could be due to genetic defects.

The presence of novel PD-associated variants in just a single or few PD patients complicates the performance of statistical tests (not sufficient numbers to obtain significance) or genetic replication (due to the rarity of the variant presumably require thousands of samples). To overcome the requirement for genetic replication, **chapter 6** includes a study that aims to biologically validate newly suggested PD genes. By focusing on LoF variants that follow an autosomal recessive inheritance pattern, this study imitates the effect of these variants by knock-down of the corresponding gene or orthologue in human neuronal cell culture, *Drosophila* and *C.elegans* animal models. Establishing the

involvement of the candidate genes in PD-related phenotypes, such as mitochondrial dysfunction and α -synuclein toxicity, provides biological insights possibly strengthening the relation of several candidate genes to PD.

Chapter 7 covers the general discussion where the chapters are summarized and discussed in the context of the current literature. Furthermore, the genomic research techniques develop at an incredible pace. Therefore, future approaches for excellent genomics research are also considered.

REFERENCES

1. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Movement disorders : official journal of the Movement Disorder Society* 2014; 29(13): 1583-90.
2. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *Journal of neurology, neurosurgery, and psychiatry* 1988; 51(6): 745-52.
3. Kalia LV, Lang AE. Parkinson's disease. *Lancet* 2015; 386(9996): 896-912.
4. Duncan GW, Khoo TK, Yarnall AJ, et al. Health-related quality of life in early Parkinson's disease: the impact of nonmotor symptoms. *Movement disorders : official journal of the Movement Disorder Society* 2014; 29(2): 195-202.
5. Postuma RB, Aarsland D, Barone P, et al. Identifying prodromal Parkinson's disease: pre-motor disorders in Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2012; 27(5): 617-26.
6. Khoo TK, Yarnall AJ, Duncan GW, et al. The spectrum of nonmotor symptoms in early Parkinson disease. *Neurology* 2013; 80(3): 276-81.
7. Tolosa E, Gaig C, Santamaria J, Compta Y. Diagnosis and the premotor phase of Parkinson disease. *Neurology* 2009; 72(7 Suppl): S12-20.
8. Lang AE. In pursuit of prodromal Parkinson's disease. *Lancet neurology* 2015; 14(1): 27-8.
9. Keller MF, Saad M, Bras J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Human molecular genetics* 2012; 21(22): 4996-5009.
10. Schrag A, Schott JM. Epidemiological, clinical, and genetic characteristics of early-onset parkinsonism. *Lancet neurology* 2006; 5(4): 355-63.
11. Spica V, Pekmezovic T, Svetel M, Kostic VS. Prevalence of non-motor symptoms in young-onset versus late-onset Parkinson's disease. *Journal of neurology* 2013; 260(1): 131-7.
12. Jankovic J, Kapadia AS. Functional decline in Parkinson disease. *Archives of neurology* 2001; 58(10): 1611-5.
13. Alves G, Wentzel-Larsen T, Aarsland D, Larsen JP. Progression of motor impairment and disability in Parkinson disease: a population-based study. *Neurology* 2005; 65(9): 1436-41.
14. Kostic V, Przedborski S, Flaster E, Sternic N. Early development of levodopa-induced dyskinesias and response fluctuations in young-onset Parkinson's disease. *Neurology* 1991; 41(2 (Pt 1)): 202-5.
15. Schrag A, Ben-Shlomo Y, Brown R, Marsden CD, Quinn N. Young-onset Parkinson's disease revisited--clinical features, natural history, and mortality. *Movement disorders : official journal of the Movement Disorder Society* 1998; 13(6): 885-94.
16. Tandberg E, Larsen JP, Aarsland D, Laake K, Cummings JL. Risk factors for depression in Parkinson disease. *Archives of neurology* 1997; 54(5): 625-30.
17. Negre-Pages L, Grandjean H, Lapeyre-Mestre M, et al. Anxious and depressive symptoms in Parkinson's disease: the French cross-sectionnal DoPaMiP study. *Movement disorders : official journal of the Movement Disorder Society* 2010; 25(2): 157-66.
18. Pontone GM, Williams JR, Anderson KE, et al. Prevalence of anxiety disorders

and anxiety subtypes in patients with Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2009; 24(9): 1333-8.
19. Wickremaratchi MM, Ben-Shlomo Y, Morris HR. The effect of onset age on the clinical features of Parkinson's disease. *European journal of neurology : the official journal of the European Federation of Neurological Societies* 2009; 16(4): 450-6.
20. Calne SM, Kumar A. Young onset Parkinson's disease. Practical management of medical issues. *Parkinsonism & related disorders* 2008; 14(2): 133-42.
21. Gibb WR, Lees AJ. Anatomy, pigmentation, ventral and dorsal subpopulations of the substantia nigra, and differential cell death in Parkinson's disease. *Journal of neurology, neurosurgery, and psychiatry* 1991; 54(5): 388-96.
22. Damier P, Hirsch EC, Agid Y, Graybiel AM. The substantia nigra of the human brain. II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. *Brain : a journal of neurology* 1999; 122 (Pt 8): 1437-48.
23. Hornykiewicz O. Basic research on dopamine in Parkinson's disease and the discovery of the nigrostriatal dopamine pathway: the view of an eyewitness. *Neuro-degenerative diseases* 2008; 5(3-4): 114-7.
24. Greffard S, Verny M, Bonnet AM, et al. Motor score of the Unified Parkinson Disease Rating Scale as a good predictor of Lewy body-associated neuronal loss in the substantia nigra. *Archives of neurology* 2006; 63(4): 584-8.
25. Tissingh G, Bergmans P, Booij J, et al. Drug-naive patients with Parkinson's disease in Hoehn and Yahr stages I and II show a bilateral decrease in striatal dopamine transporters as revealed by [123I]beta-CIT SPECT. *Journal of neurology* 1998; 245(1): 14-20.
26. Braak H, Ghebremedhin E, Rub U, Bratzke H, Del Tredici K. Stages in the development of Parkinson's disease-related pathology. *Cell and tissue research* 2004; 318(1): 121-34.
27. Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of aging* 2003; 24(2): 197-211.
28. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. Alpha-synuclein in Lewy bodies. *Nature* 1997; 388(6645): 839-40.
29. Doherty KM, Silveira-Moriyama L, Parkkinen L, et al. Parkin disease: a clinicopathologic entity? *JAMA neurology* 2013; 70(5): 571-9.
30. Poulopoulos M, Levy OA, Alcalay RN. The neuropathology of genetic Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2012; 27(7): 831-42.
31. Bendor JT, Logan TP, Edwards RH. The function of alpha-synuclein. *Neuron* 2013; 79(6): 1044-66.
32. Goedert M, Spillantini MG, Del Tredici K, Braak H. 100 years of Lewy pathology. *Nature reviews Neurology* 2013; 9(1): 13-24.
33. Cremades N, Cohen SI, Deas E, et al. Direct observation of the interconversion of normal and toxic forms of alpha-synuclein. *Cell* 2012; 149(5): 1048-59.
34. Kalia LV, Kalia SK, McLean PJ, Lozano AM, Lang AE. alpha-Synuclein oligomers and clinical implications for Parkinson disease. *Annals of neurology* 2013; 73(2): 155-69.
35. Dickson DW, Braak H, Duda JE, et al. Neuropathological assessment of Parkinson's disease: refining the diagnostic

criteria. *Lancet neurology* 2009; 8(12): 1150-7.
36. Selikhova M, Williams DR, Kempster PA, Holton JL, Revesz T, Lees AJ. A clinico-pathological study of subtypes in Parkinson's disease. *Brain : a journal of neurology* 2009; 132(Pt 11): 2947-57.
37. Irwin DJ, White MT, Toledo JB, et al. Neuropathologic substrates of Parkinson disease dementia. *Annals of neurology* 2012; 72(4): 587-98.
38. Kordower JH, Chu Y, Hauser RA, Freeman TB, Olanow CW. Lewy body-like pathology in long-term embryonic nigral transplants in Parkinson's disease. *Nature medicine* 2008; 14(5): 504-6.
39. Li JY, Englund E, Holton JL, et al. Lewy bodies in grafted neurons in subjects with Parkinson's disease suggest host-to-graft disease propagation. *Nature medicine* 2008; 14(5): 501-3.
40. Luk KC, Kehm VM, Zhang B, O'Brien P, Trojanowski JQ, Lee VM. Intracerebral inoculation of pathological alpha-synuclein initiates a rapidly progressive neurodegenerative alpha-synucleinopathy in mice. *The Journal of experimental medicine* 2012; 209(5): 975-86.
41. Luk KC, Kehm V, Carroll J, et al. Pathological alpha-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science (New York, NY)* 2012; 338(6109): 949-53.
42. Masuda-Suzukake M, Nonaka T, Hosokawa M, et al. Prion-like spreading of pathological alpha-synuclein in brain. *Brain : a journal of neurology* 2013; 136(Pt 4): 1128-38.
43. Recasens A, Dehay B, Bove J, et al. Lewy body extracts from Parkinson disease brains trigger alpha-synuclein pathology and neurodegeneration in mice and monkeys. *Annals of neurology* 2014; 75(3): 351-62.
44. Rey NL, George S, Brundin P. Review: Spreading the word: precise animal models and validated methods are vital when evaluating prion-like behaviour of alpha-synuclein. *Neuropathology and applied neurobiology* 2016; 42(1): 51-76.
45. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2015; 30(12): 1591-601.
46. Hughes AJ, Daniel SE, Lees AJ. Improved accuracy of clinical diagnosis of Lewy body Parkinson's disease. *Neurology* 2001; 57(8): 1497-9.
47. Rao G, Fisch L, Srinivasan S, et al. Does this patient have Parkinson disease? *Jama* 2003; 289(3): 347-53.
48. Brooks DJ, Pavese N. Imaging biomarkers in Parkinson's disease. *Progress in neurobiology* 2011; 95(4): 614-28.
49. Marek K, Seibyl J, Eberly S, et al. Longitudinal follow-up of SWEDD subjects in the PRECEPT Study. *Neurology* 2014; 82(20): 1791-7.
50. Hawkes CH, Shephard BC. Selective anosmia in Parkinson's disease? *Lancet* 1993; 341(8842): 435-6.
51. Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. *Archives of neurology* 1999; 56(1): 33-9.
52. Berg D, Postuma RB, Bloem B, et al. Time to redefine PD? Introductory statement of the MDS Task Force on the definition of Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2014; 29(4): 454-62.
53. Connolly BS, Lang AE. Pharmacological treatment of Parkinson disease: a review. *Jama* 2014; 311(16): 1670-83.
54. Fox SH, Katzenschlager R, Lim SY, et al. The Movement Disorder Society

Evidence-Based Medicine Review Update: Treatments for the motor symptoms of Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2011; 26 Suppl 3: S2-41.
55. Seppi K, Weintraub D, Coelho M, et al. The Movement Disorder Society Evidence-Based Medicine Review Update: Treatments for the non-motor symptoms of Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2011; 26 Suppl 3: S42-80.
56. Kalia SK, Sankar T, Lozano AM. Deep brain stimulation for Parkinson's disease and other movement disorders. *Current opinion in neurology* 2013; 26(4): 374-80.
57. Fasano A, Daniele A, Albanese A. Treatment of motor and non-motor features of Parkinson's disease with deep brain stimulation. *Lancet neurology* 2012; 11(5): 429-42.
58. Marras C, Lang A. Parkinson's disease subtypes: lost in translation? *Journal of neurology, neurosurgery, and psychiatry* 2013; 84(4): 409-15.
59. Trinh J, Farrer M. Advances in the genetics of Parkinson disease. *Nature reviews Neurology* 2013; 9(8): 445-54.
60. Hamza TH, Payami H. The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. *Journal of human genetics* 2010; 55(4): 241-3.
61. Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiology of aging* 2011; 32(10): 1923.e1-8.
62. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science (New York, NY)* 1997; 276(5321): 2045-7.
63. Klein C, Westenberger A. Genetics of Parkinson's disease. *Cold Spring Harbor perspectives in medicine* 2012; 2(1): a008888.
64. Zimprich A, Biskup S, Leitner P, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* 2004; 44(4): 601-7.
65. Dachsel JC, Farrer MJ. LRRK2 and Parkinson disease. *Archives of neurology* 2010; 67(5): 542-7.
66. Healy DG, Falchi M, O'Sullivan SS, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet neurology* 2008; 7(7): 583-90.
67. Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 1998; 392(6676): 605-8.
68. Valente EM, Abou-Sleiman PM, Caputo V, et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science (New York, NY)* 2004; 304(5674): 1158-60.
69. Bonifati V, Rizzu P, Squitieri F, et al. DJ-1(PARK7), a novel gene for autosomal recessive, early onset parkinsonism. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 2003; 24(3): 159-60.
70. Terwilliger JD, Ott J. Handbook of human genetic linkage. Baltimore: John Hopkins University Press; 1994.
71. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science (New York, NY)* 2005; 308(5720): 385-9.
72. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nature*

reviews Genetics 2005; 6(2): 109-18.
73. Valente EM, Salvi S, Ialongo T, et al. PINK1 mutations are associated with sporadic early-onset parkinsonism. *Annals of neurology* 2004; 56(3): 336-41.
74. Tan EK, Yew K, Chua E, et al. PINK1 mutations in sporadic early-onset Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2006; 21(6): 789-93.
75. Chaudhary S, Behari M, Dihana M, et al. Parkin mutations in familial and sporadic Parkinson's disease among Indians. *Parkinsonism & related disorders* 2006; 12(4): 239-45.
76. Periquet M, Latouche M, Lohmann E, et al. Parkin mutations are frequent in patients with isolated early-onset parkinsonism. *Brain : a journal of neurology* 2003; 126(Pt 6): 1271-8.
77. Macedo MG, Verbaan D, Fang Y, et al. Genotypic and phenotypic characteristics of Dutch patients with early onset Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 2009; 24(2): 196-203.
78. Klein C, Lohmann-Hedrich K. Impact of recent genetic findings in Parkinson's disease. *Current opinion in neurology* 2007; 20(4): 453-64.
79. Vilarino-Guell C, Wider C, Ross OA, et al. VPS35 mutations in Parkinson disease. *American journal of human genetics* 2011; 89(1): 162-7.
80. Zimprich A, Benet-Pages A, Struhal W, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *American journal of human genetics* 2011; 89(1): 168-75.
81. Vilarino-Guell C, Rajput A, Milnerwood AJ, et al. DNAJC13 mutations in Parkinson disease. *Human molecular genetics* 2014; 23(7): 1794-801.
82. Funayama M, Ohe K, Amo T, et al. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. *Lancet neurology* 2015.
83. Jansen IE, Bras JM, Lesage S, et al. CHCHD2 and Parkinson's disease. *Lancet neurology* 2015; 14(7): 678-9.
84. Pankratz N, Wilk JB, Latourelle JC, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Human genetics* 2009; 124(6): 593-605.
85. Simon-Sanchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics* 2009; 41(12): 1308-12.
86. Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nature genetics* 2009; 41(12): 1303-7.
87. Hamza TH, Zabetian CP, Tenesa A, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nature genetics* 2010; 42(9): 781-5.
88. International Parkinson's Disease Genomics Consortium, Wellcome Trust Case Control Consortium2. A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS genetics* 2011; 7(6): e1002142.
89. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nature reviews Genetics* 2005; 6(2): 95-108.
90. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics*

2014; 46(9): 989-93.
91. Neudorfer O, Giladi N, Elstein D, et al. Occurrence of Parkinson's syndrome in type I Gaucher disease. *QJM : monthly journal of the Association of Physicians* 1996; 89(9): 691-4.
92. Tayebi N, Callahan M, Madike V, et al. Gaucher disease and parkinsonism: a phenotypic and genotypic characterization. *Molecular genetics and metabolism* 2001; 73(4): 313-21.
93. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *The New England journal of medicine* 2009; 361(17): 1651-61.
94. Lashuel HA, Overk CR, Oueslati A, Masliah E. The many faces of alpha-synuclein: from structure and toxicity to therapeutic target. *Nature reviews Neuroscience* 2013; 14(1): 38-48.
95. Shin N, Jeong H, Kwon J, et al. LRRK2 regulates synaptic vesicle endocytosis. *Experimental cell research* 2008; 314(10): 2055-65.
96. Helton TD, Otsuka T, Lee MC, Mu Y, Ehlers MD. Pruning and loss of excitatory synapses by the parkin ubiquitin ligase. *Proceedings of the National Academy of Sciences of the United States of America* 2008; 105(49): 19492-7.
97. Winslow AR, Chen CW, Corrochano S, et al. alpha-Synuclein impairs macroautophagy: implications for Parkinson's disease. *The Journal of cell biology* 2010; 190(6): 1023-37.
98. Orenstein SJ, Kuo SH, Tasset I, et al. Interplay of LRRK2 with chaperone-mediated autophagy. *Nature neuroscience* 2013; 16(4): 394-406.
99. Cuervo AM, Stefanis L, Fredenburg R, Lansbury PT, Sulzer D. Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. *Science (New York, NY)* 2004; 305(5688): 1292-5.
100. Kaushik S, Cuervo AM. Chaperone-mediated autophagy: a unique way to enter the lysosome world. *Trends in cell biology* 2012; 22(8): 407-17.
101. Miura E, Hasegawa T, Konno M, et al. VPS35 dysfunction impairs lysosomal degradation of alpha-synuclein and exacerbates neurotoxicity in a Drosophila model of Parkinson's disease. *Neurobiology of disease* 2014; 71: 1-13.
102. Osellame LD, Duchen MR. Defective quality control mechanisms and accumulation of damaged mitochondria link Gaucher and Parkinson diseases. *Autophagy* 2013; 9(10): 1633-5.
103. Siebert M, Sidransky E, Westbroek W. Glucocerebrosidase is shaking up the synucleinopathies. *Brain : a journal of neurology* 2014; 137(Pt 5): 1304-22.
104. Hruska KS, LaMarca ME, Scott CR, Sidransky E. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Human mutation* 2008; 29(5): 567-83.
105. Pickrell AM, Youle RJ. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. *Neuron* 2015; 85(2): 257-73.
106. Narendra D, Tanaka A, Suen DF, Youle RJ. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *The Journal of cell biology* 2008; 183(5): 795-803.
107. Matsuda N, Sato S, Shiba K, et al. PINK1 stabilized by mitochondrial depolarization recruits Parkin to damaged mitochondria and activates latent Parkin for mitophagy. *The Journal of cell biology* 2010; 189(2): 211-21.
108. Kahle PJ, Waak J, Gasser T. DJ-1 and prevention of oxidative stress in

Parkinson's disease and other age-related disorders. *Free radical biology & medicine* 2009; 47(10): 1354-61.
109. Nalls MA, Plagnol V, Hernandez DG, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* 2011; 377(9766): 641-9.
110. Ahmed I, Tamouza R, Delord M, et al. Association between Parkinson's disease and the HLA-DRB1 locus. *Movement disorders : official journal of the Movement Disorder Society* 2012; 27(9): 1104-10.
111. McGeer PL, Itagaki S, Boyes BE, McGeer EG. Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains. *Neurology* 1988; 38(8): 1285-91.
112. Jorde LB. Linkage disequilibrium and the search for complex disease genes. *Genome research* 2000; 10(10): 1435-44.
113. Macgregor S, Zhao ZZ, Henders A, Nicholas MG, Montgomery GW, Visscher PM. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. *Nucleic acids research* 2008; 36(6): e35.
114. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; 511(7510): 421-7.
115. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research* 2014; 42(Database issue): D1001-6.
116. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010; 141(2): 210-7.
117. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; 461(7265): 747-53.
118. Visscher PM. Sizing up human height variation. *Nature genetics* 2008; 40(5): 489-90.
119. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics* 2014; 46(11): 1173-86.
120. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* 2010; 42(7): 565-9.
121. Metzker ML. Sequencing technologies - the next generation. *Nature reviews Genetics* 2010; 11(1): 31-46.
122. Chen K, Pachter L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS computational biology* 2005; 1(2): 106-12.
123. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2009; 25(14): 1754-60.
124. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]* 2013; 11(1110): 11.0.1-.0.33.
125. Wetterstand KA. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). www.genome.gov/sequencingcosts (accessed February 23 2016.
126. Harismendy O, Ng PC, Strausberg RL, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome biology* 2009; 10(3): R32.
127. Beaudoin M, Goyette P, Boucher G, et al. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis.

PLoS genetics 2013; 9(9): e1003723.
128. Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011; 72(2): 257-68.
129. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 2011; 72(2): 245-56.
130. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Human molecular genetics* 2010; 19(R2): R145-51.
131. Nelson MR, Wegmann D, Ehm MG, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science (New York, NY)* 2012; 337(6090): 100-4.
132. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science (New York, NY)* 2012; 337(6090): 64-9.
133. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research* 2010; 20(1): 110-21.
134. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; 489(7414): 57-74.
135. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research* 2003; 31(13): 3812-4.
136. Boyle AP, Davis S, Shulha HP, et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 2008; 132(2): 311-22.
137. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM. A general framework for estimating the relative pathogenicity of human genetic variants. 2014; 46(3): 310-5.
138. Shihab HA, Rogers MF, Gough J, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics (Oxford, England)* 2015; 31(10): 1536-43.
139. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics* 2010; 42(1): 30-5.
140. Zhang X. Exome sequencing greatly expedites the progressive research of Mendelian diseases. *Frontiers of medicine* 2014; 8(1): 42-57.
141. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, NY)* 2012; 335(6070): 823-8.
142. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467(7319): 1061-73.
143. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. *Human molecular genetics* 2010; 19(R2): R131-6.
144. Kimchi-Sarfaty C, Oh JM, Kim IW, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science (New York, NY)* 2007; 315(5811): 525-8.
145. Chevance FF, Le Guyon S, Hughes KT. The effects of codon context on in vivo translation speed. *PLoS genetics* 2014; 10(6): e1004392.
146. Kiezun A, Garimella K, Do R, et al. Exome sequencing and the genetic basis of complex traits. *Nature genetics* 2012; 44(6): 623-30.
147. Keinan A, Mullikin JC, Patterson N, Reich D. Measurement of the human allele

frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature genetics* 2007; 39(10): 1251-5.

148. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. *Proceedings of the National Academy of Sciences of the United States of America* 2014; 111(4): E455-64.
149. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *American journal of human genetics* 2014; 95(1): 5-23.
150. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; 506(7487): 185-90.
151. Cirulli ET, Lasseigne BN, Petrovski S, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science (New York, NY)* 2015; 347(6229): 1436-41.
152. Guo Y, He J, Zhao S, et al. Illumina human exome genotyping array clustering and quality control. *Nature protocols* 2014; 9(11): 2643-62.
153. Huyghe JR, Jackson AU, Fogarty MP, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nature genetics* 2013; 45(2): 197-201.
154. Wessel J, Chu AY, Willems SM, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nature communications* 2015; 6: 5897.
155. Richards AL, Leonenko G, Walters JT, et al. Exome arrays capture polygenic rare variant contributions to schizophrenia. *Human molecular genetics* 2016; 25(5): 1001-7.
156. Nalls MA, Bras J, Hernandez DG, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. *Neurobiology of aging* 2014.