SUMMARY

Within this thesis we have established the contribution of known PD genes in more detail and searched for novel ones. To further explore the known genes, two IPDGC exome datasets were accessed for validation of the novel Mendelian PD gene *CHCHD2* and for determining the effect of rare genetic variants within the published PD risk loci. For *CHCHD2* (**chapter 2**), a gene involved in mitochondrial function which was first observed in a PD cohort from Japan,[5] no association signal of common variants was detected in our European population. One identical common variant to a significant SNP in the original Asian study was not associated in our NeuroX dataset, suggesting Asian-specificity for that particular variant. Although we did not observe an association of common variants, 3 rare nonsynonymous variants (MAF ≤ 0.0008) were observed in 4 cases while not a single rare variant was observed in healthy controls. These results suggest that *CHCHD2* could also be a rare risk factor in the European population, but more extensive sequencing research involving larger datasets should be performed to enable any definitive conclusions.

Multiple genome-wide association studies have linked dozens of common loci to the risk of developing PD. By using our exome datasets, we tested whether rare coding variants within these known risk loci might also influence disease risk (**chapter 3**). For the analysis of single rare variants large datasets with tens of thousands of individuals are required to reach sufficient statistical power. To increase this power we therefore aggregated rare variants per gene and gene-set to test for their joint effect of rare variants on PD. As GWAS risk loci indicate regions in which variants are in linkage disequilibrium (LD) that frequently encompass multiple genes of which the causal gene is undefined, we used a strategy (PrixFixe) that selected the most probable causal gene based on underlying functional similarities. We confirmed the association of the well-established *LRRK2* concerning both common and rare variant associations. Furthermore, *STBD1* and *SPATA19* show an association to PD in the WES and NeuroX datasets, respectively. The gene-set aggregation approach detected, as anticipated, a moderate association of common variants. However, there was no effect of aggregated rare variants when testing the gene-set. It is complicated to conclude whether this is due to an incorrect selection of causal genes by the PrixFixe approach or whether rare variants in PD risk loci are genuinely not contributing to PD risk. This chapter clearly shows the need for improved functional prioritization tools, besides PrixFixe, that are able to adequately call the most probably gene(s) while taking into account the biological features of the disease of interest.

Our search for novel PD genes was guided by focusing on groups of genes that share a functional property related to PD. *GBA* is a gene with a lysosomal function that has been related to autosomal recessive Gaucher disease, a rare lysosomal storage disorder (LSD), and has been shown to strongly increase PD risk. We therefore studied the involvement of all known LSD-related genes in PD susceptibility (**chapter 4**). By using two independent WES datasets and a third exome array dataset, we tested for the joint effect of rare variants in LSD genes with an emphasize on the effect of the total gene-set to increase statistical power to detect an association signal as the WES datasets had relatively low sample sizes. We identified a significant joint association of rare deleterious (predicted by CADD) variants within the total set of 54 LSD genes in the original IPDGC dataset, which we were able to replicate in the additional independent PPMI WES and IPDGC NeuroX datasets. The association signal remained in all datasets, when excluding *GBA*. Zooming in on the association of individual LSD genes, we observed significant associations for *GBA*, which confirms the strength of our study design. Our results furthermore suggest that multiple variants in various LSD genes within a single individual, the so called oligogenic effect, might increase the risk for PD. Finally, *GBA, SMPD1, CTSD* and *SLC17A5* show independent significant effects on PD, of which the first 2 genes have already been discussed in the context of PD risk in literature. Overall, we conclude that additional non-*GBA* variants in LSD genes contribute to PD susceptibility where the effects might be consistent with a model of oligogenic risk factors. Additional sequencing in larger sample sizes and functional follow-up studies should further investigate the exact role of these LSD genes in PD pathogenesis.

A second approach based gene-sets of interest on biological processes that are affected in PD pathobiology (**chapter 5**), either based on literature or on gene-expression research. The latter involved a genome-wide transcriptomics study that identified specific biological pathways based on differences in gene-expression levels at distinct levels of Braak stage α-synuclein pathology by using post-mortem brain samples. By specifically investigating the affected molecular processes of Braak stages 1 and 2, the pre-symptomatic stage, we aimed to determine whether these observed gene expression changes could be caused by genetic defects. Three biological pathways (*mitochondrial dysfuntion, caveolar-mediated endocytosis signaling* and *renin-angiotensin signaling*) showed an aggregated association signal in all three independent WES datasets. The literature-based pathway *mitochondrial dysfunction* showed significant association to PD in the IPDGC WES discovery dataset which was replicated in the smaller PPMI and Merck WES datasets. Although replication was not observed for

individual genes within this pathway, multiple genes belonging to the NADH:ubiquinone oxidoreductase family, also called mitochondrial complex I subunits, were enriched for pathogenic variants in the distinct datasets. As mitochondrial complex I has an important function in ATP synthesis and a major contributor to the production of reactive oxygen species, these results imply that this specific molecular process might be affected in PD pathogenesis. Functional follow-up could specifically target these cellular events and potentially link them to PD-related phenotypes. The transcriptomics-based pathway *renin-angiotensin signaling* showed enrichment for common coding variants in PD cases compared to controls. This data suggests that the detected expression changes for genes belonging to these pathways at the beginning of PD pathology progression, are caused by genetic variation rather than by other intrinsic biological changes or environmental influences.

A last strategy to identify novel genetic risk factors for PD was applied in **chapter 6.** All coding regions of the genome were explored for high-impact loss of function variants showing a higher frequency in PD cases compared to healthy controls. Only homozygous and compound heterozygous variants were considered as the case sample set has a young average age of onset which has often been linked to an autosomal recessive disease inheritance pattern.[11] We reported 27 genes encompassing LoF variants with a higher frequency in our PD cohort, of which *GPATCH2L, UHRF1BP1L, PTPRH, ARSB* and *VPS13C* genes show convincing PD-related phenotypes in functional experiments and genetic confirmation in independent PD datasets. The identification of the PD gene *VPS13C* emphasizes the strength of our study design. We strongly believe that these 5 genes are promising PD genes and encourage the PD research field to perform additional variant screening and functional validation studies to confirm their pathogenic contribution to PD.