


UNIVERSITÀ
DI PAVIA

**Dipartimento di Biologia e Biotecnologie "L. Spallanzani"**

Laurea Magistrale in Biologia Sperimentale e Applicata

# Studio delle basi genetiche della malattia di Parkinson mediante l'analisi di varianti rare: dai casi sporadici alle coppie di gemelli discordanti

# Dissecting the genetic architecture of Parkinson's disease through rare-variant association analysis: from unrelated cases to discordant twins

Supervisor:
*Prof Rossella Tricarico*

Co-supervisors:
*Dr Letizia Straniero*
*Dr Claudia Anna Giudici*

Experimental thesis by
*Ilaria Mattarozzi*

Academic Year 2025/2026

## ABSTRACT

Parkinson's disease (PD) is a neurodegenerative movement disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra. Rare pathogenic variants have been identified in several genes, and more than 90 common *loci* have been associated with PD susceptibility, with *GBA1* representing the most common genetic risk factor. However, most *GBA1* carriers do not develop PD and the contribution of rare coding variants to the architecture of sporadic PD, as well as the factors modulating *GBA1*-associated risk, remain poorly understood. Moreover, monozygotic (MZ) twins discordant for PD offer a unique model to investigate post-zygotic and regulatory differences while controlling for the same genetic background.

This study aimed to dissect the genetic architecture of PD through rare-variant association analysis, from unrelated cases to discordant twins. The unrelated cohorts comprised 1,055 *GBA1* wild-type PD patients, 314 *GBA1* variant-carrier PD patients, and 1,587 healthy controls. The twin cohort included 23 MZ pairs discordant for PD and one triplet, with matched whole-blood RNA sequencing available. Rare, potentially deleterious coding variants (gnomAD allele frequency < 1% and CADD score ≥ 20) were collapsed at gene level and tested with SKAT-O in the unrelated cohorts and with SMMAT in the twins. Functional enrichment was performed with IPA, Enrichr and Metascape, while allele-specific expression (ASE) was assessed within twin pairs.

In *GBA1* wild-type patients *versus* controls, 25 of 723 genes remained significant after FDR correction and enrichment analysis highlighted immune-related pathways, in particular interferon-γ and JAK and IL-6 signalling. By contrast, *GBA1* variant-carrier patients showed signatures related to mitochondrial homeostasis, tetrahydrobiopterin biosynthesis and amyloid-related disease terms, with no gene surviving correction. The direct comparison between *GBA1* variant-carriers and non-carriers patients suggested differences in energy metabolism, GABA catabolism and central nervous system development, along with dementia- and Alzheimer's disease-related terms driven by *APP* and *SNCA*.

In the twin cohort, 969 genes carried putative post-zygotic variants; none was significant in gene-based testing, due to limited statistical power. Targeted analysis identified variants in eight PD-associated genes, including two different *MCCC1* variants in two affected twins, while *MFSD5* was the only gene shared with the FDR-significant set of the *GBA1* wild-type cohort. Genes carrying variants in affected twins, but not in unaffected co-twins, were enriched for nervous system, central nervous system and neurodegenerative diseases, a descriptive observation that may suggest a contribution of post-zygotic events to discordance. ASE analysis revealed 111 discordant variants

across 16 twin pairs, with enrichment of calcium-related pathways driven by *RYR1* and *RYR3* and a stop-gain *MC1R* variant expressed only in an affected twin.

Overall, these findings suggest potentially distinct molecular patterns in *GBA1* wild-type and *GBA1* variant-carrier PD and illustrate the value of integrating DNA and RNA data in discordant twins. Given the small number of carriers, the limited power of the twin cohort and the exploratory nature of the enrichment analyses, the results are hypothesis-generating and require replication and functional validation.

**TABLE OF CONTENTS**

**ABSTRACT** ..... **1**
**TABLE OF CONTENTS** ..... **1**
**1. INTRODUCTION** ..... **3**
1.1 PARKINSON'S DISEASE ..... 3
1.1.1 Epidemiology ..... 3
1.1.2 Pathogenesis ..... 4
1.1.3 Etiology ..... 5
1.1.4 Therapeutics strategies ..... 7
1.2 GENETIC LANDSCAPE OF PD ..... 8
1.2.1 Autosomal dominant PD loci and genes ..... 10
1.2.2 Autosomal recessive PD loci and genes ..... 12
1.2.3 Mitochondrial and lysosomal genes involved in PD ..... 13
1.3 GBA1 ..... 15
1.4 Monozygotic twins as a model for genetic studies ..... 17
**2. AIMS OF THE STUDY** ..... **19**
**3. MATERIAL AND METHODS** ..... **21**
3.1 STUDY COHORTS ..... 21
3.1.1 Unrelated case-control cohorts ..... 21
3.1.2 Twin cohort ..... 22
3.2 SEQUENCING ..... 23
3.2.1 WES sequencing ..... 23
3.2.2 RNA sequencing ..... 24
3.3 DATA PROCESSING ..... 25
3.3.1 Quality control and alignment ..... 25
3.3.2 WES variant calling and joint genotyping ..... 26
3.3.3 RNA-seq variant calling ..... 27
3.3.4 Variant annotation and filtering ..... 27
3.4 STATISTICAL ANALYSIS ..... 28
3.4.1 Study comparison and principal components analysis ..... 28
3.4.2 Rare-variant association analyses in the main and GBA1 cohorts ..... 30
3.4.3 Rare-variant association analyses in the twin cohort ..... 31
3.4.4 Odds ratio calculation ..... 32
3.4.5 Enrichment analysis ..... 33
**4. RESULTS** ..... **34**
4.1 RARE-VARIANT ASSOCIATION ANALYSES IN UNRELATED COHORTS ..... 34
4.1.1 GBA1 wild-type PD patients versus healthy controls ..... 34
Gene-based association results ..... 34
Functional enrichment analysis ..... 36
4.1.2 GBA1 variant-carrier PD patients versus healthy controls ..... 38
Gene-based association results ..... 38

Enrichment analysis ..... 40
4.1.3 GBA1 variant-carrier versus GBA1 wild-type PD patients ..... 42
Gene-based association results ..... 42
Functional enrichment analysis ..... 44
4.2 ANALYSIS OF MONOZYGOTIC TWINS DISCORDANT FOR PD ..... 48
4.2.1 Identification of within-pair DNA-discordant variants ..... 48
Functional enrichment analysis ..... 50
4.2.2 Allele-specific expression analysis of shared variants ..... 51
**5. DISCUSSION ..... 53**
**6. CONCLUSIONS ..... 60**
**References ..... 61**
**Web sites ..... 67**

# 1.INTRODUCTION

## *1.1 PARKINSON'S DISEASE*

*Parkinson's disease* (PD) is a progressive multisystem neurodegenerative disorder characterized by both motor and non-motor symptoms *(Sveinbjornsdottir, 2016)*. It was formally described for the first time by James Parkinson in his "An Essay on a Shaking Palsy", published in 1817, in which he mainly focused on the motor manifestations such as resting tremor, slowness or absence of voluntary movements, stooped posture and festinating gait *(Kulcsarova et al., 2024)*.

*Non-motor symptoms* (NMS), including hyposmia or anosmia, REM sleep behavior disorder, dementia, depression, anxiety, and cardiac sympathetic denervation were recognized and characterized only recently. In the early stages of the disease, motor symptoms usually appear asymmetrically and subsequently progress to affect both sides of the body. At late stages, disease progression may be accompanied by cognitive and psychiatric manifestations, such as hallucinations, depression, sleep and speech disturbances, memory loss, and dementia *(Costa et al., 2022)*.

The diagnosis of PD is primarily clinical, although definitive confirmation requires postmortem neuropathological examination. Several diagnostic guidelines have been developed to support the clinical diagnostic process. The most recent one, the *Movement Disorder Society Clinical Diagnostic Criteria* (MDS-CD), uses a two-step process in which motor symptoms remain the core defining feature. First, parkinsonism is diagnosed based on the presence of bradykinesia in combination with either rest tremor, rigidity, or both. The subsequent evaluation of supporting criteria, absolute exclusion criteria, and red flags is required to determine whether parkinsonism can be classified as clinically established or clinically probable PD *(Postuma et al., 2015)*. Supportive criteria include a clear response to dopaminergic therapy, the presence of levodopa-induced dyskinesia, and either documented olfactory loss or cardiac sympathetic denervation. NMS manifestations are also considered in the diagnostic assessment, as some support the presence of PD, whereas others may represent red flags when they occur unusually early or with disproportionate severity *(Costa et al., 2022)*.

### *1.1.1 Epidemiology*

PD is the second most common neurodegenerative disease worldwide, after Alzheimer's disease, and its incidence and prevalence are increasing as a result of population ageing. In industrialised countries, the prevalence of PD is estimated to be approximately 0.3% in the general population and 1% among individuals over 60 years of age. Prevalence progressively increases with advancing age in both men and women, reaching approximately 3.5% among individuals aged 85–89 years in Europe *(Sveinbjornsdottir, 2016)*. According to the *Global Burden of Disease study*, PD is the fastest growing neurological disorder. Between 1990 and 2015, the number of individuals living with PD

worldwide increased by 118%, reaching 6.2 million, and this figure is projected to exceed 12 million by 2040 *(Dorsey et al., 2018)*.

Disease onset typically occurs between 65 and 70 years of age. *Early-onset* (EOPD), generally defined as disease developing before the age of 50, and juvenile-onset, developing before the age of 20, are both relatively rare, with fewer than 5% of patients presenting symptoms before the age of 40 *(Tysnes & Storstein, 2017)*. Nevertheless, EOPD is characterized by specific motor and non-motor features that distinguish it from the more common *late-onset PD* (LOPD), *(Mehanna & Jankovic, 2019)*.

### *1.1.2 Pathogenesis*

PD is characterized by selective and progressive neurodegeneration of dopaminergic (DA) neurons localized in the *substantia nigra* (SN) *pars compacta*, a midbrain region that plays a key role in modulating movement and reward-related functions. This neurodegeneration is commonly associated with the presence of *Lewy bodies* (LBs) in surviving DA neurons *(Xiao et al., 2025)*. These inclusions represent one of the main neuropathological hallmarks of PD, together with the progressive loss of nigrostriatal dopaminergic neurons. LBs are intracellular aggregates primarily composed of misfolded and fibrillized α-synuclein (α-syn), which can accumulate both as LBs within neuronal cell bodies and as *Lewy neurites* (LNs) within neuronal processes. Post-translational modifications and misfolding of α-syn contribute to its aggregation and may disrupt cellular processes, ultimately promoting neuronal dysfunction and death. LBs are found in several regions of the *central nervous system* (CNS), including basal ganglia, dorsal motor nucleus of the *vagus*, olfactory bulb, *locus coeruleus*, and intermediolateral nucleus of the spinal cord. They have also been detected in the peripheral nervous system and the enteric nervous system of PD patients *(Raza et al., 2019)*. α-Syn aggregates have also been observed at synaptic terminals in human neuropathological studies, suggesting that early synaptic dysfunction may represent an important contributor to disease pathogenesis *(Morris et al., 2024)*.

In addition to α-syn aggregation and synaptic dysfunction, several interconnected cellular alterations contribute to PD pathogenesis, such as mitochondria impairment, dysregulation of calcium and *reactive oxygen species* (ROS) homeostasis, defects in protein quality-control and degradation systems, lysosomal impairment and alterations in autophagic and apoptotic pathways (***Figure 1***) *(Antony et al., 2013)*.

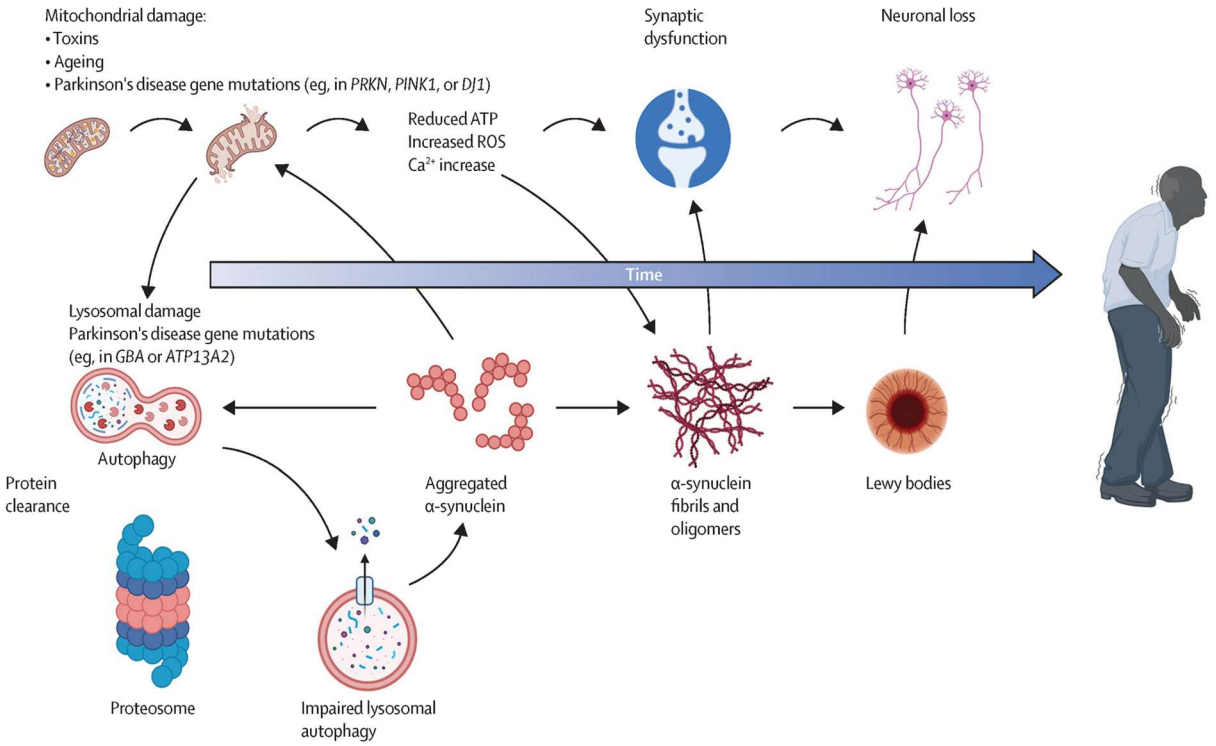



***Figure 1: Molecular mechanisms contributing to Parkinson's Disease (Morris et al., 2024)***

Mitochondrial dysfunction, for example, is considered an early event in PD pathogenesis. Alterations in mitochondrial structure and dynamics are linked to increased production of ROS and abnormal intracellular calcium levels, which may, in turn, promote α-syn misfolding and aggregation.

The autophagy-lysosomal and the ubiquitin-proteasome systems are also strongly implicated in PD pathobiology. Since α-syn clearance depends substantially on lysosomal degradation pathways, lysosomal dysfunction can impair α-syn turnover, increasing its intracellular levels and promoting its accumulation. In particular, reduced activity of the lysosomal enzyme β-glucocerebrosidase (GCase) disrupts lysosomal and lipid homeostasis and is associated with impaired α-syn clearance. The resulting accumulation of α-syn may further compromise GCase trafficking and activity, generating a bidirectional pathogenic loop that promotes progressive α-syn aggregation *(Morris et al., 2024)*.

### *1.1.3 Etiology*

PD is currently considered a multifactorial disorder arising from a complex interaction among genetic susceptibility, environmental exposures, and ageing-related cellular changes *(Costa et al., 2022)*. Although familial forms account for approximately 15% of PD cases and 5–10% show a clear Mendelian inheritance pattern, most cases are sporadic and have no single identifiable cause *(Costa et al., 2022; Bandres-Ciga et al., 2020)*. Accordingly, several demographic, environmental, and lifestyle factors have been associated with either an increased or a reduced risk of developing PD.

Advanced age represents the strongest known risk factor for PD. The incidence of LOPD is five- to tenfold higher than that of EOPD, with the latter characterized by slower disease progression, a lower

risk of dementia and other non-motor manifestations, and a higher frequency of levodopa-induced dyskinesia. *Age on set* (AAO) is therefore an important contributor to the clinical heterogeneity observed among patients with PD, however the mechanisms underlying these differences remain incompletely understood *(Huang et al., 2024).*

Male sex is also associated with an increased risk of PD and may influence disease progression. Men show a higher incidence and slightly higher mortality, whereas women generally have a later onset of motor symptoms and a more benign motor disease course. However, depression, anxiety and fatigue are more frequently reported in women, while cognitive impairment and severe sleep disturbances appear to be more common in men *(Marinus et al., 2018)*. These differences may reflect both sex-specific characteristics of dopaminergic pathways and the effects of sex hormones, Estradiol seems to have a neuroprotective effect by a pleiotropic action on the CNS, however, although their contribution to PD remains incompletely understood *(Cerri et al., 2019; Meoni et al., 2020)*.

Several environmental and occupational exposures have been associated with an increased risk of PD, including pesticides, selected heavy metals (Al, Cu, Fe, Hg, Mn, Pb), rural living, and agricultural or mining occupations. Some pesticides can interfere with complex I function in the mitochondrial respiratory chain, therefore inducing mitochondrial dysfunction and oxidative stress (*Brolin et al., 2024; Costa et al., 2022)*. Although altered iron and zinc homeostasis has also been investigated in relation to oxidative stress and neurodegeneration, studies comparing their levels between patients with PD and healthy controls have produced inconsistent results. Therefore, the contribution of individual metals to PD risk remains incompletely understood *(Bergsland et al., 2019; Sun et al., 2017)*.

Conversely, physical activity, cigarette smoking and coffee consumption have been linked to a lower risk of developing PD. Physical activity may support mitochondrial biogenesis and antioxidant defence, whereas nicotine and caffeine have been proposed to modulate α-syn aggregation, oxidative stress, neuroinflammation, and dopaminergic neurotransmission *(Boulos et al., 2019; Costa et al., 2022; Kolahdouzan & Hamadeh, 2017)*. However, these epidemiological associations do not necessarily demonstrate a direct causal or protective effect, and their underlying biological mechanisms remain under investigation.

### *1.1.4 Therapeutics strategies*

Current therapeutic interventions remain primarily symptomatic, aiming to reduce motor and non-motor symptoms and improve patients' quality of life rather than arresting or reversing neurodegeneration *(Stocchi et al., 2024)*.

*Levodopa* (L-DOPA), the metabolic precursor of dopamine, remains the most effective treatment for motor symptoms. Unlike dopamine, L-DOPA can cross the blood–brain barrier and it is subsequently converted into dopamine within the CNS, partially compensating for the loss of nigrostriatal DA neurons *(Cramb et al., 2023)*. It is generally administered together with a peripheral dopa-decarboxylase inhibitor, such as carbidopa or benserazide, to reduce its peripheral conversion and increase its availability in the brain. Although L-DOPA remains the gold-standard treatment, long-term therapy is frequently associated with motor fluctuations and levodopa-induced dyskinesia as the disease progresses *(Stocchi et al., 2024)*. *Deep brain stimulation* (DBS), most commonly targeting the subthalamic nucleus or globus pallidus internus, represents an established surgical alternative for selected patients with levodopa-responsive motor complications or medically refractory tremor *(Ashkan et al., 2017)*.

As no approved treatment is currently able to prevent or slow PD progression, several potential disease-modifying strategies are under investigation (***Figure 2***). These include approaches aimed at reducing α-syn production or aggregation, enhancing its clearance, modulating neuroinflammation, restoring mitochondrial and lysosomal function, and replacing or protecting dopaminergic neurons. Cell-based therapies using dopamine-producing neurons derived from embryonic stem cells or induced pluripotent stem cells are also being explored, although their long-term efficacy and safety remain to be established. In parallel, viral vector-mediated gene therapy is being investigated to deliver neurotrophic factors, such as GDNF, or enzymes involved in dopamine synthesis, with the aim of protecting residual neurons or restoring dopaminergic function *(Elkouzi et al., 2019; Raza et al., 2019; Ling et al., 2023)*.

Lysosomal dysfunction represents an additional therapeutic target, particularly in patients carrying *GBA1* variants. Reduced GCase activity impairs glycosphingolipid metabolism and lysosomal homeostasis and is associated with α-syn accumulation. Experimental strategies therefore aim to increase GCase activity by enhancing protein folding and trafficking, reducing the accumulation of its lipid substrates, or delivering a functional copy of *GBA1* using *adeno-associated viral* (AAV) vectors. Although these approaches may potentially restore lysosomal function and improve α-syn clearance, they remain under preclinical or clinical investigation and their disease-modifying efficacy has not yet been established *(Szunyogh et al., 2025)*.

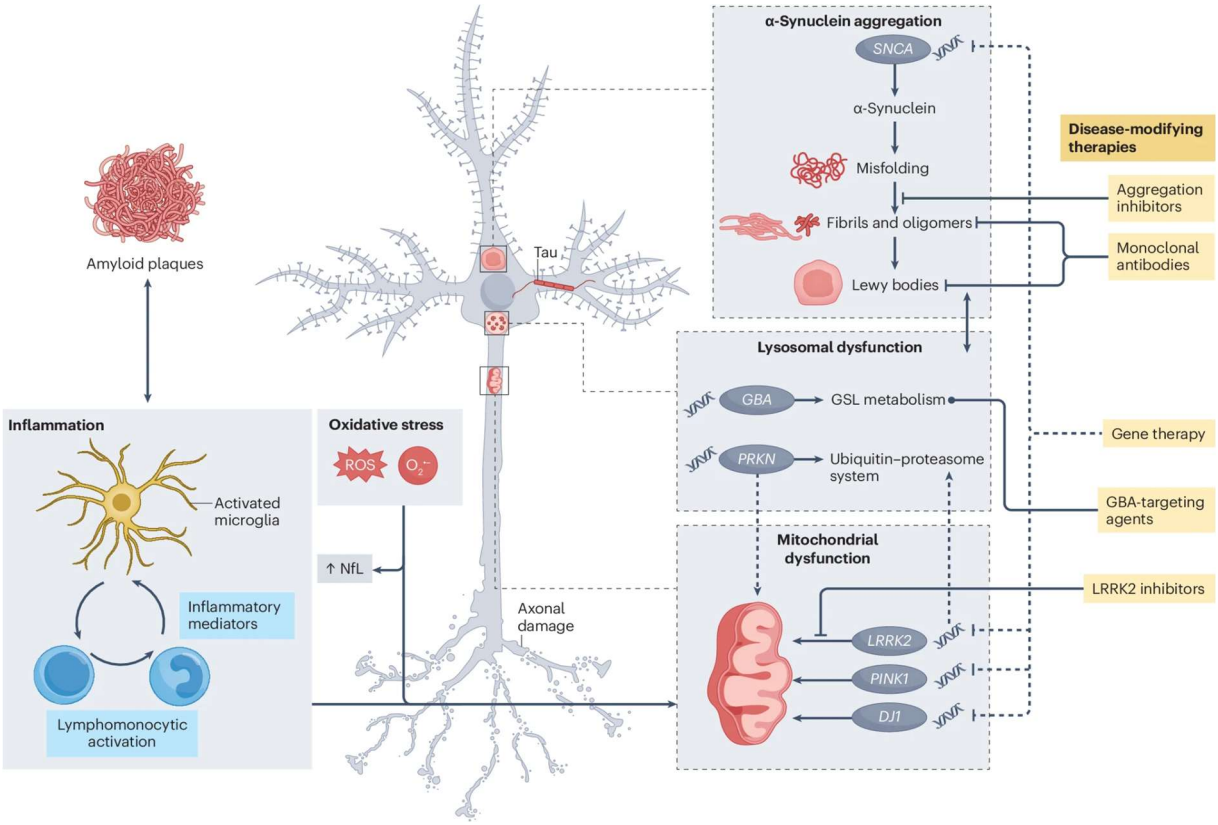



***Figure 2: Pathophysiological mechanisms underlying Parkinson disease and possible disease-modifying therapies (Stocchi et al., 2024).***

### *1.2 GENETIC LANDSCAPE OF PD*

For decades, PD was considered a predominantly non-genetic disease. However, following the identification of the first causative mutation in the *SNCA* gene in 1997, several other genes and genetic risk loci have been associated with the disease, and PD is now recognized as a complex disorder with a substantial genetic component *(Bandres-Ciga et al., 2020)*.

The genetic architecture of PD spans a continuous spectrum of allele frequencies and effect sizes (***Figure 3***). At one end of this spectrum, rare, highly penetrant variants with large effects responsible for monogenic forms of PD and are generally observed in familial cases. At the other end, common variants individually confer only small increases in disease risk but collectively contribute to susceptibility to sporadic PD. Between these extremes, rare or uncommon variants with intermediate and variable effects, such as variants in *GBA1* and selected variants in *LRRK2*, can substantially increase disease risk without being fully penetrant *(Lunati et al., 2018)*.

Rare variants (*Minor Allele Frequency*, MAF <1%) were historically identified using linkage analysis in large families showing Mendelian inheritance, followed by Sanger sequencing of the linked regions. More recently, *whole-exome sequencing* (WES) and *whole-genome sequencing* (WGS) have enabled the systematic identification of rare coding and non-coding variants in both familial and apparently sporadic cases. In contrast, common risk variants (MAF >5%), have mostly been identified

through *Genome-Wide Association Studies* (GWASs). These studies have revealed numerous independent risk *loci*, supporting the highly polygenic nature of PD and highlighting biological pathways related to lysosomal function, mitochondrial homeostasis, vesicular trafficking, and immune regulation *(Nalls et al., 2018)*.

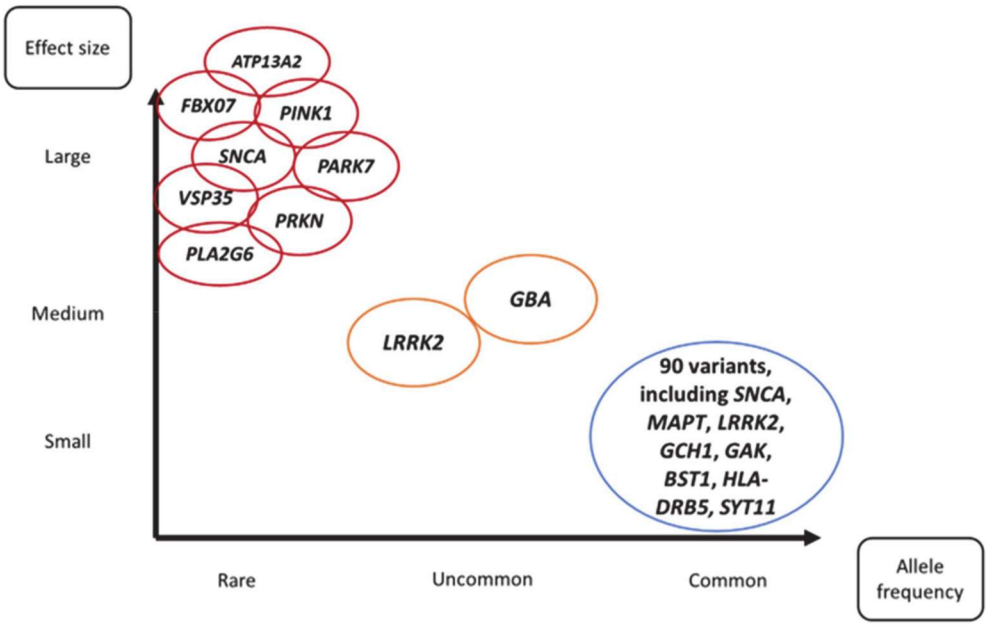



***Figure 3. Summary of genetic variants in Parkinson's disease grouped according to allele frequency and associated risk of Parkinson's disease (Day & Mullin, 2021).***

Although monogenic forms represent only 5% of PD cases, their discovery has been fundamental to shed light on disease pathophysiology. Linkage studies performed on families with Mendelian or quasi-Mendelian transmission of PD led to the identification of 23 *loci* involved in both autosomal dominant and autosomal recessive forms of the disease (***Table 1***, *Deng et al., 2018*). Well-established causes include pathogenic variants in *SNCA*, *LRRK2*, and *VPS35* in autosomal dominant PD and biallelic pathogenic variants in *PRKN*, *PINK1*, and *PARK7* in autosomal recessive PD. Additional genes have been proposed; however, the evidence supporting some of these associations remains limited because of insufficient segregation, replication or functional validation *(Blauwendraat et al., 2020)*.

| Locus (OMIM) | Location | Full Gene Name Approved by HGNC | HGNC Approved Gene Symbol (OMIM) | Inheritance | Disease onset | Lewy bodies |
|---|---|---|---|---|---|---|
| *PARK1* (168601) | 4q22.1 | synuclein alpha | *SNCA* (163890) | AD | Early-onset, late-onset* | C |
| *PARK2* (600116) | 6q26 | parkin RBR E3 ubiquitin protein ligase | *PRKN* (602544) | AR | Early-onset | NC |
| *PARK3* (602404) | 2p13 | Parkinson disease 3 | *PARK3* (Unclear) | AD | Late-onset | NC |
| *PARK4* (605543) | 4q22.1 | synuclein alpha | *SNCA* (163890) | AD | Early-onset | C |
| *PARK5* (613643) | 4p13 | ubiquitin C-terminal hydrolase L1 | *UCHL1* (191342) | AD | Early-onset, late-onset | NC |
| *PARK6* (605909) | 1p36 | PTEN induced putative kinase 1 | *PINK1* (608309) | AR | Early-onset | NC |
| *PARK7* (606324) | 1p36.23 | parkinsonism associated deglycase | *PARK7* (602533) | AR | Early-onset | NC |
| *PARK8* (607060) | 12q12 | leucine rich repeat kinase 2 | *LRRK2* (609007) | AD | Late-onset | C |
| *PARK9* (606693) | 1p36.13 | ATPase 13A2 | *ATP13A2* (610513) | AR | Early-onset | NC |
| *PARK10* (606852) | 1p32 | Parkinson disease 10 | *PARK10* (Unclear) | Unclear | Late-onset | NC |
| *PARK11* (607688) | 2q37.1 | GRB10 interacting GYF protein 2 | *GIGYF2* (612003) | AD | Late-onset | NC |
| *PARK12* (300557) | Xq21-q25 | Parkinson disease 12 | *PARK12* (Unclear) | X-linked inheritance | Late-onset | NC |
| *PARK13* (610297) | 2p13.1 | HtrA serine peptidase 2 | *HTRA2* (606441) | AD | Late-onset, early-onset* | NC |
| *PARK14* (612593) | 22q13.1 | phospholipase A2 group VI | *PLA2G6* (603604) | AR | Early-onset | NC |
| *PARK15* (260300) | 22q12.3 | F-box protein 7 | *FBXO7* (605648) | AR | Early-onset | NC |
| *PARK16* (613164) | 1q32 | Parkinson disease 16 | *PARK16* (Unclear) | Unclear | Late-onset | NC |
| *PARK17* (614203) | 16q11.2 | VPS35, retromer complex component | *VPS35* (601501) | AD | Late-onset | NC |
| *PARK18* (614251) | 3q27.1 | eukaryotic translation initiation factor 4 gamma 1 | *EIF4G1* (600495) | AD | Late-onset | NC |
| *PARK19* (615528) | 1p31.3 | DnaJ heat shock protein family (Hsp40) member C6 | *DNAJC6* (608375) | AR | Early-onset | NC |
| *PARK20* (615530) | 21q22.1 | synaptojanin 1 | *SYNJ1* (604297) | AR | Early-onset | NC |
| *PARK21* (616361) | 20p13 | transmembrane protein 230 | *TMEM230* (617019) | AD | Late-onset, early-onset* | C |
| *PARK22* (616710) | 7p11.2 | coiled-coil-helix-coiled-coil-helix domain containing 2 | *CHCHD2* (616244) | AD | Late-onset, early-onset* | NC |
| *PARK23* (616840) | 15q22.2 | vacuolar protein sorting 13 homolog C | *VPS13C* (608879) | AR | Early-onset | NC |
| | 11p15.4 | RIC3 acetylcholine receptor chaperone | *RIC3* (610509) | AD | Late-onset, early-onset* | NC |

***Table 1: Gene locus and disease-causing genes of Parkinson disease (Deng et al., 2018).***

### *1.2.1 Autosomal dominant PD loci and genes*

Pathogenic variants in *LRKK2*, *SNCA* and *VPS35* are well-established causes of autosomal dominant PD. These genetic forms show variable penetrance and clinical expression but have provided important insights into the molecular mechanisms underlying both familial and sporadic disease.

The *LRRK2* gene (leucine rich repeat kinase 2) encodes for a large multidomain protein kinase involved in several cellular processes such as neuronal plasticity, autophagy, lysosomal homeostasis, and vesicular trafficking. Among its best-characterized substrates are RAB GTPases, important regulators of intracellular membrane trafficking *(Trevisan et al., 2024)*. Pathogenic *LRRK2* variants represent one of the most common causes of PD (*PARK8*), accounting for approximately 1-2% of all PD cases, although their frequency varies considerably across populations *(Xiong & Yu, 2024)*. Although numerous *LRRK2* variants have been reported, only few variants (i.e. NM_198578 p.N1437H, p.R1441C/G/S/H, p.Y1699C, p.G2019S and p.I2020T) have been classified as pathogenic so far *(Trevisan et al., 2024)*. These variants are mainly located within the enzymatically active ROC-COR and kinase domains of the protein. The p.G2019S, and p.I2020T variants directly affect the kinase domain whereas variants such p.R1441C/G/H and p.Y1699C alter GTPase activity and indirectly increase kinase activity *(Xiong & Yu, 2024)*. Increased LRRK2 kinase activity dysregulates RAB phosphorylation and may impair cytoskeletal organization, vesicle trafficking, and inflammatory responses, ultimately contributing to neuronal dysfunction *(Trevisan et al., 2024)*.

Consequently, LRRK2 kinase inhibitors are currently being investigated as potential disease-modifying treatments *(Xiong & Yu, 2024)*.

The *SNCA* gene encodes α-syn, a protein involved in SNARE complex assembly, synaptic-vesicle dynamics, and neurotransmitter release. As previously described, the abnormal accumulation of α-syn in LB and LN is a major neuropathological feature of PD. Autosomal dominant PD associated with *SNCA* can result from missense variants or multiplication of the entire gene, including duplications and triplications *(Lunati et al., 2018).* A clear gene-dosage effect has been observed: duplications generally cause a phenotype resembling *idiopathic PD* (IPD), whereas triplications are associated with an earlier onset, more rapid disease progression and a higher frequency of cognitive and autonomic manifestations *(Book et al., 2018)*. Several pathogenic missense variants have also been described, including NM_000345.3 p.A30P, p.E46K, p.G51D, p.A53E, and p.A53T. These variants differ in their effects on α-syn structure, aggregation, and cellular toxicity, resulting in considerable clinical variability *(Jia et al., 2022)*. The p.A53T mutation, which is the most frequent *SNCA* mutation, causes EOPD (before age 50 years), sometimes resembling IPD but often associated with dementia, while p.A30P and p.E46K mutations are responsible for a later age at onset, around 60 years, and high prevalence of dementia and hallucinations. In contrast, the mutations p.G51D and p.A53E seemed to be more deleterious, with early onset parkinsonism, dementia, and multiple system atrophy, such as autonomic impairment and rapid disease progression *(Deng et al., 2018; Lunati et al., 2018).* Patients carrying pathogenic *SNCA* variants may show a clinically meaningful response to dopaminergic treatment, although treatment response and disease progression vary according to the underlying variant. *(Lunati et al., 2018).*

Finally, *VPS35 (*Vacuolar Protein Sorting 35) encodes for a core component of the retromer complex, involved in the transport of proteins from endosomes to the trans-Golgi network and contributes to endosomal and lysosomal homeostasis *(Trevisan et al., 2024)*. The NM_018206.6 p.D620N missense variant is currently the only VPS35 variant confirmed as the cause of autosomal dominant PD *(Lunati et al., 2018)*. This variant impairs retromer-mediated cargo sorting and has been associated with altered endolysosomial maturation, receptor recycling, mitochondrial-derived vesicle formation, and α-syn accumulation *(Day & Mullin, 2021)*. Interaction with other genes causing monogenic PD (such as *SNCA*, *LRRK2,* and *PRKN*) has been suggested, but still unexplored. Patients carrying *VPS35* mutation show a disease form that resembles IPD, with a median age at onset of approximately 49 years, good responsiveness to levodopa, and relatively slow disease progression. Cognitive impairment appears to be uncommon, even after a long disease duration *(Jia et al., 2022).*

### *1.2.2 Autosomal recessive PD loci and genes*

The main genes associated with autosomal recessive forms of typical PD are *PRKN*, *PINK1*, and *PARK7,* the latter encoding the DJ-1 protein. These forms share a broadly similar clinical phenotype, characterized by early onset, slow disease progression, a good response to levodopa, and a relatively low frequency of cognitive decline. Although the proteins encoded by these genes have distinct functions, they all contribute to mitochondrial homeostasis and cellular responses to stress *(Lunati et al., 2018)*.

The *PRKN* gene encodes Parkin, an E3 ubiquitin ligase involved in protein ubiquitination and mitochondrial quality control. Together with *PINK1*, Parkin participates in mitophagy, the selective autophagic removal of defective mitochondria. Loss-of-function variants in *PRKN* impair this process, leading to the accumulation of dysfunctional mitochondria, increased oxidative stress, and progressive neuronal damage *(Trevisan et al., 2024)*. Different pathogenic variants have been identified, including large deletions or duplications of one or more exons (43.2%); small deletions/insertions (16.5%), nonsense and missense mutations (22.3%) *(Jia et al., 2022)*. Biallelic *PRKN* variants represent the most commonly known cause of autosomal recessive EOPD, accounting for up to 49% EOPD patients and 19% sporadic EOPD *(Deng et al., 2018)*. *PRKN*-associated PD is generally characterized by slow progression, generally sustained responsiveness to levodopa, frequent treatment-related motor complications, and relatively preserved cognitive function *(Jia et al., 2022)*.

The *PINK1* gene (PTEN induced kinase 1) encodes a mitochondrial serine/threonine kinase with a central role in sensing mitochondrial damage and initiating *PINK1*–Parkin-mediated mitophagy. *Loss-of-function* (LoF) variants impair the recognition and removal of dysfunctional mitochondria, thereby increasing susceptibility to oxidative stress and has mitochondrial homeostasis dysfunction *(Deng et al., 2018)*. Biallelic pathogenic variants in *PINK1* represent another important cause of autosomal recessive PD. To date, around 60 mutations have been described; the most predominant variant is the NM_032409 p.L347P missense mutation *(Jia et al., 2022)*. The associated clinical phenotype is largely similar to that observed in *PRKN*-associated disease, with early onset, slow progression, and a generally good response to levodopa. Lower-limb dystonia may occur at disease onset, whereas psychosis and cognitive impairment have occasionally been reported at later stages *(Trevisan et al., 2024)*.

Finally, *PARK7* encodes DJ-1, a multifunctional protein expressed in the brain, and in particular in astrocytes. Together with PINK1 and parkin, DJ-1 acts as a sensor for oxidative stress, and contributes to the regulation of mitochondria morphology, as well as the degradation of unfolded proteins

*(Wilhelmus et al., 2012)*. Although *gain-of-function* (GoF) mutations in *DJ-1* have been related to various types of cancer, biallelic LoF variants in *PARK7* are a rare cause of autosomal recessive EOPD. The clinical presentation generally resembles that associated with *PRKN* and *PINK1,* although non-motor manifestations, including depression, cognitive decline, psychosis, and anxiety have been reported in some patients *(Lunati et al., 2018)*.

### *1.2.3 Mitochondrial and lysosomal genes involved in PD*

Beyond the well-characterized roles *PRKN*, *PINK1* and *PARK7* in mitochondrial quality control and protection against oxidative stress, an increasing number of genetic findings support the involvement of mitochondrial and lysosomal pathways in PD pathogenesis. These findings include both rare pathogenic variants responsible for monogenic forms of PD or atypical parkinsonism and common or rare variants that modify susceptibility to sporadic disease *(Navarro-Romero et al., 2020; Lunati et al., 2018)*.

Mitochondrial involvement in PD was initially supported by the observation that exposure to MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) can induce parkinsonism in humans and animal models MPTP is converted into MPP+, which inhibits complex I of the mitochondrial respiratory chain leading to impaired energy production and increased oxidative stress. Reduced complex I activity has also been observed in the substantia nigra of sporadic PD patients, suggesting that DA neurons appear particularly sensitive to mitochondrial dysfunction. (*Vos, 2022; Malpartida et al., 2020*). Furthermore, high levels of somatic *mitochondrial DNA* (mtDNA) mutations (particularly deletions) have been detected in *substantia nigra* neurons from patients with PD patients, although similar alterations also accumulate during normal ageing *(Pickrell & Youle, 2015)*.

Several additional PD-associated genes are involved in mitochondrial biology. For example, *CHCHD2* encodes a protein located in the mitochondrial intermembrane space that contributes to oxidative phosphorylation and the cellular response to mitochondrial stress. Pathogenic variants in *CHCHD2* have been associated with autosomal dominant PD, particularly in families of Asian ancestry, although these variants are rare. Loss of *CHCHD2* may impair mitochondrial respiration and increase oxidative stress in dopaminergic neurons. In addition, *FBXO7* codes for an F-box-containing protein that interacts functionally with the PINK1-Parkin pathway and contributes to mitochondrial quality control. Biallelic pathogenic variants cause *a rare* juvenile or early-onset parkinsonian-pyramidal syndrome (also called PARK15), characterized by parkinsonism accompanied by pyramidal and, in some cases, psychiatric manifestations *(Jia et al., 2022)*.

Genetic evidence also strongly implicates the autophagy-lysosomal system in PD. Lysosomes are the main cellular compartments responsible for the degradation and recycling of proteins, lipids, and

damaged organelles. Since α-syn is partly cleared through lysosomal pathways, impaired lysosomal function may promote its intracellular accumulation like LBs. Conversely, α-syn aggregates may further compromise lysosomal trafficking and degradation, generating a self-reinforcing pathogenic process *(Bourdenx & Dehay, 2016)*.

Studies investigating lysosomal genes have identified an increased burden of rare potentially damaging variants in some PD cohorts, further supporting the contribution of this pathway. However, the pathogenicity and effect size of individual variants often remain uncertain. *(Milenkovic et al., 2022; Robak et al., 2017)*.

Among the genes involved in lysosomial function, *ATP13A2* and *TMEM175* illustrate two distinct types of genetic contribution. *ATP13A2* is a lysosomal P5-type ATPase, required for normal lysosomal acidification and polyamine homeostasis. Biallelic pathogenic variants in *ATP13A2* cause a complex form of parkinsonism known as *Kufor-Rakeb syndrome* (KRS), characterized by juvenile onset and cognitive impairment *(Jia et al., 2022)*. In contrast, *TMEM175* was identified as a PD risk *locus* through GWASs. It encodes an endolysosomal proton channel that regulates lysosomal pH and proteolytic activity. PD-associated variants reduce channel function, potentially impairing lysosomal degradative capacity, and promoting α-syn accumulation *(Hu et al., 2022)*.

Taken together, the convergence of independently identified PD-associated genes and risk variants on mitochondrial and lysosomal functions indicated that the homeostasis of these organelles is central to PD pathogenesis. Among the genes connecting lysosomal dysfunction to PD, *GBA1* has a particularly prominent role, as its variants represent the most common genetic risk factor for the disease. The structure and function of *GBA1* and its contribution to PD susceptibility are therefore discussed in detail in the following section.

### *1.3 GBA1*

Variants in *GBA1* represent the most common risk factor for PD. The *GBA1* gene encodes β-glucosylceramidase (GCase), a lysosomal enzyme responsible for the hydrolysis of glucosylceramides and glucosylsphingosines, into ceramides or sphingosine and glucose, respectively. This enzyme plays a central role in glycosphingolipid homeostasis and cellular membrane turnover *(Trevisan et al., 2024).* The *GBA1* gene is located on chromosome 1 (1q21) and spans approximately 7.6 kb, with 11 exons in its conventionally described transcript structure ***(Figure 4)***.

A highly homologous pseudogene, commonly referred to as *GBAP1*, is located downstream of *GBA1* and shares approximately 96% sequence identity with the functional gene, particularly within the coding regions. This high degree of homology complicates genetic analysis and promotes

recombination and gene-conversion events that can generate complex pathogenic alleles *(Gan-Or et al., 2018)*.

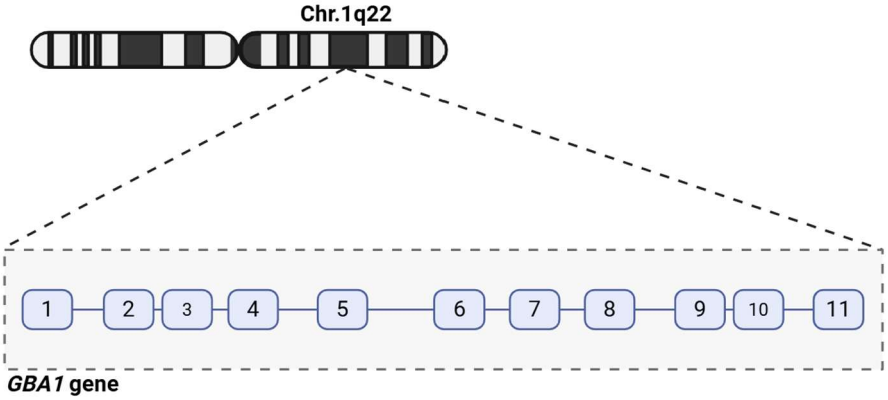



***Figure 4: GBA1 gene structure and its position on chromosome 1, created with Biorender***

GCase is synthesized as a precursor protein and processed into a mature lysosomal enzyme of approximately 497 amino acids.

The mature GCase protein has three structural domains ***(Figure 5)***. Domain I (residues 1–27 and 383–414), contains two disulfide, domain II (residues 30–75 and 431–497) has an immunoglobulin-like structure, and domain III (residues 76–381 and 416–430) forms a TIM (triosephosphate isomerase) barrel domain with catalytic residues Glu235 and Glu340, which are essential for enzyme activity *(Gan-Or et al., 2018)*.

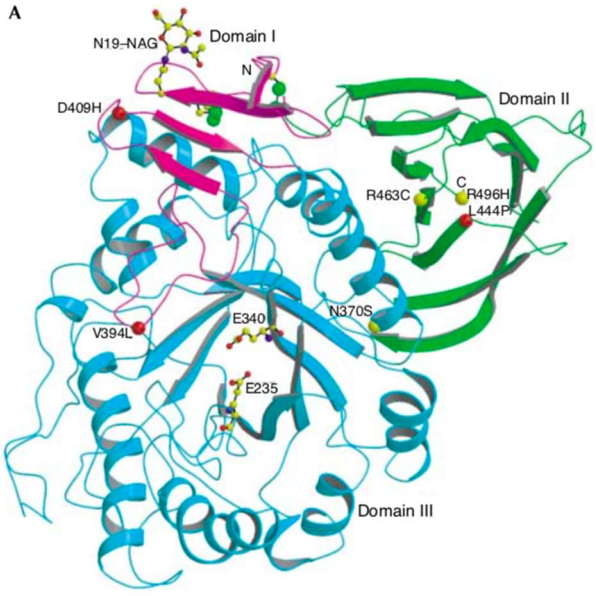



***Figure 5. X-ray structure of β-glucocerebrosidase****. Domain I in pink, domain II in red and domain III, with the active-site residues E235 and E340 (ball and stick model), in blue. Selected common GBA1 variants are also indicated (Smith & Schapira, 2022)*

Biallelic pathogenic variants in *GBA1* cause *Gaucher disease* (GD), one of the most common lysosomal storage disorders. In contrast, individuals carrying a single variant do not display any overt

disease symptoms *(Smith & Schapira, 2022)*. For a long time *GBA1* was linked to GD, but it has been recently correlated to an increased risk of PD. The association between *GBA1* and parkinsonism was first suggested by clinical observations of patients with GD and their relatives and subsequently confirmed in large genetic studies. Depending on the population examined, *GBA1* variants are detected in approximately 5–15% of PD patients with substantially higher frequencies in some populations, including individuals of Ashkenazi Jewish ancestry *(Trevisan et al., 2024)*.

The clinical presentation of *GBA1-associated PD* (*GBA1*-PD) largely overlaps with that of IPD, making two forms indistinguishable at the individual level. However, at the group level, carriers of *GBA1* variants tend to show an earlier age at onset, more rapid motor and cognitive progression, and higher frequency of non-motor symptoms such as cognitive impairment, REM sleep behavior disorder, hallucinations, and autonomic dysfunction. The severity of the clinical phenotype and the magnitude of PD risk vary according to the underlying variant, with severe GD-causing variants generally associated with greater risk and more rapid progression than mild or non-GD-causing risk variants *(Gan-Or et al., 2018; Smith & Schapira, 2022)*.

Several complementary mechanisms have been proposed to explain the association between *GBA1* variants and PD. Reduced GCase activity alters glycosphingolipid metabolism and lysosomal homeostasis, potentially impairing α-syn clearance and promoting its accumulation. In turn, accumulated α-syn interferes with the trafficking of GCase from *endoplasmic reticulum* (ER) to the lysosome and further reduce its lysosomal, generating a bidirectional pathogenic loop *(Mazzulli et al., 2011)*. In addition, some *GBA1* variants lead to the production of misfolded GCase that is retained in the ER, activating ER stress and the unfolded protein response. These mechanisms, including loss of enzymatic activity, altered lipid metabolism, lysosomal dysfunction and toxic effects related to protein misfolding, are not mutually exclusive and may contribute differently according to the specific variant *(Trevisan et al., 2024)*.

A wide spectrum of *GBA1* variants has been described, including missense, nonsense, frameshift, and splice-site variants, as well as deletions, gene conversions, and complex recombinant alleles, arising from the high sequence homology between *GBA1* and *GBAP1*. Missense variants are the most frequently reported *(Smith & Schapira, 2022)*. Importantly, *GBA1* variants show incomplete and age-dependent penetrance for PD. Their effect cannot therefore be summarized by a single risk estimate: mild, severe, and non-GD-causing variants confer different degrees of susceptibility. While the overall increase in PD risk among carriers has often been estimated at approximately 5-fold, variant-specific estimates vary considerably. *(Trevisan et al., 2024)*.

Most carriers never develop PD, and substantial clinical variability is observed even among individuals carrying the same variant and within the same family. This suggests that additional

genetic, environmental, and age-related factors modify disease penetrance and phenotype. In particular, *GBA1*-associated PD may reflect the cumulative burden of dysfunction across the broader lysosomal gene network, rather than the effect of *GBA1* alone *(Straniero et al., 2022).*

### *1.4 Monozygotic twins as a model for genetic studies*

Monozygotic twins discordant for a complex trait provide a particularly informative model for investigating genetic and regulatory factors associated with a complex disease *(Van Dongen et al., 2012)*. Because co-twins share nearly the same inherited genetic background, as well as age, sex, and much of their early-life environment, within-pair comparisons reduce several sources of variability that commonly affect studies of unrelated cases and controls. However, monozygotic twins are not necessarily genetically identical throughout life. Post-zygotic variants may arise during embryonic development or later somatic cell divisions and may therefore be present in only one member of a twin pair. Recent studies have revealed an average of 5.2 early post-zygotic de novo variants per MZ twins *(Vadgama et al., 2019; Dal et al., 2014; Jonsson et al., 2021)*, proving that they are not entirely genetically identical. The identification of variants detected in the affected twin but not in the unaffected co-twin may consequently reveal candidate genetic events potentially contributing to disease discordance.

Genetic differences are not the only potential source of phenotypic discordance between monozygotic twins. Heterozygous variants shared by both co-twins may exhibit allele-specific expression, whereby the two alleles of the same gene are expressed at different levels. Allelic imbalance may result from cis-regulatory variants, epigenetic mechanisms, or other allele-specific regulatory processes. Comparing *allele-specific expression* (ASE) between affected and unaffected co-twins may therefore reveal differential regulatory effects that are not detectable from DNA sequence alone. The combined investigation of within-pair DNA-discordant variants and differential allele-specific expression offers a complementary approach for studying genetic mosaicism and allelic regulation in diseases and in particular in PD.

## 2. AIMS OF THE STUDY

PD is a multifactorial neurodegenerative disorder resulting from the interplay between genetic susceptibility and non-genetic factors, including aging, male sex, environmental exposures, and occupational factors. Rare pathogenic variants have been identified in several genes, while numerous common variants across more than 90 *loci* have been associated with PD susceptibility. Nevertheless, the contribution of rare variants to the broader genetic architecture of sporadic PD remains incompletely understood.

In light of these considerations, the primary aim of this thesis was to investigate the role of rare coding variants to PD susceptibility in a large cohort of unrelated sporadic PD cases and healthy controls. Gene-based association analyses were performed using the optimized *Sequence Kernel Association Test* (SKAT-O) to identify genes showing a differential burden of qualifying rare variants between cases and controls. For prioritized genes, a gene-level odds ratio was also calculated to provide a descriptive indication of the direction of enrichment.

Most individuals carrying *GBA1* variants, the main genetic risk factor of PD, do not develop PD, and substantial variability in disease onset and clinical presentation is observed even among carriers of the same variant. This incomplete penetrance and phenotypic heterogeneity suggest that the effect of *GBA1* may be influenced by additional genetic factors. Moreover, whether PD patients carrying *GBA1* variants are characterized by a rare-variant architecture distinct from that of *GBA1* wild-type patients remains poorly understood. In this frame, the second aim of this thesis was to specifically investigate *GBA1* and potential genetic modifiers. To this end, PD patients carrying *GBA1* variants were compared both with healthy controls and with *GBA1* wild-type PD patients. These analyses were designed to identify genes that could act as potential additional genetic modulators of *GBA1*-associated risk and that may contribute to the heterogeneity observed among patients.

Finally, the study was extended to a cohort of MZ twins discordant for PD as a complementary model for investigating allele-specific regulatory genetic differences associated with disease discordance. Specifically, we aimed to identify candidate post-zygotic variants differing between affected and unaffected co-twins and to determine whether variants shared at the DNA level exhibited differential allele-specific expression. Together, these analyses sought to assess whether genetic mosaicism or differential allelic regulation could contribute to the development of PD in genetically matched individuals.

This study has been conducted in the Laboratory of Medical Genetics and RNA biology at Humanitas University, Pieve Emanuele (Milan), under the supervision of Doctor Letizia Straniero.

# 3. MATERIAL AND METHODS

## *3.1 STUDY COHORTS*

### *3.1.1 Unrelated case-control cohorts*

We studied a total cohort of 1,370 patients affected by PD and 1629 healthy controls, all enrolled at the *Parkinson Institute of Milan* (ASST G. Pini-CTO, Italy) between 2002 and 2023. From this initial population, specific subgroups were defined according to the study design.

This study was approved by the ethics committees of the *"Centro per la Malattia di Parkinson e i Disturbi del Movimento"* and was conducted according to the Declaration of Helsinki and the Italian legislation on sensitive personal data recording. Blood samples were collected during clinical visits after participants had signed written informed consent and PD was diagnosed according to the *UK Brain Bank Criteria (Hughes et al., 2001).* For each participant, the following demographic and clinical data were collected: sex, age at onset (the age at which the patient noticed the first motor symptom), disease duration (calculated on the basis of the last examination), dementia diagnosis (according to the criteria of the *Diagnostic and Statistical Manual of Mental Disorders, 4th edition,* DSM-IV), and family history (patients were classified as familial if at least one relative among their first- or second-degree relatives had a diagnosis of PD).

*Main cohort*

Within the framework of the study population, two main sub-cohorts were established. The primary cohort comprises 1,056 patients affected by PD and 1,588 healthy control subjects, after the exclusion of subjects carrying *GBA1* variants. The demographic and clinical characteristics of the cohort, including sex, age at sample collection, age at onset and other clinical variables, are summarized in ***Table 2***.

| | **Sex** *(male%)* | **Age at sample collection** *(mean±SD)* | **Age at onset** *(mean±SD)* | **Familiar PD** *(%)* |
|---|---|---|---|---|
| **Affected (PD patients) n=1056** | 65% | 54.0±8.5 | 45.0±6.0 | 26.7% |
| **Healthy controls n=1588** | 88% | 39.6±4.9 | - | - |

***Table 2:*** *Characteristics of the main cohort, in which all the GBA carriers have been removed. SD is the abbreviation for standard deviation.*

*GBA1 cohort*

In parallel, a dedicated *GBA1* cohort was created, containing 314 patients carrying *GBA1* mutations (*GBA*-PD) and 41 healthy control carriers. Both groups followed the identical recruitment, clinical evaluation, and data collection protocols described above. Cases and controls were genotyped for

*GBA1* variants using two different approaches. For most subjects, *GBA* variants were identified through an in-house developed allele-specific PCR assay that allows the simultaneous detection of the 4 most common GBA PD-related variants (legacy names: p.E326K, p.T369M, p.N370S, and p.L444P). *(Straniero et al., 2020)*. The residual patients underwent a diagnostic *Next-Generation Sequencing* (NGS) gene panel, while the remaining healthy controls were analyzed by *Whole-Exome Sequencing* (WES) *(Straniero et al., 2022)*. The demographic and clinical characteristics of this cohort are summarized in ***Table 3*** below.

| | Sex (male%) | Age at onset (mean±SD) | Age at sample collection (mean±SD) |
|---|---|---|---|
| **GBA carriers, affected (PD) n=314** | 55% | 54.7±10.9 | 62.0±10.3 |
| **GBA carriers, unaffected (healthy) n=41** | 80% | - | 39.1±4.8 |

***Table 3: Characteristics of the GBA1 cohort, divided by healthy GB1A carriers and GBA1-PD patients.*** *SD is the abbreviation for standard deviation.*

### *3.1.2 Twin cohort*

The twin cohort was recruited at the *Parkinson Institute of Milan* by experts in movement disorders. All 48 participants underwent a careful neurological examination and provided peripheral whole blood samples for DNA extraction, in accordance with the same ethical standards, consent procedures, and diagnostic criteria detailed above.

Data collection followed the same standardized framework of the previous cohorts, with the addition of specific variables such as smoking status, coffee consumption and anxiety. Dementia status was evaluated according to the same DSM-IV criteria, supplemented by *Movement Disorder Society* recommended diagnostic criteria when full neuropsychological testing was available *(Barton et al., 2012)*.

The cohort comprised:

- 23 monozygotic twin pairs discordant for PD;
- one monozygotic triplet composed of one PD patient and two healthy individuals.

***Table 4*** shows the main demographic and clinical characteristics of this particular cohort.

| | | Sex (male%) | Age at sample collection (mean±SD) | Age at onset (mean±SD) | Disease duration (mean±SD) | Smoking (%) | Coffee (%) | Anxiety (%) |
|---|---|---|---|---|---|---|---|---|
| **All twin cohort** | **Affected twins** | 67% | 62.0±11.5 | 54.4±11.5 | 7.6±5.3 | 30% | 86% | 46% |
| | **Healthy twins** | 67% | 62.0±11.5 | - | - | 35% | 86% | 38% |

***Table 4: Characteristics of the twin cohort, divided by healthy individuals and PD patients.*** *SD is the abbreviation for standard deviation.*

## *3.2 SEQUENCING*

### *3.2.1 WES sequencing*

DNA was extracted from peripheral whole blood samples using the *Chemagic Star* workstation (Hamilton, Reno, NV, USA). WES libraries were performed starting from 50 ng of genomic DNA and using the *Agilent SureSelect All Exon v7* (Agilent, Santa Clara, CA, USA) for the twin cohort and *Twist Comprehensive Exome* (Twist Bioscience, San Francisco, CA, USA) for the main cohort, following the manufacturer's instructions. Briefly, genomic DNA was enzymatically fragmented, end-repaired, adenylated, and ligated to sequencing adapters. The resulting libraries were then hybridized with biotinylated probes targeting the protein-coding regions of the genome, captured using streptavidin-coated magnetic beads, PCR-amplified, and purified before sequencing.

Paired-end 150-bp libraries were sequenced on a *NovaSeq6000* platform (Illumina, San Diego, CA, USA). Illumina sequencing protocol involves solid-phase clonal amplification: the target DNA fragments, characterized by two different adapters at the ends, are immobilized on a glass slide to which PCR primers, that are complementary to the adapters, are previously ligated. A bridge PCR is performed: neosynthesized molecules bend in a way that they can pair to primers in the proximity, working as template for the next elongation cycle.

The bridge PCR generates groups of identical molecules ("clusters") obtained from clonal amplification of a single fragment, that will be sequenced. The sequencing reaction is based on reversible terminators, which are dNTPs labeled with fluorochromes and blocked at the 3'. The addition of primer, reversible terminators and polymerase at every cycle permits the incorporation of one nucleotide in the growing strand; the subsequent laser excitation and the acquisition of the emitted fluorescence allows the identification and registration of the first base of every cluster.

At each new sequencing cycle, the blocked 3' of each polymerized nucleotide is unlocked, through removal of the chemical group that blocks the polymerization, and the previously described steps are repeated for a total of 150 cycles *(Illumina, 2010)* ***(Figure 6)***. The method and workflow described above were applied to process and sequence both the main and the twin cohort.

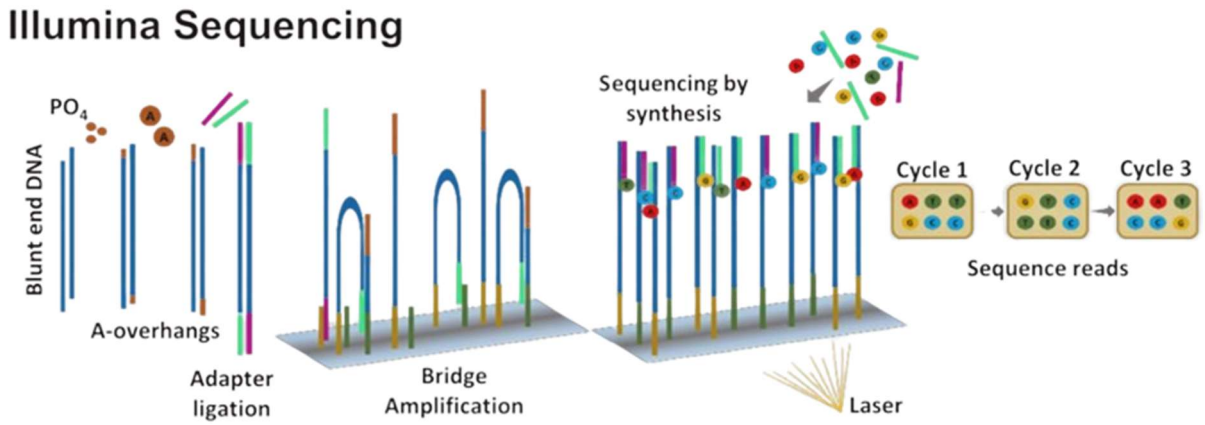



***Figure 6: Overview of the Illumina sequencing-by-synthesis workflow, based on reversible terminators. (Richa Bharti & Grimm, 2019)***

### *3.2.2 RNA sequencing*

RNA samples were extracted from peripheral whole blood of the 22 twin pairs and the trio using the *PAXgene kit* (Qiagen, Venlo, NL), followed by removal of residual genomic DNA via DNase treatment. RNA integrity and concentration were assessed using high sensitivity RNA *ScreenTape assay* with a *4200 Tapestation System* (Agilent Technologies) with RNA Integrity Number (RIN) values used to establish a quality threshold for library preparation eligibility.

Sequencing libraries were subsequently prepared using the *TruSeq Stranded Total RNA Library Prep Kit* (Illumina), following the manufacturer's recommended protocol. Firstly, ribosomal RNA was depleted from total RNA input using *Ribo-Zero rRNA* removal reagents, allowing for the enrichment of both coding and non-coding transcripts.

Following purification, the resulting RNA is fragmented and reverse-transcribed into first-strand cDNA using random primers, followed by second-strand cDNA synthesis incorporating dUTP in place of dTTP to preserve strand specificity.

Double-stranded cDNA fragments underwent end-repair, adenylation of 3' ends and ligation of indexed adapters to allow hybridization onto the flow cell.

Because the polymerase employed in the assay is unable to incorporate past dUTP, the second strand is effectively quenched during the amplification step, allowing only the first strand to be enriched via PCR *(TruSeq Stranded Total RNA Reference Guide, 2017)*.

The resulting products are then purified to create the final cDNA library ***(Figure 7)***. Final libraries were then sequenced using *NovaSeq6000 Platform* (Illumina), generating 150bp paired-end reads necessary for the downstream transcriptomic analyses.

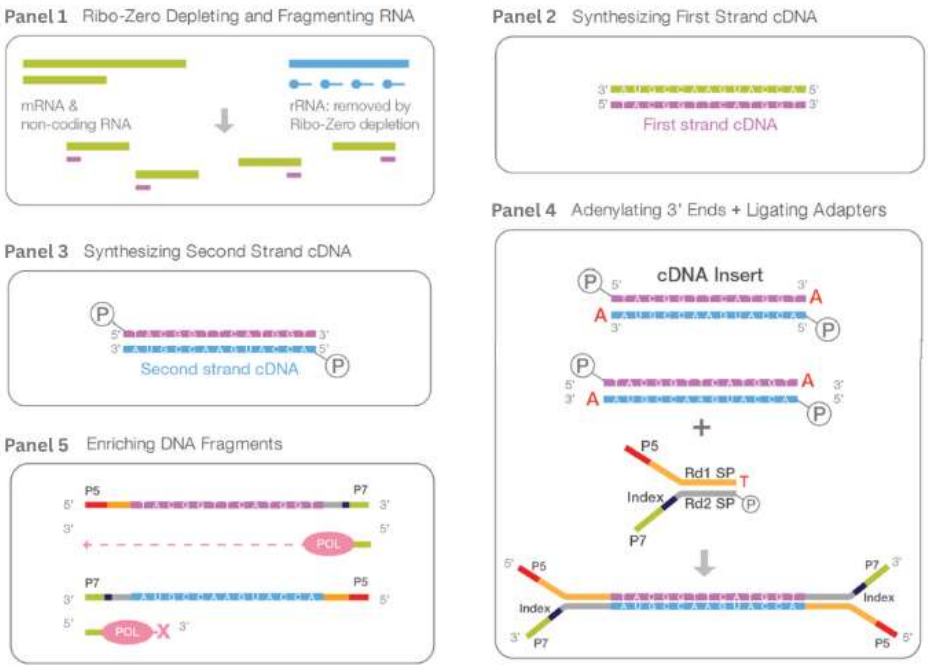



***Figure 7: Overview of RNA-sequencing library preparation using TruSeq Stranded Total RNA Library Prep Kit.***

## *3.3 DATA PROCESSING*

### *3.3.1 Quality control and alignment*

Base call files were converted using the *bcl-convert tool v. 3.8.2* (Illumina) with default parameters. On the resulting FASTQ files, quality control was performed running the *FastQC software (Ewels et al. 2016)*. Next, a multiQC report was created combining and summarizing all the FastQC output files across samples. The trimming of the adapters was performed using Trimmomatic *(Bolger et al., 2014)* and subsequently reads were aligned to the reference genome (GRCh38) based on the nature of the nucleic acid.

Concerning WES, alignment was carried out using *BWA-mem v 0.7.17 (Li et al., 2009)* according to the guidelines for harmonization in genome sequencing analysis *(Regier et al., 2018)*. *BWA-mem* is an algorithm designed for mapping low-divergence sequences against a large reference genome, efficiently handling gapped alignments for paired-end reads *(Li et al., 2009)*. PCR and optical duplicates were tagged using *MarkDuplicates* (Picard tools, http://broadinstitute.github.io/picard). Reads and mapping quality measures were assessed using *Alignment Metrics* (Picard tools).

Conversely, for the RNA sequencing data, after QC and trimming, alignment was performed using *STAR v 2.7.9* (*Spliced Transcripts Alignment to a Reference*), a software designed to specifically address many of the challenges of RNA-seq data mapping, that uses a novel strategy for spliced alignments. The algorithm consists of two major steps: the seed searching step and the clustering/stitching/scoring step; both necessary to handle splice junctions by aligning RNA-seq reads

accurately to the reference genome and mapping them onto exons *(Dobin et al., 2012)*. The resulting BAM files were sorted and indexed using *Samtools v 1.13* Quality control and performance metrics of the RNA-seq alignment were assessed using *RNA-SeQC v 4.0.0.*

### *3.3.2 WES variant calling and joint genotyping*

The *Genome Analysis ToolKit* (GATK) *v 4.2.2.0* best practices were followed, with dataset-specific adaptations *(McKenna et al., 2010; DePristo et al. 2011)*.

Variant calling of WES data from the main and the twin cohort was performed using *GATK HaplotypeCaller (Poplin et al., 2018)*. This algorithm identifies genomic regions showing evidence of variation and performs local de novo assembly of the aligned reads to reconstruct candidate haplotypes. This dynamic graph-based realignment allows to accurately identify both *single nucleotide variants* (SNVs) and small *insertion-deletion* polymorphisms (indels).

*HaplotypeCaller* was run separately for each sample in GVCF mode, generating one genomic VCF (GVCF) file per individual. Unlike a standard VCF file, which reports only positions where a variant is detected, a GVCF also contains information on non-variant genomic regions. This allows homozygous-reference positions to be distinguished from sites for which no reliable genotype can be assigned because of insufficient sequencing coverage or poor data quality.

The GVCF files were then imported using *GATK GenomicsDBImport* and jointly genotyped using *GATK GenotypeGVCFs*. Joint genotyping was performed separately for the main cohort and the twin cohort, generating an independent multisample VCF for each cohort. Briefly, site-level variant information was extracted from each multisample VCF using *GATK MakeSitesOnlyVcf*. *Variant Quality Score Recalibration* (VQSR) was then performed using *GATK VariantRecalibrator*, which applies a gaussian mixture model to estimate the probability that each variant represents a true genetic variant rather than a technical artefact. *HapMap*, *Omni 2.5M SNP array*, *1000 Genomes*, and *dbSNP* datasets were used as reference resources for model training and variant annotation (*The International HapMap Consortium 2003*; Omni 25M SNP array: Omni2.5MSNParray.website; *The 1000 Genomes Project Consortium 2015; Sayers et al. 2024*). The recalibrated probabilities were subsequently applied using *GATK ApplyVQSR*, with a sensitivity threshold of 0.95, to remove variants likely to represent false-positive calls. Multiallelic variants were decomposed into separate biallelic records using *BCFtools*.

### *3.3.3 RNA-seq variant calling*

Variant calling from RNA-sequencing data requires specific preprocessing steps to account for RNA splicing, as conventional callers are primarily designed for DNA-sequencing data. The RNA

alignment files were therefore processed using *GATK SplitNCigarReads.* This tool identifies reads containing an 'N' operator in their CIGAR string, corresponding regions (introns) to be skipped because of splicing, and splits them into individual exon-aligned segments. Intronic overhangs and mismatching regions are hard-clipped and mapping quality scores are adjusted to make the resulting alignments compatible with the downstream variant-calling workflow. Subsequently, SNV variant calling was performed separately for each sample using *GATK HaplotypeCaller* in GVCF mode. Default values were used for base-quality mapping-quality, and phred-scaled confidence thresholds. The individual GVCF files were subsequently imported using *GATK GenomicsDBImport* and jointly genotyped across the twin RNA-seq samples using *GATK GenotypeGVCFs*, generating a multisample VCF.

Given the specific characteristics of RNA-seq data, hard filtering was applied to jointly genotyped using *GATK VariantFiltration*. Variants were flagged when they met at least one of the following criteria: variant *quality* below 30 (QUAL < 30.0), *quality by depth* below 2.0 (QD < 2.0), root mean square *mapping quality* below 40 (MQ < 40.0), *mapping-quality rank sum* below −12.5 (MQRankSum < −12.5), *strand odds ratio* above 3.0 (SOR > 3.0), *Fisher strand* bias above 30.0 (FS > 30.0), or *read-position rank sum* below −8.0 (ReadPosRankSum < −8.0). Clusters of at least three SNVs occurring within a 35-bp window were also flagged as potential false-positive variant hotspots. At the genotype level, calls supported by fewer than 10 reads (DP < 10) were excluded. Only variants and genotypes passing all the specified filters were retained for downstream annotation and analysis.

### *3.3.4 Variant annotation and filtering*

SNVs and indels identified from both DNA- and RNA-sequencing data were annotated using the *ANNOtate VARiation* software (ANNOVAR*, Wang, 2023*). The same annotation workflow was applied separately to the DNA- and RNA-derived variant datasets. Gene-based annotation was performed using the "*ref.Gene*" database, which maps to genes and transcripts and classified them according to their genomic location. and predicted functional consequences, including exonic, splicing, intronic, and intergenic variants. For variants located within coding regions, ANNOVAR was also used to determine the corresponding nucleotide and amino acid changes.

Population allele frequencies were retrieved from the *Genome Aggregation Database* (gnomAD v4.1), considering both genome and exome datasets. Frequencies were evaluated in the overall gnomAD population and in the *non-Finnish European* (NFE) subpopulation.

The NFE-specific frequency was included to provide an ancestry-matched estimate of variant frequency and to improve the identification of variants that are rare in the population most closely resembling the study cohorts. *(Atkinson et al., 2023 and Schmitz et al., 2025)*.

Finally, the potential deleteriousness of each variant was assessed using *Combined Annotation Dependent Depletion* (CADD) score, which integrates multiple genomic annotations into a single score reflecting the relative likelihood that a variant has a functionally deleterious effect.

For downstream analyses, a conservative filtering approach was applied to identify rare potentially deleterious coding variants. In particular, variants were retained when they met all the following criteria: (i) they were classified as "exonic" by the ANNOVAR's *func.refGene* annotation; (ii) their *allele frequency* (AF) was either missing (NA) or strictly below 1% (AFs <0.01) in all the considered gnomAD genome and exome populations, including the overall and NFE groups; and (iii) they had a CADD PHRED score of at least 20. A CADD score of 20 indicates that a variant is among the top 1% of the most deleterious substitutions predicted across the human genome. Variants classified as “unknown” or “synonymous” were excluded from subsequent analyses.

### *3.4 STATISTICAL ANALYSIS*

#### *3.4.1 Study comparison and principal components analysis*

To better address the aims of this study, the initial populations (unrelated cases and *GBA1*, as described above) were analyzed through three specific case-control comparisons. First, *GBA1 wild-type* (wt) PD patients were compared against healthy controls. Second, to investigate the potential role of genetic modifiers *GBA1*-related disease risk, *GBA1* variant-carrier PD patients were compared against the entire control population, rather than against *GBA1*-mutated healthy carriers alone; this choice was dictated by the limited sample size of the latter group (n = 41), which did not provide sufficient statistical power for a dedicated comparison.

Finally, to investigate whether the genetic architecture of PD differs according to *GBA1* mutation status, *GBA1*-mutated PD patients were compared against *GBA1* wt PD patients. This analysis aimed to identify differences in the burden and distribution of rare variants between the two groups independently of comparisons with the healthy controls.

*Principal components analysis*

Once the comparison cohorts were defined, PCA was performed to assess the overall structure of the filtered variant dataset and detect potential sample outliers. PCA is a dimensionality-reducing technique that transforms a high-dimensional dataset into a smaller set of orthogonal variables, known as *principal components* (PCs), while retaining as much of the original variance as possible. The PCs are derived directly from the data rather than defined a priori *(Jolliffe & Cadima, 2016)*.

When applied to genetic data, the first PCs typically capture population structure and may also reflect technical batch effects. Samples that deviate markedly from the main cluster may therefore represent ancestry outliers, sample mix-up, or sequencing and genotyping quality issues *(Privé et al., 2020)*.

PCA was performed separately for each of the three analytical datasets using the filtered variant set obtained as described in Section 3.3.4 and the *prcomp* package implemented in R.

The first two PCs were visualized in scatter plots according to the phenotype considered in each analytical comparison, namely PD status in the case-control datasets and *GBA1* variant status in the comparison between PD subgroups (***Figure 8***).

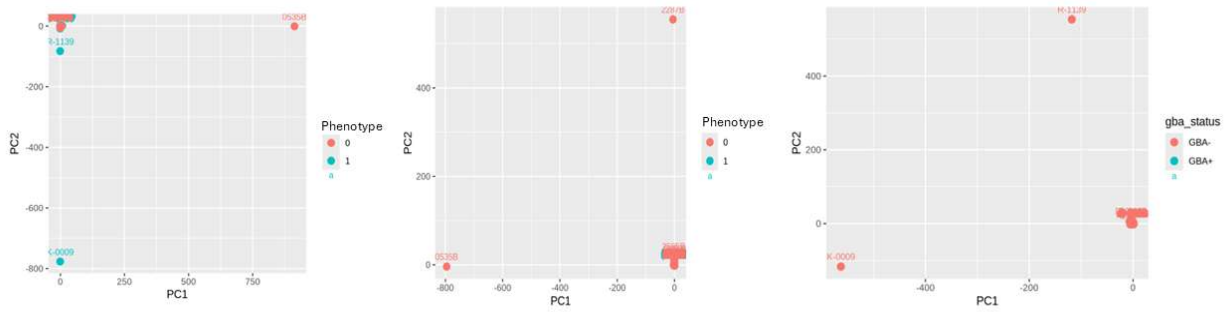



***Figure 8: Principal components analysis.*** *The image shows the PC1 and PC2 of the datasets (A. GBA1 wt PD vs, ctrl, B. GBA1-mutated PD vs ctrl and C. GBA1-mutated cases against GBA1 wt cases. Samples were differentially colored based on their status. Phenotypes: 0=controls; 1=cases. GBA-: GBA1 wt cases; GBA+: GBA-mutated cases.*

The samples that showed a marked deviation from the main cluster along the first two principal components were excluded from the downstream analyses. Specifically, four samples (K-0009, R-1139, 0535B and 2287B) were excluded because they clearly deviated from the main cluster along the first two principal components. After this quality-control step, the *GBA1* wt case-control cohort comprised 1,055 PD cases and 1,587 healthy controls. The *GBA1* variant-carrier case-control cohort comprised 314 cases and 1586 healthy individuals. Finally, the direct comparison between *GBA1* variant carriers and non-carriers included 314 carriers and 1,054 *GBA1* wt patients.

### *3.4.2 Rare-variant association analyses in the main and GBA1 cohorts*

Rare-variant association analyses in the main exome and *GBA1* cohorts were performed using the burden test and the *optimized Sequence Kernel Association Test* (SKAT-O), implemented in the R package SKAT.

The burden test evaluates whether rare genetic variants within a defined genomic region, such as a gene, are collectively associated with a disease or trait. In this approach, multiple variants are collapsed into a single genetic score and the frequency of individuals carrying rare, potentially deleterious variants in each gene is compared between cases and controls. Aggregating variants at gene or pathway level increases statistical power compared with single-variant analyses, which are generally underpowered when variants are rare *(Guo et al., 2018)*. To perform the burden test, the R package SKAT was employed, using the "*r.corr*" parameter equal to 1, as shown in ***Figure 9***.

```
> #rho=1, Burden test
> SKAT(Z, obj, r.corr=1)$p.value
```

***Figure 9: Example of a basic script to run a Burden test using SKAT package on Rstudio (Lee, 2023)***

As depicted in the figure above, the p-value was the selected parameter for evaluating statistical significance, as it estimates the probability that the relationship between the burden of rare variants and the investigated trait occurred by chance, thus ensuring that significant genes enriched with deleterious mutations are easily identifiable without false positives.

SKAT test is a kernel-based test that, rather than collapsing multiple variants into a single score, evaluates their combined association with the phenotype through a kernel matrix. It is particularly powerful when variants within the same genomic region have effects of different magnitudes or directions, or when a large proportion of the analyzed variants are non-causal. The SKAT test is formulated as a variance-component test within a mixed-model framework, in which the regression coefficients representing the effects of individual variants are assumed to be independent and to follow a distribution with a common variance component *(S. Lee et al., 2012)*.
The optimal test, SKAT-O, adaptively combines the burden test and the standard SKAT test, as illustrated in ***Figure 10***. This approach allows the test to maintain high statistical power across various genetic architectures, including scenarios in which most rare variants have effects in the same direction, where burden tests are generally more powerful, or scenarios involving a mixture of risk, protective, or neutral variants within the same gene, where SKAT is expected to perform better *(Lee et al., 2014)*.

$$Q_\rho = (1-\rho)Q_S + \rho Q_B$$

***Figure 10: SKAT-O, a combined test of burden test and SKAT.*** *The formula illustrates how the SKAT-O test combines the burden and SKAT test statistics. QS is a test statistic of SKAT, and QB is a score test statistic of the burden test. The parameter ρ determines the relative contribution of the two components, with ρ =0 corresponding to SKAT and ρ =1 corresponding to the burden test (Lee, 2023)*

SKAT-O was implemented using the SKAT R package, setting the method parameter to "*SKATO*" (method="SKATO"). As described above for the Burden test, the p-value was used to evaluate the evidence of an association between the cumulative effect of rare variants within each gene and the phenotype under investigation.

For the main and *GBA1* cohorts, the analyses were performed using both the burden test and the SKAT-O. Each analysis was conducted under four distinct covariate specifications: (i) no covariate adjustment, (ii) adjustment for sex alone, (iii) adjustment for age alone, and (iv) joint adjustment for both sex and age. For each covariate specification, a null model was fitted using the

"*SKAT_Null_Model"* function and the resulting null model object was subsequently used to test each gene set with both the burden and SKAT-O tests. The p-values obtained under each covariate specification were corrected for multiple testing across gene sets using the Benjamini-Hochberg *False Discovery Rate* (FDR) method. The correction was applied separately within each covariate specification, yielding both unadjusted and FDR-adjusted p-values for downstream interpretation.

### *3.4.3 Rare-variant association analyses in the twin cohort*

Before performing the rare-variant association analyses, variants shared by both members of each monozygotic twin pair were excluded. Only within-pair discordant variants, defined as variants detected in one twin but not in the corresponding co-twin, were retained for subsequent analyses. This filtering step was applied to focus the analysis on genetic differences potentially contributing to phenotypic discordance between co-twins.

Rare-variant association analyses in the twin cohort were performed using the *Set Mixed Model Association Tests* (SMMAT). Unlike standard burden and SKAT approaches, which assume that individuals are unrelated, uses a generalized linear mixed-model framework to account for population structure and familial or cryptic relatedness. In this framework, correlations among samples are modeled through a genetic relationship matrix included in the model as a random effect. This approach reduces the risk of false-positive associations caused by sample relatedness while assessing the joint effect of rare variants within specific regions of the genome *(Chen et al., 2019)*.

SMMAT includes four different test strategies: SMMAT-B, a burden test that assumes consistent directions of effect among variants; SMMAT-S, a variance-component test analogous to SKAT; SMMAT-O, an optimally adaptive test that combines burden test and SKAT; and SMMAT-E, an efficient hybrid test combining burden and SKAT statistics. All four approaches were applied to account for different possible distributions and directions of causal variant effects. The resulting p-values were used to evaluate the evidence of an association between each variant set and the phenotype under investigation *(Chen et al., 2019)*.

The correlation between monozygotic co-twins was accounted for through a kinship matrix included in the mixed model as a random effect. This approaches the within-pair correlation to be modeled through the variance-covariance structure of the data. As for the main and *GBA1* cohorts, four covariate specifications listed above were considered for testing and the obtained p-values were adjusted separately within each covariate specifications, using the FDR method.

### *3.4.4 Odds ratio calculation*

To complement the SKAT-O results with a descriptive estimate of the direction and the magnitude of variants' enrichment, an allele-based *odds ratio* (OR) was calculated for each prioritized gene.

Although SKAT-O evaluates the aggregated association of multiple variants within a gene, its p-value does not directly indicate whether qualifying alternative alleles are more frequent in one group than in the other.

For each gene, the total number of alternative alleles was calculated separately in the two groups and each of them divided by the total number of alleles in a specific group (2n). The OR for each gene was then defined as the ratio between these two quantities:

$$odds\ ratio = \frac{odds\ of\ \ group\ 1}{odds\ of\ \ group\ 2}$$

***Figure 11: OR formula.*** *The formula illustrates how the OR was calculated for each prioritized gene from the SKAT-O analysis. Group 1 is represented by PD patients, while group 2 by their healthy counterparts.*

An OR greater than 1 indicates that variants an enrichment of alternative alleles in group 1, whereas an OR below 1 indicates their relative depletion. $Log_2$-transformed OR values were used for visualization to represent enrichment and depletion symmetrically around zero.

### *3.4.5 Enrichment analysis*

Exploratory functional enrichment analyses were performed on the gene sets generated from both the unrelated cohorts and the monozygotic twin cohort.

For each comparison involving unrelated individuals, genes showing nominal significance evidence of association in the SKAT-O analysis were selected.

When 100 or fewer genes reached an unadjusted ($p < 0.05$), all nominally associated genes were included; when more than 100 genes met this criterion, the 100 genes with the lowest SKAT-O p-values were retained. For the monozygotic twin cohort, enrichment analyses were performed separately on: (i) genes harbouring qualifying DNA variants detected in affected twins but not in their unaffected co-twins; (ii) genes harbouring qualifying variants detected in unaffected co-twins but not in the corresponding affected twins; and (iii) genes containing shared heterozygous variants showing differential allele-specific expression between affected and unaffected co-twins.

Functional enrichment was performed using three different platforms: QIAGEN Ingenuity Pathway Analysis (IPA; ipa.qiagen.website*; Krämer et al. 2014*), Enrichr (enrichr.website; *Chen et al. 2013*; *Kuleshov et al. 2016; Xie et al. 2021*), and Metascape (metascape.website*; Zhou et al. 2019*). IPA was employed to identify curated canonical pathways and disease-associated interaction networks based on the *Ingenuity Knowledge Base.*

Enrichr was utilized to perform over-representation analysis across diverse annotation libraries, including *Gene Ontology*, *KEGG* and *Reactome*. Lastly, Metascape was used to combine functional annotation from multiple databases and to cluster enriched terms into functional groups and biological modules *(E. Y. Chen et al., 2013; Zhou et al., 2019)*. All analytical tools were executed under default algorithmic settings.

# 4. RESULTS

## *4.1 RARE-VARIANT ASSOCIATION ANALYSES IN UNRELATED COHORTS*

### *4.1.1 GBA1 wild-type PD patients versus healthy controls*

*Gene-based association results*

Rare-variant association analysis was performed using the SKAT-O test to compare *GBA1 wild-type* PD patients with healthy controls, adjusting for sex and age. A total of 14,817 genes were tested. The genome-wide distribution of the resulting gene-level p-values is shown in the Manhattan plot in ***Figure 12***. Overall, 723 genes showed nominal evidence of association ($p < 0.05$), as indicated by the yellow line. After correction for multiple testing, 25 genes remained significant at an FDR-adjusted (p-value < 0.05; ***Table 5***), of which 11 also exceeded the more stringent Bonferroni threshold indicated by the red dashed line.

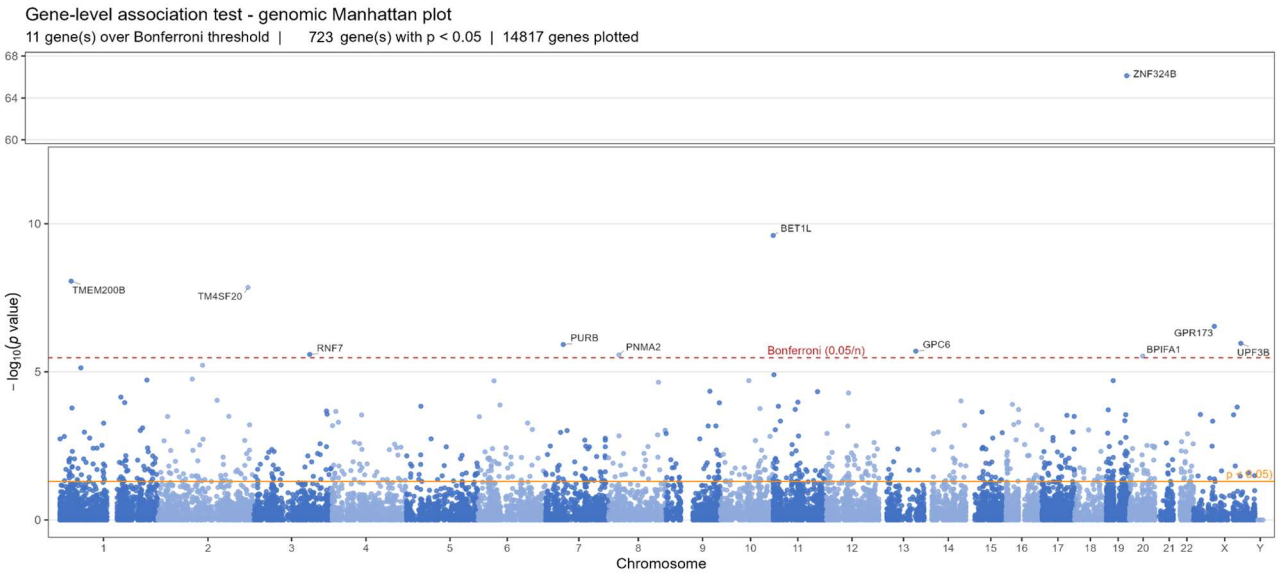



***Figure 12: Genomic Manhattan plot of the gene-based SKAT-O analysis comparing GBA1 wild-type PD patients with healthy controls.*** *Each point represents a gene plotted according to its genomic position (x-axis) and the -log₁₀(p value) of the association test (y-axis). The solid yellow line indicates the nominal significance threshold (p=0.05), whereas the dashed red line indicates the Bonferroni-corrected threshold (p=0.05/n, corresponding to -log₁₀(0.05/n) on the plotted scale), where (n) is the total number of genes tested. Genes exceeding the Bonferroni threshold are labelled. The y-axis is discontinuous to accommodate the markedly stronger association signal observed for ZNF324B.*

| Gene | Chr | p-value | Adj. p-value |
|---|---|---|---|
| *ZNF324B* | chr19 | $7.51 \times 10^{-67}$ | $1.12 \times 10^{-62}$ |
| *BET1L* | chr11 | $2.51 \times 10^{-10}$ | $1.25 \times 10^{-6}$ |
| *TMEM200B* | chr1 | $8.68 \times 10^{-9}$ | $3.24 \times 10^{-5}$ |
| *TM4SF20* | chr2 | $1.42 \times 10^{-8}$ | $4.24 \times 10^{-5}$ |
| *GPR173* | chrX | $2.92 \times 10^{-7}$ | $7.26 \times 10^{-4}$ |
| *UPF3B* | chrX | $1.09 \times 10^{-6}$ | 0.0022 |
| *PURB* | chr7 | $1.19 \times 10^{-6}$ | 0.0022 |
| *GPC6* | chr13 | $2.01 \times 10^{-6}$ | 0.0033 |
| *RNF7* | chr3 | $2.59 \times 10^{-6}$ | 0.0036 |
| *PNMA2* | chr8 | $2.68 \times 10^{-6}$ | 0.0036 |
| *BPIFA1* | chr20 | $2.95 \times 10^{-6}$ | 0.0037 |
| *CKAP2L* | chr2 | $6.02 \times 10^{-6}$ | 0.0069 |
| *YIPF1* | chr1 | $7.34 \times 10^{-6}$ | 0.0078 |

| Gene | Chr | p-value | Adj. p-value |
|---|---|---|---|
| *INS* | chr11 | $1.25 \times 10^{-5}$ | 0.0123 |
| *MFSD5* | chr12 | $1.32 \times 10^{-5}$ | 0.0123 |
| *CHMP3* | chr2 | $1.76 \times 10^{-5}$ | 0.0144 |
| *MTARC2* | chr1 | $1.89 \times 10^{-5}$ | 0.0144 |
| *ASCC1* | chr10 | $1.98 \times 10^{-5}$ | 0.0144 |
| *NIBAN3* | chr19 | $1.98 \times 10^{-5}$ | 0.0144 |
| *LRFN2* | chr6 | $2.02 \times 10^{-5}$ | 0.0144 |
| *LRATD2* | chr8 | $2.25 \times 10^{-5}$ | 0.0153 |
| *ZNF483* | chr9 | $4.53 \times 10^{-5}$ | 0.0291 |
| *PTS* | chr11 | $4.68 \times 10^{-5}$ | 0.0291 |
| *RNF41* | chr12 | $5.19 \times 10^{-5}$ | 0.0310 |
| *ZBTB7B* | chr1 | $7.13 \times 10^{-5}$ | 0.0409 |

***Table 5: Genes remaining significant after multiple-testing correction.*** *For each gene, the table reports: HUGO gene symbol, chromosome, nominal SKAT-O p-value, and FDR-adjusted p value.*

Among the genes remaining significant after FDR correction, *ZNF324B* showed by far the strongest association signal (FDR-adjusted p = $7.51 \times 10^{-7}$). This was followed by *BET1L*, *TMEM200B*, and *TM4SF20*, which also displayed highly significant gene-level associations.

The significant signals were distributed across multiple chromosomes, rather than being confined to a single genomic region.

Several of the identified genes are involved in intracellular membrane trafficking and protein-homeostasis-related processes, including *BET1L*, *YIPF1*, *CHMP3*, *RNF41*, and *RNF7*, whereas others are implicated in transcriptional or post-transcriptional regulation, including *ZNF324B*, *UPF3B*, *PURB*, and *ASCC1*. These potential functional relationships were further investigated through downstream enrichment analyses.

Subsequently, for each gene an OR was calculated to determine whether qualifying alternative alleles were relatively enriched in PD patients or healthy controls. The ratios were $\log_2$-transformed to represent enrichment and depletion symmetrically around zero.

Among this gene subset, the $\log_2$(OR) values ranged from 4.05 (*GPR173*) to -3.11 (*ATG16L2*), indicating a wide range of effect directions across the gene set. Overall, 43 genes showed a positive $\log_2$(OR), consistent with an enrichment of rare variants in cases compared to controls, while 57 genes showed a negative $\log_2$(OR) (***Figure 13***).

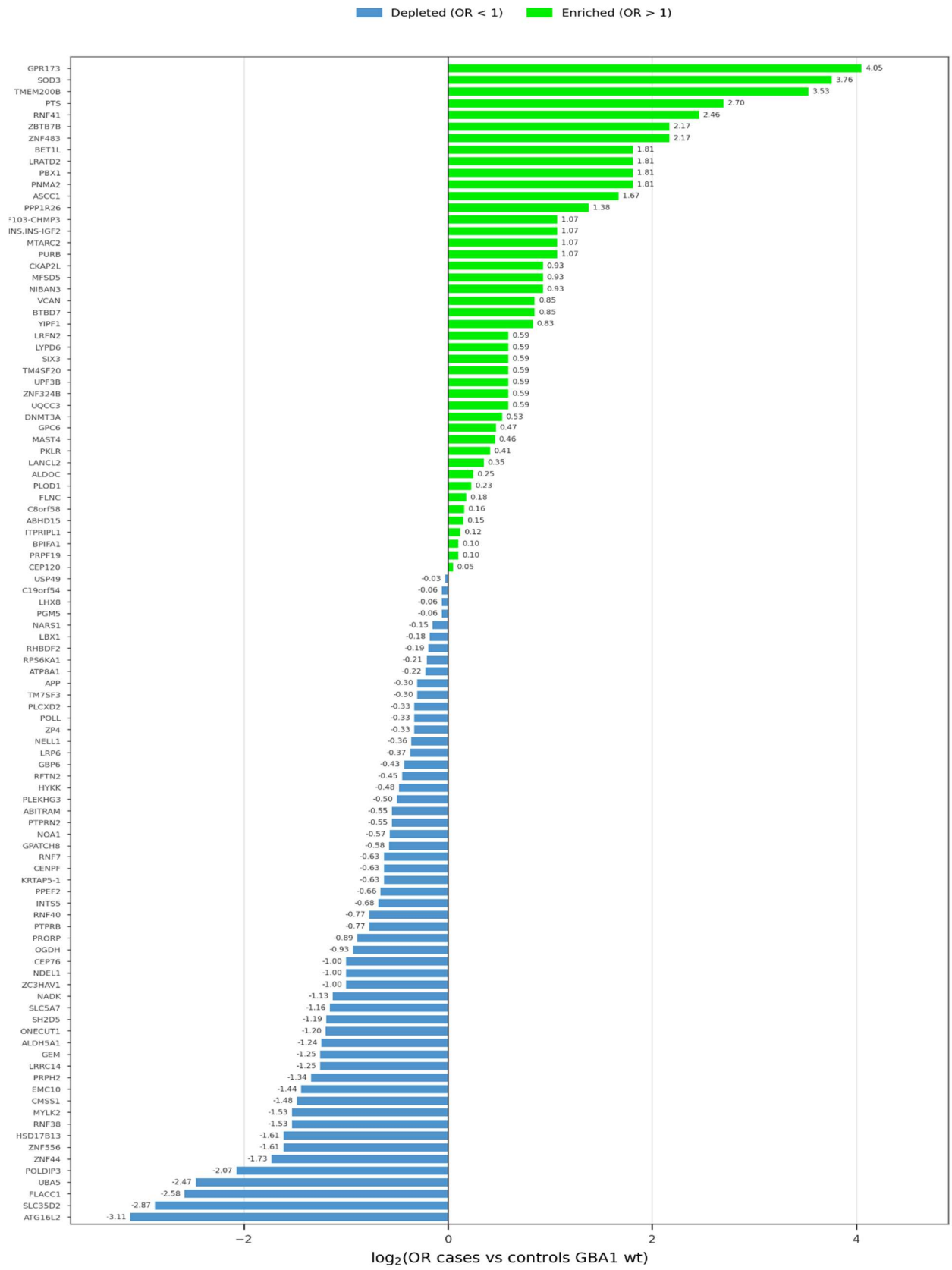



***Figure 13: Log₂(OR) values for the 100 genes with lowest p values, ranked in decreasing order.*** *Bars in green represent genes enriched for rare variants in cases (OR > 1), while bars in blue represent a variants' burden stronger in controls rather than in cases (OR < 1).*

This distribution illustrates a minor tendency of rare variants to be more frequent, or to carry greater weight, in controls rather than in cases, as reflected by the slightly higher number of genes showing a negative $\log_2$(OR) rather than a positive one. On the other hand, as it is shown by ***Figure 14***, the 25 genes that survived FDR correction show a $\log_2$(OR) values distribution clearly shifted toward an enrichment in cases rather than in controls.

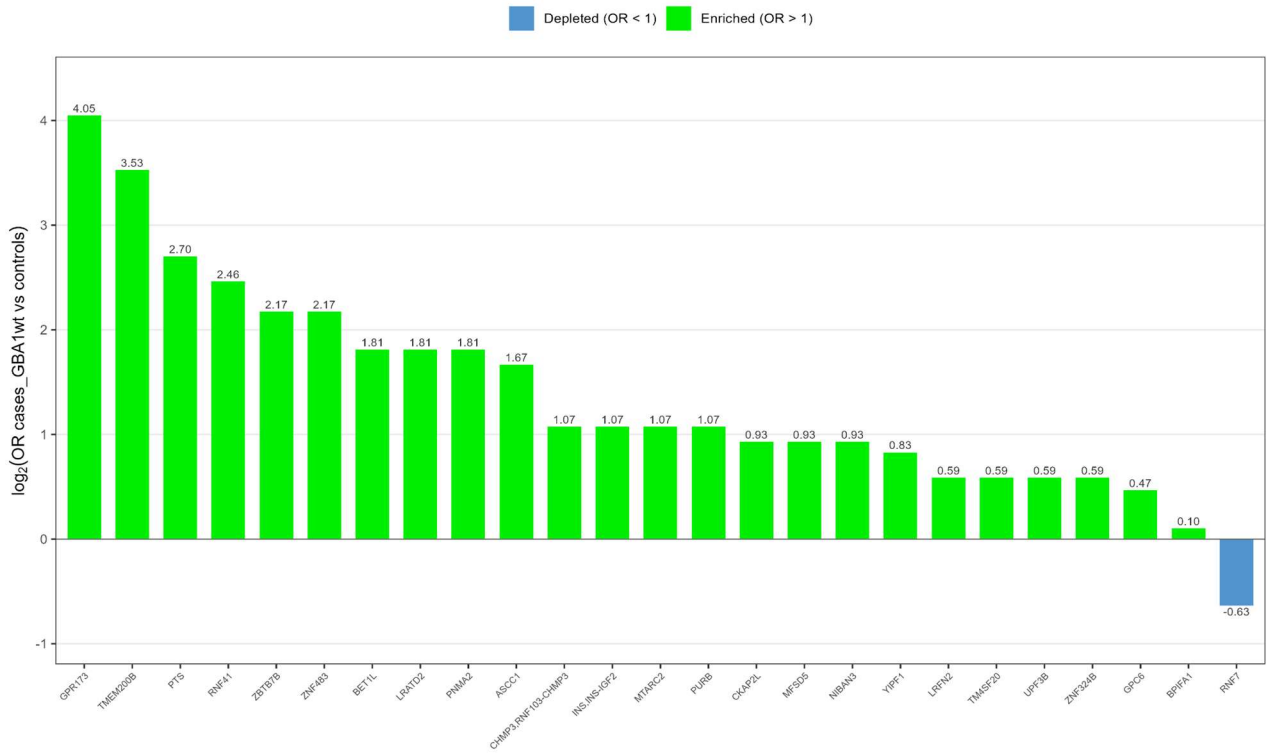



***Figure 14: Log$_2$(OR) values for the 25 that remained significant after FDR correction.*** *Genes are ranked by decreasing log$_2$OR. Positive values, shown in green, indicate relative enrichment of qualifying alternative alleles in case, while negative values, shown in blue, indicate relative enrichment in controls.*

The genes with the strongest enrichment in cases were *GPR173*, *TMEM200B, PTS, RNF41, ZBTB7B* and *ZNF483* ($\log_2$(OR) equal to 4.05, 3.53, 2.70, 2.46 and 2.17, respectively), whereas the gene showing depletion was *RNF7* ($\log_2$(OR) equal to -0.63); while the majority of genes display a $\log_2$(OR) value between 0 and +1.

*Functional enrichment analysis*

To ensure methodological consistency and comparability across all cohort comparisons, functional enrichment analysis was performed using the 100 top-ranked genes based on their SKAT-O p-values (***Table 6***).

| Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *ZNF324B* | $7.51 \times 10^{-67}$ | $1.12 \times 10^{-62}$ | *LYPD6* | $9.14 \times 10^{-5}$ | 0.0506 | *SESTD1* | $3.18 \times 10^{-4}$ | 0.0888 | *GOLT1A* | $9.54 \times 10^{-4}$ | 0.1839 |
| *BET1L* | $2.51 \times 10^{-10}$ | $1.25 \times 10^{-6}$ | *BTBD7* | $9.56 \times 10^{-5}$ | 0.0510 | *RFNG* | $3.20 \times 10^{-4}$ | 0.0888 | *VSTM2A* | $9.60 \times 10^{-4}$ | 0.1839 |
| *TMEM200B* | $8.68 \times 10^{-9}$ | $3.24 \times 10^{-5}$ | *UQCC3* | $1.06 \times 10^{-4}$ | 0.0536 | *PFN4* | $3.22 \times 10^{-4}$ | 0.0888 | *SHARPIN* | $1.02 \times 10^{-3}$ | 0.1920 |
| *TM4SF20* | $1.42 \times 10^{-8}$ | $4.24 \times 10^{-5}$ | *PBX1* | $1.09 \times 10^{-4}$ | 0.0536 | *BPHL* | $3.27 \times 10^{-4}$ | 0.0888 | *PCGF1* | $1.05 \times 10^{-3}$ | 0.1957 |
| *GPR173* | $2.92 \times 10^{-7}$ | $7.26 \times 10^{-4}$ | *PPP1R26* | $1.11 \times 10^{-4}$ | 0.0536 | *FKRP* | $4.47 \times 10^{-4}$ | 0.1191 | *RALGAPA1* | $1.07 \times 10^{-3}$ | 0.1957 |
| *UPF3B* | $1.09 \times 10^{-6}$ | 0.0022 | *CEP20* | $1.26 \times 10^{-4}$ | 0.0587 | *UEVLD* | $4.60 \times 10^{-4}$ | 0.1198 | *SOCS3* | $1.08 \times 10^{-3}$ | 0.1957 |
| *PURB* | $1.19 \times 10^{-6}$ | 0.0022 | *GFRAL* | $1.31 \times 10^{-4}$ | 0.0595 | *WAS* | $4.65 \times 10^{-4}$ | 0.1198 | *ANGPTL3* | $1.09 \times 10^{-3}$ | 0.1957 |
| *GPC6* | $2.01 \times 10^{-6}$ | 0.0033 | *BTBD10* | $1.45 \times 10^{-4}$ | 0.0621 | *TAPT1* | $5.04 \times 10^{-4}$ | 0.1262 | *STARD3NL* | $1.11 \times 10^{-3}$ | 0.1971 |
| *RNF7* | $2.59 \times 10^{-6}$ | 0.0036 | *SLC1A3* | $1.45 \times 10^{-4}$ | 0.0621 | *ARMC5* | $5.07 \times 10^{-4}$ | 0.1262 | *DUSP14* | $1.12 \times 10^{-3}$ | 0.1973 |
| *PNMA2* | $2.68 \times 10^{-6}$ | 0.0036 | *CHRDL1* | $1.55 \times 10^{-4}$ | 0.0642 | *TPD52L1* | $5.38 \times 10^{-4}$ | 0.1307 | *SLCO3A1* | $1.14 \times 10^{-3}$ | 0.1979 |
| *BPIFA1* | $2.95 \times 10^{-6}$ | 0.0037 | *SNRNP40* | $1.66 \times 10^{-4}$ | 0.0671 | *CHIA* | $5.43 \times 10^{-4}$ | 0.1307 | *MFSD12* | $1.17 \times 10^{-3}$ | 0.2010 |
| *CKAP2L* | $6.02 \times 10^{-6}$ | 0.0069 | *PKD2L1* | $1.74 \times 10^{-4}$ | 0.0683 | *TMC5* | $6.01 \times 10^{-4}$ | 0.1426 | *CACNA1C* | $1.20 \times 10^{-3}$ | 0.2018 |
| *YIPF1* | $7.34 \times 10^{-6}$ | 0.0078 | *OR5L1* | $1.86 \times 10^{-4}$ | 0.0699 | *NMUR1* | $6.16 \times 10^{-4}$ | 0.1437 | *THTPA* | $1.21 \times 10^{-3}$ | 0.2018 |
| *INS* | $1.25 \times 10^{-5}$ | 0.0123 | *COX6A2* | $1.88 \times 10^{-4}$ | 0.0699 | *ZNF839* | $6.37 \times 10^{-4}$ | 0.1443 | *DMRT1* | $1.22 \times 10^{-3}$ | 0.2018 |
| *MFSD5* | $1.32 \times 10^{-5}$ | 0.0123 | *ARRDC5* | $1.92 \times 10^{-4}$ | 0.0699 | *VAT1L* | $6.38 \times 10^{-4}$ | 0.1443 | *RBFOX2* | $1.23 \times 10^{-3}$ | 0.2020 |
| *CHMP3* | $1.76 \times 10^{-5}$ | 0.0144 | *HTR3D* | $2.11 \times 10^{-4}$ | 0.0751 | *UVSSA* | $6.56 \times 10^{-4}$ | 0.1443 | *MTSS1* | $1.34 \times 10^{-3}$ | 0.2178 |
| *MTARC2* | $1.89 \times 10^{-5}$ | 0.0144 | *DRD5* | $2.18 \times 10^{-4}$ | 0.0757 | *ZBTB34* | $6.73 \times 10^{-4}$ | 0.1443 | *ADRA1A* | $1.45 \times 10^{-3}$ | 0.2331 |
| *ASCC1* | $1.98 \times 10^{-5}$ | 0.0144 | *PPP1R14D* | $2.27 \times 10^{-4}$ | 0.0770 | *NFE2* | $6.74 \times 10^{-4}$ | 0.1443 | *VEGFB* | $1.47 \times 10^{-3}$ | 0.2331 |
| *NIBAN3* | $1.98 \times 10^{-5}$ | 0.0144 | *SENP2* | $2.69 \times 10^{-4}$ | 0.0872 | *RAD23B* | $6.76 \times 10^{-4}$ | 0.1443 | *UROS* | $1.51 \times 10^{-3}$ | 0.2367 |
| *LRFN2* | $2.02 \times 10^{-5}$ | 0.0144 | *RAI2* | $2.76 \times 10^{-4}$ | 0.0872 | *IRF6* | $7.81 \times 10^{-4}$ | 0.1643 | *NPPA* | $1.53 \times 10^{-3}$ | 0.2377 |
| *LRATD2* | $2.25 \times 10^{-5}$ | 0.0153 | *SNRNP70* | $2.81 \times 10^{-4}$ | 0.0872 | *IFNGR1* | $8.83 \times 10^{-4}$ | 0.1815 | *PRRG2* | $1.63 \times 10^{-3}$ | 0.2497 |
| *ZNF483* | $4.53 \times 10^{-5}$ | 0.0291 | *ARMCX2* | $2.82 \times 10^{-4}$ | 0.0872 | *MC1R* | $8.87 \times 10^{-4}$ | 0.1815 | *UNC119* | $1.64 \times 10^{-3}$ | 0.2497 |
| *PTS* | $4.68 \times 10^{-5}$ | 0.0291 | *PARM1* | $2.86 \times 10^{-4}$ | 0.0872 | *OR52B6* | $9.02 \times 10^{-4}$ | 0.1818 | *ST6GALNAC3* | $1.71 \times 10^{-3}$ | 0.2568 |
| *RNF41* | $5.19 \times 10^{-5}$ | 0.0310 | *KCNH6* | $2.94 \times 10^{-4}$ | 0.0879 | *ZNF397* | $9.13 \times 10^{-4}$ | 0.1818 | *PIAS1* | $1.72 \times 10^{-3}$ | 0.2568 |
| *ZBTB7B* | $7.13 \times 10^{-5}$ | 0.0409 | *C2orf50* | $3.09 \times 10^{-4}$ | 0.0888 | *ZNF707* | $9.30 \times 10^{-4}$ | 0.1829 | *ABCB8* | $1.75 \times 10^{-3}$ | 0.2584 |

***Table 6: Table of the 100 top-ranking genes in the GBA1 wt cohort.*** *The table illustrates the 100 genes with the lowest nominal p-value within the cohort. For each gene the HUGO name, the nominal p-value, the p-value after FDR correction are shown.*

Enrichment analysis was performed using the QIAGEN IPA and EnrichR, with the subset of genes described above as input. The IPA analysis (***Figure 15***) identified 13 significantly enriched pathways ($-\log_{10}$(p-value) > 1.3). Among these, interferon γ signaling was the most significantly enriched pathway (p=0.0005). Additionally, another immunity related pathway, the Role of JAK family kinases in IL-6-type Cytokine Signaling, was identified as the second most significant one.

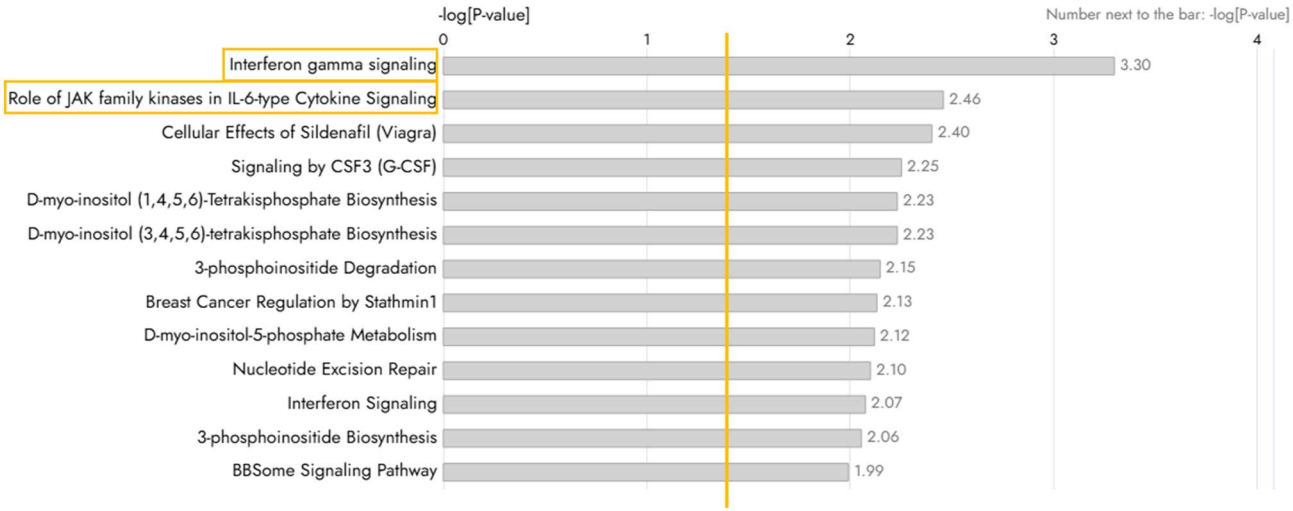
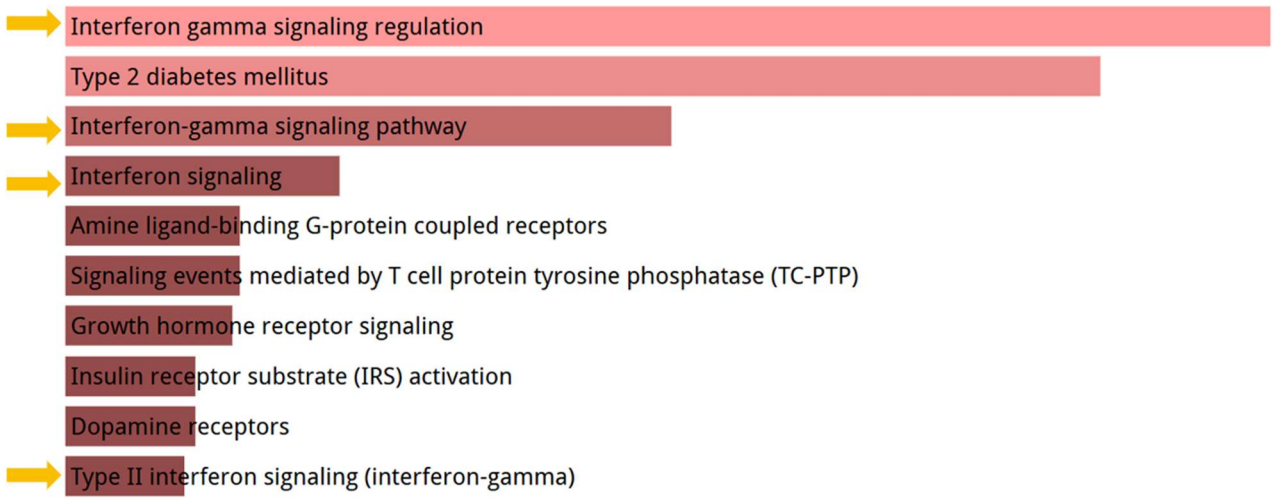



***Figure 15: IPA enrichment analysis.*** *The figure illustrates the 13 most statistically significant enriched pathways among the 100 top-ranked genes from the SKAT-O analysis. The yellow line indicates the significance threshold, corresponding to a $-\log_{10}$(p-value) of 1.3. Immunity related pathways are highlighted.*

Similar results were obtained using EnrichR (***Figure 16***). Regulation of INF-γ signalling from the BioPlanet 2019 database was significantly enriched after multiple-testing correction (q-value < 0.05).

Interferon γ signaling was also identified, although its enrichment was significant only at the nominal p-value level. Among the 100 genes included in the analysis, four were associated with these pathways: *SOCS3*, *IFNGR1*, *PIAS1* and *IRF6,* whereas *IRF6* was annotated only in the interferon γ signaling pathway.

(A)




(B)

| Pathway | p value | q value | Genes involved |
|---|---|---|---|
| Interferon gamma signaling regulation | 0.000045 | 0.01035 | *SOCS3, IFNGR1, PIAS1* |
| Interferon gamma signalling pathway | 0.001536 | 0.1178 | *SOCS3, IFNGR1, PIAS1, IRF6* |
| Interferon signaling | 0.01079 | 0.3893 | *SOCS3, IFNGR1, PIAS1, IRF6* |
| Type II interferon signaling (interferon gamma) | 0.02692 | 0.383 | *SOCS3, IFNGR1* |

***Figure 16: EnrichR enrichment analysis.*** *(A) graph representing the enriched pathways among the 100 top-ranked genes identified in the GBA1 wt cohort, highlighting the most statistically significant pathway, the Interferon γ signaling regulation. (B) Table summarizing the results for the immunological pathway, including the pathway name, associated p-values, q-values, and the genes contributing to each pathway.*

### *4.1.2 GBA1 variant-carrier PD patients versus healthy controls*

*Gene-based association results*

Rare-variant association analysis was performed in the cohort comprising *GBA1* variant-carrier PD patients and healthy controls, following the same approach described for the *GBA1* wt cohort.

The distribution of the analyzed genes across chromosomes according to their $-\log_{10}$(p-value) is displayed in the Manhattan plot below (***Figure 17***). The analysis identified 404 genes with a nominal significant p-value (p-value < 0.05); however, none remained significant after multiple-testing correction.

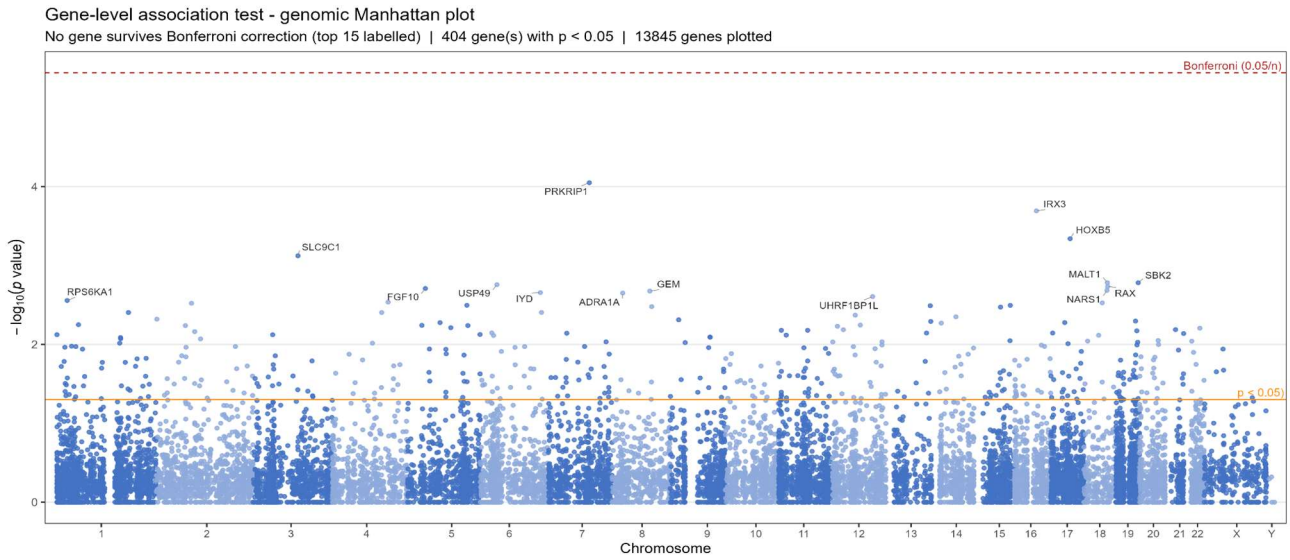



***Figure 17: Manhattan plot of the gene-based SKAT-O analysis comparing GBA1 variant-carrier PD cases with healthy controls.*** *Genes are plotted according to their genomic position and the -log₁₀(p-value). The yellow and dashed red lines indicate the nominal ($p < 0.05$) and Bonferroni-corrected significance thresholds, respectively. No gene reached Bonferroni-corrected significance. The 15 genes with the lowest nominal p-values are labelled.*

Since no gene remained significant after FDR correction, the 100 top-ranked genes based on their nominal SKAT-O p-values were selected for further analysis ***(Table 7)***.

This also ensured methodological consistency and comparability with the other cohort comparisons. For each gene, the OR was calculated. Unlike the distribution observed for the *GBA1* wt cohort, the $\log_2$(OR) values showed a marked predominance of positive values (***Figure 18***).

| Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *PRKRIP1* | $8.92 \times 10^{-5}$ | 0.9709 | *TDRD5* | $3.94 \times 10^{-3}$ | 0.9709 | *MINK1* | $7.18 \times 10^{-3}$ | 0.9709 | *FCRL3* | $9.66 \times 10^{-3}$ | 0.9709 |
| *IRX3* | $2.03 \times 10^{-4}$ | 0.9709 | *ATP5F1B* | $4.26 \times 10^{-3}$ | 0.9709 | *ITGBL1* | $7.18 \times 10^{-3}$ | 0.9709 | *NBR1* | $9.80 \times 10^{-3}$ | 0.9709 |
| *HOXB5* | $4.57 \times 10^{-4}$ | 0.9709 | *SIX1* | $4.47 \times 10^{-3}$ | 0.9709 | *IGFBP3* | $7.21 \times 10^{-3}$ | 0.9709 | *SULF2* | $1.00 \times 10^{-2}$ | 0.9709 |
| *SLC9C1* | $7.53 \times 10^{-4}$ | 0.9709 | *PXDN* | $4.79 \times 10^{-3}$ | 0.9709 | *COL6A2* | $7.28 \times 10^{-3}$ | 0.9709 | *ZNF610* | $1.02 \times 10^{-2}$ | 0.9709 |
| *MALT1* | $1.65 \times 10^{-3}$ | 0.9709 | *KLHL9* | $4.88 \times 10^{-3}$ | 0.9709 | *ZFP62* | $7.48 \times 10^{-3}$ | 0.9709 | *KMT5A* | $1.02 \times 10^{-2}$ | 0.9709 |
| *SBK2* | $1.65 \times 10^{-3}$ | 0.9709 | *MAMSTR* | $5.06 \times 10^{-3}$ | 0.9709 | *HES4* | $7.53 \times 10^{-3}$ | 0.9709 | *NAE1* | $1.02 \times 10^{-2}$ | 0.9709 |
| *USP49* | $1.75 \times 10^{-3}$ | 0.9709 | *SOX1* | $5.12 \times 10^{-3}$ | 0.9709 | *QRICH1* | $7.55 \times 10^{-3}$ | 0.9709 | *DNALI1* | $1.05 \times 10^{-2}$ | 0.9709 |
| *RAX* | $1.85 \times 10^{-3}$ | 0.9709 | *LIG3* | $5.29 \times 10^{-3}$ | 0.9709 | *LDHAL6A* | $7.63 \times 10^{-3}$ | 0.9709 | *GLG1* | $1.06 \times 10^{-2}$ | 0.9709 |
| *FGF10* | $1.96 \times 10^{-3}$ | 0.9709 | *RASGRF2* | $5.30 \times 10^{-3}$ | 0.9709 | *KIAA1328* | $7.66 \times 10^{-3}$ | 0.9709 | *CDK14* | $1.06 \times 10^{-2}$ | 0.9709 |
| *NARS1* | $2.07 \times 10^{-3}$ | 0.9709 | *CARMIL3* | $5.37 \times 10^{-3}$ | 0.9709 | *PHF1* | $7.71 \times 10^{-3}$ | 0.9709 | *PLCL1* | $1.06 \times 10^{-2}$ | 0.9709 |
| *GEM* | $2.11 \times 10^{-3}$ | 0.9709 | *PARS2* | $5.63 \times 10^{-3}$ | 0.9709 | *NR4A3* | $8.08 \times 10^{-3}$ | 0.9709 | *RPF2* | $1.07 \times 10^{-2}$ | 0.9709 |
| *IYD* | $2.21 \times 10^{-3}$ | 0.9709 | *LYZ* | $5.69 \times 10^{-3}$ | 0.9709 | *STX17* | $8.11 \times 10^{-3}$ | 0.9709 | *SLC5A9* | $1.07 \times 10^{-2}$ | 0.9709 |
| *ADRA1A* | $2.23 \times 10^{-3}$ | 0.9709 | *SPEF2* | $5.73 \times 10^{-3}$ | 0.9709 | *CRP* | $8.17 \times 10^{-3}$ | 0.9709 | *TAMALIN* | $1.07 \times 10^{-2}$ | 0.9709 |
| *UHRF1BP1L* | $2.48 \times 10^{-3}$ | 0.9709 | *NDST1* | $5.77 \times 10^{-3}$ | 0.9709 | *DCAF8* | $8.46 \times 10^{-3}$ | 0.9709 | *SH2D5* | $1.09 \times 10^{-2}$ | 0.9709 |
| *RPS6KA1* | $2.77 \times 10^{-3}$ | 0.9709 | *SPR* | $5.77 \times 10^{-3}$ | 0.9709 | *ACOXL* | $8.52 \times 10^{-3}$ | 0.9709 | *RTKN* | $1.09 \times 10^{-2}$ | 0.9709 |
| *SETD7* | $2.91 \times 10^{-3}$ | 0.9709 | *LRP6* | $5.89 \times 10^{-3}$ | 0.9709 | *SLC2A10* | $8.91 \times 10^{-3}$ | 0.9709 | *HTR1E* | $1.09 \times 10^{-2}$ | 0.9709 |
| *SIGLEC15* | $2.97 \times 10^{-3}$ | 0.9709 | *FBXL17* | $6.14 \times 10^{-3}$ | 0.9709 | *NGRN* | $8.99 \times 10^{-3}$ | 0.9709 | *SLC35D2* | $1.10 \times 10^{-2}$ | 0.9709 |
| *THNSL2* | $3.01 \times 10^{-3}$ | 0.9709 | *GRAP2* | $6.23 \times 10^{-3}$ | 0.9709 | *TWSG1* | $9.07 \times 10^{-3}$ | 0.9709 | *UQCC3* | $1.10 \times 10^{-2}$ | 0.9709 |
| *SPINK5* | $3.19 \times 10^{-3}$ | 0.9709 | *APP* | $6.49 \times 10^{-3}$ | 0.9709 | *GP1BB* | $9.08 \times 10^{-3}$ | 0.9709 | *INF2* | $1.11 \times 10^{-2}$ | 0.9709 |
| *FES* | $3.20 \times 10^{-3}$ | 0.9709 | *TM7SF3* | $6.50 \times 10^{-3}$ | 0.9709 | *RILPL1* | $9.27 \times 10^{-3}$ | 0.9709 | *ZPR1* | $1.13 \times 10^{-2}$ | 0.9709 |
| *ANKRD10* | $3.23 \times 10^{-3}$ | 0.9709 | *ARFIP2* | $6.62 \times 10^{-3}$ | 0.9709 | *OR2A25* | $9.28 \times 10^{-3}$ | 0.9709 | *NUAK1* | $1.13 \times 10^{-2}$ | 0.9709 |
| *VPS13B* | $3.32 \times 10^{-3}$ | 0.9709 | *ANAPC15* | $6.62 \times 10^{-3}$ | 0.9709 | *NINJ2* | $9.32 \times 10^{-3}$ | 0.9709 | *SNX18* | $1.14 \times 10^{-2}$ | 0.9709 |
| *LCTL* | $3.36 \times 10^{-3}$ | 0.9709 | *ZNF613* | $6.72 \times 10^{-3}$ | 0.9709 | *CNOT3* | $9.39 \times 10^{-3}$ | 0.9709 | *UBA1* | $1.15 \times 10^{-2}$ | 0.9709 |
| *RGS17* | $3.94 \times 10^{-3}$ | 0.9709 | *STARD7* | $6.89 \times 10^{-3}$ | 0.9709 | *TOMM5* | $9.49 \times 10^{-3}$ | 0.9709 | *JAK1* | $1.15 \times 10^{-2}$ | 0.9709 |
| *SPRY1* | $3.94 \times 10^{-3}$ | 0.9709 | *TRIM27* | $7.14 \times 10^{-3}$ | 0.9709 | *DDIT4L* | $9.65 \times 10^{-3}$ | 0.9709 | *GPR150* | $1.15 \times 10^{-2}$ | 0.9709 |

***Table 7: Table displaying the 100 top-ranking genes in the analyzed controls cohort.*** *The table illustrates the 100 genes with the lowest nominal p-value within the cohort. For each gene the HUGO name, the nominal p-value, the p-value after FDR correction are shown.*

In particular, 96 genes displayed a positive $\log_2$(OR), indicating a higher relative frequency of rare variants in *GBA1* variant-carrier PD patients.

Only four genes retained negative values, indicating relative enrichment in controls: *ZPR1*, *VPS13B*, *SOX1*, *RTKN,* with $\log_2$ ratios of −0.04, −0.16, and −0.66, respectively.

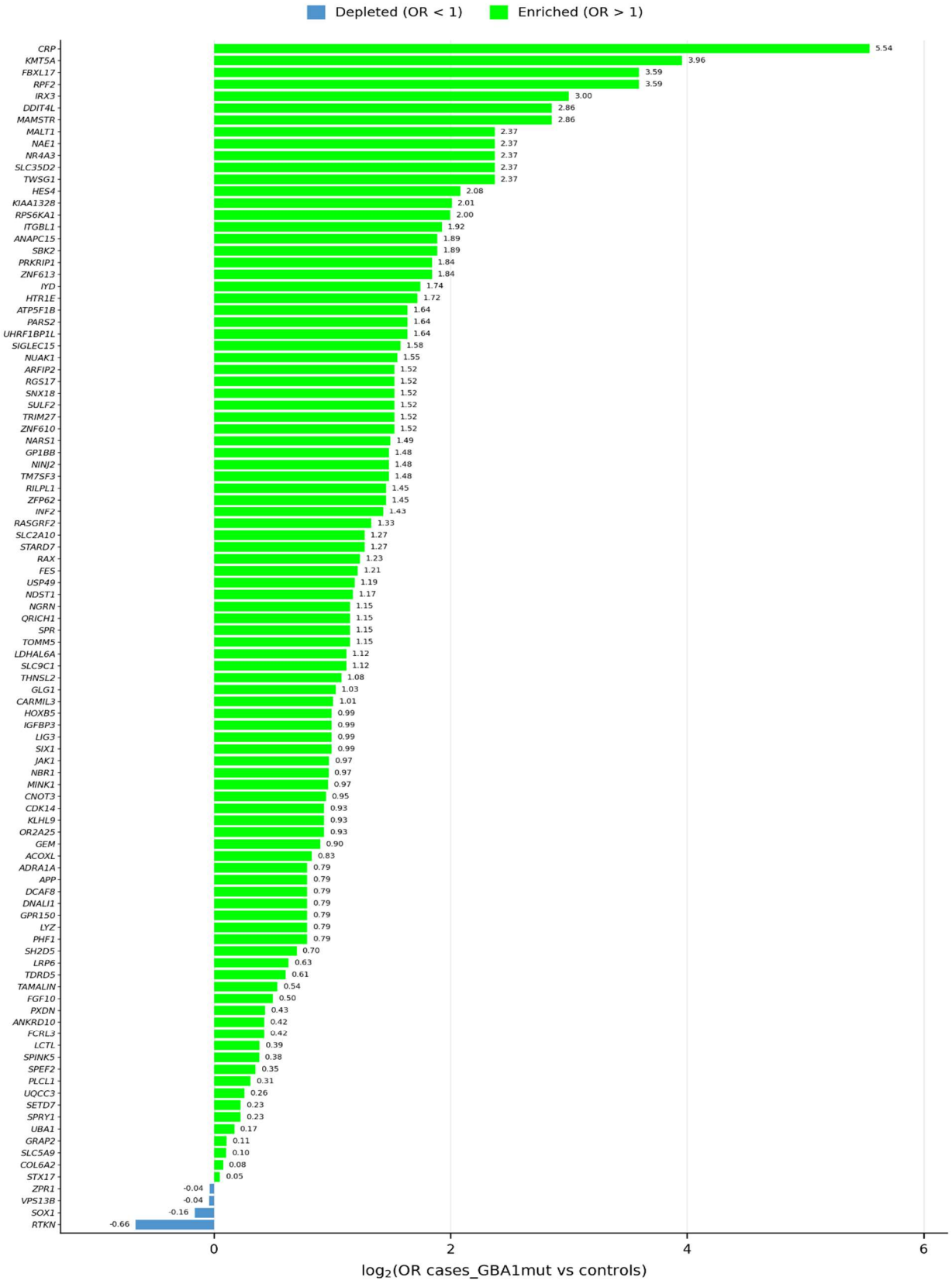



***Figure 18: Log₂(OR) values for the 100 top-ranked genes in the comparison between GBA1 variant-carrier PD patients and healthy controls.*** *Genes are ranked by decreasing log₂OR. Positive values, shown in green, indicate relative enrichment of qualifying alternative alleles in case, while negative values,shown in blue, indicate relative enrichment in controls.*

*Enrichment analysis*

Subsequently, enrichment analysis was performed on the 100 top-ranked genes, following the same approach used for the *GBA1* wt cohort. The QIAGEN IPA analysis identified several nominally significant pathways ($-\log_{10}$(p-value)>1.3), of which the 13 most significant are shown in ***Figure 19***. Cellular hexose transport presented the strongest enrichment, with a $-\log_{10}$(p-value) of 2.54.

Interestingly, Mitochondrial protein degradation and Mitochondrial division signaling were also among the top enriched pathways, with a $-\log_{10}$(p-value) of 2.25 and 1.64, respectively. Tetrahydrobiopterin (BH4) biosynthesis (I and II) was also enriched, each with a $-\log_{10}$ (p-value) of 1.96.

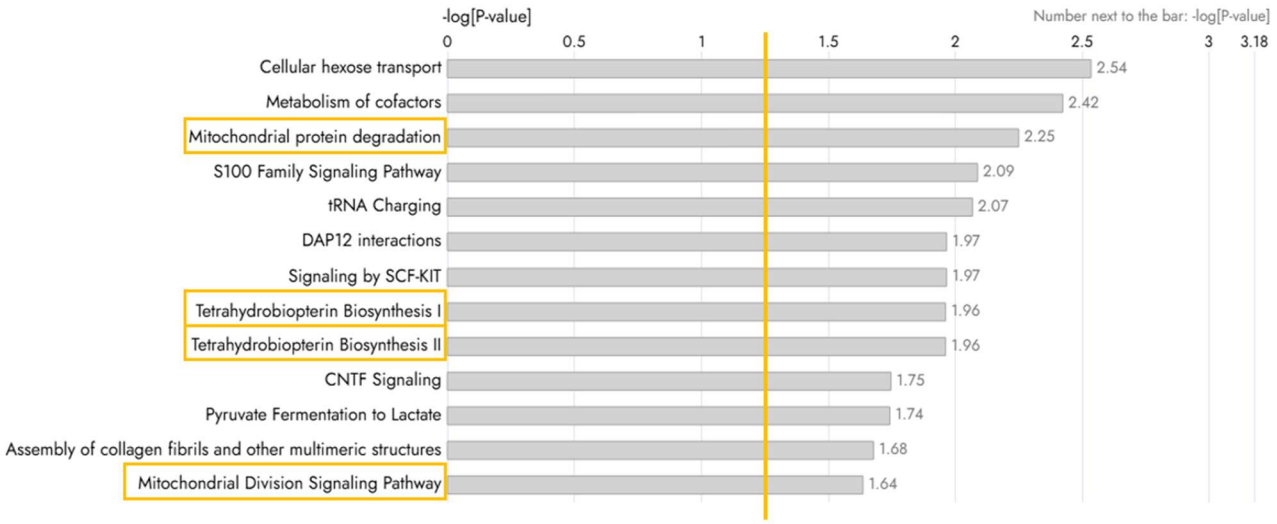
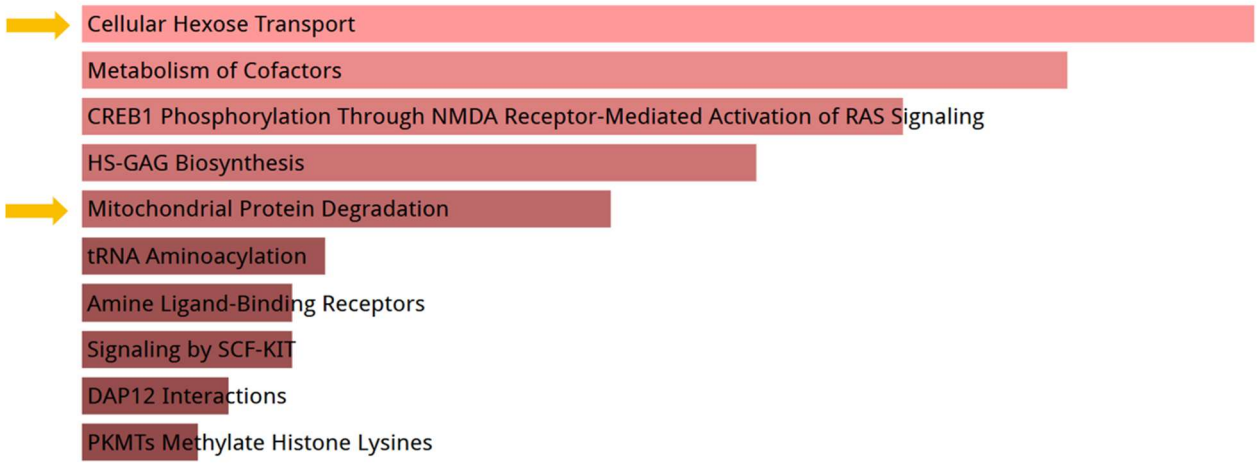



***Figure 19: IPA enrichment analysis.*** *The figure illustrates the 13 mostly significantly enriched pathways among the 100 top-ranked genes from the SKAT-O analysis. The threshold is defined by the yellow line, corresponding to a $-\log_{10}$(p-value) of 1.3. Pathways related to mitochondrial homeostasis and BH4 biosynthesis pathways are highlighted.*

The same subset of 100 genes was also used as input for the enrichment analysis using EnrichR. Results from the Reactome 2024 database were consistent with those obtained using IPA, with Hexose Transport identified as the top-ranked pathway and Mitochondrial Protein Degradation also among the enriched terms (***Figure 20***). However, these enrichments were significant only at a nominal level and did not remain significant after multiple-testing correction.

(A)




(B)

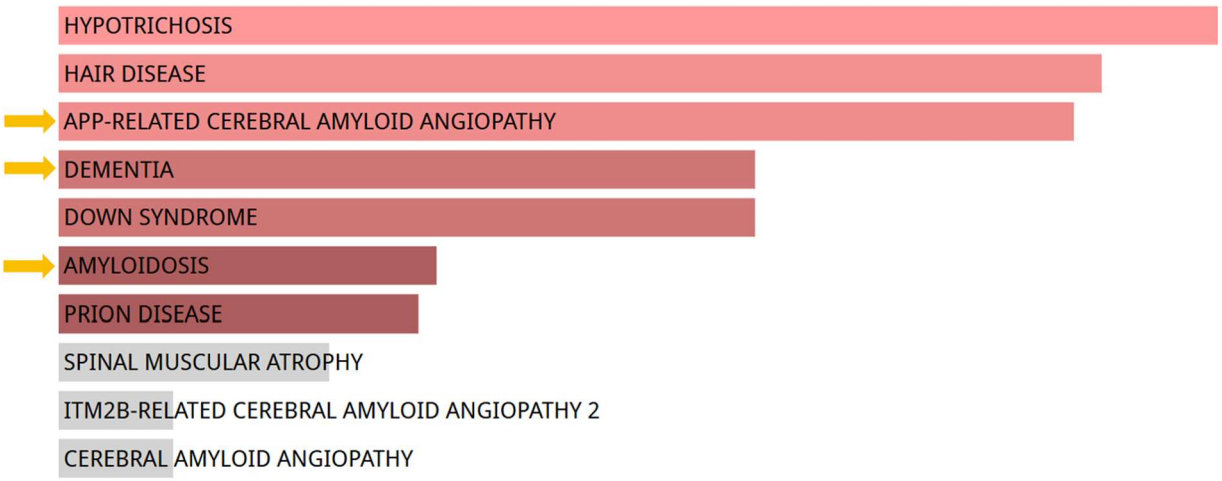
| Pathway | p value | q value | Genes involved |
|---|---|---|---|
| Cellular hexose transport | 0.0054 | 0.3489 | *SLC5A9, SLC2A10* |
| mitochondria proteln degradation | 0.0105 | 0.3489 | *APP, ATP5F1B. STARD7* |

***Figure 20: EnrichR enrichment analysis using the Reactome Pathway 2024 library.*** *(A) graph representing the enriched pathway terms among the specific gene subset from the GBA1 variant-carrier PD patients and healthy controls cohort, highlighting the Cellular hexose transport and Mitochondrial protein degradation pathways. (B) Table summarizing enrichment analysis results, including pathway-term name, associated p-values, q-values, and the contributing genes.*

An additional result was obtained from the Jensen DISEASES-curated 2025 database (***Figure 21)***, which showed nominal significant enrichment for *APP*-related cerebral amyloid angiopathy, dementia and amyloidosis diseases.

All three disease terms were related and driven by the presence of *APP* in the analyzed gene set (except for amyloidosis which was also associated with *LYZ*); therefore, they represent closely related annotations rather than independent enrichment signals.

(A)




(B)

| Disease | p value | q value | Genes involved |
|---|---|---|---|
| *APP* -related cerebral amyloid angiopathy | 0.02475 | 0.425 | *APP* |
| Dementia | 0.3448 | 0.425 | *APP* |
| Amyloidosis | 0.04798 | 0.425 | *APP, LYZ* |

***Figure 21: EnrichR enrichment analysis using the Jensen DISEASES Curated 2025 library.*** *(A) graph representing the enriched disease terms among the specific gene subset from the analyzed. (B) Table summarizing enrichment analysis results, including disease-termname, associated p-values, q-values, and the contributing genes.*

To investigate whether the functional enrichment results differed according to the direction of the relative OR, the initial subset of genes was divided into two separate groups. Subsequently, separate over-rapresentation analyses were performed using the two gene sets using EnrichR. In the gene set enriched in *GBA1* variant-carrier PD cases *APP*-related cerebral amyloid angiopathy, dementia and amyloidosis were identified as the top-ranked disease terms (***Table 8***). Conversely, the four genes enriched in healthy controls display significant enrichment for hypotrichosis, hair and nutrition diseases (q-value = 0.03630).

| Disease | p value | q value | Genes involved |
|---|---|---|---|
| *APP* -related cerebral amyloid angiopathy | 0.02377 | 0.4268 | *APP* |
| Dementia | 0.03312 | 0.4268 | *APP* |
| Amyloidosis | 0.0446 | 0.4268 | *APP, LYZ* |

***Table 8: EnrichR enrichment analysis of the gene set of GBA1 variant-carrier PD patients.*** *Table summarizing enrichment results of the separate subset of genes associated with GBA1 variant-carrier PD patients, including the disease-term names, nominal p-values, q-values, and contributing genes.*

This indicates that the hair-related enrichment observed when analyzing the complete set of 100 genes was mainly driven by genes with a higher alternative allele burden in controls. Therefore, the direction-stratified analysis separated this control-related signal from the *APP*-related terms identified among genes enriched in *GBA1* variant-carrier PD patients. However, the *APP*-related results should be considered exploratory, as they reached only nominal significance and were driven by a single gene.

### *4.1.3 GBA1 variant-carrier versus GBA1 wild-type PD patients*

*Gene-based association results*

Rare-variant association analysis was performed by directly comparing *GBA1 variant-carrier* and *GBA1* wt PD patients. This comparison aimed to investigate whether the rare-variant genetic architecture of PD differs according to the presence or absence of a *GBA1* variant, independently of the healthy control group.

The chromosomal distribution of the analyzed genes, according to their -$\log_{10}$(p-value) is shown in the Manhattan plot below (***Figure 22***). The analysis identified 77 genes with nominally significant p-value (p-value < 0.05) (***Table 9***); however, none remained significant after FDR correction for multiple testing.

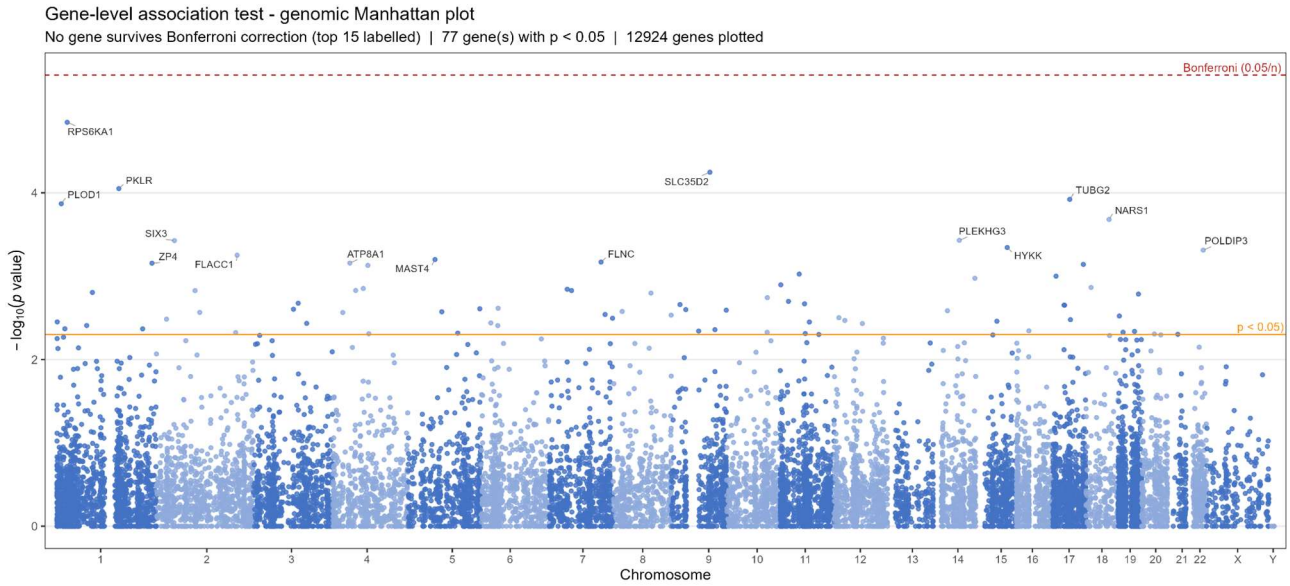



***Figure 22: Manhattan plot of the gene-based SKAT-O analysis comparing GBA1 variant-carrier and GBA1 wt PD patients.*** *Genes are plotted according to their genomic position and -$\log_{10}$(p-value). The yellow and dashed red lines indicate the nominal (p < 0.05) and FDR-corrected significance thresholds, respectively. No gene reached FDR-corrected significance. The 15 genes with the lowest nominal p-values are labelled.*

| Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted | Gene | p-value | p-value adjusted |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *PRKRIP1* | $8.92 \times 10^{-5}$ | 0.9709 | *TDRD5* | $3.94 \times 10^{-3}$ | 0.9709 | *MINK1* | $7.18 \times 10^{-3}$ | 0.9709 | *FCRL3* | $9.66 \times 10^{-3}$ | 0.9709 |
| *IRX3* | $2.03 \times 10^{-4}$ | 0.9709 | *ATP5F1B* | $4.26 \times 10^{-3}$ | 0.9709 | *ITGBL1* | $7.18 \times 10^{-3}$ | 0.9709 | *NBR1* | $9.80 \times 10^{-3}$ | 0.9709 |
| *HOXB5* | $4.57 \times 10^{-4}$ | 0.9709 | *SIX1* | $4.47 \times 10^{-3}$ | 0.9709 | *IGFBP3* | $7.21 \times 10^{-3}$ | 0.9709 | *SULF2* | $1.00 \times 10^{-2}$ | 0.9709 |
| *SLC9C1* | $7.53 \times 10^{-4}$ | 0.9709 | *PXDN* | $4.79 \times 10^{-3}$ | 0.9709 | *COL6A2* | $7.28 \times 10^{-3}$ | 0.9709 | *ZNF610* | $1.02 \times 10^{-2}$ | 0.9709 |
| *MALT1* | $1.65 \times 10^{-3}$ | 0.9709 | *KLHL9* | $4.88 \times 10^{-3}$ | 0.9709 | *ZFP62* | $7.48 \times 10^{-3}$ | 0.9709 | *KMT5A* | $1.02 \times 10^{-2}$ | 0.9709 |
| *SBK2* | $1.65 \times 10^{-3}$ | 0.9709 | *MAMSTR* | $5.06 \times 10^{-3}$ | 0.9709 | *HES4* | $7.53 \times 10^{-3}$ | 0.9709 | *NAE1* | $1.02 \times 10^{-2}$ | 0.9709 |
| *USP49* | $1.75 \times 10^{-3}$ | 0.9709 | *SOX1* | $5.12 \times 10^{-3}$ | 0.9709 | *QRICH1* | $7.55 \times 10^{-3}$ | 0.9709 | *DNALI1* | $1.05 \times 10^{-2}$ | 0.9709 |
| *RAX* | $1.85 \times 10^{-3}$ | 0.9709 | *LIG3* | $5.29 \times 10^{-3}$ | 0.9709 | *LDHAL6A* | $7.63 \times 10^{-3}$ | 0.9709 | *GLG1* | $1.06 \times 10^{-2}$ | 0.9709 |
| *FGF10* | $1.96 \times 10^{-3}$ | 0.9709 | *RASGRF2* | $5.30 \times 10^{-3}$ | 0.9709 | *KIAA1328* | $7.66 \times 10^{-3}$ | 0.9709 | *CDK14* | $1.06 \times 10^{-2}$ | 0.9709 |
| *NARS1* | $2.07 \times 10^{-3}$ | 0.9709 | *CARMIL3* | $5.37 \times 10^{-3}$ | 0.9709 | *PHF1* | $7.71 \times 10^{-3}$ | 0.9709 | *PLCL1* | $1.06 \times 10^{-2}$ | 0.9709 |
| *GEM* | $2.11 \times 10^{-3}$ | 0.9709 | *PARS2* | $5.63 \times 10^{-3}$ | 0.9709 | *NR4A3* | $8.08 \times 10^{-3}$ | 0.9709 | *RPF2* | $1.07 \times 10^{-2}$ | 0.9709 |
| *IYD* | $2.21 \times 10^{-3}$ | 0.9709 | *LYZ* | $5.69 \times 10^{-3}$ | 0.9709 | *STX17* | $8.11 \times 10^{-3}$ | 0.9709 | *SLC5A9* | $1.07 \times 10^{-2}$ | 0.9709 |
| *ADRA1A* | $2.23 \times 10^{-3}$ | 0.9709 | *SPEF2* | $5.73 \times 10^{-3}$ | 0.9709 | *CRP* | $8.17 \times 10^{-3}$ | 0.9709 | *TAMALIN* | $1.07 \times 10^{-2}$ | 0.9709 |
| *UHRF1BP1L* | $2.48 \times 10^{-3}$ | 0.9709 | *NDST1* | $5.77 \times 10^{-3}$ | 0.9709 | *DCAF8* | $8.46 \times 10^{-3}$ | 0.9709 | *SH2D5* | $1.09 \times 10^{-2}$ | 0.9709 |
| *RPS6KA1* | $2.77 \times 10^{-3}$ | 0.9709 | *SPR* | $5.77 \times 10^{-3}$ | 0.9709 | *ACOXL* | $8.52 \times 10^{-3}$ | 0.9709 | *RTKN* | $1.09 \times 10^{-2}$ | 0.9709 |
| *SETD7* | $2.91 \times 10^{-3}$ | 0.9709 | *LRP6* | $5.89 \times 10^{-3}$ | 0.9709 | *SLC2A10* | $8.91 \times 10^{-3}$ | 0.9709 | *HTR1E* | $1.09 \times 10^{-2}$ | 0.9709 |
| *SIGLEC15* | $2.97 \times 10^{-3}$ | 0.9709 | *FBXL17* | $6.14 \times 10^{-3}$ | 0.9709 | *NGRN* | $8.99 \times 10^{-3}$ | 0.9709 | *SLC35D2* | $1.10 \times 10^{-2}$ | 0.9709 |
| *THNSL2* | $3.01 \times 10^{-3}$ | 0.9709 | *GRAP2* | $6.23 \times 10^{-3}$ | 0.9709 | *TWSG1* | $9.07 \times 10^{-3}$ | 0.9709 | *UQCC3* | $1.10 \times 10^{-2}$ | 0.9709 |
| *SPINK5* | $3.19 \times 10^{-3}$ | 0.9709 | *APP* | $6.49 \times 10^{-3}$ | 0.9709 | *GP1BB* | $9.08 \times 10^{-3}$ | 0.9709 | *INF2* | $1.11 \times 10^{-2}$ | 0.9709 |
| *FES* | $3.20 \times 10^{-3}$ | 0.9709 | *TM7SF3* | $6.50 \times 10^{-3}$ | 0.9709 | *RILPL1* | $9.27 \times 10^{-3}$ | 0.9709 | *ZPR1* | $1.13 \times 10^{-2}$ | 0.9709 |
| *ANKRD10* | $3.23 \times 10^{-3}$ | 0.9709 | *ARFIP2* | $6.62 \times 10^{-3}$ | 0.9709 | *OR2A25* | $9.28 \times 10^{-3}$ | 0.9709 | *NUAK1* | $1.13 \times 10^{-2}$ | 0.9709 |
| *VPS13B* | $3.32 \times 10^{-3}$ | 0.9709 | *ANAPC15* | $6.62 \times 10^{-3}$ | 0.9709 | *NINJ2* | $9.32 \times 10^{-3}$ | 0.9709 | *SNX18* | $1.14 \times 10^{-2}$ | 0.9709 |
| *LCTL* | $3.36 \times 10^{-3}$ | 0.9709 | *ZNF613* | $6.72 \times 10^{-3}$ | 0.9709 | *CNOT3* | $9.39 \times 10^{-3}$ | 0.9709 | *UBA1* | $1.15 \times 10^{-2}$ | 0.9709 |
| *RGS17* | $3.94 \times 10^{-3}$ | 0.9709 | *STARD7* | $6.89 \times 10^{-3}$ | 0.9709 | *TOMM5* | $9.49 \times 10^{-3}$ | 0.9709 | *JAK1* | $1.15 \times 10^{-2}$ | 0.9709 |
| *SPRY1* | $3.94 \times 10^{-3}$ | 0.9709 | *TRIM27* | $7.14 \times 10^{-3}$ | 0.9709 | *DDIT4L* | $9.65 \times 10^{-3}$ | 0.9709 | *GPR150* | $1.15 \times 10^{-2}$ | 0.9709 |

***Table 9: Table displaying the 100 top-ranking genes in the GBA1 variant-carrier PD patients and healthy controls cohort.*** *The table illustrates the 100 genes with the lowest nominal p-value within the cohort. For each gene the HUGO name, the nominal p-value, the p-value after FDR correction are shown.*

Since the number of genes with nominally significant p-value was comparable to the gene sets selected for the other comparisons, all of 77 genes were retained for subsequent analyses. Consistent with the previously used approach, the OR was calculated for each gene, using *GBA1* variant-carrier PD patients as the numerator group and *GBA1* wild-type PD patients as the denominator group. The distribution of the $\log_2$(OR) values showed a marked predominance of genes with higher relative frequency of variants in *GBA1* variant-carrier patients (***Figure 23***). In particular, 73 genes showed a positive $\log_2$(OR), while only four genes showed negative values indicating relative enrichment in *GBA1* wild-type patients: *PITPNM2*, *FLNC*, *VCAN*, *MAST4* with a $\log_2$OR of -0.04, -1.52, -1.53, and -2.14, respectively. Among the genes with positive OR, *APP* and *SNCA*, both well-established neurodegeneration-related genes, displayed a $\log_2$OR of 1.06 and 1.75, respectively.

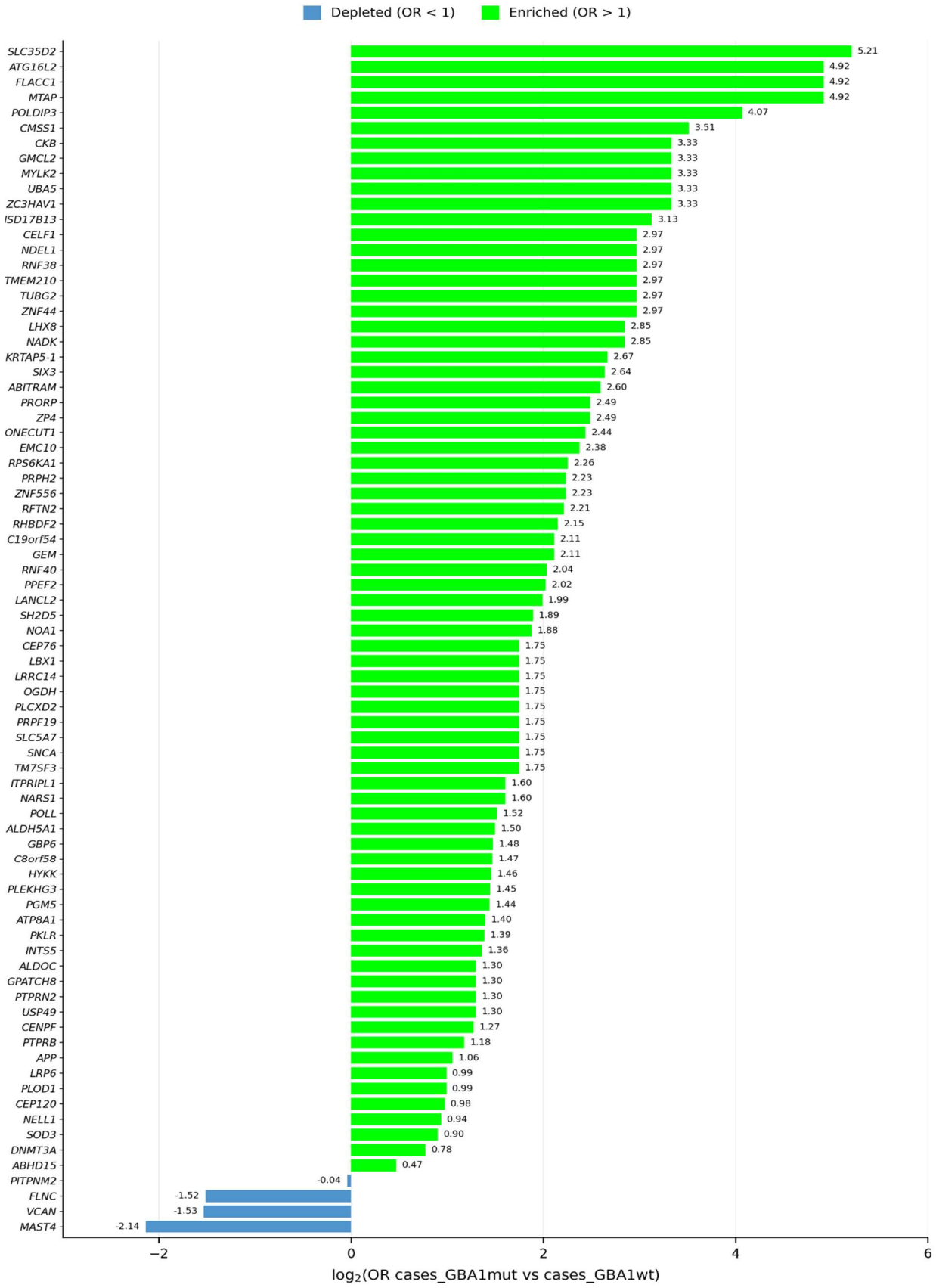



***Figure 23: Log₂(OR) values for the 77 genes with nominal significant p-value.*** *Genes are ranked by decreasing log₂OR. Positive values, shown in green, indicate relative enrichment of qualifying alternative alleles in GBA1variant-carriers, while negative values, shown in blue, indicate relative enrichment in GBA1 wt cases.*

*Functional enrichment analysis*

Over-rapresentation analysis was performed on the 77 genes with nominally significant SKAT-O p-values. QIAGEN IPA, EnrichR and Metascape were used to obtain a complementary functional characterization of this gene set, with Metascape being included in addition to the tools used for the previous cohort comparisons.

IPA enrichment analysis identifies several nominally enriched pathways ($-\log_{10}$(p-value)>1.3), of which the 13 top-ranked pathways are shown in ***Figure 24***. Glycogen degradation II and III displayed the strongest enrichment signals, with $-\log_{10}$(p-value) of 3.39 and 3.23, respectively.

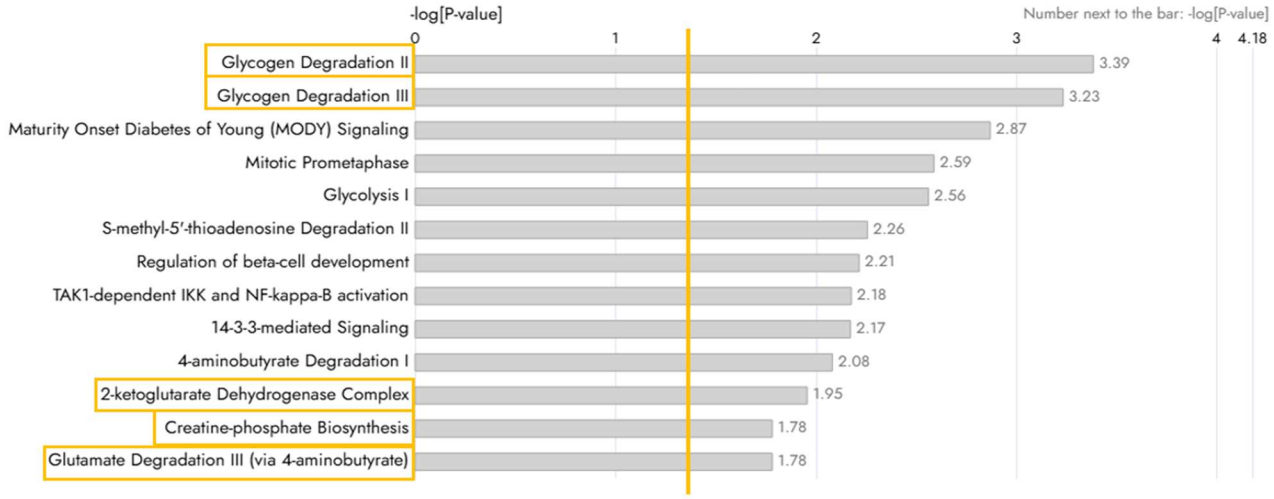



***Figure 24: IPA enrichment analysis.*** *The figure illustrates the 13 most significantly enriched pathways among the 77 genes with nominally significant SKAT-O p-values. The yellow line indicates the nominal significance threshold, corresponding to $-\log_{10}$(p-value) = 1.3 (p-value = 0.05). Pathways related to mitochondrial energy production and glutamate and GABA degradation are highlighted.*

Interestingly, several nominally enriched pathways were to cellular energy metabolism, including glycolysis, glycogen metabolism, Krebs cycle (2-ketoglutarate dehydrogenase complex), and creatine-phosphate biosynthesis. These results were consistent with the enrichment of pathways related to mitochondrial metabolism and homeostasis previously observed in comparison between *GBA1* variant-carrier PD patients and healthy controls cohort.

Pathways involved in neurotransmitter catabolism were also nominally enriched. In particular, glutamate and 4-aminobutyrate degradation $-\log_{10}$(p-value) of 1.78 and 2.08, respectively. Both pathways contribute to GABA shunt, a metabolic route that links GABA catabolism to the Krebs cycle through the production of succinate.

Analysis of the same gene set using Metascape provided additional information on the relationships among the enriched biological processes (***Figure 25A***) revealed a large, densely interconnected cluster centered on central nervous system development, indicating strong representation of this biological process among the analyzed genes.

Smaller clusters were associated with the regulation of protein localization, mitotic prometaphase, the *Toll-like receptor 10* (TLR10) cascade, and amino acid metabolism. Consistently, the ranking based on statistical significance (***Figure 25B***) identifies central nervous system development (GO:0007417) as the most significantly enriched term (q-value = 0.0055).

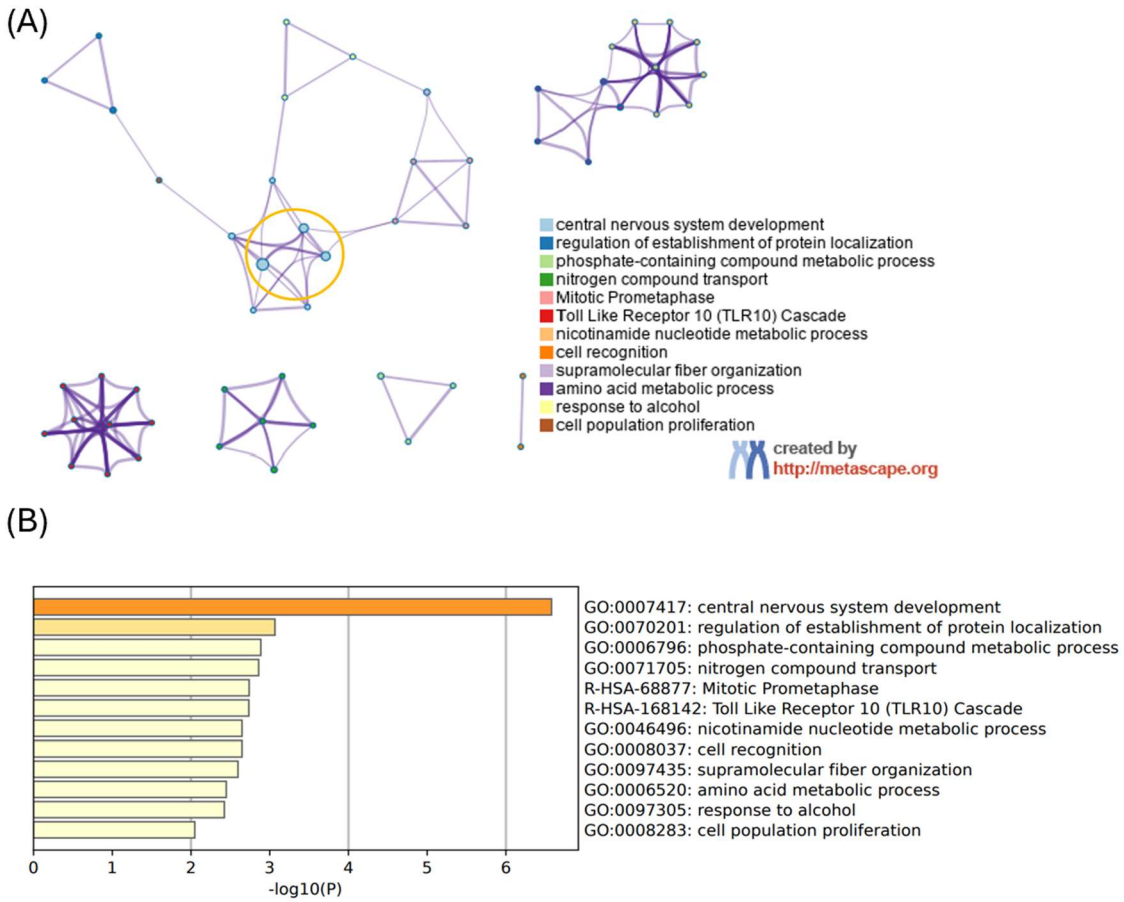


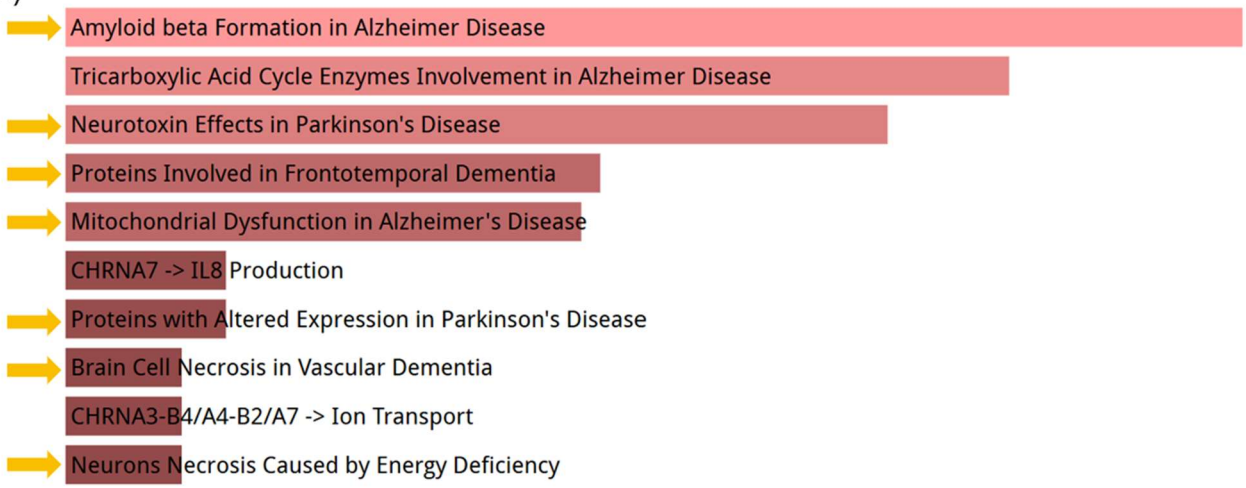

***Figure 25: Metascape enrichment analysis of the 77 nominally significant genes identified by comparing GBA1 variant-carrier and GBA1 wild-type PD patients.*** *(A) Network of enriched terms. Node colors indicate the corresponding functional clusters, while edge thickness reflects the degree of gene overlap between terms. The orange circle highlights the largest cluster, centered on central nervous system development. (B) Top enriched terms ranked by their -log10(p-value). Central nervous system development (GO:0007417) showing the most statistically significant term.*

The enrichment of terms related to central nervous system development was particularly relevant because the underlying analysis directly compared two PD subgroups defined by their *GBA1* status, rather than cases and healthy controls. Over-representation analysis using EnrichR showed thematically consistent results, identifying several terms related to dementia and AD in the Elsevier Pathway Collection and Jensen DISEASE-curated 2025 databases.

In the Elsevier Pathway Collection (***Figure 26***), Amyloid beta formation in AD remained significant after multiple-testing correction (q-value = 0.045). Additional terms reached nominal significance but did not remain significant after correction, including Neurotoxin effects in PD (p-value = 0.0013), Proteins involved in frontotemporal dementia (p-value = 0.0041), Mitochondrial dysfunction in AD (p-value = 0.0045), Proteins with altered expression in PD (p-value = 0.019), and Brain cell necrosis

in vascular dementia (p-values = 0.023). The genes contributing to these related enrichment signals included *SNCA*, *APP* and *OGDH*, with partial overlap across the different terms. Neurons necrosis caused by energy deficiency was also nominally enriched, further connecting these findings with the energy-metabolism pathways identified by IPA.

(A)




(B)

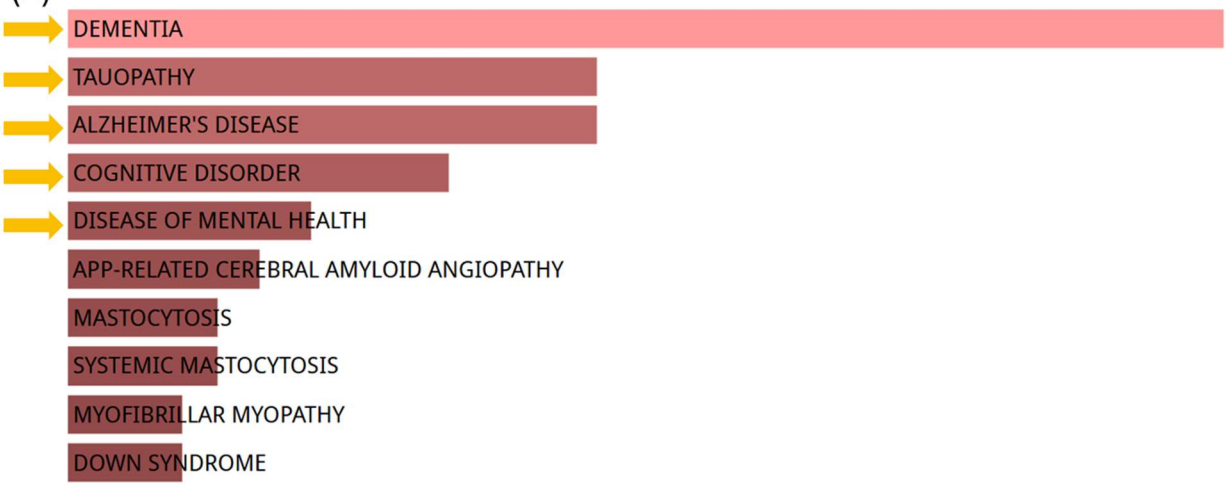
| Pathway | p value | q value | Genes involved |
|---|---|---|---|
| Amyloid beta formation in Alzheimer disease | 0.0003034 | 0.04551 | *APP, SNCA* |
| Neurotoxin effects in Parkinson's disease | 0.001292 | 0.05901 | *SNCA, OGDH* |
| Proteins involved in frontotemporal dementia | 0.004144 | 0.1344 | *APP, SNCA* |
| Mitochondrial dysfunction in Alzheimer's disease | 0.004479 | 0.1344 | *APP, OGDH* |
| Proteins with altered expression in Parkinson's disease | 0.0191 | 0.2099 | *OGDH* |
| Brain cell necrosis in vascular dementia | 0.02288 | 0.2099 | ATP8A1 |
| Neurons necrosis caused by enrgy deficiency | 0.02288 | 0.2099 | ATP8A1 |

***Figure 26: EnrichR enrichment analysis using the Elvensier Pathway collection.*** *(A) graph representing the top enriched pathways associated among the 77 nominally significant genes. (B) Table summarizing enrichment results, including pathway names, nominal p-values, q-values, and contributing genes.*

Consistent findings were obtained using the Jensen DISEASE curated 2025 database. Dementia was significantly enriched among the analyzed genes and remained significant after multiple-testing correction (q-value = 0.028).

Several related disease terms reached nominal significance but did not remain significant after correction, including Tauopathy (p-value = 0.0045), AD p-value = 0.0045), Cognitive disorder (p-value = 0.0085), Disease of mental health (p-value = 0.015), and *APP*-related cerebral amyloid angiopathy (p-value = 0.019) (***Figure 27***).

(A)




(B)

| Disease | p value | q value | Genes involved |
|---|---|---|---|
| Dementia | 0.0003034 | 0.02791 | *APP, SNCA* |
| Tauopathy | 0.004479 | 0.1373 | *APP, SNCA* |
| Alzheimer's disease | 0.004479 | 0.1373 | *APP, SNCA* |
| Cognitive disorder | 0.008468 | 0.1948 | *APP, SNCA* |
| Disease of mental health | 0.0153 | 0.2321 | *APP, DNMT3A, SNCA* |
| *APP* -related cerebral amyloid angiopathy | 0.0191 | 0.2321 | *APP, LYZ* |

***Figure 27: EnrichR enrichment analysis using the Jensen DISEASE curated 2025 database.*** *(A) graph representing the top enriched disease terms among the 77 nominally significant genes identified by comparing GBA1 variant-carrier and GBA1 wild-type PD patients. (B) Table summarizing enrichment results, including the disease-term names, nominal p-values, q-values, and contributing genes.*

The disease-enrichment results were mainly driven by the presence of *APP* and *SNCA* in the analyzed gene set. To further characterize these gene-level signals, the individual *APP* and *SNCA* variants were examined (***Table 10***).

Five distinct *SNCA* variants were found in six patients. Three variants were reported in *GBA1* variant-carrier patients, occurring in two individuals. Notably, the only *SNCA* variant classified as likely pathogenic in Clinvar was found in two *GBA1* wt patients. For *APP* nine distinct variants were found in ten *GBA1* wt cases, while five variants were detected in six *GBA1* variant-carrier patients. Among the latter, *APP* p.A673V (NM_000484.4:c.2287C>T) was present in two individuals. Because the two PD subgroups differed considerably in sample size, these raw counts should be interpreted together with the corresponding allele frequencies and gene-level burden ratios.

| Variant | Gene | Numerosity | GBA status of the patient | Predicted pathogenicity | Clinvar validated |
|---|---|---|---|---|---|
| NM_000345.4:c.631A>G | *SNCA* | 1 | GBA+ | vus | no |
| NM_000345.4:c.589T>C | *SNCA* | 1 | GBA+ | likely benign | no |
| NM_000345.4:c.520G>C | *SNCA* | 1 | GBA- | likely benign | no |
| **NM_000345.4:c.269A>G** | *SNCA* | 2 | GBA- | likely pathogenic | yes |
| NM_000345.4:c.238T>C | *SNCA* | 1 | GBA+ | likely benign | no |
| **NM_000484.4:c.2287C>T** | *APP* | 2 | GBA+ | likely pathogenic | yes |
| NM_000484.4:c.2093C>A | *APP* | 1 | GBA- | vus | no |
| NM_000484.4:c.1975C>A | *APP* | 1 | GBA- | likely benign | no |
| **NM_000484.4:c.1945C>T** | *APP* | 1 | GBA- | likely benign | yes |
| **NM_000484.4:c.1925G>A** | *APP* | 1 | GBA+ | likely pathogenic | yes |
| **NM_000484.4:c.1273C>T** | *APP* | 2 | GBA- | likely benign | yes |
| **NM_000484.4:c.1240G>A** | *APP* | 1 | GBA- | likely benign | yes |
| NM_000484.4:c.1153C>T | *APP* | 1 | GBA- | vus | no |
| NM_000484.4:c.1051G>A | *APP* | 1 | GBA- | vus | no |
| **NM_000484.4:c.854G>A** | *APP* | 1 | GBA+ | likely benign | yes |
| **NM_000484.4:c.724C>T** | *APP* | 1 | GBA+ | likely benign | yes |
| **NM_000484.4:c.607C>T** | *APP* | 1 | GBA- | likely benign | yes |
| **NM_000484.4:c.448G>A** | *APP* | 1 | GBA+ | likely benign | yes |
| **NM_000484.4:c.401T>C** | *APP* | 1 | GBA- | likely benign | yes |

***Table 10: SNCA and APP variants in the GBA1 variant-carrier and GBA1 wt PD patients.*** *For each variant, the reports the affected gene, variant annotation, GBA1 status, predicted functional effect, Clinvar classification, when available (variants validated in Clinvar are highlighted in bold).*

## *4.2 ANALYSIS OF MONOZYGOTIC TWINS DISCORDANT FOR PD*

### *4.2.1 Identification of within-pair DNA-discordant variants*

To complement the analyses performed in the unrelated PD cohorts, gene-based rare-variant association analysis was conducted in the cohort of MZ twins discordant for PD. The analysis included 969 genes harboring at least one putative post-zygotic variant, defined as a variant detected in one twin but not in the corresponding co-twin. Association testing was performed using the four tests implemented in the SMMAT framework: SMMAT-B, SMMAT-S, SMMAT-O and SMMAT-E. None of the tested genes reached either nominal significance or significance after correction for multiple-testing. The low number of variants and limited number of twin pairs resulted in low statistical power to detect gene-level associations.

Given these limitations, a targeted candidate-gene analysis was subsequently performed. First, genes carrying postzygotic variants were cross-referenced with a curated list of genes previously associated with PD (***Table 11***).

| Gene | Association with Parkinson's disease | Gene | Association with Parkinson's disease |
|---|---|---|---|
| *ACMSD* | GWAS risk locus. | *PODXL* | Candidate gene. |
| *ATP13A2* | Recessive juvenile parkinsonism (Kufor-Rakeb syndrome). | *POLG* | Mitochondrial disease with parkinsonism. |
| *ATP6AP2* | Splicing variants cause X-linked parkinsonism with spasticity. | *PPP2R5D* | Early-onset parkinsonism reported in patients. |
| *ATP6V1B2* | Candidate PD gene (lysosomal pathway). | *PRKN* | Most frequent cause of recessive early-onset PD. |
| *BST1* | GWAS risk locus. | *PTPA* | Candidate gene. |
| *CCDC62* | GWAS risk locus. | *PTRHD1* | Recessive intellectual disability with early-onset parkinsonism. |
| *CCN3* | GWAS risk locus. | *RAB32* | p.S71R variant reported in dominant PD. |
| *CHCHD2* | Autosomal dominant PD. | *RAB39B* | X-linked early-onset PD with intellectual disability. |
| *DCTN1* | Perry syndrome (parkinsonism, depression, hypoventilation). | *RIC3* | Candidate gene. |
| *DNAJC13* | Reported in dominant PD. | *SCARB2* | GWAS risk locus. |
| *DNAJC6* | Recessive juvenile parkinsonism. | *SIPA1L1* | GWAS risk locus. |
| *FBXO7* | Recessive parkinsonian-pyramidal syndrome. | *SLC9A6* | Christianson syndrome, parkinsonism features in adults. |
| *FGF20* | Reported PD risk modifier. | *SMPD1* | Variants increase PD risk. |
| *GAK* | GWAS risk locus. | *SNCA* | Autosomal dominant PD. |
| *GALC* | GWAS risk locus. | *SPG11* | Juvenile parkinsonism reported. |
| *GBA1* | Strongest common genetic risk factor for PD. | *STBD1* | GWAS risk locus. |
| *GCH1* | Variants increase PD risk; dopa-responsive dystonia. | *STK3* | GWAS risk locus. |
| *GPNMB* | GWAS risk locus. | *STX1B* | GWAS risk locus. |
| *HIP1R* | GWAS risk locus. | *SYNJ1* | Recessive early-onset parkinsonism. |
| *HLA-DRB5* | GWAS locus, supports immune involvement. | *SYT11* | GWAS risk locus. |
| *LAMP1* | Candidate gene; marker of lysosomal dysfunction in PD. | *TBC1D24* | Candidate PD gene. |
| *LAMP3* | GWAS risk locus (3q27). | *TGM6* | Candidate PD gene. |
| *LRP10* | Variants reported in familial PD and dementia with Lewy bodies. | *TH* | Deficiency causes dopa-responsive parkinsonism. |
| *LRRK2* | Most common cause of dominant PD (e.g. G2019S). | *TMEM175* | GWAS risk locus. |
| *MCCC1* | GWAS risk locus. | *TMEM230* | Reported in familial PD. |
| *MECP2* | Parkinsonism reported in case studies. | *TRIM41* | Candidate gene. |
| *NR4A2* | Rare variants in familial PD. | *UCHL1* | Variant reported in a family with PD. |
| *NRXN2* | Candidate gene. | *UQCRC1* | Rare variant reported in familial PD. |
| *NUS1* | Loss-of-function variants reported in PD. | *VPS13C* | Recessive early-onset PD. |
| *PANK2* | PKAN (iron accumulation) with parkinsonism. | *VPS35* | p.D620N causes dominant PD. |
| *PARK7* | Recessive early-onset PD. | *WARS2* | Variants reported in levodopa-responsive parkinsonism. |
| *PINK1* | Recessive early-onset PD. | *WASL* | Candidate gene. |
| *PLA2G6* | Recessive dystonia-parkinsonism. | *WDR45* | BPAN with parkinsonism-dystonia in adults. |
| *PLXNA4* | Candidate gene. | | |

***Table 11: Curated list of genes associated with PD pathogenesis or PD-risk.*** *For each gene the HUGO name and the association with PD are shown.*

This analysis highlighted the presence of eight PD genes. Five genes carried variants specific to the affected twins (*DNAJC13*, *GAK*, *MCCC1*, *TBC1D24* and *UQCRC1*), whereas three carried variants specific to the unaffected co-twins (*RIC3*, *TMEM175* and *TRIM41*) (***Table 12***).

| PD genes | Number of variants | Twin's phenotype | Pair |
|---|---|---|---|
| *DNAJC13* | 1 | case | 3 |
| *GAK* | 1 | case | 15 |
| *MCCC1* | 2 | case | 2, 14 |
| *RIC3* | 1 | control | 7 |
| *TBC1D24* | 1 | case | 8 |
| *TMEM175* | 1 | control | 7 |
| *TRIM41* | 1 | control | 9 |
| *UQCRC1* | 1 | case | 16 |

***Table 12: Table of variants in PD-associated genes found in the twin cohort.*** *The table shows in which PD-related gene the variants were detected, their numerosity, the phenotype of the twin showcasing the mutation and its pair.*

Two different *MCCC1* variants were found in the affected twins from pair 2 and 14. The p.L598M (NM_020166.5:c.1848G>T) variant found in pair 2 was classified as VUS in Clinvar, while the p.R281* (NM_020166.5:c.968G>A) variant identified in pair 14 was annotated as pathogenic. Pathogenic variants in *MCCC1* are associated with 3-Methylcrotonylglycinuria, an autosomal recessive disorder of leucine metabolism. Therefore, the identification of a single putative post-zygotic *MCCC1* variant should not be interpreted as evidence of this metabolic disorder, but its

occurrence in two independent affected twins supported the prioritization of *MCCC1* for further investigation.

The genes harboring putative post-zygotic variants in the discordant twin cohort were subsequently compared with the gene subset selected from the *GBA1* wt PD case-control analysis. This cohort was selected for this comparison because the affected twins had also sporadic PD and did not carry disease-associated *GBA1* variants, making it the most closely matched PD subgroup among the unrelated cohorts examined in this study.

Four genes were shared between the two analyses: *ABCB8*, *MFSD5*, *TAPT1* and *UVSSA* (***Table 13***). Among these, *MFSD5* was the only gene that remained significant after FDR correction in the *GBA1* wt cohort, while the other three genes reached only nominal significance. The direction of the findings was also compared descriptively between the two datasets. Noticeably, postzygotic variants in *ABCB8* were found in the healthy twin of pair 29 and the *MFSD5* and *TAPT1* variants were detected in the affected twins of pair 24 and 5, respectively. These observations were directionally consistent with the corresponding OR in the *GBA1* wt case-control cohort.

Conversely, the *UVSSA* gene was characterized by an OR below 1, indicating relative enrichment of alternative alleles in controls, whereas its within-pair discordant variant was identified in the affected twin of pair 24.

| ***GBA1* wt genes** | **Biological function** | **Twin's phenotype** | **Pair** | ***GBA1* wt enrichment** |
|---|---|---|---|---|
| *ABCB8* | mitochondrial | control | 29 | controls |
| ***MFSD5*** | molybdate anion | case | 24 | cases |
| *TAPT1* | cilium transmembrane | case | 5 | cases |
| *UVSSA* | DNA repair linked to | case | 24 | controls |

***Table 13: Genes shared between the within-pair DNA-discordant variant analysis and the GBA1 wild-type PD case-control analysis.*** *For each gene, the table reports its biological function, the twin pair in which the discordant variant was identified, whether the variant was detected in the affected or unaffected co-twin, and the direction of the enrichment in the GBA1 wt case-control cohort. MFSD5 is shown in bold as it was the only gene that resulted significant even after FDR correction.*

*Functional enrichment analysis*

An initial over-representation analysis was performed using EnrichR on the complete set of 969 genes harboring putative postzygotic variants in one member of the twin pair. Analysis with the Jensen DISEASE curated 2025 database identified a broad range of significantly enriched disease terms, including monogenic disease, anemia, myopathy, and various cancer types. This heterogeneous pattern was consistent with the large and functionally diverse input gene set.

Given the focus of this study, the enrichment results were subsequently examined for terms related to the nervous system. Three such terms remained significant after multiple-testing correction: Central nervous system disease (q-value = 0.0056), Nervous system disease (q-value = 0.0073), and

Neurodegenerative disease (q-value = 0.024). These terms were supported by partially overlapping subsets of genes, as summarized in ***Table 14***.

| Disease | p value | q value | Genes involved |
|---|---|---|---|
| Central nervous system disease | 0.0001824 | 0.005577 | *SAMD9L, SLC2A1, LMNB2, SYNE1, TBK1, ATIC, SACS, NUP133, VPS13D, TWNK, VPS13A, ANXA11, WDR73, DNAJC3, MED25, AOPEP, CYP2U1, POLR1A, POLR1C, UQCRC1, TBC1D24, RAPGEF2, CEP104, GRN, AARS1, KARS1, TRAK1, CACNA1E, GIGYF2, RNF213, KLC2, PITRM1, RNF170, POC1B, ARSB, VPS16, REPS1, ARFGEF1, WWOX, SURF1, FBXO38, GOT2, DNAJC13, SUOX, GARS1, AHI1, KIAA0586, PI4KA, EIF4G1* |
| Nervous system disease | 0.0003455 | 0.007256 | *SAMD9L, SLC2A1, SLC4A4, SIPA1L3,RLBBP1, LMNB2, SYNE1, TBK1, ATIC, ZNF408, IDH3B, SACS,EYS, WDR36, NUP133, VPS13D, TWNK, VPS13A, ANXA11, WDR73, DNAJC3, VWA8, MED25, VCAN, AOPEP, CYP2U1, UQCRC1, TBC1D24, RAPGEF2, CEP104, GNR, AARS1, KARS1, TRAK1, CACNA1E, GIGYF2, CAMSAP1, RNF213, KLC2, TCOF1, PITRM1, RNF170, NLRP3, POC1B, NPHP4, ARSB, VPS16, REPS1, ARFGEF1, WWOX, SURF1, FBXO38, CRYBB1, GOT2, DNAJC13, SUOX, GARS1, MYO9A, AHI1,TTLL5, IMPDH1, KIAA0586, ERCC2, FBN1, EIF4G1* |
| Neurodegenerative disease | 0.001453 | 0.02409 | *GRN, AARS1, SAMD9L, KARS1, GIGYF2, LMNB2, SYNE1, TBK1, KLC2, PITRM1, SACS, RNF170, REPS1, WWOX, SURF1, FBXO38, VPS13D, TWNK, VPS13A, ANXA11, DNAJC13, DNAJC3, GRAS1, MED25, CYP2U1, UQCRC1, PI4KA, EIF4G1* |

***Table 14: EnricR table of statistically significant diseases in affected twins.*** *The table illustrates the significant disease found in the Jensen DISEASE-curated 2025 database, considering the genes mutated in PD twins within the analyzed cohort. The table reports the name of the disease, the corresponding p-value, q-value and which genes are directly involved.*

To investigate whether the enrichment results differ according to the phenotype of the twin carrying the discordant variants, genes were subsequently divided into two groups. The first one comprised all the genes in which postzygotic variants were majorly found in affected twins (n = 655), while the second included genes showing the opposite distribution (n = 279 genes). Genes carrying the same number of discordant variants in cases and controls were excluded from this direction-stratified analysis (n = 35). Because most genes harbored only one within-pair discordant variant, their assignment to the affected-enriched or unaffected-enriched gene set generally reflected a single twin-specific event rather than a recurrent difference across multiple pairs.

Separate over-representation analyses were then performed for the affected-enriched and unaffected-enriched gene sets in EnrichR. The affected-enriched gene set showed significant enrichment for Central nervous system disease (q-value = 0.0056), Nervous system disease (q-value = 0.0073), and Neurodegenerative disease (q-value = 0.024) (as in ***Table 14***). By contrast, the unaffected-enriched gene set showed only nominal enrichment for terms related to male fertility and the reproductive and urogenital systems, none of which remained significant after multiple-testing correction.

These findings indicate that nervous system-related terms were retained in the affected-enriched gene set but not in the unaffected-enriched set; however, this directional difference was descriptive and was not formally tested.

### *4.2.2 Allele-specific expression analysis of shared variants*

The availability of matched DNA- and RNA-sequencing data allowed ASE to be investigated within the monozygotic twin pairs discordant for PD. The analysis focused on variants that were heterozygous in DNA of both twins but showed opposite allelic expression patterns in RNA. Specifically, the alternative allele was predominantly or exclusively expressed in one twin, whereas the corresponding co-twin predominantly or exclusively expressed the reference allele.

Across the cohort, 111 variants in 109 distinct genes showed discordant allelic expression within the twin pairs. For 56 variants, the alternative allele was expressed in the affected twin and the reference allele in the healthy co-twin. The opposite pattern was observed for the remaining 55 variants. Overall, discordant ASE events were identified in 16 of the 24 twin pairs (***Table 15***).

(A)

| ID | Pair ID | Variant | Gene | PD RNA status | Healthy RNA status | ID | Pair ID | Variant | Gene | PD RNA status | Healthy RNA status |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C3457 | pair7 | NM_003459.5:c.892C>T (p.R298C) | *SLC30A3* | homo | WT | N0500 | pair25 | NM_001512.4:c.299T>C (p.L100P) | *GSTA4* | homo | WT |
| C3457 | pair7 | NM_182706.5:c.433G>A (p.G145R) | *SCRIB* | homo | WT | N0500 | pair25 | NM_001286496.2:c.880C>T (p.R294W) | *PIF1* | homo | WT |
| C4099 | pair8 | NM_001201427.2:c.2968C>T (p.R990C) | *DAAM2* | homo | WT | N0500 | pair25 | NM_002386.4:c.456C>A (p.Y152*) | *MC1R* | homo | WT |
| C4099 | pair8 | NM_022071.4:c.860C>T (p.P287L) | *SH2D4A* | homo | WT | N0500 | pair25 | NM_000540.3:c.418G>A (p.A140T) | *RYR1* | homo | WT |
| C4232 | pair9 | NM_001391957.1:c.2884G>A (p.E962K) | *FHAD1* | homo | WT | N0500 | pair25 | NM_178820.5:c.764A>C (p.H255P) | *FBXO27* | homo | WT |
| L1133 | pair16 | NM_173666.4:c.200C>G (p.P67R) | *DTWD2* | homo | WT | N0500 | pair25 | NM_080823.4:c.304G>A (p.V102M) | *SRMS* | homo | WT |
| L1133 | pair16 | NM_057095.3:c.1313G>A (p.G438E) | *CYP3A43* | homo | WT | P2436 | pair26 | NM_001004316.3:c.1976C>A (p.P659H) | *LEKR1* | homo | WT |
| L1133 | pair16 | NM_020647.4:c.1142C>T (p.T381I) | *JPH1* | homo | WT | P2436 | pair26 | NM_017570.5:c.2396C>T (p.T799M) | *OPLAH* | homo | WT |
| M2997 | pair18 | NM_001723.7:c.7135G>A (p.A2379T) ‡ | *DST* | homo | WT | P2436 | pair26 | NM_015392.4:c.500C>T (p.P167L) | *NPDC1* | homo | WT |
| M2997 | pair18 | NM_003931.3:c.1456C>T (p.R486C) | *WASF1* | homo | WT | P2436 | pair26 | NM_022772.4:c.2145C>G (p.S715R) | *EPS8L2* | homo | WT |
| M2997 | pair18 | NM_002208.5:c.1745C>T (p.A582V) | *ITGAE* | homo | WT | P2436 | pair26 | NM_020798.4:c.1822C>T (p.R608C) | *USP35* | homo | WT |
| M2997 | pair18 | NM_001051.5:c.1006C>T (p.R336C) | *SSTR3* | homo | WT | P2436 | pair26 | NM_004274.5:c.2663C>A (p.T888N) | *AKAP6* | homo | WT |
| M3322 | pair20 | NM_014675.5:c.5654T>C (p.V1885A) | *CROCC* | homo | WT | T0224 | pair29 | NM_025179.4:c.2510A>G (p.H837R) | *PLXNA2* | homo | WT |
| M3322 | pair20 | NM_001408.3:c.1259G>C (p.R420P) | *CELSR2* | homo | WT | T0224 | pair29 | NM_015662.3:c.3401G>T (p.R1134L) | *IFT172* | homo | WT |
| M3322 | pair20 | NM_005527.4:c.158T>C (p.I53T) | *HSPA1L* | homo | WT | T0224 | pair29 | NM_001277115.2:c.2783A>T (p.D928V) | *DNAH11* | homo | WT |
| M3322 | pair20 | NM_001360.3:c.719A>G (p.N240S) | *DHCR7* | homo | WT | T0224 | pair29 | NM_018351.4:c.4247C>T (p.S1416L) | *FGD6* | homo | WT |
| M3322 | pair20 | NM_173483.4:c.109C>T (p.R37C) | *CYP4F22* | homo | WT | T0224 | pair29 | NM_004756.5:c.1163C>G (p.T388S) | *NUMBL* | homo | WT |
| M1907 | pair22 | NM_005524.4:c.578G>A (p.G193D) | *HES1* | homo | WT | T1035 | pair30 | NM_001395010.1:c.2990A>G (p.N997S) | *DAB2IP* | homo | WT |
| M1907 | pair22 | NM_001423.3:c.32T>G (p.V11G) | *EMP1* | homo | WT | T1035 | pair30 | NM_001363514.2:c.769C>T (p.R257*) | *DUSP13B* | homo | WT |
| M1907 | pair22 | NM_138440.3:c.872G>A (p.R291Q) | *VASN* | homo | WT | T1035 | pair30 | NM_000392.5:c.842G>A (p.S281N) | *ABCC2* | homo | WT |
| M3114 | pair23 | NM_004672.5:c.2837C>T (p.P946L) | *MAP3K6* | homo | WT | T1035 | pair30 | NM_000260.4:c.905G>A (p.R302H) | *MYO7A* | homo | WT |
| M3114 | pair23 | NM_001364857.2:c.3256G>A (p.V1086I) | *ADGRB2* | homo | WT | T1035 | pair30 | NM_001284230.2:c.2552G>A (p.R851H) | *MAP3K9* | homo | WT |
| M3114 | pair23 | NM_001036.6:c.1235G>A (p.R412Q) | *RYR3* | homo | WT | V1178 | pair31 | NM_001371928.1:c.2672G>A (p.R891Q) | *AHDC1* | homo | WT |
| M3114 | pair23 | NM_031421.5:c.433C>G (p.L145V) | *ODAD4* | homo | WT | V1178 | pair31 | NM_152366.5:c.917C>T (p.P306L) | *KLHDC9* | homo | WT |
| M3114 | pair23 | NM_001365103.2:c.665G>A (p.R222H) | *ERFL* | homo | WT | V1178 | pair31 | NM_014176.4:c.106G>A (p.A36T) | *UBE2T* | homo | WT |
| N0442 | pair24 | NM_144978.3:c.749A>T (p.E250V) | *CCDC138* | homo | WT | V1178 | pair31 | NM_014935.5:c.1097C>T (p.P366L) | *PLEKHA6* | homo | WT |
| N0500 | pair25 | NM_145697.3:c.472G>A (p.A158T) | *NUF2* | homo | WT | V1178 | pair31 | NM_001039763.4:c.1426C>T (p.R476*) | *TMEM232* | homo | WT |
| N0500 | pair25 | NM_018983.4:c.83A>T (p.H28L) | *GAR1* | homo | WT | V1178 | pair31 | NM_012244.4:c.280G>A (p.A94T) | *SLC7A8* | homo | WT |

(B)

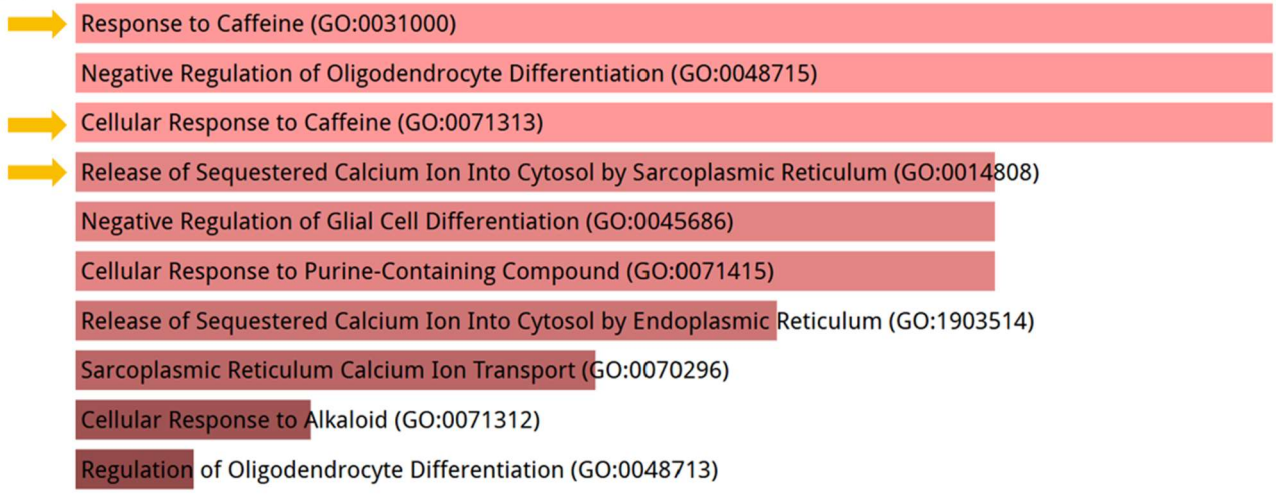
| ID | Pair ID | Variant | Gene | PD RNA status | Healthy RNA status | ID | Pair ID | Variant | Gene | PD RNA status | Healthy RNA status |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A1187 | pair1 | NM_170707.4:c.1930C>T (p.R644C) | *LMNA* | WT | homo | M3321 | pair20 | NM_032450.3:c.3932C>T (p.A1311V) | *MROH1* | WT | homo |
| A1187 | pair1 | NM_001370.2:c.7975C>T (p.R2659W) | *DNAH6* | WT | homo | M3321 | pair20 | NM_022337.3:c.583C>A (p.P195T) | *RAB38* | WT | homo |
| A1187 | pair1 | NM_177531.6:c.9808G>A (p.G3270S) | *PKHD1L1* | WT | homo | M2141 | pair22 | NM_001080504.3:c.2531C>A (p.A844D) | *RBM44* | WT | homo |
| A1187 | pair1 | NM_000606.3:c.268C>T (p.R90*) | *C8G* | WT | homo | M2141 | pair22 | NM_001291303.3:c.3402A>T (p.E1134D) | *FAT4* | WT | homo |
| A1187 | pair1 | NM_002216.3:c.529G>A (p.V177M) | *ITIH2* | WT | homo | M2141 | pair22 | NM_001010924.2:c.731C>T (p.A244V) | *FAM171A1* | WT | homo |
| A1187 | pair1 | NM_194325.3:c.1762C>T (p.R588W) | *ZNF30* | WT | homo | M2141 | pair22 | NM_020845.3:c.1768G>A (p.V590M) | *PITPNM2* | WT | homo |
| C3458 | pair7 | NM_025179.4:c.365C>A (p.S122Y) | *PLXNA2* | WT | homo | M2141 | pair22 | NM_144686.4:c.533T>C (p.M178T) | *TMC4* | WT | homo |
| C3458 | pair7 | NM_001365276.2:c.9467C>T (p.P3156L) | *TNXB* | WT | homo | M3113 | pair23 | NM_003637.5:c.3499C>T (p.Q1167*) | *ITGA10* | WT | homo |
| C4098 | pair8 | NM_006952.4:c.527G>A (p.R176Q) | *UPK1B* | WT | homo | M3113 | pair23 | NM_002772.3:c.2704C>G (p.Q902E) | *TMPRSS15* | WT | homo |
| C4098 | pair8 | NM_022463.5:c.802C>T (p.R268W) | *NXN* | WT | homo | N0443 | pair24 | NM_002705.5:c.1420C>T (p.R474W) | *PPL* | WT | homo |
| L1134 | pair16 | NM_001368882.1:c.478G>A (p.A160T) | *COL13A1* | WT | homo | N0501 | pair25 | NM_017412.4:c.1414A>G (p.M472V) | *FZD3* | WT | homo |
| L1134 | pair16 | NM_030665.4:c.5254G>A (p.G1752R) | *RAI1* | WT | homo | N0501 | pair25 | NM_177531.6:c.10792A>C (p.S3598R) | *PKHD1L1* | WT | homo |
| L1134 | pair16 | NM_004653.5:c.1090G>A (p.A364T) | *KDM5D* | WT | homo | N0501 | pair25 | NM_001130448.3:c.224C>A (p.T75N) | *C15orf62* | WT | homo |
| M2998 | pair18 | NM_001164508.2:c.23510C>T (p.T7837M) | *NEB* | WT | homo | N0501 | pair25 | NM_002373.6:c.5573C>G (p.S1858C) | *MAP1A* | WT | homo |
| M2998 | pair18 | NM_020707.4:c.1150G>T (p.D384Y) | *HHATL* | WT | homo | N0501 | pair25 | NM_001384474.1:c.2557C>T (p.R853C) | *LOXHD1* | WT | homo |
| M2998 | pair18 | NM_000428.3:c.1295C>T (p.P432L) | *LTBP2* | WT | homo | N0501 | pair25 | NM_021785.6:c.773C>T (p.S258F) | *RAI2* | WT | homo |
| M2998 | pair18 | NM_015719.4:c.3148C>T (p.P1050S) | *COL5A3* | WT | homo | P2437 | pair26 | NM_001277285.4:c.3943C>T (p.L1315F) | *IGSF9B* | WT | homo |
| M2998 | pair18 | NM_025176.6:c.1631A>T (p.E544V) | *NINL* | WT | homo | T1019 | pair29 | NM_006589.3:c.1463T>C (p.L488P) | *ENTREP3* | WT | homo |
| M3235 | pair19 | NM_152492.3:c.1258C>T (p.Q420*) | *CCDC27* | WT | homo | T1019 | pair29 | NM_003383.5:c.902G>A (p.R301Q) | *VLDLR* | WT | homo |
| M3235 | pair19 | NM_001267046.2:c.200A>G (p.H67R) | *FRMD6* | WT | homo | T1019 | pair29 | NM_001007122.4:c.1739C>T (p.S580F) | *FSD2* | WT | homo |
| M3321 | pair20 | NM_014640.5:c.1132C>T (p.R378W) | *TTLL4* | WT | homo | T1019 | pair29 | NM_001297436.2:c.1240G>A (p.V414M) | *HAS1* | WT | homo |
| M3321 | pair20 | NM_178496.4:c.68C>T (p.P23L) | *MB21D2* | WT | homo | T1034 | pair30 | NM_016239.4:c.10391C>T (p.T3464M) | *MYO15A* | WT | homo |
| M3321 | pair20 | NM_182982.3:c.641C>T (p.A214V) | *GRK4* | WT | homo | T1034 | pair30 | NM_001136191.3:c.473C>T (p.S158F) | *KANK2* | WT | homo |
| M3321 | pair20 | NM_003643.4:c.1211C>T (p.P404L) | *GCM1* | WT | homo | V1177 | pair31 | NM_001165946.2:c.1834C>T (p.R612C) | *IDE* | WT | homo |
| M3321 | pair20 | NM_005715.3:c.31G>C (p.G11R) | *UST* | WT | homo | V1177 | pair31 | NM_002605.3:c.160G>T (p.E54*) | *PDE8A* | WT | homo |
| M3321 | pair20 | NM_012454.4:c.485T>G (p.L162R) | *TIAM2* | WT | homo | V1177 | pair31 | NM_002666.5:c.902C>T (p.T301M) | *PLIN1* | WT | homo |
| M3321 | pair20 | NM_052924.3:c.1718C>T (p.T573M) | *RHPN1* | WT | homo | V1177 | pair31 | NM_024320.4:c.41G>A (p.R14Q) | *PRR15L* | WT | homo |
| M3321 | pair20 | NM_017570.5:c.2558G>A (p.R853Q) | *OPLAH* | WT | homo | | | | | | |

***Table 15: Tables illustrating the 111 variants showing allelic discordant expression.*** *Table A) illustrates the 56 variants where the alternative allele is expressed in the affected co-twin; Table B) illustrates the 55 variants where the alternative allele is expressed only by the unaffected co-twins. In each table the twin ID, Pair ID, Variant annotation, Gene HUGO name, RNA status in the affected and unaffected twin are shown.*

To evaluate potential biological relevance of these findings, the genes were divided into two subsets according to the direction of the allelic expression pattern. Separate over-representation analyses were then performed using EnrichR. The gene set characterized by expression of the alternative allele in affected twins showed statistically significant enrichment for eight Gene Ontology Biological

Processes terms. Among these, Response to caffeine and Cellular response to caffeine both remained significant after multiple-testing correction (q-value = 0.026 for both) (***Figure 28***). Both terms were supported by *RYR1* and *RYR3* which also contributed to the Release of sequestrated calcium ion into cytosol by sarcoplasmic reticulum (q-value = 0.026), consistent with the role of ryanodine receptors in intracellular calcium homeostasis in cells.

(A)




(B)

| Pathway | p value | q value | Genes involved |
|---|---|---|---|
| Response to caffeine | 0.0001603 | 0.02564 | *RYR1, RYR3* |
| Cellular response to caffeine | 0.0001603 | 0.02564 | *RYR1, RYR3* |
| Release of sequestrated calcium ion into cytosol by sarcoplasmic reticulum | 0.0002737 | 0.02564 | *RYR1, RYR3* |

***Figure 28: EnrichR analysis of genes showing differential allele-specific expression between affected and unaffected twins.*** *The analysis included genes harboring heterozygous DNA variants for which the alternative allele was expressed in the affected twin, whereas only the reference allele was detected in the unaffected co-twin. (A) Enriched GO Biological Process terms. (B) Table summarizing enrichment results, including the term name, nominal p-values, q-values, and contributing genes.*

Conversely, the reciprocal analysis performed on genes for the alternative allele was detected in the RNA of unaffected twins but not in their affected co-twins, identifying no significantly enriched terms.

Among the genes showing differential ASE, *MC1R* was considered of particular interest. In pair 25, *MC1R* stop-gain variant p.Y152* (NM_002386.4:c.853C>A) was detected at the RNA level only in the affected twin, whereas only the reference allele was observed in the healthy twin. The variant is currently classified as VUS in Clinvar. Recent longitudinal studies have associated LoF mutations in this gene with faster motor decline in PD patients.

## 5. DISCUSSION

This study investigates the contribution of rare variants to the genomic landscape of PD through three comparisons involving unrelated PD patients and healthy controls, together with a smaller cohort of MZ twins discordant for PD. Gene-based association analyses identified different patterns across the unrelated cohorts. In the *GBA1* wt comparison, 25 genes remained at a significant level after FDR correction, whereas no gene reached FDR-corrected significance in either comparison involving *GBA1* variant-carriers. Consequently, the subsequent analyses of OR-ranked or nominally significant genes should be regarded as exploratory and interpreted primarily in terms of biological convergence rather than as evidence of definitive gene-disease associations.

The ORs provided an additional descriptive measure of the direction of the observed differences. A predominance of positive ratios was observed in *GBA1* variant-carrier PD patients compared with both wild-type patients and healthy controls. Conversely, the distribution was more balanced in the *GBA1* wt case-control comparison. Although these ratios do not provide an independent test of association, they suggest that qualifying rare alternative alleles may be relatively more frequent among *GBA1* variant carriers within the analyzed gene sets.

The enrichment analyses suggested potentially distinct biological signatures according to the *GBA1* status. In the *GBA1* wt cohort, over-representation analysis revealed a marked enrichment of immune-related pathways, particularly IFN-γ signaling and the role of JAK family kinases in IL-6 signaling. The contributing genes included *SOCS3*, *IFNGR1*, *PIAS1*, and *IRF6*. *IFNGR1* encodes a subunit of the IFN-γ receptor, while *SOCS3* and *PIAS1* act as negative regulators of JAK-STAT and interferon signaling; *IRF6* belongs to the interferon regulatory factors family. Together, this finding is consistent with the growing evidence implicating immune dysregulation and neuroinflammation in PD pathophysiology. Particularly, higher levels of INF-γ and IL-6 and JAK-STAT activation have been found in sporadic PD cases and associated with pro-inflammatory microglial responses and dopaminergic neurotoxicity *(Isik et al. 2023)*. Moreover, in animal models, IFN-γ deficiency attenuated dopaminergic neuron loss, while pharmacological inhibition of JAK kinases reduced neuroinflammatory transcriptional programs and α-syn inclusions (*Hong et al. 2024*). Lastly, a recent study demonstrated that IFN-γ induces coordinated changes in LRRK2, GCase, and cathepsin activity in monocytes of PD patients, supporting a functional interplay between interferon signaling and PD-relevant lysosomal enzymes independently of *GBA1* mutations *(L. P. Hughes et al. 2025)*. Although it remains unclear whether neuroinflammation represents a primary driver of dopaminergic degeneration or a secondary response to neuronal damage, the present results support further investigation of immune pathways in PD pathogenesis.

A different pattern emerged from the comparison between *GBA1* variant-carriers PD patients and healthy controls, in which pathways related to mitochondrial homeostasis and BH4 biosynthesis were among the most relevant findings. Mitochondrial dysfunction is considered a central component of PD pathogenesis, and DA neurons are particularly vulnerable to impaired mitochondrial activity. The enrichment of mitochondrial processes in *GBA1* variant-carriers is consistent previous evidence connecting GCase deficiency with altered mitochondrial quality control. *GBA1* mutations may disrupt mitochondria-lysosome contact sites, impair mitophagy and reduce mitochondrial bioenergetic function, thus promoting the accumulation of damaged mitochondria in neuronal models and post-mortem brain tissue from *GBA1* variant-carrier PD patients *(Sheshadri et al. 2026; Yuan et al. 2022)*. The parallel enrichment of BH4 biosynthesis pathways provides a complementary connection between mitochondrial function and dopamine metabolism. BH4 is an essential cofactor for tyrosine hydroxylase, the rate-limiting enzyme in dopamine synthesis, and has also been implicated in ATP production and ROS clearance in the mitochondria. Reduced plasma BH4 levels have been reported in sporadic PD together with decreased dopamine and serotonin levels, while congenital BH4 deficiency can produce parkinsonian manifestations through striatal dopamine depletion *(Cronin et al., 2023)*. Therefore, the simultaneous representation of mitochondrial and BH4-related pathways may indicate convergence on processes relevant to dopaminergic neuronal function, although this interpretation requires confirmation in independent cohorts and functional models.

Disease-term enrichment also identified terms related to *APP*, dementia and amyloidosis. These findings are of potential interest because *GBA1*-associated PD has been linked to a greater frequency of non-motor symptoms, cognitive decline, and elevated risk of progression to dementia with Lewy bodies *(Vieira and Schapira 2022)*. It raises the hypothesis that variants affecting genes involved in cognitive and amyloid-related processes may contribute to the clinical heterogeneity of *GBA1*-PD. Direct integration with cognitive and longitudinal clinical data will be required to test this possibility.

Biologically related findings emerged from the direct comparison of the two different PD subgroups defined by their *GBA1* status. IPA highlighted several terms related to mitochondrial energy metabolism, including glycolysis, glycogen degradation, and the Krebs cycle. DA neurons have a continuous and substantial energy requirement, and alterations in glucose metabolism and mitochondrial oxidative phosphorylation have been repeatedly implicated in PD. Postmortem transcriptomic studies have shown metabolic changes compatible with increased reliance on aerobic glycolysis, while glycolytic dysregulation may influence neuronal survival and response to oxidative stress *(Flores-Ponce and Velasco 2024; Shen et al. 2025)*. In the present study, the over-representation of genes involved in these processes is therefore consistent with a greater contribution

of energy-metabolism-related mechanisms in the *GBA1* variant-carrier subgroup, although pathway activity was not directly measured.

The concurrent enrichment of glutamate and GABA degradation pathways also pointed to the GABA shunt, a metabolic route that links neurotransmitter catabolism to the Krebs cycle through succinate production. This pathway is particularly important in nervous tissue because it contributes both to neurotransmitter turnover and cellular energy metabolism *(Antoni et al. 2011)*. Altered GABA/glutamate homeostasis has been associated with oxidative stress and mitochondrial dysfunction in PD models *(Alharbi et al. 2024)*. Taken together, the enrichment of canonical energy pathways and GABA-shunt-related processes on convergence on metabolic functions that may distinguish *GBA1* variant-carriers from *GBA1* wt PD patients.

Metascape analysis on the same gene set identified central nervous system development (GO:0007417) as the most significantly enriched term. Since both groups included patients affected by PD, his result cannot simply be attributed to a case-control distinction. It suggests that genes involved in nervous-system-related processes may be different according to *GBA1* status. Nevertheless, the comparison does not demonstrate altered gene expression or pathway activity, and the biological relevance of this signal should be tested in independent datasets.

The disease-term enrichment was consistent with some of the findings observed in the comparison between *GBA1* variant carriers and healthy controls. Terms related to amyloid formations, dementia, mitochondrial dysfunction, and PD were represented in both the Elsevier Pathway Collection and the Jensen disease database. These results were mainly driven by *SNCA*, *APP*, *OGDH,* and *LYZ*. *OGDH* encodes a component of the mitochondrial α-ketoglutarate dehydrogenase complex, and impaired activity of this complex has been linked to progressive neurodegeneration *(Yoon et al. 2017)*. *LYZ* encodes lysozyme C, which can interact with amyloid-β species and reduce their aggregation, while missense mutations in this gene could cause hereditary systemic amyloidosis *(Helmfors et al. 2015)*. Their presence provides biological context for the enriched terms but does not establish a direct role in the clinical phenotype of the analyzed patients.

Particular attention was given to *SNCA* and *APP,* as both genes contributed to several disease-related terms. Although SNCA showed a positive OR, the only variant classified as likely pathogenic in Clinvar was identified in two *GBA1* wild-type patients. In these individuals, the *SNCA* variant may represent a major genetic determinant of PD rather than a modifier of *GBA1*-associated risk. Conversely, the rare *SNCA* variants identified in *GBA1* variant-carriers were not classified as pathogenic and may therefore represent candidate modifiers of the genetic background associated with *GBA1*. However, their functional relevance and any effect on disease risk or phenotypic

expression remain to be established. This observation also illustrates that the direction of a gene-level burden ratio does not necessarily reflect the distribution or clinical relevance of individual variants.

*For APP*, more distinct variants were detected in *GBA1* wild-type patients, however the p.Ala673Val (NM_000484.4:c.2287C>T) variant, classified as likely pathogenic in Clinvar and previously associated with early-onset AD, was found in two PD patients carrying a *GBA1* variant. This finding is noteworthy and is consistent with the disease-term enrichment related to amyloid and dementia. However, because it involves only two individuals and no direct cognitive comparison was performed, it should be considered hypothesis-generating rather than evidence that *GBA1* status shifts PD toward an AD-like phenotype.

The analysis of MZ twins discordant for PD provided a complementary approach for investigating candidate postzygotic and allele-specific events against a largely shared genetic background. Gene-based testing identified no nominally significant or multiple-testing-corrected associations, most likely reflecting the limited sample size and the small number of within-pair DNA-discordant variants. This negative result is important and indicates that the subsequent overlap and enrichment analyses should be interpreted as descriptive and exploratory.

Among genes harboring within-pair DNA-discordant variants, eight overlapped with a curated list of PD-associated genes. Two different variants were identified in the *MCCC1* gene in two affected twins, including p.Arg281Ter (NM_020166.5:c.968G>A), which is classified as pathogenic or likely pathogenic in relation to 3-methylcrotonylglycinuria. *MCCC1* encodes a mitochondrial enzyme involved in leucine catabolism and has also emerged as a PD susceptibility locus in genetic studies *(Nalls et al. 2014)*. Proposed links with PD include mitochondrial dysfunction, autophagy, and inflammatory signaling, as *MCCC1* has been shown to enhance NF-kB signaling *(Cao et al. 2016)*. Nevertheless, the clinical classification of p.Arg281Ter (NM_020166.5:c.968G>A) relates to an established metabolic disorder rather than to PD, and its possible contribution to the phenotype of the affected twin remains uncertain.

Comparison with the genes selected in the *GBA1* wt cohort identified four overlapping genes, with *MFSD5* being the only one that had remained significant after FDR correction in the unrelated case-control analysis. In the twin cohort, DNA-discordant *MFSD5* variant was detected in the affected twin of pair 24, in agreement with the direction of the OR observed in the unrelated cohort (OR = 1.90). *MFSD5*, also known as *SLC61A1*, encodes a molybdate transporter involved in the cellular uptake of molybdate required for molybdenum cofactor-dependent enzymes. Altered activity of these enzymes may affect redox metabolism and sulfite detoxification *(Atwal and Scaglia 2016)*. However,

a role for *MFSD5* in PD has not been established, and its overlap between the two analyses should be regarded as a candidate signal requiring replication and functional investigation.

Enrichment analyses of genes carrying DNA-discordant variants identified nervous-system-related disease terms both in the complete gene set and in which variants observed predominantly in affected twins. In contrast, the subset associated predominantly with unaffected twins showed no comparable nervous-system-related enrichment. This difference is compatible with the possibility that postzygotic events contribute to phenotypic discordance within some pairs. However, the two subsets were not compared through a formal association test, and factors such as gene size, sequence coverage, variant callability, and the small number of observations per gene may have influenced the results. The enrichment should therefore be viewed as a source of candidate genes and processes rather than as evidence that postzygotic variants explain PD discordance.

The integration of DNA and RNA sequencing data made it possible to examine differential ASE within the pairs. Overall, 111 variants showed discordant allelic expression, with the alternative allele detected in the affected twin and the reference allele in the unaffected co-twin for 56 variants, and the reciprocal pattern for 55 variants. Over-representation analysis of genes belonging to the first subgroup highlighted caffeine-response pathways and pathways related to calcium release from intracellular stores. Conversely, the reciprocal gene set produced no significantly enriched terms.

The caffeine-response and calcium-related terms were driven by *RYR1* and *RYR3*, which encode ryanodine receptors responsible for intracellular calcium homeostasis. Interestingly, altered calcium handling has been implicated in the selective vulnerability of dopaminergic neurons, partly through its effects on mitochondrial and oxidative stress *(Savio et al. 2009; Surmeier et al., 2011)*. Sustained calcium influx places a substantial buffering demand on the endoplasmic reticulum and mitochondria, and excessive calcium release through ryanodine receptors may further promote mitochondrial stress and reactive oxygen species production. Ryanodine receptors have consequently been proposed as potential therapeutic targets in PD and other neurodegenerative diseases (*L. Huang et al. 2017; L. Sun and Wei 2021*). Nevertheless, the enrichment observed here was based on only two contributing genes and does not demonstrate calcium dyshomeostasis in the affected twins.

Caffeine intake has a well-established epidemiological association with reduced PD risk, generally attributed to antagonism of the adenosine A2A receptor and the resulting modulation of dopaminergic neurotransmission *(Schwarzschild et al. 2003; Kalda et al. 2006)*. However, the GO caffeine-response terms identified in this analysis were driven by *RYR1* and *RYR3* and should not be interpreted

as evidence of differential caffeine exposure or A2A receptor activity in the twin cohort. Instead, they provide an additional annotation linking the same genes to calcium-regulatory processes.

Another potentially relevant observation involved *MC1R.* In pair 25, the alternative allele of the stop-gain variant p.Tyr152Ter (NM_002386.4:c.853C>A) was detected at the RNA level in the affected twin, whereas only the reference allele was observed in the healthy co-twin. *MC1R* encodes the melanocortin 1 receptor, best known for its role in pigmentation but has also recently been investigated as a genetic modifier of PD risk and progression. Experimental activation of *MC1R* has been reported to protect DA neurons from α-syn toxicity through Nrf2 antioxidant activity *(Cai et al. 2022)*. Additionally, a recent study proved that *MC1R* variant-carriers display a 30% faster motor decline compared to non-carriers, along with an increased disease progression *(Schumacher et al. 2026)*. This makes the finding of a stop-gain *MC1R* variant, present exclusively in the PD-affected twin and absent in the unaffected co-twin, particularly noteworthy. However, p.Tyr152Ter (NM_002386.4:c.853C>A) is currently classified as a variant of uncertain significance, it was observed in a single pair, and its allele-specific expression was measured in peripheral blood. Its relevance to PD progression consequently remains to be established.

This study presents several strengths. First, the unrelated case-control cohorts provided sufficient statistical power to identify gene-level signals, including 25 genes that remained significant after FDR correction in the *GBA1* wild-type comparison. The direct comparison between *GBA1* variant-carrier and *GBA1* wild-type PD patients also enabled the rare-variant landscape to be examined according to *GBA1* status independently of healthy controls. Second, the integration of whole-exome and RNA sequencing data in the twin cohort allowed the identification of candidate within-pair differences in allelic expression that would not have been captured through genotype-based analyses alone. Third, the discordant MZ twin design provided a naturally controlled experimental system in which co-twins share germline genetic background, thereby reducing genetic heterogeneity and facilitating the investigation of candidate postzygotic and allele-specific events. Finally, the large unrelated cohorts were recruited over two decades at a single, specialized movement disorder center, using consistent diagnostic criteria and data collection protocol, reducing the diagnostic heterogeneity that can affect multi-center studies.

Several limitations should also be acknowledged. First, the twin cohort was underpowered: with only 23 discordant MZ pairs and one triplet, rare-variant association analysis did not identify significant associations, limiting the generalizability of the findings. Second, only 41 healthy *GBA1* carriers were available, preventing a sufficiently powered comparison between *GBA1* variant-carrier PD patients

and healthy carriers. Such a comparison would have helped separate effects associated with *GBA1* status from those associated with PD diagnosis. Third, RNA was extracted from peripheral whole blood rather than brain or dopaminergic tissue. Although blood is one of the few accessible tissues in living patients, allele-specific expression may be tissue- and cell-type-dependent, and differences in blood-cell composition may influence the observed patterns. Fourth, the cohorts originated from a single center and a relatively homogeneous Italian population, which may limit the transferability of the findings to populations with different genetic backgrounds. Finally, some of the most interesting findings in this study, including *SNCA* and *APP* variants and the *RYR1*/*RYR3*/*MC1R* allele-specific events, were based on one or few carriers and were not evaluated in an independent replication cohort. These observations should be interpreted as candidate signals that require technical confirmation, replication, integration with longitudinal clinical data, and functional validation.

## 6. CONCLUSIONS

This study contributes to the characterization of the genetic and biological heterogeneity underlying PD by combining rare-variant analyses in unrelated cohorts stratified according to *GBA1* status with the study of MZ twins discordant for PD. The results suggest that the rare-variant background and the biological processes represented among candidate genes may differ according to *GBA1* status, supporting the view of PD as a spectrum of molecular heterogeneous conditions rather than a single homogeneous disorder.

The integration of DNA and RNA sequencing data in discordant MZ twins provided a complementary perspective by enabling the identification of candidate postzygotic variants and within-pair differences in allele-specific expression. Although these observations do not explain the phenotypic discordance between co-twins, they demonstrate the potential value of combining population-based and genetically matched study designs to investigate molecular events that cannot be captured through conventional case-control analyses alone. If replicated, the molecular patterns identified across the different cohorts could contribute to the future stratification of PD patients and support the identification of biomarkers associated with specific genetic or biological subgroups.

These findings should nevertheless be considered exploratory and hypothesis-generating. Their biological and clinical relevance will require confirmation in larger and independent cohorts, technical and functional validation of candidate variants, more comprehensive analysis of allele-specific expression, and integration with longitudinal clinical data. Overall, this study provides a foundation for future investigations into how *GBA1* status, the broader rare-variant background, postzygotic events, and allele-specific regulation may jointly contribute to PD susceptibility, disease expression, and clinical variability.

**References**


1. Alharbi B, Al-kuraishy HM, Al-Gareeb AI, et al. Role of GABA pathway in motor and non-motor symptoms in parkinson's disease: A bidirectional circuit. *European journal of medical research*. 2024;29(1). doi:10.1186/s40001-024-01779-7

2. American Psychiatric Association. DSM-4 Task Force. *Diagnostic and Statistical Manual of Mental Disorders*. 2017.

3. Antoni K, Żylińska L, Boczek T, Rębas E. Gaba-shunt enzymes activity in GH3 cells with reduced level of PMCA2 or PMCA3 isoform. *Biochemical and Biophysical Research Communications*. 2011;411(4):815-820. doi:10.1016/j.bbrc.2011.07.043

4. Antony PMA, Diederich NJ, Krüger R, Balling R. The hallmarks of parkinson's disease. *FEBS Journal*. 2013;280(23):5981-5993. doi:10.1111/febs.12335

5. Ashkan K, Rogers P, Bergman H, Ughratdar I. Insights into the mechanisms of deep brain stimulation. *Nature reviews Neurology*. 2017;13(9):548-554. doi:10.1038/nrneurol.2017.105

6. Atkinson EG, Artomov M, Loboda AA, et al. Discordant calls across genotype discovery approaches elucidate variants with systematic errors. *Genome Research*. 2023;33(6):999-1005. doi:10.1101/gr.277908.123

7. Atwal PS, Scaglia F. Molybdenum cofactor deficiency. *Molecular Genetics and Metabolism*. 2016;117(1):1-4. doi:10.1016/j.ymgme.2015.11.010

8. Bandres-Ciga S, Diez-Fairen M, Kim JJ, Singleton AB. Genetics of parkinson's disease: An introspection of its journey towards precision medicine. *Neurobiology of Disease*. 2020;137:104782. doi:10.1016/j.nbd.2020.104782

9. Barton B, Grabli D, Bernard B, et al. Clinical validation of movement disorder society–recommended diagnostic criteria for parkinson's disease with dementia. *Movement Disorders*. 2011;27(2):248-253. doi:10.1002/mds.24059

10. Bergsland N, Zivadinov R, Schweser F, Hagemeier J, Lichter D, Guttuso T. Ventral posterior substantia nigra iron increases over 3 years in parkinson's disease. *Movement Disorders*. 2019;34(7):1006-1013. doi:10.1002/mds.27730

11. *Best Practices Workflows*. 2025. Accessed July 28, 2026. https://gatk.broadinstitute.org/

12. Blauwendraat C, Heilbron K, Vallerga CL, et al. Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and α-synuclein mechanisms. *Movement Disorders*. 2019;34(6):866-875. doi:10.1002/mds.27659

13. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.

14. Book A, Guella I, Candido T, et al. A meta-analysis of α-synuclein multiplication in familial parkinsonism. *Frontiers in Neurology*. 2018;9. doi:10.3389/fneur.2018.01021

15. Boulos C, Yaghi N, El Hayeck R, Heraoui GN, Fakhoury-Sayegh N. Nutritional risk factors, microbiota and parkinson's disease: What is the current evidence? *Nutrients*. 2019;11(8):1896. doi:10.3390/nu11081896

16. Bourdenx M, Dehay B. What lysosomes actually tell us about parkinson's disease? *Ageing Research Reviews*. 2016;32:140-149. doi:10.1016/j.arr.2016.02.008

17. Brolin KA, Schaeffer E, Kuri A, et al. Environmental risk factors for parkinson's disease: A critical review and policy implications. *Movement Disorders*. Published online November 27, 2024. doi:10.1002/mds.30067

18. *CADD - Combined Annotation Dependent Depletion*. 2024. Accessed July 29, 2026. https://cadd.bihealth.org/

19. Cai W, Srivastava P, Feng D, et al. Melanocortin 1 receptor activation protects against alpha-synuclein pathologies in models of parkinson's disease. *Molecular Neurodegeneration*. 2022;17(1). doi:10.1186/s13024-022-00520-4

20. Cao Z, Xia Z, Zhou Y, et al. Methylcrotonoyl-CoA carboxylase 1 potentiates rlr-induced NF-κB signaling by targeting MAVS complex. *Scientific Reports*. 2016;6(1). doi:10.1038/srep33557

21. Cerri S, Mus L, Blandini F. Parkinson's disease in women and men: What's the difference? *Journal of Parkinson's Disease*. 2019;9(3):501-515. doi:10.3233/jpd-191683

22. Chen EY, Tan CM, Kou Y, et al. ENRICHR: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14(1):128. doi:10.1186/1471-2105-14-128

23. Chen H, Huffman JE, Brody JA, et al. Efficient variant set mixed model association tests for continuous and binary traits in large-scale whole-genome sequencing studies. *The American Journal of Human Genetics*. 2019;104(2):260-274. doi:10.1016/j.ajhg.2018.12.012

24. Costa HN, Esteves AR, Empadinhas N, Cardoso SM. Parkinson's disease: A multisystem disorder. *Neuroscience Bulletin*. 2022;39(1):113-124. doi:10.1007/s12264-022-00934-6

25. Cramb KML, Beccano-Kelly D, Cragg SJ, Wade-Martins R. Impaired dopamine release in parkinson's disease. *Brain*. 2023;146(8). doi:10.1093/brain/awad064

26. Cronin SJF, Yu W, Hale A, et al. Crucial neuroprotective roles of the metabolite BH4 in dopaminergic neurons. *bioRxiv*. Published online May 8, 2023:2023.05.08.539795. doi:10.1101/2023.05.08.539795

27. Dal GM, Bekir Ergüner, Sağıroğlu MS, et al. Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. *Journal of Medical Genetics*. 2014;51(7):455-459. doi:10.1136/jmedgenet-2013-102197

28. Day JO, Mullin S. The genetics of parkinson's disease and implications for clinical practice. *Genes*. 2021;12(7):1006.

29. Deng H, Wang P, Jankovic J. The genetics of parkinson disease. *Ageing Research Reviews*. 2018;42:72-85. doi:10.1016/j.arr.2017.12.007

30. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*. 2011;43(5):491-498. doi:10.1038/ng.806

31. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal rna-seq aligner. *Bioinformatics*. 2012;29(1):15-21. doi:10.1093/bioinformatics/bts635

32. Dorsey ER, Sherer T, Okun MS, Bloem BR. The emerging evidence of the parkinson pandemic. *Journal of Parkinson's Disease*. 2018;8(S1):S3-S8. doi:10.3233/jpd-181474

33. Elkouzi A, Vedam-Mai V, Eisinger RS, Okun MS. Emerging therapies in parkinson disease — repurposed drugs and new approaches. *Nature Reviews Neurology*. 2019;15(4):204-223. doi:10.1038/s41582-019-0155-7

34. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048. doi:10.1093/bioinformatics/btw354

35. Fabrizio Stocchi, Bravi D, Emmi A, Antonini A. Parkinson disease therapy: Current strategies and future research priorities. *Nature Reviews Neurology*. 2024;20. doi:10.1038/s41582-024-01034-x

36. Flores-Ponce X, Velasco I. Dopaminergic neuron metabolism: Relevance for understanding parkinson's disease. *Metabolomics*. 2024;20(6). doi:10.1007/s11306-024-02181-4

37. Gan-Or Z, Liong C, Alcalay RN. Gba-associated parkinson's disease and other synucleinopathies. *Current Neurology and Neuroscience Reports*. 2018;18(8). doi:10.1007/s11910-018-0860-4

38. Guo MH, Plummer L, Chan YM, Hirschhorn JN, Lippincott MF. Burden testing of rare variants identified through exome sequencing via publicly available control data. *The American Journal of Human Genetics*. 2018;103(4):522-534. doi:10.1016/j.ajhg.2018.08.016

39. *GVCF - Genomic Variant Call Format*. 2024. Accessed July 28, 2026. https://gatk.broadinstitute.org/hc/en-us/articles/360035531812-GVCF-Genomic-Variant-Call-Format

40. Hassane Lakhloufi. Gene-based SKAT-O test. MetricGate. April 28, 2025. Accessed July 31, 2026. https://metricgate.com/docs/gene-based-skat-o-test/

41. Helmfors L, Boman A, Civitelli L, et al. Protective properties of lysozyme on B-amyloid pathology: Implications for alzheimer disease. *Neurobiol Dis*. 2015;83:122-133. doi:10.1016/j.nbd.2015.08.024

42. Hong H, Wang Y, Menard M, et al. Suppression of the JAK/STAT pathway inhibits neuroinflammation in the line 61-PFF mouse model of parkinson's disease. *Res Sq*. Published online May 7, 2024:rs.3.rs-4307273. doi:10.21203/rs.3.rs-4307273/v1

43. Hu M, Li P, Wang C, et al. Parkinson's disease-risk protein TMEM175 is a proton-activated proton channel in lysosomes. *Cell*. 2022;185(13):2292-2308.e20. doi:10.1016/j.cell.2022.05.021

44. Huang L, Xue Y, Feng D, et al. Blockade of RyRs in the ER attenuates 6-ohda-induced calcium overload, cellular hypo-excitability and apoptosis in dopaminergic neurons. *Front Cell Neurosci*. 2017;11:52. doi:10.3389/fncel.2017.00052

45. Huang Y, Chen Q, Wang Z, et al. Risk factors associated with age at onset of parkinson's disease in the UK biobank. *npj Parkinson's Disease*. 2024;10(1):1-8. doi:10.1038/s41531-023-00623-9

46. Hughes AJ, Ben-Shlomo Y, Daniel SE, Lees AJ. What features improve the accuracy of clinical diagnosis in parkinson's disease: A clinicopathologic study. 1992. *Neurology*. 2001;57(10 Suppl 3):S34-8. Accessed July 28, 2026. https://pubmed.ncbi.nlm.nih.gov/11775598/

47. Hughes LP, Wallings RL, Agin-Liebes J, et al. Interferon gamma stimulates coordinated changes in LRRK2, gcase, and cathepsin activities in idiopathic and genetic parkinson's disease monocytes. *NPJ Parkinsons Dis*. 2025;11(1):337. doi:10.1038/s41531-025-01185-8

48. Illumina. Technology spotlight: Illumina ® sequencing illumina sequencing technology highest data accuracy, simple workflow, and a broad range of applications. In: 2010. Accessed July 28, 2026. https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

49. Isik S, Yeman Kiyak B, Akbayir R, Seyhali R, Arpaci T. Microglia mediated neuroinflammation in parkinson's disease. *Cells*. 2023;12(7):1012. doi:10.3390/cells12071012

50. Jia F, Fellner A, Kumar KR. Monogenic parkinson's disease: Genotype, phenotype, pathophysiology, and genetic testing. *Genes*. 2022;13(3):471. doi:10.3390/genes13030471

51. Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2016;374(2065):20150202. doi:10.1098/rsta.2015.0202

52. Jonsson H, Magnusdottir E, Eggertsson HP, et al. Differences between germline genomes of monozygotic twins. *Nature Genetics*. 2021;53(1):27-34. doi:10.1038/s41588-020-00755-1

53. Kalda A, Yu L, Oztas E, Chen JF. Novel neuroprotection by caffeine and adenosine A2A receptor antagonists in animal models of parkinson's disease. *Journal of the Neurological Sciences*. 2006;248(1-2):9-15. doi:10.1016/j.jns.2006.05.003

54. Kolahdouzan M, Hamadeh MJ. The neuroprotective effects of caffeine in neurodegenerative diseases. *CNS Neuroscience & Therapeutics*. 2017;23(4):272-290. doi:10.1111/cns.12684

55. Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*. 2014;30(4):523-530. doi:10.1093/bioinformatics/btt703

56. Kulcsarova K, Matej Skorvanek, Postuma RB, Berg D. Defining parkinson's disease: Past and future. *Journal of Parkinson's disease/Journal of Parkinson's disease (Online)*. Published online March 14, 2024:1-15. doi:10.3233/jpd-230411

57. Kuleshov MV, Jones MR, Rouillard AD, et al. ENRICHR: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016;44(W1):W90-W97. doi:10.1093/nar/gkw377

58. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: Study designs and statistical tests. *The American Journal of Human Genetics*. 2014;95(1):5-23. doi:10.1016/j.ajhg.2014.06.009

59. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *The American Journal of Human Genetics*. 2012;91(2):224-237. doi:10.1016/j.ajhg.2012.06.007

60. Lee SS. SKAT package . Published online January 15, 2023.

61. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics (oxford, England)*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324

62. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and samtools. *Bioinformatics*. 2009;25(16):2078-2079.

63. Ling Q, Herstine JA, Bradbury A, Gray SJ. Aav-based in vivo gene therapy for neurological disorders. *Nature Reviews Drug Discovery*. 2023;22(10):789-806. doi:10.1038/s41573-023-00766-7

64. Lunati A, Lesage S, Brice A. The genetic landscape of parkinson's disease. *Revue Neurologique*. 2018;174(9):628-643. doi:10.1016/j.neurol.2018.08.004

65. Malpartida AB, Williamson M, Narendra DP, Wade-Martins R, Ryan BJ. Mitochondrial dysfunction and mitophagy in parkinson's disease: From mechanism to therapy. *Trends in Biochemical Sciences*. 2020;46(4). doi:10.1016/j.tibs.2020.11.007

66. Marinus J, Zhu K, Marras C, Aarsland D, van Hilten JJ. Risk factors for non-motor symptoms in parkinson's disease. *The Lancet Neurology*. 2018;17(6):559-568. doi:10.1016/s1474-4422(18)30127-3

67. Mazzulli JR, Xu YH, Sun Y, et al. Gaucher disease glucocerebrosidase and α-synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell*. 2011;146(1):37-52. doi:10.1016/j.cell.2011.06.001

68. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. *Genome research*. 2010;20(9):1297-1303. doi:10.1101/gr.107524.110

69. Mehanna R, Jankovic J. Young-onset parkinson's disease: Its unique features and their impact on quality of life. *Parkinsonism & Related Disorders*. 2019;65:39-48. doi:10.1016/j.parkreldis.2019.06.001

70. Meoni S, Macerollo A, Moro E. Sex differences in movement disorders. *Nature Reviews Neurology*. 2020;16(2):84-96. doi:10.1038/s41582-019-0294-x

71. Milenkovic I, Blumenreich S, Futerman AH. GBA mutations, glucosylceramide and parkinson's disease. *Current Opinion in Neurobiology*. 2022;72:148-154. doi:10.1016/j.conb.2021.11.004

72. Morris HR, Spillantini MG, Sue CM, Williams-Gray CH. The pathogenesis of parkinson's disease. *Lancet (London, England)*. 2024;403(10423):293-304. doi:10.1016/S0140-6736(23)01478-2

73. 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):989-993. doi:10.1038/ng.3043

74. 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):989-993. doi:10.1038/ng.3043

75. Navarro-Romero A, Montpeyó M, Martinez-Vicente M. The emerging role of the lysosome in parkinson's disease. *Cells*. 2020;9(11):2399. doi:10.3390/cells9112399

76. Odds ratios: The odd one out? Stats By Slough. August 20, 2015. Accessed September 22, 2026. https://statsbyslough.com/2015/08/20/odds-ratios-the-odd-one-out/

77. Parkinson J. An essay on the shaking palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 1817;14(2):223-236. doi:10.1176/jnp.14.2.223

78. *Picard Tools - by Broad Institute*. Accessed July 26, 2026. http://broadinstitute.github.io/picard

79. Pickrell AM, Youle RJ. The roles of PINK1, parkin, and mitochondrial fidelity in parkinson's disease. *Neuron*. 2015;85(2):257-273. doi:10.1016/j.neuron.2014.12.007

80. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*. 2018;2(3):158-164. doi:10.1038/s41551-018-0195-0

81. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for parkinson's disease. *Movement Disorders*. 2015;30(12):1591-1601. doi:10.1002/mds.26424

82. Privé F, Luu K, Blum MGB, McGrath JJ, Vilhjálmsson BJ. Efficient toolkit implementing best practices for principal component analysis of population genetic data. , ed.Schwartz R. *Bioinformatics*. Published online May 16, 2020. doi:10.1093/bioinformatics/btaa520

83. Raza C, Anjum R, Shakeel NA. Parkinson's disease: Mechanisms, translational models and management strategies. *Life Sciences*. 2019;226:77-90. doi:10.1016/j.lfs.2019.03.057

84. Regier AA, Farjoun Y, Larson DE, et al. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nature communications*. 2018;9(1):4038. doi:10.1038/s41467-018-06159-4

85. Richa Bharti, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics*. 2019;22(1). Accessed July 24, 2026.

https://www.researchgate.net/publication/338048326_current_challenges_and_best-practice_protocols_for_microbiome_analysis

86. Robak LA, Jansen IE, van Rooij J, et al. Excessive burden of lysosomal storage disorder gene variants in parkinson's disease. *Brain : a journal of neurology*. 2017;140(12):3191-3203. doi:10.1093/brain/awx285

87. Savio CC, Gertler TS, James SD. Calcium homeostasis, selective vulnerability and parkinson's disease. *Trends in Neurosciences*. 2009;32(5):249-256. doi:10.1016/j.tins.2009.01.006

88. Sayers EW, Beck J, Bolton EE, et al. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*. 2024;53(D1). doi:10.1093/nar/gkae979

89. Schmitz MJ, Bashar A, Soman V, et al. Leveraging diverse genomic data to guide equitable carrier screening: Insights from gnomad v.4.1.0. *The American Journal of Human Genetics*. 2025;112(1):181-195. doi:10.1016/j.ajhg.2024.11.004

90. Schumacher JG, Zhang X, Wang J, et al. Common MC1R variants and parkinson disease progression. *JAMA Neurol*. Published online September 8, 2026:e262998. doi:10.1001/jamaneurol.2026.2998

91. Schwarzschild MA, Xu K, Oztas E, et al. Neuroprotection by caffeine and more specific A2A receptor antagonists in animal models of parkinson's disease. *Neurology*. 2003;61(11 Suppl 6):S55-61. doi:10.1212/01.wnl.0000095214.53646.72

92. Shen J, Yao E, Tian W, He J, Gu Y, Zhao D. Glycolytic pathways: The hidden regulators in parkinson's disease. *Heliyon*. 2025;11(3):e41831. doi:10.1016/j.heliyon.2025.e41831

93. Sheshadri P, Costa-Besada MA, Fisher A, et al. Targeting lysosomal pH restores mitochondrial quality control in gba1-mutant parkinson's disease. *Transl Neurodegener*. 2026;15(1):27. doi:10.1186/s40035-026-00559-z

94. Skou LD, Johansen SK, Okarmus J, Meyer M. Pathogenesis of dj-1/park7-mediated parkinson's disease. *Cells*. 2024;13(4):296. doi:10.3390/cells13040296

95. Smith L, Schapira AHV. GBA variants and parkinson disease: Mechanisms and treatments. *Cells*. 2022;11(8):1261. doi:10.3390/cells11081261

96. Straniero L, Asselta R, Bonvegna S, et al. The spid-gba study: Sex distribution, penetrance, incidence, and dementia in gba-pd. *Neurology Genetics*. 2020;6(6):e523. doi:10.1212/NXG.0000000000000523

97. Straniero L, Rimoldi V, Monfrini E, et al. Role of lysosomal gene variants in modulating <scp>*gba*</scp>-associated parkinson's disease risk. *Movement Disorders*. 2022;37(6):1202-1210. doi:10.1002/mds.28987

98. Sun H, Liu X, Ge H, Wang T, Wang Y, Li W. Association between serum zinc levels and the risk of parkinson's disease: A meta-analysis. *Biological Trace Element Research*. 2017;179(1):45-51. doi:10.1007/s12011-017-0941-2

99. Sun L, Wei H. Ryanodine receptors: A potential treatment target in various neurodegenerative disease. *Cellular and Molecular Neurobiology*. 2021;41(8):1613-1624. doi:10.1007/s10571-020-00936-w

100. Surmeier, D.J, Guzman, J.N, Sanchez-Padilla J, Schumacker, P.T. The role of calcium and mitochondrial oxidant stress in the loss of substantia nigra pars compacta dopaminergic neurons in parkinson's disease. *Function and Dysfunction of the Basal Ganglia*. 2011;198:221-231. doi:10.1016/j.neuroscience.2011.08.045

101. Sveinbjornsdottir S. The clinical symptoms of parkinson's disease. *Journal of Neurochemistry*. 2016;139(S1):318-324. doi:10.1111/jnc.13691

102. Szunyogh S, Carroll E, Wade-Martins R. Recent developments in gene therapy for parkinson's disease. *Molecular Therapy*. Published online March 2025. doi:10.1016/j.ymthe.2025.03.030

103. *Technology Spotlight: Illumina ® Sequencing*.

104. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

105. The International HapMap Consortium. The international HapMap project. *Nature*. 2003;426(6968):789-796. doi:10.1038/nature02168

106. Trevisan L, Gaudio A, Monfrini E, Avanzino L, Di Fonzo A, Mandich P. Genetics in parkinson's disease, state-of-the-art and future perspectives. *British Medical Bulletin*. 2024;149(1):ldad035. doi:10.1093/bmb/ldad035

107. TruSeq stranded total RNA reference guide. In: 2017. Accessed July 31, 2026. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseq-stranded-total-rna-workflow/truseq-stranded-total-rna-workflow-reference-1000000040499-00.pdf

108. Tysnes OB, Storstein A. Epidemiology of parkinson's disease. *Journal of Neural Transmission*. 2017;124(8):901-905. doi:10.1007/s00702-017-1686-y

109. Vadgama N, Pittman A, Simpson M, et al. De novo single-nucleotide and copy number variation in discordant monozygotic twins reveals disease-related genes. *European journal of human genetics: EJHG*. 2019;27(7):1121-1133. doi:10.1038/s41431-019-0376-7

110. van Dongen J, Slagboom PE, Draisma HHM, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. *Nature Reviews Genetics*. 2012;13(9):640-653. doi:10.1038/nrg3243

111. Vieira SRL, Schapira AHV. Glucocerebrosidase mutations and parkinson disease. *Journal of Neural Transmission*. 2022;129(9):1105-1117. doi:10.1007/s00702-022-02531-3

112. Vos M. Mitochondrial complex i deficiency: Guilty in parkinson's disease. *Signal Transduction and Targeted Therapy*. 2022;7(1). doi:10.1038/s41392-022-00983-3

113. Wang K. *Gene-Based Annotation - ANNOVAR Documentation*. 2023. Accessed July 29, 2026. https://annovar.openbioinformatics.org/en/latest/user-guide/gene/

114. Wilhelmus MMM, Nijland PG, Drukarch B, de Vries HE, van Horssen J. Involvement and interplay of parkin, PINK1, and DJ1 in neurodegenerative and neuroinflammatory disorders. *Free Radical Biology and Medicine*. 2012;53(4):983-992. doi:10.1016/j.freeradbiomed.2012.05.040

115. Xiao B, Zhou Z, Chao Y, Tan EK. Pathogenesis of parkinson's disease. *Neurologic Clinics*. 2025;43(2):185-207. doi:10.1016/j.ncl.2024.12.003

116. Xie Z, Bailey A, Kuleshov MV, et al. Gene set knowledge discovery with ENRICHR. *Current Protocols*. 2021;1(3). doi:10.1002/cpz1.90

117. Xiong Y, Yu J. LRRK2 in parkinson's disease: Upstream regulation and therapeutic targeting. *Trends in Molecular Medicine*. 2024;30(10):982-996. doi:10.1016/j.molmed.2024.07.003

118. Ye H, Robak LA, Yu M, Cykowski M, Shulman JM. Genetics and pathogenesis of parkinson's syndrome. *Annual Review of Pathology: Mechanisms of Disease*. 2023;18(1):95-121. doi:10.1146/annurev-pathmechdis-031521-034145

119. Yoon WH, Sandoval H, Nagarkar-Jaiswal S, et al. Loss of nardilysin, a mitochondrial co-chaperone for α-ketoglutarate dehydrogenase, promotes mTORC1 activation and neurodegeneration. *Neuron*. 2017;93(1):115-131. doi:10.1016/j.neuron.2016.11.038

120. Yuan Y, Ma X, Song N, Xie J. Expanding views of mitochondria in parkinson's disease: Focusing on PINK1 and GBA1 mutations. *Neurosci Bull*. 2022;38(7):825-828. doi:10.1007/s12264-022-00867-0

121. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*. 2019;10(1):1523. doi:10.1038/s41467-019-09234-6

122. Zhu H, Hixson P, Ma W, Sun J. Pharmacology of LRRK2 with type i and II kinase inhibitors revealed by cryo-em. *Cell discovery*. 2024;10(1):10. doi:10.1038/s41421-023-00639-8

***Web sites***

1. Picard tools - by Broad institute. http://broadinstitute.github.io/picard
2. Best Practices Workflows - *Broadinstitute.org*. https://gatk.broadinstitute.org/
3. GVCF - genomic variant call format - *GATK*. https://gatk.broadinstitute.org/hc/en-us/articles/360035531812-GVCF-Genomic-Variant-Call-Format
4. HapMap, *The International HapMap Consortium 2003*: https://www.genome.gov/genetics-glossary/hapmap
5. Omni 25M SNP array: https://emea.illumina.com/library-prep-array-kit-selector/kits-and-arrays/humanomni25-8_8-sample.html
6. The 1000 Genomes Project Consortium 2015: https://www.internationalgenome.org/
7. dbSNP: https://www.ncbi.nlm.nih.gov/snp/

8. ANNOVAR Documentation: https://annovar.openbioinformatics.org/en/latest/user-guide/gene/
9. CADD - Combined Annotation Dependent Depletion. In *Bihealth.org*. https://cadd.bihealth.org/
10. SKAT Package https://cran.r-project.org/web/packages/SKAT/vignettes/SKAT.pdf.
11. *TruSeq Stranded Total RNA Reference Guide*. (2017). https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseq-stranded-total-rna-workflow/truseq-stranded-total-rna-workflow-reference-1000000040499-00.pdf
12. OR formula https://statsbyslough.com/2015/08/20/odds-ratios-the-odd-one-out/
13. Ipa QIAGEN: https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/
14. EnrichR: https://maayanlab.cloud/Enrichr/enrich
15. Metascape: https://metascape.org/gp/index.html#/main/step1
16. Biorender: ***https://www.biorender.com/***
17. Odd Ratio calculation: *(https://statsbyslough.com/2015/08/20/odds-ratios-the-odd-one-out/)*