

Article

Not peer-reviewed version

---

# Novel Variants Linked to the Prodromal Stage of Parkinson's Disease (PD) Patients

---

[Marwa T. Badawy](#), Aya Salama, [Mohamed Salama](#)\*

Posted Date: 4 April 2024

doi: 10.20944/preprints202404.0303.v1

Keywords: Parkinson's Disease (PD); The Parkinson's Progression Markers Initiative (PPMI); VCF; neurodegenerative diseases



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Novel Variants Linked to the Prodromal Stage of Parkinson's Disease (PD) Patients

Marwa T. Badawy <sup>1</sup>, Aya A. Salama <sup>2</sup> and Mohamed Salama <sup>3,4,5</sup>

<sup>1</sup> Biology Department, School of Sciences and Engineering, The American University in Cairo, New Cairo, Egypt

<sup>2</sup> Applied Science, Windows and Web Experience, Microsoft, Egypt

<sup>3</sup> Institute of Global Health and Human Ecology (IGHHE), The American University in Cairo, New Cairo, Egypt

<sup>4</sup> Toxicology Department, Faculty of Medicine, Mansoura University, Mansoura, Egypt

<sup>5</sup> Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

**Abstract: Background and objective:** Symptoms of most neurodegenerative diseases, including Parkinson's disease (PD), usually do not occur until substantial neuronal loss occurs. This makes the process of early diagnosis very challenging. Hence, this research used variant call format (VCF) analysis to detect variants and novel genes that could be used as prognostic indicators in the early diagnosis of prodromal PD. **Materials and Methods:** Data were obtained from the Parkinson's Progression Markers Initiative (PPMI), and we analyzed prodromal patients with gVCF data collected in the 2021 cohort. A total of 304 participants were included, including 100 healthy controls, 146 prodromal genetic individuals, 21 prodromal hyposmia individuals, and 37 prodromal RBDs. A pipeline was developed to process the samples from gVCF to reach variant annotation and pathway and disease association analysis. **Results:** Novel variant percentages were detected in the analyzed prodromal subgroups. The prodromal subgroup analysis revealed novel variations of 1.0%, 1.2%, 0.6%, 0.3%, 0.5%, and 0.4% for the genetic male, genetic female, hyposmia male, hyposmia female, RBD male, and RBD female, respectively. Interestingly, 12 potentially novel loci (MTF2, PIK3CA, ADD1, SYBU, IRS2, USP8, PIGL, FASN, MYLK2, USP25, EP300 and PPP6R2) that were recently detected in PD patients were detected in the prodromal stage of PD. **Conclusion:** Genetic biomarkers are crucial for the early detection of Parkinson's disease and its prodromal stage. The novel PD genes detected in prodromal patients could aid in the use of gene biomarkers for early diagnosis of the prodromal stage without relying only on phenotypic traits.

**Keywords:** Parkinson's Disease (PD); The Parkinson's Progression Markers Initiative (PPMI); VCF; and neurodegenerative diseases

---

## Plain Language Summary

Early detection of Parkinson's Disease is important factor in successful intervention. One of the areas worth studying is the prodromal stage of PD. In this paper we studied the Parkinson's Progression Markers Initiative (PPMI) genetic data trying to identify novel variants that are more prevalent in this stage of the disease. Several new genes were identified that may be related to prodromal stage of PD that can help as early markers of the disease.

## Introduction

Parkinson's disease (PD) is a neurodegenerative disease that is pathologically defined as the death of dopaminergic neurons in the midbrain and the inclusion of Lewy bodies in the brain [1]. It has become obvious that PD has a prodromal stage, which is the period before the beginning of neurodegeneration without detecting motor signs by classical diagnosis. The basis of the nonmotor prodromal stage is that the pathological process could not yet occur in the substantia nigra pars compacta (SNpc) [2]. The classical

diagnosis of PD relies on the loss of mesodiencephalic dopaminergic (mdDA) neurons in the substantia nigra pars compacta (SNpc) and the development of Lewy bodies in some surviving neurons [3].

Early investigations focused on the role of genetic factors in Parkinson's disease to identify rare mutations linked to familial disease [4]. Moreover, the past decade has shown the great role of genetics in sporadic disease [5]. The identification of novel variants and genes for the early diagnosis of the prodromal and/or PD stages is receiving increased amounts of attention [6]. Variant Call Format (VCF) is becoming a community standard for reporting variations in genetic data acquired from medical genetic diagnostics and research [7].

In this study, we analyzed gVCF data acquired from the Parkinson's Progression Markers Initiative (PPMI) for healthy participants and prodromal PD patients. The prodromal PD subgroups involved in this study were genetic, RBD, and hyposmia. Sleep behavior disorder (RBD), rapid eye movement (REM), is a parasomnia condition characterized by complex abnormal motor movements in the REM state during sleep [8]. RBD is mainly associated with abnormal movement behaviors, nightmares, and loss of normal skeletal atonia in the REM state [9]. Moreover, hyposmia is an olfactory dysfunction that leads to a loss of smell ability and is the most common nonmotor symptom of PD [10].

This study was based on annotation, biological pathway, and disease association analyses of prodromal patients with PD. We detected the percentages and types of variation in each population with their percentile on each chromosome. The gene lists were generated and are presented as gene-annotation network clusters and bar charts, which enabled us to detect novel genes with counts in prodromal stages identified recently in PD patients, as illustrated in Table 2.

## Materials and Methods

### *PPMI—Data Collection*

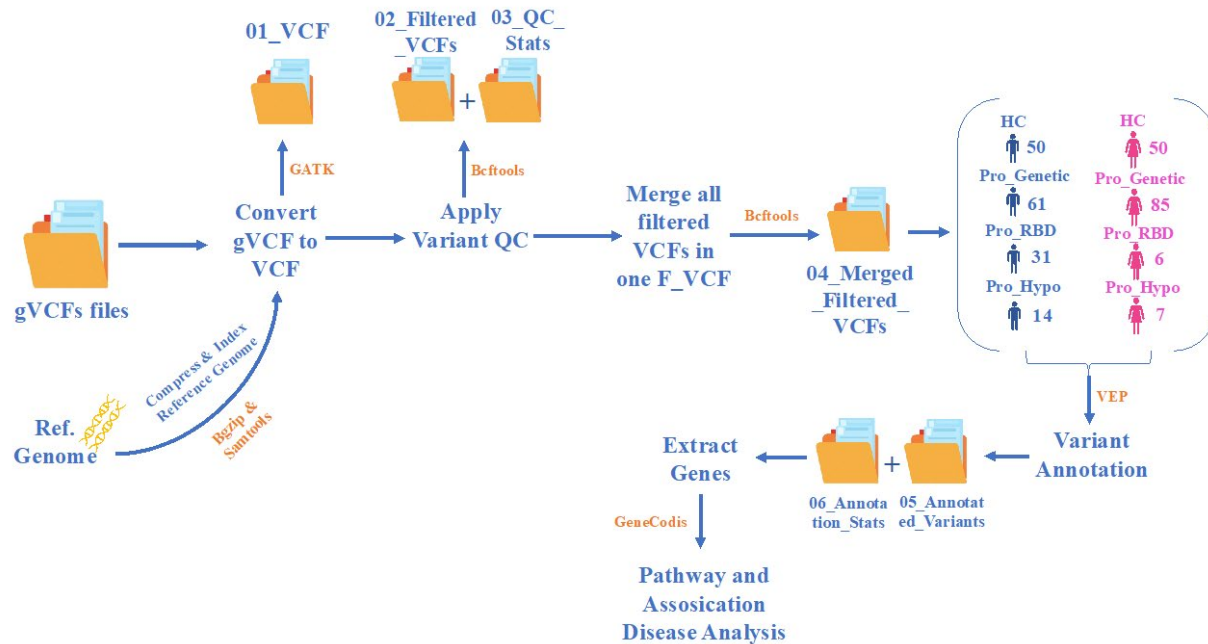
The data were obtained from the Parkinson Progression Markers Initiative (PPMI). The PPMI is an international multisite prospective observational study investigating biomarkers for Parkinson's disease (PD). The specific study methodology can be found at [ppmi-info.org](http://ppmi-info.org).

### *PPMI—Study Participants*

The participants in the PPMI study met these criteria: were at least 30 years of age, presented two of three cardinal symptoms (bradykinesia, rigidity, or resting tremor), were diagnosed within 2 years before entering the study, were untreated for PD when entering the study, and had a deficit in dopamine transporters [11]. Written informed consent was obtained from all the participants in the PPMI study, including the genetic research part. The PPMI study was conducted under the ethical standards of the Helsinki Declaration of 1975.

### *Study Design*

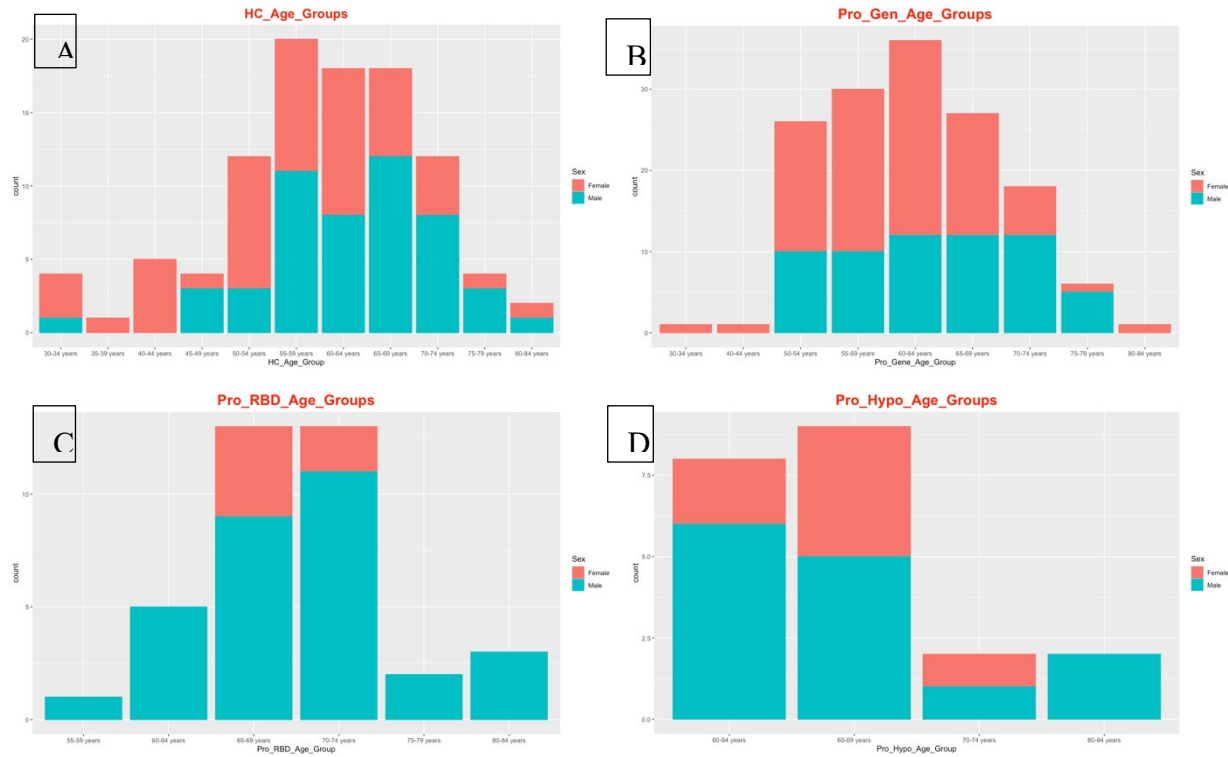
In this study, we considered the very recent cohort from the PPMI, April 2021 cohort. A pipeline was developed for this work; gVCF files were used as input, and analyses were performed to convert gVCF files to VCF files, apply variant quality control, extract summaries of all variants in each chromosome, merge filtered VCF files and annotating variants, and finally carry out pathway analysis and disease association, as illustrated in Figure 1.



**Figure 1.** Pipeline workflow diagram for the genomic variation analysis. This pipeline takes gVCF files as input and performs the following analyses: Convert gVCF to VCF, Variants quality control, Extract summary of all variants in each chromosome, Merge VCF filtered files, Variants annotation, Pathway analysis and disease association.

Installation of the *Genome Analysis Toolkit (GATK)* [12], *Variant Call Format tools (VCFtools)* [7], *Binary Calling Format tools (BCFtools)*, *Sequence Alignment/Map tools (SAMtools)* [13], *Burrows–Wheeler Aligner (BWA)* [14], *Picard tools*, and *Variant Effect Predictor (VEP)* [15] were used to perform the specified analyses within the pipeline.

These gVCF data consisted of 100 healthy controls (HCs) and 165 prodromal individuals, comprising 146 genetic, 21 hyposmia, and 37 RBDs. The HCs were divided into 50 males and 50 females. The prodromal genetic group included 61 males and 85 females, the prodromal RBD group included 31 males and 6 females, and the prodromal hyposmia group included 14 males and 7 females. Participants were classified based on their age, and each age group consisted of 5 years, as shown in Figure 2.



**Figure 2.** Bar plots for the age groups of the study participants. (A) Healthy control (HC) age groups ranging from 30-85 years. (B) Prodromal genetic (Pro\_Gen) age groups ranging from 30-85 years. (C) Prodromal RBD (Pro\_RBD) age groups ranging from 55-85 years. (D) Prodromal hyposmia (Pro\_Hypo) age groups ranging from 60-85 years. Figure generated with R software version 4.3.2.

### High Percentage of Intronic and Intergenic Variants

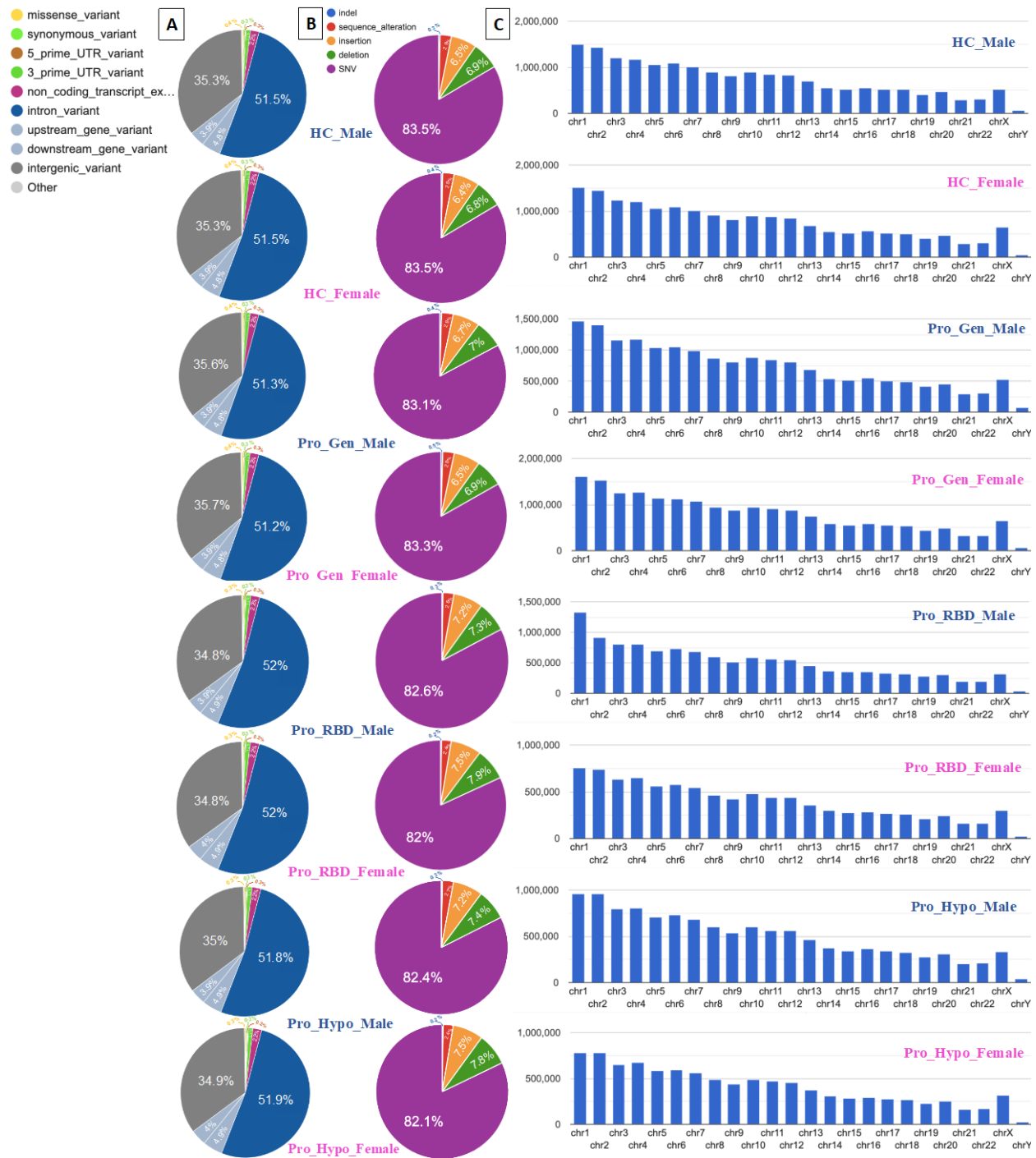
The Ensembl Variant Effect Predictor (VEP) is a powerful tool for analyzing genomic variation in coding and noncoding regions. It also provides access to a very extensive collection of genomic annotations [15].

In this study, VEP was used for annotation analyses and detection of genomic variations and their associations. The highest two percentages of variation were detected for intronic and intergenic variants, as illustrated in Figure 3A. Similarly, 51.5% and 35.3% of the healthy control male and female populations, respectively, exhibited the same percentages of intron variants and intergenic variants. Among those with a prodromal genetic male population, 51.3% had intron variants and 35.6% had intergenic variants, while among those with a prodromal genetic female population, 51.2% had intron variants and 35.7% had intergenic variants. Prodromal RBD males and females had the same percentages of intron variants and intergenic variants (52% and 34.8%, respectively). Among the males with prodromal hyposmia, 51.8% had intron variants and 35% had intergenic variants; moreover, among the females with prodromal hyposmia, 51.9% had intron variants and 34.9% had intergenic variants.

The detection of variants in intronic and intergenic regions is common across the entire human genome, as these noncoding regions make up half of the human noncoding genome and can play important regulatory roles [16]. The presence of intronic and intergenic variants in the studied healthy population and prodromal populations suggests that these variants are not specifically associated with prodromal PD. However, these findings likely represent background genetic variation.

A single-nucleotide variant (SNV), also called a single-nucleotide polymorphism (SNP), is a variant of a specific single nucleotide and occurs at a specific position in the genome. Moreover, SNVs are the most common type of genetic variation [17]. This clearly explains the high percentages of SNVs detected in all the healthy and prodromal populations presented in Figure 3B. All the populations presented  $\geq 82\%$  of the SNVs.

Deletion mutations were detected as the second highest number of elements after SNVs. The two highest percentages of deletions were detected in the RBD females and hyposmia females (7.9% and 7.8%, respectively). These deletion mutations may be associated with genetic factors involved in the development of hyposmia and RBD in individuals, specifically females. Moreover, these deletions could contain genes or regulatory regions relevant to olfactory function and sleep regulation. A sex difference could be related to the association between high deletion percentages and prodromal symptoms in females. This could suggest potential sex-specific genetic risk factors for PD. Notably, all female participants in this study had a Y chromosome. It is suspected to be due to androgen insensitivity syndrome (AIS), which is characterized by evidence of feminization of the external genitalia at birth and abnormal sexual development with a 46,XY karyotype [18]. Consequently, the two most common insertion mutations were also detected in 7.9% of the RBD females and 7.8% of the Hyposmia females, as shown in Figure 2b. The detection of insertions in the RBD and Hyposmia populations suggested genetic variability within these groups. These high percentages of insertions could be associated with disease susceptibility or progression. Additionally, depending on their location within the genome, these insertions can have various functional consequences.



**Figure 3.** (A) Pie chart visualization of the most severe variant consequences, (B) pie chart visualization of the variant classes, and (C) bar plot visualization for variants by chromosome for each population. Pie charts and bar plots were generated using the ensemble VEP tool.

Chromosomes 1 and 2 are among the largest chromosomes in the human genome and contain a large number of genes and regulatory elements [19]. Therefore, these genes may represent a greater number of variants simply because of their size and gene density. High percentages of variants were detected in all the analyzed samples, including those of the healthy controls. Healthy control males had variant counts of

1,500,086 on chr1 and 1,423,662 on chr2, while healthy control females had variant counts of 1,509,369 on chr1 and 1,441,362 on chr2 (Figure 3C). Consequently, it could be normal to find these variants in prodromal populations. However, chromosomes 1 and 2 may influence biological processes relevant to PD, such as mitochondrial function, protein aggregation, the oxidative stress response, and neuroinflammation. Understanding how variants in these genomic regions affect molecular pathways associated with PD is crucial for providing insights into disease mechanisms. Gene-annotation network cluster and pathway analyses are shown in Figures 4, 5, and 6.

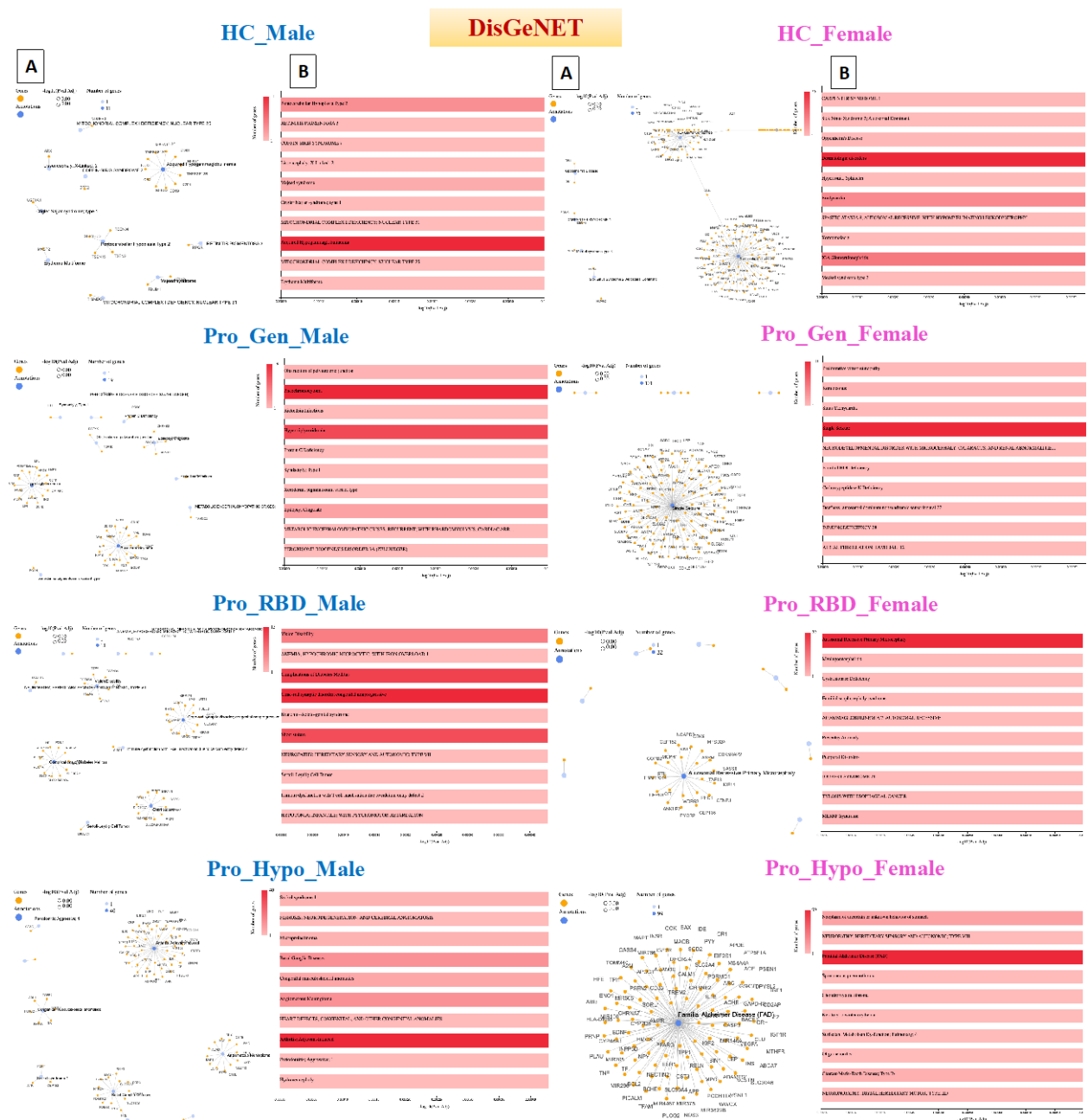
**Table 1.** General statistics of the variation analysis results for the healthy control (HC) (male and female), prodromal PD (genetic male and female), prodromal PD (RBD male and female), and prodromal PD (hyposmia male and female) populations.

Population	No. of samples in Population	Lines of input read	Processed Variants	Novel/Existing Variants	Overlapped Genes	Overlapped Transcripts
HC_Male	50	18191327	18191327	168623 (0.9%)/18022704 (99.1%)	61987	250277
HC_Female	50	18387115	18387115	182458 (1.0%)/18204657 (99.0%)	61663	249831
Pro_Gen_Male	61	17763968	17763968	179814 (1.0%)/17584154 (99.0%)	62009	250316
Pro_Gen_Female	85	19371326	19371326	228369 (1.2%)/19142957 (98.8%)	61655	249834
Pro_RBD_Male	31	12264624	12264624	63331 (0.5%)/12201293 (99.5%)	61898	250173
Pro_RBD_Female	6	9568588	9568588	40093 (0.4%)/9528495 (99.6%)	61542	249619
Pro_Hypo_Male	14	12081647	12081647	69229 (0.6%)/12012418 (99.4%)	61924	250185
Pro_Hypo_Female	7	9942790	9942790	29510 (0.3%)/9913280 (99.7%)	61565	249627

All the input lines of the variants were processed, and the number of filtered variants was zero in all the populations. We used the filtered VCF file that we produced during the pipeline after the QC step. This table shows that novel variation percentages were detected in all the tested populations. The detection of novel genetic variations in healthy populations is a natural consequence of genetic diversity and the complexity of the human genome. As a result, the novel variation detected in the tested population may be natural compared to the percentage of healthy controls. However, the chromosomal location and type of variant, whether Indels or SNVs, could provide a deeper justification of whether these variants could be potentially related to prodromal PD or whether they are just natural variants.

## Disease Gene Network (DisGeNET) detection in prodromal PD populations

The disease gene network, known as DisGeNET, is a comprehensive knowledge base that integrates information on human disease-associated genes and variants from multiple sources [20]. This database was accessed through the GeneCodis website, and annotation was carried out through this tool [21].



**Figure 4.** Network and bar chart plots for the disease-gene network (DisGeNET). (A) Network cluster connections based on gene pathways. (B) Bar charts for the pathways detected with the number of genes involved in each pathway. Networks and charts were generated with GeneCodis.4.

Acquired hypogammaglobulinemia was detected in the healthy male population, in which 11 genes were associated with this disease. Acquired hypogammaglobulinemia is also known as secondary hypogammaglobulinemia and is a condition characterized by low levels of immunoglobulins (antibodies) in the blood. This condition can increase the risk of infections and can occur due to various factors, such as certain medications, underlying medical conditions, or

environmental exposures [22]. Therefore, the detection of this disease in the healthy male population could be due to the age of the participants, as 49 participants were aged  $\geq 45$  years, as shown in Figure 4. On the other hand, 75 genes were detected to be related to dermatological disorders in the healthy female population. Many dermatological disorders, including eczema, psoriasis, acne, and others, involve multiple genes and have complex genetic architectures. Variations in these genes can influence susceptibility to these conditions, and the involvement of 75 genes may indicate a polygenic inheritance pattern. Each gene may have a small effect on the overall risk of developing dermatological disorders. Environmental factors also contribute to disease susceptibility, and exposures to allergens, irritants, pollutants, UV radiation, and microbial agents can interact with genetic predispositions to trigger or exacerbate skin conditions.

Pheochromocytoma and hypertriglyceridemia were detected in the prodromal genetic male population with 19 and 15 genes, respectively. Pheochromocytoma (PCC) is a rare neuroendocrine tumor that arises from the adrenal glands and can also occur elsewhere in the sympathetic nervous system. It is characterized by the excessive production of catecholamines, such as noradrenaline and adrenaline, which can cause a variety of symptoms, including palpitations, hypertension, headache, anxiety, and sweating [23]. The detection of pheochromocytoma in the prodromal genetic PD population is rare and unusual but possible. While there is no known direct genetic link between PCC and PD, it is important to note that both conditions can be influenced by genetic predispositions, environmental factors, and complex interactions between various biological pathways. Additionally, several genes associated with PD may have other roles in different cellular processes beyond the central nervous system.

Hypertriglyceridemia is an elevated level of triglycerides in the blood and is a lipid abnormality associated with an increased risk of cardiovascular disease [24]. Hypertriglyceridemia is primarily influenced by lifestyle factors such as physical activity, diet, and obesity. Genetic factors can also play a role in lipid metabolism and contribute to elevated triglyceride levels. Investigating potential shared genetic predispositions between HTG and PD may provide insights into overlapping biological pathways or susceptibility genes. Furthermore, dysregulation of lipid metabolism and glucose homeostasis has been implicated in the pathogenesis of PD. Emerging evidence suggests potential links between metabolic dysfunction, insulin resistance, and neurodegeneration in PD patients. Detecting hypertriglyceridemia in individuals with prodromal genetic PD in the male population may raise questions about underlying metabolic disturbances and their implications for disease progression.

Among the prodromal genetic female population, 101 patients had a single seizure. These seizures can be triggered by fever (febrile seizures), head injury, metabolic disturbances, sleep deprivation, stress, alcohol, or drug withdrawal [25]. While PD primarily affects dopaminergic neurons in the brain, there is evidence to suggest that individuals with PD, or in the prodromal stage, may have an increased susceptibility to seizures compared to the general population. Genetic factors, including mutations in genes associated with both PD and epilepsy, could contribute to this increased risk. The involvement of 101 genes may indicate a polygenic or multifactorial basis for the seizure phenotype, with variations in multiple genes contributing to the risk of seizures.

Cone-rod synaptic disorder (CRSD) was detected in 13 genes of the prodromal RBD male population. CRSD is a rare genetic disorder characterized by dysfunction of the synaptic connections between cone and rod photoreceptor cells in the retina. This leads to visual impairment, particularly affecting color vision, central vision, and visual acuity [26]. Additionally, RBD is a rapid eye movement behavior disorder characterized by the loss of muscle atonia during REM sleep, leading to the enactment of dreams through vocalizations and movements. While CRSD primarily affects the retina, several genes associated with retinal function may have broader roles in neuronal health and function. Variants in these genes could contribute to neurodegenerative processes in conditions such as PD.

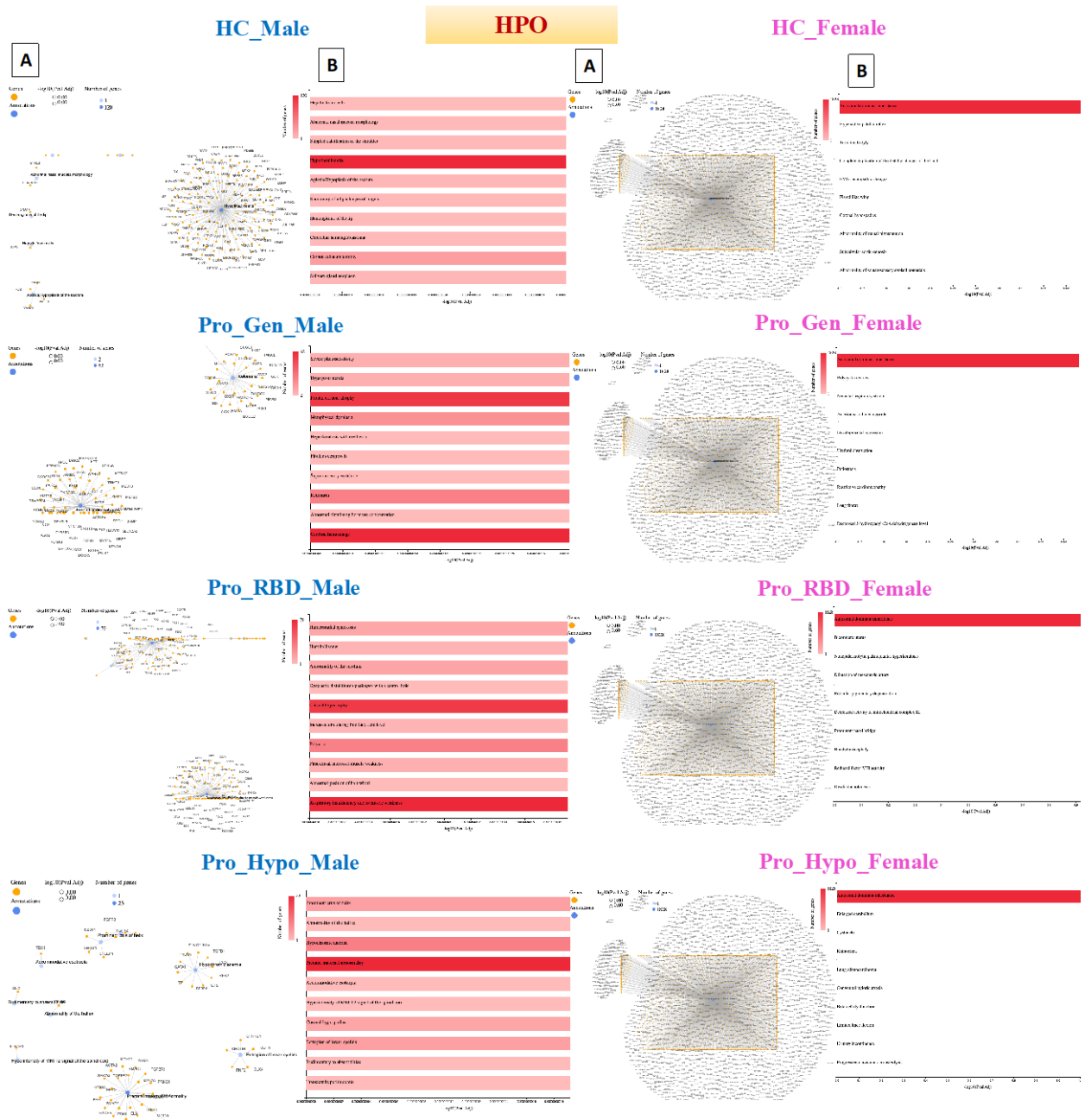
Autosomal recessive primary microcephaly (MCPH) was detected in the prodromal RBD female population with 22 genes. MCPH is a rare neurodevelopmental disorder characterized by significantly reduced head size (microcephaly) and intellectual disability. It is inherited in an

autosomal recessive manner, meaning that both copies of the affected gene (one from each parent) must be mutated for the condition to manifest [27]. While MCPH primarily affects brain development, several genes associated with neurodevelopmental disorders may have broader roles in neuronal health and function. Genetic variations in these genes may contribute to neurodegenerative processes in conditions such as prodromal RBD.

Adjuvant arthritis was detected in a prodromal hyposmia male population with 40 genes. In humans, a type of reactive arthritis occurs when the immune system reacts to a triggering event, such as an infection or exposure to certain substances. It typically presents with joint pain, swelling, and stiffness, similar to other forms of inflammatory arthritis. PD and autoimmune disorders such as rheumatoid arthritis (RA) have distinct etiologies, and there is a growing recognition of shared genetic susceptibility and environmental factors that may contribute to both conditions. However, when Arthritis was detected, adjuvant-induced Arthritis in a male population with prodromal hyposmia involving 40 genes suggested a complex interplay of genetic and environmental factors. More importantly, this arthritis occurs at older ages, and all the hyposmic male patients were aged  $\geq 60$  years. On the other hand, familial Alzheimer disease (FAD) was detected in 99 prodromal hyposmia females. FAD is associated with mutations in specific genes, including amyloid precursor protein (APP), presenilin 1 (PSEN1), and presenilin 2 (PSEN2) [28]. These mutations are typically inherited in an autosomal dominant manner, meaning that a single copy of the mutated gene is sufficient to cause the disease. While some genetic mutations may be associated with both Alzheimer's disease and Parkinson's disease, detecting FAD in a prodromal hyposmia female population involving 99 genes would need further in-depth investigations.

#### **Detection of Human Phenotype Ontology (HPO) Data in Prodromal PD Populations**

The HPO dataset is the Human Phenotype Ontology, which consists of phenotypic abnormalities encountered in human disease [29].



**Figure 5.** Network and bar chart plots for the Human Phenotype Ontology (HPO). (A) Network cluster connections based on gene pathways. (B) Bar charts for the pathways detected with the number of genes involved in each pathway. Networks and charts were generated with GeneCodis.4.

Hyperinsulinemia was detected in a healthy male population with 120 genes. Hyperinsulinemia is a condition characterized by higher-than-normal levels of insulin in the blood. Insulin is a hormone produced by pancreatic beta cells that helps regulate glucose levels by facilitating the uptake of glucose into cells for energy or storage [30]. The detection of hyperinsulinemia in the male population of HCs may suggest underlying metabolic abnormalities or insulin resistance. Moreover, hyperinsulinemia can occur as a compensatory response to insulin resistance and can be influenced by various factors, such as diet, physical activity, genetics, and medications. This could also be justified by the older ages of the healthy male population, as they are  $\geq 45$  years old.

The HPO results revealed that all female populations, including the healthy population, had the same gene network of autosomal dominant inheritance, with 1828 genes being involved. Autosomal dominant inheritance is a pattern of inheritance for a trait or disorder determined by genes located

on autosomal chromosomes (nonsex chromosomes). In other words, a single copy of the mutated gene, inherited from one parent, is sufficient to express the trait or disorder. This means that individuals who inherit the mutated gene from either parent will exhibit the trait or disorder. Examples of disorders with autosomal dominant inheritance include Parkinson's disease, Huntington's disease, familial hypercholesterolemia, Marfan syndrome, neurofibromatosis type 1, and some other forms of familial Alzheimer's disease [31]. As mentioned in the chromosome analysis, all female samples had a Y chromosome (Figure 3C).

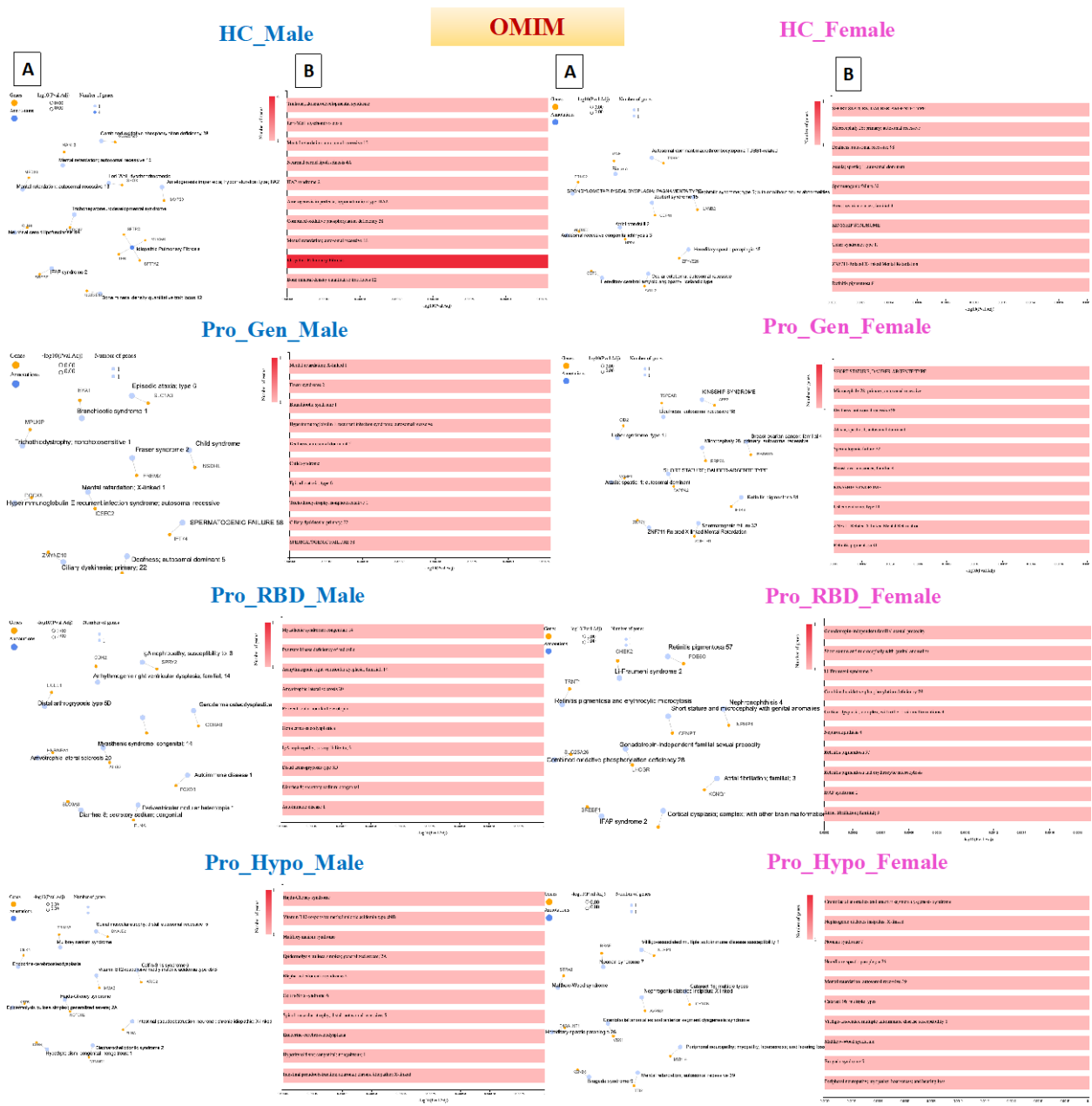
Cerebral hemorrhage was detected in a prodromal genetic male population with 62 genes. Cerebral hemorrhage is a medical condition characterized by bleeding within brain tissues. It occurs when a blood vessel within the brain ruptures, leading to the leakage of blood into the surrounding brain tissue. This bleeding can cause damage to brain cells and disrupt normal brain function [32]. In general, cerebral hemorrhage is not a common feature of prodromal PD. However, its detection in the prodromal genetic PD male population suggests a potential overlap or interaction between genetic factors predisposing patients to PD and cerebrovascular events.

Respiratory insufficiency due to muscle weakness was detected in 79 genes in the prodromal RBD male population. In this condition, the muscles involved in breathing are unable to adequately perform their function, leading to impaired respiratory function. This can occur because of various underlying causes, including neurological conditions, neuromuscular disorders, or muscular dystrophies [33]. Detecting respiratory insufficiency due to muscle weakness in the prodromal RBD PD male population with 79 genes may suggest genetic predispositions or variants associated with neuromuscular or respiratory function. Moreover, respiratory insufficiency in PD patients is more commonly associated with factors such as upper airway obstruction, aspiration pneumonia, or respiratory muscle rigidity.

Prenatal maternal abnormalities were detected in the prodromal hyposmia male population with 23 genes. Prenatal maternal abnormalities are not linked to maternal health, but they may also occur because of genetic conditions or mutations carried by the father, which can be transmitted to the fetus and influence prenatal development and health outcomes. Epigenetic modifications, such as DNA methylation patterns or histone modifications, can reflect prenatal environmental exposures or maternal health conditions [34]. Epigenetic changes could influence gene expression and neurodevelopmental processes relevant to PD risk.

### **Online Mendelian Inheritance in Man (OMIM) Detection in Prodromal PD Populations**

**The** OMIM database contains Mendelian Inheritance in Man. It is a comprehensive and authoritative compendium of human genes, genetic disorders, syndromes, and traits [35].



**Figure 6.** Network and bar chart plots for Online Mendelian Inheritance in Man (OMIM). (A) Network cluster connections based on gene pathways. (B) Bar charts for the pathways detected with the number of genes involved in each pathway. Networks and charts were generated with GeneCodis.4.

The Online Mendelian Inheritance in Man was the third phenotypic dataset to be used in this study to obtain a wider view of the whole of the 3 available phenotypic databases. Notably, none of the populations exhibited significant pathway or significant gene-network cluster annotations. However, one to four genes were detected from each detected pathway. In the HC male population, idiopathic pulmonary fibrosis (IPF) was detected with 4 genes. IPF is a chronic and progressive lung disease characterized by scarring (fibrosis) of the lungs, leading to impaired lung function and difficulty breathing. The exact cause of IPF is unknown; however, IPF is believed to result from a combination of environmental factors, genetic predispositions, and abnormal wound healing processes in the lungs [36]. This could also be explained by the older age of the HC male population, as mentioned previously.

The reason why there was no significant pathway or gene network detected could be justified due to the rare Mendelian forms of PD. Known Mendelian forms of PD include certain monogenic

forms caused by mutations in genes such as PARK2, SNCA, and LRRK2; these forms represent a small proportion of all PD cases, particularly those in the prodromal stage [37].

### Novel Gene Detection in Prodromal PD Populations

Interestingly, 12 potentially novel PD loci recently detected by Kim [38] were found to be present in prodromal populations. The 12 potentially novel loci were MTF2, PIK3CA, ADD1, SYBU, IRS2, USP8, PIGL, FASN, MYLK2, USP25, EP300 and PPP6R2. Table 2 shows the 12 novel genes and their counts in each prodromal population. This detection could lead to the use of this gene as a potential genetic biomarker for the early detection of prodromal PD patients.

**Table 2.** Recently, 12 potentially novel PD genes were detected in the populations with prodromal PD (genetic male and female), prodromal PD (adolescent and female), and prodromal PD (hyposmia male and female).

Population Gene Name	Pro_Gen_Mal Gene Count	Pro_Gen_Femal Gene Count	Pro_RBD_Mal Gene Count	Pro_RBD_Femal Gene Count	Pro_Hypo_Ma le Gene Count	Pro_Hypo_Fema le Gene Count
MTF2	2985	3458	3065	1397	2035	1606
ADD1	9095	10457	5535	3471	4404	3722
PIK3CA	2958	3072	1874	1390	1606	1830
SYBU	13897	15607	7991	5943	8012	6030
IRS2	237	291	192	138	210	147
USP8	5997	6524	4432	3358	4953	1953
PIGL	8277	9021	4350	2819	3902	2923
FASN	1184	1224	691	552	775	593
MYLK2	265	296	183	132	111	107
USP25	4081	4952	2672	753	2796	534
EP300	3146	3510	1704	1256	1790	1398
PPP6R2	5652	6393	3090	2693	3723	2561

### Conclusion

Genetic composition plays a crucial role in the development of Parkinson's disease and its prodromal stage subgroups. The novel PD genes detected in prodromal patients could aid in the use of gene biomarkers for early diagnosis of the prodromal stage without relying only on phenotypic traits. The network clusters of the prodromal populations showed how prodromal PD subgroups may be linked to a wide range of diseases and complications. This is mainly because PD-related genetic factors may have other functions beyond the nervous system that can result in other complications and illnesses throughout the human body. However, further clinical research is needed to provide in-depth information about the representative genetic results.

**Author contributions:** M.B: Conceptualization, data curation, writing–review; A.S: data curation, writing–review; M.S: conceptualization, supervision, writing–review & editing. All the authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the American University in Cairo (AUC) research grant for PhD candidates.

**Conflicts of interest:** The authors declare that they have no competing interests.

**Acknowledgments:** We would like to acknowledge the Michael J. Fox Foundation (MJF) for providing access to all the patient data from the Stevens Neuroimaging and Informatics Institute - Keck School of Medicine - University of Southern California (USC). Use of artificial intelligence tools: During the preparation of this work, AI tools were used to improve the readability and language of the manuscript or to generate images, and subsequently, the authors revised and edited the content produced by the AI tools as necessary, taking full responsibility for the ultimate content of the present manuscript.

## References

1. Kim, J.J., Vitale, D., Otani, D. V., Lian, M. M., Heilbron, K., Iwaki, H., ... & Mata, I., *Multi-ancestry genome-wide association meta-analysis of Parkinson's disease*. *Nature Genetics*, 2023: P. 1-10.
2. Postuma, R.B., Aarsland, D., Barone, P., Burn, D. J., Hawkes, C. H., Oertel, W., & Ziemssen, T., *Identifying prodromal Parkinson's disease: Pre-motor disorders in Parkinson's disease*. *Movement Disorders*, 2012. **27**(5): P. 617-626.
3. Langston, J.W., Schüle, B., Rees, L., Nichols, R. J., & Barlow, C., *Multisystem Lewy body disease and the other parkinsonian disorders*. *Nature genetics*, 2015. **47**(12): P. 1378-1384.
4. Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., ... & Gwinn-Hardy, K.,  *$\alpha$ -Synuclein locus triplication causes Parkinson's disease*. *Science signaling*, 2003. **302**(5646): P. 841-841.
5. Chang, D., Nalls, M. A., Hallgrímsdóttir, I. B., Hunkapiller, J., Van Der Brug, M., Cai, F., ... & Graham, R. R., *A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci*. *Nature genetics*, 2017. **49**(10): P. 1511-1516.
6. Rajan, R., Divya, K. P., Kandadai, R. M., Yadav, R., Satagopam, V. P., Madhusoodanan, U. K., ... & Lux-GIANT Consortium, *Genetic architecture of Parkinson's disease in the Indian population: Harnessing genetic diversity to address critical gaps in Parkinson's disease research*. *Frontiers in neurology*, 2020. **11**.
7. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group., *The variant call format and VCFtools*. *Bioinformatics*, 2011. **27**(15): P. 2156-2158.
8. Diaconu, Ș., Falup-Pecurariu, O., Ținț, D., & Falup-Pecurariu, C., *REM sleep behaviour disorder in Parkinson's disease*. *Experimental and Therapeutic Medicine*, 2021. **22**(2): P. 1-5.
9. Tekriwal, A., Kern, D. S., Tsai, J., Ince, N. F., Wu, J., Thompson, J. A., & Abosch, A., *REM sleep behaviour disorder: Prodromal and mechanistic insights for Parkinson's disease*. *Journal of Neurology, Neurosurgery & Psychiatry*, 2017. **88**(5): P. 445-451.
10. Haehner, A., Masala, C., Walter, S., Reichmann, H., & Hummel, T., *Incidence of Parkinson's disease in a large patient cohort with idiopathic smell and taste loss*. *Journal of neurology*, 2019(266): P. 339-345.
11. Nalls, M.A., McLean, C. Y., Rick, J., Eberly, S., Hutten, S. J., Gwinn, K., ... & Singleton, A. B., *Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: A population-based modelling study*. *The Lancet Neurology*, 2015. **14**(10): P. 1002-1009.
12. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A., *The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data*. *Genome research*, 2010. **20**(9): P. 1297-1303.
13. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., ... & Li, H., *Twelve years of SAMtools and BCFtools*. *Gigascience*, 2021. **10**(2).
14. Li, H., *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. 2013.
15. McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... & Cunningham, F., *The ensembl variant effect predictor*. *Genome biology*, 2016(17): P. 1-14.
16. Rigau, M., Juan, D., Valencia, A., & Rico, D., *Intronic CNVs and gene expression variation in human populations*. *PLoS Genetics*, 2019. **15**(1).

17. Human-Genome-Structural-Variation-Working-Group, E., E. E., Nickerson, and A. D. A., D., Bowcock, A. M., Brooks, L. D.; et al., *Completing the map of human genetic variation*. *Nature*, 2007: P. 161–165.
18. Mongan, N.P., Tadokoro-Cuccaro, R., Bunch, T., & Hughes, I. A., *Androgen insensitivity syndrome*. *Best practice & research Clinical endocrinology & metabolism*, 2015. **29**(4): P. 569-580.
19. Rosa, A., & Everaers, R., *Structure and dynamics of interphase chromosomes*. *PLoS computational biology*, 2008. **4**(8).
20. Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., ... & Furlong, L. I., *DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants*. *Nucleic acids research*, 2016.
21. García-Moreno, A., López-Domínguez, R., Ramírez-Mena, A., Pascual-Montano, A., Aparicio-Puerta, E., Hackenberg, M., & Carmona-Saez, P. , *GeneCodis 4: Expanding the modular enrichment analysis to regulatory elements*. *bioRxiv*, 2021: P. 2021-04.
22. Jaffe, E.F., Lejtenyi, M. C., Noya, F. J., & Mazer, B. D., *Secondary hypogammaglobulinemia*. *Immunology and allergy clinics of North America*, 2001. **21**(1): P. 141-163.
23. Tsirlin, A., Oo, Y., Sharma, R., Kansara, A., Gliwa, A., & Banerji, M. A., *heochromocytoma: A review*. *Maturitas*, 2014. **77**(3): P. 229-238.
24. Brunzell, J.D., *Hypertriglyceridemia*. *New England Journal of Medicine*, 2007. **357**(10): P. 1009-1017.
25. Kim, L.G., Johnson, T. L., Marson, A. G., & Chadwick, D. W. , *Prediction of risk of seizure recurrence after a single seizure and early epilepsy: Further results from the MESS trial*. *The Lancet Neurology*, 2006. **5**(4): P. 317-322.
26. Mechaussier, S., Almoallem, B., Zeitz, C., Van Schil, K., Jeddawi, L., Van Dorpe, J., ... & Perrault, L., *Loss of function of RIMS2 causes a syndromic congenital cone-rod synaptic disease with neurodevelopmental and pancreatic involvement*. *The American Journal of Human Genetics*, 2020. **106**(6): P. 859-871.
27. Woods, C.G., Bond, J., & Enard, W., *Autosomal recessive primary microcephaly (MCPH): A review of clinical, molecular, and evolutionary findings*. *The American Journal of Human Genetics*, 2005. **76**(5): P. 717-728.
28. Wu, L., Rosa-Neto, P., Hsiung, G. Y. R., Sadovnick, A. D., Masellis, M., Black, S. E., ... & Gauthier, S., *Early-onset familial Alzheimer's disease (EOFAD)*. *Canadian Journal of Neurological Sciences*, 2012. **39**(4): P. 436-445.
29. Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J. O. B., Danis, D., Gourdine, J. P., ... & Robinson, P. N., *Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources*. *Nucleic acids research*, 2019. **47**(D1).
30. Shanik, M.H., Xu, Y., Skrha, J., Dankner, R., Zick, Y., & Roth, J., *Insulin resistance and hyperinsulinemia: Is hyperinsulinemia the cart or the horse?* *Diabetes care*, 2008. **31**: P. S262-S268.
31. de Vries, L., Kauschansky, A., Shohat, M., & Phillip, M., *Familial central precocious puberty suggests autosomal dominant inheritance*. *The Journal of clinical endocrinology & metabolism*, 2004. **89**(4): P. 1794-1800.
32. Magid-Bernstein, J., Girard, R., Polster, S., Srinath, A., Romanos, S., Awad, I. A., & Sansing, L. H., *Cerebral hemorrhage: Pathophysiology, treatment, and future directions*. *Circulation research*, 2022. **130**(8): P. 1204-1229.
33. Mauro, A.L., & Aliverti, A., *Physiology of respiratory disturbances in muscular dystrophies*. *Breathe*, 2016. **12**(4): P. 318-327.
34. Wagner, R., Tse, W. H., Gosemann, J. H., Lacher, M., & Keijzer, R., *Prenatal maternal biomarkers for the early diagnosis of congenital malformations: A review*. *Pediatric Research*, 2019. **86**(5): P. 560-566.
35. Amberger, J.S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., & Hamosh, A., *OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders*. *Nucleic acids research*, 2015. **43**(D1): P. D789-D798.
36. Barratt, S.L., Creamer, A., Hayton, C., & Chaudhuri, N., *Idiopathic pulmonary fibrosis (IPF): An overview*. *Journal of clinical medicine*, 2018. **7**(8).

37. Nuytemans, K., Theuns, J., Cruts, M., & Van Broeckhoven, C., *Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: A mutation update*. *Human mutation*, 2010. **31**(7): P. 763-780.
38. Kim, J.J., Vitale, D., Otani, D. V., Lian, M. M., Heilbron, K., Iwaki, H., ... & Mata, I., *Multi-ancestry genome-wide association meta-analysis of Parkinson's disease*. *Nature genetics*, 2024. **56**(1): P. 27-36.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.