Study Protocol

Study Title: Omics sciences for the identification of pathogenic mechanisms and biomarkers in neurodegenerative diseases

Study Code: NeurOmics

Sponsor: IRCCS Ospedale Policlinico San Martino

Co-Sponsor: Fondazione Istituto Italiano di Tecnologia (IIT)

Principal Investigator: Dr. Paola Mandich, Director of the Unit of Medical Genetics, IRCCS Ospedale Policlinico San Martino, Largo Rosanna Benzi 10, Genoa, Italy

Principal Investigator: Prof. Stefano Gustincich, Scientific Director CMP3VdA, IIT, Via Lavoratori - Vittime del Col du Mont 28, Aosta, Italy

Principal Investigator: Dr. Alessio Di Fonzo, Senior Neurologist, Department of Neurology, IRCCS Ca' Granda Ospedale Maggiore Policlinico, Via Sforza 28, Milan, Italy

Version and Date: V. 1 30.11.2023

1. INTRODUCTION

Neurodegenerative diseases are a heterogeneous group of disorders characterized by the progressive deterioration of structures and functions of the nervous system. The most relevant conditions in terms of incidence, impact on patients' quality of life, and burden on the national healthcare system are Alzheimer's disease (AD), Parkinson's disease (PD), and Amyotrophic Lateral Sclerosis (ALS)[1,2,3]. The prevalences of AD, PD, and ALS are estimated at 742, 174, and 3 cases per 100,000, respectively[4,5,6]. Overall, approximately 1,000,000 individuals in Italy are affected by neurodegenerative diseases, largely represented by dementias and parkinsonian syndromes[7].

Neurodegenerative diseases are caused by the progressive loss of neurons within the nervous system and are generally associated with protein deposits or pathological protein conformations, leading to biochemical alterations, loss of function, and apoptosis in neurons and glial cells[8]. They are chronic, progressive, and disabling disorders. At present, only symptomatic treatments are available, while no curative or disease-modifying therapy has yet been established.

These conditions may arise from acquired and environmental factors, genetic determinants, or—more likely—from their combined effect. Molecular genetics and pathology have played a pivotal role in identifying genes and proteins involved in the intracellular aggregation processes that lead to neurodegeneration. The most common proteins implicated in these pathological aggregates are amyloid-β, tau, α-synuclein, and TDP-43[8,9].

Currently, genetic diagnosis is possible only in a minority of cases, particularly in early-onset forms and/or with a positive family history. For example, in AD, approximately 1% of cases have a clear genetic etiology, whereas in about 40% a genetic and/or environmental risk factor for disease development can be identified[10]. In PD, a genetic cause can be determined in approximately 10% of cases[11], while in ALS this proportion ranges from 10% to 15%[12].

1.1 Neurodegenerative diseases: background

Alzheimer's disease (AD) is the most common form of dementia. Dementia is defined as a progressive and often irreversible decline of mental faculties, with diagnostic criteria well established by the DSM-V. From a neuropathological standpoint, AD is characterized by extracellular β-amyloid plaques and intracellular fibrillary tangles in cortical and limbic brain regions, resulting in progressive cognitive and functional decline[13]. Symptom onset generally occurs after the fifth decade of life, with variable progression. Three genes have been associated with early-onset AD (< 60 years), while numerous genomic loci have been linked to increased risk in late-onset forms[10]. Encouraging preliminary results have been reported for disease-modifying therapies that slow progression[14,15]. Consequently, a precise diagnosis and accurate patient stratification, including genetic characterization, are crucial for the implementation of personalized therapeutic approaches.

Frontotemporal dementia (FTD) is the second most common cause of early-onset dementia[16] and it encompasses a broad spectrum of neuropathological entities and clinical phenotypes. The disease primarily affects three domains: behavior, language, and motor function, due to progressive atrophy of the frontal and temporal lobes. Motor dysfunction is also observed in related disorders such as progressive

supranuclear palsy (PSP), corticobasal syndrome (CBS), and FTD with amyotrophic lateral sclerosis (FTD-ALS)[17]. Executive dysfunction, a core clinical feature, is present across all clinical variants.

All FTD phenotypes share a pathological frontotemporal lobar degeneration (FTLD), which can be classified according to the predominant protein composition of cellular inclusions[18,19]. Alongside clinical and pathological heterogeneity, a rapidly growing body of research highlights the genetic complexity of FTLD, with 20–30% of patients carrying variants in the progranulin (GRN), microtubule-associated protein tau (MAPT), or chromosome 9 open reading frame 72 (C9orf72) genes[20]. Some patients with FTD develop motor neuron disease characterized by generalized muscle atrophy, weakness, fasciculations, and bulbar symptoms. This heterogeneity, together with the lack of a strict

correlation between clinical, genetic, and neuropathological features[21,22], represents a major challenge to the development of unified pathogenic models and, consequently, effective disease-modifying strategies. The rising prevalence of FTD and its escalating public health impact highlight the critical need for therapies that can prevent the disease, delay its onset, slow its progression, and ease its symptoms.

Other forms of early-onset dementia **(Young Onset Dementia, YOD)**, with onset before 65 years of age, have an estimated prevalence of 119 cases per 100,000, corresponding to approximately 3.9 million individuals worldwide[23]. Although AD and FTD are the leading causes of YOD, this group comprises a wider range of disorders. Indeed, YOD can be associated with conditions such as Huntington's disease and vascular dementia, while the accumulation of proteins (e.g., amyloid-β) may give rise to atypical AD variants, such as posterior cortical atrophy or the behavioral/frontal variant[24]. This phenotypic heterogeneity is also observed in genetically determined forms of YOD. Biomarkers—including plasma and cerebrospinal fluid proteins, neuroimaging, and genetic markers—have shown promise for early identification of YOD and for improving understanding of overlapping psychiatric and neurodegenerative symptoms in younger patients. Therefore, clinical management of YOD should consider the specific age-related challenges of younger dementia patients and their families, and provide integrated social, healthcare, and financial support tailored to both age and type of disability[24].

Mild Cognitive Impairment (MCI) represents an intermediate condition between physiological age-related cognitive decline and dementia, characterized by cognitive impairment greater than expected for age, yet with preserved functional independence in daily life. Unlike dementia, autonomy remains intact. The most recent definition of MCI was proposed by Petersen and colleagues[25].

A major challenge in MCI is the variability of clinical trajectories depending on the underlying etiology. Three main outcomes are possible:
-stability: some individuals with MCI remain stable for extended periods without significant decline ("stable MCI").
-recovery: in certain cases, cognitive function improves and returns to normal or near-normal levels.
-progression to dementia: MCI may evolve into more severe cognitive impairment with loss of functional independence, the hallmark of dementia. The annual progression rate from MCI to dementia is estimated at 10–15%, depending on underlying causes (notably AD and FTD) and individual risk factors.
Cognitive impairment may be restricted to a single domain (single-domain MCI) or involve multiple domains (multi-domain MCI). Patients can thus be classified into four clinical subtypes: (i) amnestic single-domain MCI (aMCI-SD), (ii) amnestic multi-domain MCI (aMCI-MD), (iii) non-amnestic single-domain MCI (naMCI-SD), and (iv) non-amnestic multi-domain MCI (naMCI-MD).

The MCI construct can further be staged clinically using the Clinical Dementia Rating (CDR) scale[26], a widely used instrument to assess cognitive decline and dementia severity over time. A CDR score of 0.5 is consistent with MCI, reflecting subtle cognitive changes exceeding normal aging (CDR = 0), but not yet meeting dementia criteria. Dementia is stratified according to severity and functional impairment into mild (CDR = 1), moderate (CDR = 2), severe (CDR = 3), very severe (CDR = 4), and terminal (CDR = 5).
Clinical characterization integrates laboratory and neuroimaging findings to guide clinicians in estimating risk and timing of progression. A central concept is that, by combining clinical subtypes with presumed etiologies, it may be possible to predict the specific type of dementia likely to emerge in patients with MCI. Current research is increasingly focused on genetic risk factors to identify individuals at higher risk of short-term progression.

Parkinson's Disease (PD) is the second most common neurodegenerative disorder in terms of prevalence and incidence. Its estimated prevalence is approximately 0.3% in the general population and 1–3% in individuals over 65 years of age.
Disease progression leads to degeneration of the nigrostriatal dopaminergic pathway, with marked loss of neurons in the substantia nigra pars compacta (SNpc) and subsequent dopamine (DA) depletion. Non-motor dysfunctions such as mild cognitive impairment (MCI), dementia, hyposmia, and gastrointestinal abnormalities frequently accompany disease progression.
The pathological hallmarks of PD are intracellular α-synuclein aggregates, referred to as Lewy bodies and Lewy neurites, which are found in multiple regions of the central nervous system, including the basal ganglia, dorsal motor nucleus of the vagus, olfactory bulb, locus coeruleus, intermediolateral nucleus of the spinal cord, celiac ganglia, and enteric nervous system.[27] Genetic factors are estimated to account for approximately 25% of PD risk, with associated genetic variants differing in both frequency and relative risk[28]. Familial Parkinson's disease (also referred to as Mendelian or monogenic PD) is characterized by rare but highly penetrant variants that significantly increase disease risk. Autosomal dominant forms (e.g., *SNCA*, *VPS35*) and autosomal recessive forms (*PARK2*, *PINK1*, *DJ-1*) represent approximately 5–10% of all PD cases[29].
Genetic classification of PD may guide therapeutic approaches and thereby improve patient prognosis and quality of life, including that of caregivers. The presence of pathogenic variants, together with clinical features such as age of onset and family history, can be used to stratify patients into homogeneous subgroups.

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disorder that typically begins with muscle weakness localized to one region, progressively spreading to other muscle groups. In some cases, dysarthria, dysphonia, and dysphagia may appear[30]. A considerable proportion of patients also exhibit cognitive impairment, sometimes nonspecific, and in other cases presenting as frontotemporal dementia. Disease progression is generally rapid, although it shows substantial interindividual variability, suggesting a significant role of genetic background in disease course. Epidemiological studies have shown a slightly higher incidence of ALS in women, and a greater prevalence in Europe compared to Asia and the Middle East[31,32,33]. Median survival is approximately 2–5 years from diagnosis, with respiratory failure representing the most common cause of death. ALS can be classified as familial (10% of cases) or sporadic (90% of cases). Environmental factors have also been implicated in disease onset[34]. The genes most frequently associated with ALS include *C9orf72*, *SOD1*, *TARDBP*, *FUS*, and *TBK1*[32,34].
The complexity of ALS is underscored by the failure of more than 40 randomized controlled trials aimed at identifying effective disease-modifying therapies. Currently, riluzole remains the only approved disease-modifying drug in most European countries; its antiglutamatergic properties prolong survival by approximately six months. More recently, the FDA approved tofersen for patients carrying *SOD1* mutations.

1.2. Potential therapeutic targets

Neurodegenerative diseases remain incurable. In recent decades, extensive experimental evidence has deepened our understanding of the cellular and molecular mechanisms underlying their development and progression. However, current knowledge remains insufficient not only for prevention but also for the effective treatment of these disorders. Despite their diverse clinical manifestations, neurodegenerative diseases share common molecular and cellular mechanisms[35]. This raises the prospect of identifying intracellular targets common to specific patient subgroups, thus enabling targeted therapeutic strategies.
Shared pathological features across neurodegenerative diseases include protein aggregation, altered proteolysis, mitochondrial dysfunction, impaired intracellular transport (including neurotransmitter trafficking), and dysregulated protein expression. A key unresolved question is whether these cellular processes are causal in neurodegeneration or represent downstream consequences of disease. From a practical standpoint, identifying dysfunctional molecular mechanisms responsible for the development of multiple neurodegenerative disorders is crucial. The advent of "omics" sciences (genomics, transcriptomics, proteomics, metabolomics, etc.) holds the potential to expand our understanding in this field and to provide innovative diagnostic, prognostic, and therapeutic opportunities for patients and their families.

1.3. Omics sciences

In recent years, the substantial reduction in costs associated with omics technologies has enabled the collection of large-scale datasets, fostering the integration of multimodal data into disease-specific networks. Notably, the integration of clinical data with whole-genome sequencing and transcriptomic, proteomic, and metabolomic profiling—supported by increasingly powerful machine learning algorithms—has allowed the identification of (i) novel causal loci or risk-associated variants, thereby improving diagnostic yield, and (ii) rare variants not detectable in small-scale studies, broadening our understanding of pathogenetic mechanisms involved in neurodegeneration[36,37]. This progress has created new opportunities to identify, characterize, and stratify patient subgroups, with the aim of enhancing clinical, genetic, and prognostic assessments and advancing personalized medicine strategies tailored to each individual[38]. The ultimate aim is to optimize treatment, thereby improving both health outcomes and socio-economic conditions associated with these diseases. Nevertheless, challenges remain in the application and management of omics data. Technical issues include computational demands for big data analysis and requirements for storage and data management infrastructure, which often represent a limitation for medium-sized centers. The most significant challenges, however, are ethical. The accumulation of personal information increases the vulnerability of systems that must ensure individual privacy. Moreover, the availability of broad clinical, instrumental, and molecular data can lead to the discovery of unexpected findings. In this context, the scientific community has long been engaged in regulating the use of such data. One example is the management of secondary/incidental findings (SF/IFs) in genomic sequencing[39]. According to numerous scientific publications and guidelines from scientific societies such as the American College of Medical Genetics and the European Society of Human Genetics, incidental findings are defined as “genetic variants identified through genomic sequencing that are unrelated to the primary disease under investigation”[40]. These include numerous genetic and genomic variants present in every individual. For the purpose of this project, and in line with the current literature[41], incidental findings are defined as “information relating to an individual’s health and/or relevant to their reproductive or existential choices—such as genetic data—that emerge during the course of a biomedical research project (such as the present one), without being among its primary or secondary objectives, as a result of commonly employed practices, analyses, or methodologies within that type of research” (CNR Research Ethics and Bioethics Commission Report, 2023). Among incidental findings that may emerge in omics-based research, a list of genes associated with highly penetrant disorders has been identified, for which early detection is linked to reduced morbidity and mortality (“actionable genes,” ACMG). In the present project, ACMG recommendations for reporting incidental findings identified during clinical exome and genome sequencing will be followed. This periodically updated list will serve as the reference throughout the project[42].

2. STUDY RATIONALE

Dementias, Parkinson's disease (PD) and parkinsonisms, and motor neuron diseases are clinically and genetically highly heterogeneous disorders. To date, genomic studies have largely focused on coding regions, i.e., protein-coding genes, which explain only a limited proportion of intra- and interindividual clinical variability. Furthermore, numerous cases with a family history strongly suggestive of a genetic basis but lacking a molecular diagnosis point to the involvement of other genomic regions—yet to be identified—in the etiology of these conditions. In this context, currently available treatments are often ineffective, and the list of failed pharmacological trials is long, with significant implications for both patient care and pharmaceutical investment. In recent years, a potential revolution in medical and healthcare practice has emerged through the interconnection and integrated analysis of data derived from multiple "omics" technologies. When combined with clinical and instrumental data, these data enable statistically meaningful correlations to be identified and novel molecular contributors to be revealed, which would be impossible to detect in small-scale studies. Ultimately, the systematic collection and accessible integration of clinical and genomic data aim to develop algorithms for stratifying patients into specific subgroups, thereby optimizing disease risk prediction, prognosis, and therapeutic response.

Several examples illustrate this approach:

Alzheimer's disease (AD): Large-cohort genomic studies using Genome-Wide Association Studies (GWAS) have identified ~90 loci associated with increased AD risk, thereby expanding the list of AD-related genes[43,44]. The construction of multimodal, interconnected networks has further enabled risk stratification profiles to be calculated based on clinical, genomic, metabolomic, and other data. Recent studies have investigated polygenic risk scores (PRS) for penetrance and recurrence in familial AD, as well as PRS associated with microbiome composition or specific cognitive and functional decline trajectories. However, predictive performance has so far been limited[45,46,47].

Parkinson's disease (PD): GWAS has extended the list of PD-associated genes, identified population-specific risk variants, and highlighted differences in genetic predisposition across populations[48,49,50]. Several studies have attempted to build risk prediction algorithms based on PRS, biochemical, clinical, and demographic profiles. While statistical performance has remained modest, these efforts provide proof-of-concept for the utility of integrating clinical, genomic, and metabolomic data[51].

Amyotrophic lateral sclerosis (ALS): Large-scale studies are fewer, owing to the lower prevalence of the disease. Nevertheless, GWAS have identified new risk loci and pathogenic mechanisms. For instance, DNA methylation studies have revealed alterations in several metabolic pathways[52,53].

Despite these advances, it remains unclear which factors influence disease progression in each neurodegenerative condition. Genetic studies have identified associations between certain phenotypes (slow versus aggressive progression) and single-gene variants. However, these findings cannot yet be applied reliably at the individual level, as (i) most patients lack identifiable genetic causes and (ii) the role of other factors in the metabolic context that may shape progression speed remains poorly understood. Moreover, the contribution of structural variants (CNVs and transposable element–related changes), noncoding genes (transcribed but untranslated sequences), and regulatory DNA regions (promoters, enhancers, repeat sequences) remains to be fully explored.

In this regard, whole-genome sequencing (WGS) offers the potential to improve diagnostic yield in patients with neurodegenerative diseases who test negative through next-generation sequencing (NGS) approaches limited to the coding genome (WES, whole-exome sequencing). Comprehensive genome characterization—enabling the simultaneous analysis of monogenic and polygenic factors as well as structural variants—combined with other "omics" sciences (e.g., transcriptomics, proteomics) provides a fundamental strategy for more precise and personalized elucidation of pathogenetic mechanisms and disease progression. For example, the combined genome–transcriptome approach allows investigation of genomic variants that drive aberrant splicing events (e.g., intron retention in mRNA), thereby uncovering novel disease-causing mutations.

Data generated through this integrated "omics" approach may ultimately enable the identification of subgroups of patients with distinct clinical profiles. At present, however, integrated genomic analyses are not routine practice.

Therefore, the **NeurOmics Study aims to exploit state-of-the-art omics technologies to identify pathogenic genomic variants, altered proteins, and/or dysregulated molecular pathways in neurodegenerative diseases, thereby providing a novel and comprehensive characterization of affected individuals. By integrating genomic, transcriptomic (including epigenomic), proteomic, metabolomic, and clinical data, the study seeks to identify new biomarkers for diagnosis, prognosis (including treatment response), and disease monitoring**. These omics data will expand knowledge of the molecular basis of neurodegenerative diseases and constitute a valuable resource to be shared—fully anonymized—within the scientific community. Importantly, the omics analysis of hundreds of cases across different neurodegenerative disease categories, as planned in this study, will have a significant impact on both the scientific community and public health, given the limited number of omics studies currently available at the national level. Collecting omics and clinical data from such large cohorts will contribute to defining homogeneous patient subgroups, achieving greater diagnostic–etiological resolution, and enabling more accurate patient stratification. Ultimately, this will support improved phenotypic characterization and foster the development of precision medicine approaches for neurodegenerative disorders.

3. STUDY DESIGN

NeurOmics is a multicenter, prospective, observational study enrolling patients diagnosed with neurodegenerative diseases, including Parkinson's disease and degenerative parkinsonisms, neurodegenerative cognitive disorders at the stage of mild cognitive impairment (MCI) or early/late-onset dementia, and motor neuron diseases (ALS). The study aims to identify pathogenic genomic variants and/or dysregulated molecular pathways in these patients by performing whole-genome sequencing and integrating these data with transcriptomic, epigenomic, and/or proteomic profiles. It further aims to identify novel biomarkers for the diagnosis and monitoring of neurodegenerative diseases through the analysis of biological samples (e.g., blood and derivatives), including liquid biopsy approaches using the most advanced technologies currently available.

The study plans to enroll at least **1,000 individuals** with neurodegenerative disease (see section *7.4 Statistical Analysis*).
Thanks to the availability of large public databases containing whole-genome sequences from healthy populations, such as **gnomAD** and the **1000 Genomes Project**[54], these datasets can be used as controls for variant identification in participants diagnosed with PD and parkinsonisms, early- and late-onset dementias (including MCI), and ALS. For example, the 1000 Genomes Project provides a catalog of common human genetic variants derived from self-reported healthy individuals, which has become a widely used reference in biomedical genomics.
To confirm the association of variants identified in this study with the relevant neurodegenerative disease, their presence will be checked against curated mutation repositories, including **ALZFORUM** and the **Alzheimer's Disease Neuroimaging Initiative (ADNI)** (http://adni.loni.usc.edu/) for AD, and **PPMI** for PD[55,56]. For subjects with MCI, genomic data deposited in ADNI (including controls, AD cases, and MCI) will be used. For ALS, the study will leverage the **Project MinE** database and genomic resources.

The case series and available data from each reference database are summarized in **Table 1**, while detailed statistical methodology is provided in section *7.4 Statistical Analysis*.

Database Name	Type of Sequencing	Number of Genomes (WGS), Exomes (WES), Genes, SNPs, Indels, SVs or CNV	Number of Subjects	Type of Case
gnomAD: genome aggregation Database, versione 3.1 (GRCh38)	WGS	**76.156** WGS	**76.156**	Controls
	Structural Variants (SVs)	**10,738**	**10,738**	Controls
1000 Genomes Project (GRCh38)	WGS	**3.202** WGS (30x) contenti **602** trio padre, madre, figlio	**3.202**	Controls
Alzheimer's Disease Neuroimaging Initiative ADNI	ADNI WGS	**SNPs: ~3.7 million** **Indels: ~700.000** **SVs: ~3.500**	**818** **267** **128** **415** **8**	Total Controls AD MCI Uncertain Diagnosis
Parkinson's Progression Markers Initiative PPMI	WGS and WES	**NA**	**195** **184** **414** **287**	Controls Genetic PD Sporadic PD Prodromal
Project MinE	WGS	**6,198 WGS**	**4,366** **1,832**	ALS Controls

Table 1. List of the main databases to be used for variant analysis.
WGS, Whole Genome Sequencing; WES, Whole Exome Sequencing; AD, Alzheimer's Disease; PD, Parkinson's Disease; MCI, Mild Cognitive Impairment; ALS, Amyotrophic Lateral Sclerosis; NA, not applicable.

Various types of biological samples will be collected from patients enrolled in the study, primarily peripheral blood, saliva (when blood sampling is not feasible), and urine.
Omics analyses—mainly genomics, transcriptomics, epigenomics, etc.—will be performed on the collected biological samples and/or their derivatives. Analyses may be conducted on genetic material extracted from whole peripheral blood, saliva, PBMCs (Peripheral Blood Mononuclear Cells), plasma, serum, and/or urine. In selected subgroups of cases (e.g., patients carrying genetic variants involved in specific metabolic processes, such as lipoprotein metabolism), analyses may also include proteins and/or lipids from biological samples and their derivatives.
The study does not involve the use of experimental drugs or interventions beyond those required for routine diagnostic practice.
The study will also assess the feasibility of implementing genomic analysis into clinical practice, and the potential clinical relevance of identifying novel genetic alterations, including those in non-coding regions of the genome.

3.1 Sample Management

- **Molecular analyses** will be performed using state-of-the-art technologies, in particular next-generation sequencing (NGS) of both second- and third-generation (short-read [SR] and long-read [LR]) platforms for DNA and/or RNA sequencing at the Genoa and Aosta (CMP3VdA) sites of the Istituto Italiano di Tecnologia Foundation.
- **NGS analyses of single genes, gene panels, clinical exome, and/or whole exome (WES)**—when indicated in the clinical management of these patients—will be carried out at the laboratories of IRCCS Policlinico San Martino (Genoa) and Fondazione IRCCS Policlinico Cà Granda (Milan). When necessary, these may also be performed at the IIT sites in Genoa and Aosta (CMP3VdA).
- **Proteomics, metabolomics, and lipidomics analyses:** Laboratory selection is ongoing to identify centers able to perform these investigations in accordance with the quality standards imposed by the project. Once identified, relevant ethical committees will be notified.
- **Data management and privacy:** The study's principal investigator is responsible for ensuring participant privacy. Study data will be recorded in a validated Data Management System via an electronic Case Report Form (eCRF), provided by the Scientific Directorate of IRCCS Policlinico San Martino (Genoa) and Fondazione IRCCS Policlinico Cà Granda (Milan). The REDCap (Research Electronic Data Capture) platform will be employed. The REDCap consortium comprises >1,000 institutional partners worldwide and supports secure web-based applications for research data collection. REDCap enables rapid and secure creation/management of online databases and is currently used for >110,000 projects with ~150,000 users across multiple research fields.
  For this study, REDCap will enable:
  a) user-level identification with role-based access restrictions;
  b) real-time data integrity validation;
  c) patient de-identification prior to data export;
  d) centralized data storage with daily backup on secure servers within the IT infrastructure of IRCCS Policlinico San Martino and Fondazione IRCCS Policlinico Cà Granda.
- **Biobanking:** Within the NeurOmics project, the Biological Resource Center (CRB) of IRCCS Policlinico San Martino and the Telethon Biobank of Fondazione IRCCS Policlinico Cà Granda will coordinate the collection, processing, and short- and long-term storage of samples, as well as their redistribution.

Specifically, the CRB will:

1. Develop the workflow for biological material management at the Genoa center and harmonize all study SOPs in collaboration with the Telethon Biobank (Milan).
2. Prepare blood collection kits to trace pre-analytical variables such as vacutainer batch number, order of tube collection, fill volume, and time between sampling and processing. Each kit, assigned at enrollment, will contain:
   a) a CRF for pre-analytical and storage data (completed by phlebotomist/nurse and CRB staff);
   b) a 5 mL vacutainer with clot activator and gel separator for serum (Ref: Vacuette®, Greiner 456073);
   c) a 6 mL K3EDTA vacutainer for plasma (Ref: Vacuette®, Greiner 456036);
   d) a 9 mL K3EDTA vacutainer (Ref: Vacuette®, Greiner 455036) for PBMC isolation using Ficoll.
3. Ensure sample quality (high-quality samples) through dedicated internal and/or external quality controls.

4. Trace all pre-analytical, analytical, and storage-related information via biobank software. These data will be pseudonymized and linked to laboratory, clinical, and follow-up data (well-annotated data).

4. STUDY DURATION

The study will have an overall duration of **10 years**. Follow-up assessments will be conducted periodically, as per routine clinical practice for each condition under investigation.

5. STUDY OBJECTIVES

The study aims to collect omics characterization for at least **1,000 patients** with neurodegenerative diseases, recruited from the Neurology and Geriatrics Units of IRCCS Policlinico San Martino (Genoa), Fondazione IRCCS Policlinico Cà Granda (Milan), and other units joining after ethics committee approval. Estimated recruitment: >400 dementia/cognitive impairment cases, >450 PD/parkinsonism cases, ~150 ALS/motor neuron disease cases.
The primary goal is to identify, through whole-genome analysis, both rare and common monogenic and polygenic variants—single nucleotide variants (SNVs) and structural variants (CNVs, insertions, translocations, inversions, mobile element insertions [MEIs])—that may be causative or confer increased risk for neurodegenerative diseases. Integration with other omics sciences will enable correlation of these variants with gene, protein, and metabolic expression patterns, thereby defining distinct molecular trajectories underlying neurodegeneration and their impact on disease progression.

Primary Objectives

1. Identify genomic variants not yet recognized as causes or susceptibility factors for neurodegenerative diseases, through WGS and dedicated bioinformatic pipelines (including protein-coding genes, noncoding RNA genes, regulatory DNA sequences—promoters, enhancers, repeats—and both SNVs and SVs).
2. Assess correlations between genetic–molecular alterations and clinical phenotype (e.g., severity, comorbidities) for the most frequent variants.
3. Identify splicing variants associated with neurodegenerative diseases using genomic and transcriptomic analyses.
4. Detect SNVs and/or SVs affecting enhancer or promoter regions.
5. Identify molecular mechanisms associated with neurodegenerative diseases and their impact on disease progression via multi-omics approaches.
6. Identify novel diagnostic, prognostic, and predictive biomarkers in neurodegenerative diseases.

Secondary Objectives

1. Identify causative or susceptibility genomic variants and provide accurate genetic counseling to patients and families.
2. Investigate associations between pathogenic variants in neurodegeneration-related genes and systemic comorbidities to inform prevention and early treatment[57].
3. Identify and characterize molecular markers (genes, transcripts, proteins, pathways) commonly dysregulated across different neurodegenerative diseases (e.g., cross-disorder gene overlap).
4. Identify cross-talk between known causative/susceptibility variants and additional transcripts/proteins/pathways influencing penetrance/expressivity of neurodegenerative disorders.

This study lays the foundation for defining homogeneous patient subgroups with combined clinical and genomic features, ultimately enabling personalized therapeutic strategies.

Given the sample size, disease-specific targeted studies may be conducted on selected subgroups. For example in Parkinson's disease, transcriptomic analyses will be performed in early-onset PD cases to detect expression or splicing defects in known PD genes, thereby revealing candidate genes with cryptic pathogenic variants through genome–transcriptome integration. Genetic data from PD patients will also be used to investigate associations between excess rare variants in predefined gene clusters (e.g., protein misfolding, mitochondrial dysfunction/oxidative stress, endolysosomal defects) and clinical phenotype, stratifying patients by progression rate: rapid progression (Hoehn & Yahr ≥2.5 at 5 years) versus slow progression (Hoehn & Yahr ≤1 at 5 years).

6. STUDY POPULATION AND SUBJECT SELECTION CRITERIA

The study population consists of patients affected by neurodegenerative diseases, referred to IRCCS Ospedale Policlinico San Martino (Genoa), Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico (Milan), and any additional Units that may join the study following Ethics Committee (EC) approval. Eligible subjects will be identified during outpatient visits conducted as part of routine clinical follow-up at the participating centers. At these visits, patients will be invited to participate in the study.
The clinical diagnosis will be established according to the criteria detailed below.

Eligibility Criteria
Subjects will be enrolled only if they fulfill all inclusion criteria and none of the exclusion criteria for each neurodegenerative condition under investigation, as described below.

ALZHEIMER'S DISEASE (AD)
Inclusion criteria
- Signed written informed consent from the patient and/or legal representative (Appendix X);
- Age ≥18 years;
- Clinical diagnosis of AD in patients referred to the participating clinical centers, as follows:

  - Predominantly amnestic presentation based on primary complaint and comprehensive neuropsychological assessment;
  - MMSE >24, with instrumental activities of daily living preserved (MCI) or mildly impaired (mild dementia), corresponding to a CDR score of 0.5 and 1, respectively;
  - Absence of significant neurological/systemic comorbidities or structural abnormalities indicative of other causes of cognitive impairment on morphological imaging;
  - Absence of non-AD–related structural abnormalities on imaging review;
  - CSF biomarkers (Aβ42, Aβ40, p-Tau181, t-Tau) consistent with AD profile (A+T+) or amyloid-positive PET with fluorinated tracers, according to current diagnostic criteria for AD-related MCI and dementia[58,59];
  - Availability of longitudinal clinical assessments, including MMSE and daily living activities, for at least 24 months.

Exclusion criteria
- Failure to meet any of the above inclusion criteria.

FRONTOTEMPORAL LOBAR DEGENERATION (FTLD)
Inclusion criteria
- Signed written informed consent from the patient and/or legal representative (Appendix X);
- Age ≥18 years;
- Clinical diagnosis of FTLD in patients referred to the participating centers, as follows:

- Predominant behavioral/dysexecutive presentation or progressive language impairment based on primary complaint and full neuropsychological evaluation, according to current diagnostic criteria[60,61];
- MMSE >24, with preserved (MCI) or mildly impaired (mild dementia) instrumental daily living activities (CDR score 0.5–1);
- Absence of significant neurological/systemic comorbidities or structural abnormalities indicative of other causes of cognitive impairment;
- Imaging findings consistent with bvFTD [60][one of the following required]:
  a) Prominent frontal and/or anterior temporal atrophy on structural imaging;
  b) Frontal/anterior temporal hypoperfusion or hypometabolism on PET or SPECT;
  OR
  Imaging consistent with primary progressive aphasia (PPA) [61] [one of the following required]:
  a) Prominent left posterior fronto-insular/parietal atrophy (non-fluent variant, nfvPPA) or anterior temporal lobe atrophy (semantic variant, svPPA) on structural imaging;
  b) Hypoperfusion or hypometabolism in the same regions on PET or SPECT;
- Availability of longitudinal clinical assessments, including MMSE and daily living activities, for at least 24 months.

Exclusion criteria
- Failure to meet any of the above inclusion criteria.

DEMENTIA WITH LEWY BODIES (DLB)
Inclusion criteria
- Signed written informed consent from the patient and/or legal representative (Appendix X);
- Age ≥18 years;
- Clinical diagnosis of DLB in patients referred to the participating centers, as follows:

- Predominantly non-amnestic cognitive impairment (visuospatial, dysexecutive, attentional) based on patient report or full neuropsychological assessment, plus characteristic clinical features according to current diagnostic criteria for DLB or MCI-LB[62,63];
- MMSE >24, with preserved (MCI) or mildly impaired (mild dementia) instrumental daily living activities (CDR 0.5–1);
- Absence of significant neurological/systemic comorbidities or structural abnormalities indicative of other causes of cognitive impairment;
- Supportive instrumental or imaging findings consistent with probable DLB[62,63], at least one of:
  a) Polysomnographic confirmation of REM sleep behavior disorder without atonia;
  b) Reduced striatal dopamine transporter uptake on SPECT/PET;
  c) Abnormal cardiac [123i]-MIBG scintigraphy (reduced uptake);
- Availability of longitudinal clinical assessments, including MMSE and daily living activities, for at least 24 months.

Exclusion criteria
- Failure to meet any of the above inclusion criteria.

MILD COGNITIVE IMPAIRMENT (MCI)
Inclusion criteria
- Signed written informed consent from the patient and/or legal representative (Appendix X);
- Age ≥18 years;
- Clinical diagnosis of MCI in patients referred to the participating centers, defined as:

- Subjective memory complaint (preferably corroborated by an informant);
- Intact ability to perform daily activities (e.g., meal preparation);
- Preserved general cognitive abilities;
- Objective evidence of memory deficit;
- Absence of dementia.

Exclusion criteria
- Failure to meet any of the above inclusion criteria.

AMYOTROPHIC LATERAL SCLEROSIS (ALS)
Inclusion criteria
- Signed written informed consent from the patient and/or legal representative (Appendix X);
- Age ≥18 years;
- Clinical diagnosis of motor neuron disease in patients referred to participating centers, as follows:

- Diagnosis of definite or probable ALS according to revised El Escorial criteria[64];
- EMG must demonstrate:
  a) Active denervation (fibrillation potentials, positive sharp waves, complex fasciculations);
  b) Chronic denervation (prolonged, often polyphasic motor unit potentials, reduced recruitment with high firing frequency at maximal effort);
  c) Nerve conduction studies typically show reduced motor potential amplitude with preserved or minimally reduced motor conduction velocity, normal sensory conduction, and absence of conduction block;
- Transcranial magnetic stimulation (TMS) may demonstrate UMN involvement (reduced corticobulbar/spinal conduction velocity, decreased cortical motor neuron excitability);
- Neuroimaging and laboratory investigations must exclude compressive, infiltrative, or inflammatory conditions mimicking UMN/LMN signs.

Exclusion criteria
- Failure to meet any of the above inclusion criteria.

PARKINSON'S DISEASE (PD)
Inclusion criteria
- Signed written informed consent from the patient and/or legal representative (Appendix X);
- Age ≥18 years;
- Clinical diagnosis of PD according to revised MDS criteria, as follows:

- Predominant motor presentation with bradykinesia plus resting tremor and/or rigidity;
- MMSE >24, with preserved (MCI) or mildly impaired (mild dementia) instrumental daily living activities (CDR 0.5–1);
- Absence of significant neurological/systemic comorbidities or structural abnormalities indicative of other causes of cognitive impairment;
- Imaging findings:
  a) Normal structural brain imaging;
  b) Striatal DAT binding reduction on SPECT DAT-SCAN;
- Availability of longitudinal clinical assessments, including motor scales (e.g., UPDRS-III, Hoehn & Yahr), MMSE/MOCA, and daily living activities, for at least 24 months.

Exclusion criteria
- Failure to meet any of the above inclusion criteria.

7. LABORATORY PROCEDURES: GENETICS, OMICS, BIOMARKERS

7.1 Enrollment and Follow-up

- Patient selection
Eligible subjects will be identified during outpatient visits at IRCCS Policlinico San Martino (Genoa), Fondazione IRCCS Cà Granda (Milan), and other centers that may later join the study.

- Baseline visit and informed consent
The Investigator will invite patients to participate after verifying inclusion/exclusion criteria. Patients (or legal representatives in case of cognitive impairment) will receive detailed information on study aims, benefits, and potential risks, along with written information sheets and informed consent forms. Adequate time will be allowed for discussion with the general practitioner before consent. Patients will also be asked whether they agree to be re-contacted for future research studies.
Investigators at the Genoa and Milan centers are responsible for obtaining and archiving signed informed consent before any study-specific procedures.

- Visit 1
At this visit, informed consent will be obtained, and subjects will be enrolled with assignment of a pseudonymized code generated in REDCap using the Mainzelliste algorithm (PMID: 25656224). The algorithm processes identifying attributes (e.g., name, surname, date of birth) to generate a unique, anonymous personal identifier (PID), ensuring record linkage across multiple databases while tolerating spelling variations.
Peripheral blood (20–25 mL) will be collected for preparation of PBMCs, serum, and/or plasma, and urine will be collected for omics analyses. DNA will be extracted from whole blood, RNA preferably from PBMCs; saliva may be used as an alternative genetic source. Sampling will be integrated into routine clinical workflow.
Clinical and biological data will be collected at IRCCS Policlinico San Martino (Prof. Paola Mandich), Fondazione IRCCS Cà Granda (Dr. Alessio Di Fonzo), and other participating centers.

Clinical data collection
Following consent, clinical data will be collected at participating Units and managed through REDCap. Case report forms (CRFs) will specify disease-specific data. At Visit 1, the following will be recorded:

- Sex
- Age
- Medical history
- Neurological examination
- Neuropsychological profile
- Neuroimaging findings
- Previous treatments
- Planned treatments

- Visit 2
If consent is given, patients/legal representatives will be informed of clinically relevant genetic variants identified. Variants associated with risk of developing other conditions (subject to confirmatory testing) will be disclosed during genetic counseling with psychological support. Patients will also decide whether information should be extended to relatives.

- Follow-up visits
Subsequent follow-up visits will coincide with routine clinical visits in Neurology and Geriatrics Units at Genoa and Milan centers.

7.2 Sample Collection and Processing

- Genomic data collection
Molecular and genetic diagnostic data generated at Genoa and Milan centers will be entered into the CRF and REDCap. Genomic data generated by IIT-CMP3VdA will be stored at the Engineering D.Hub data center in Pont Saint Martin (Aosta), with backup in Vicenza.

- Biological sample processing
Samples will be collected at Neurology and Geriatrics Units of IRCCS Policlinico San Martino (Prof. Paola Mandich) and Fondazione IRCCS Cà Granda (Dr. Alessio Di Fonzo), frozen, and stored in local biobanks (Genoa Biobank; Telethon Biobank, Milan). DNA/RNA will be extracted from PBMCs and/or peripheral blood and derivatives according to omics analyses to be performed at the respective hospital laboratories. Part of the pseudonymized DNA/RNA sample will be transferred to IIT (Genoa and Aosta CMP3VdA) for subsequent genomic analyses.

- Biobanking

Serum and plasma isolation
Serum and plasma isolation will be performed following standard procedures: blood samples will be collected in dedicated tubes (K2/K3 EDTA for plasma and clot activator for serum and/or with stabilization solutions for nucleic acids) and centrifuged once or multiple times (ranging from 1600 to 3000g) at room temperature and/or +4°C, depending on whether serum or plasma is being prepared. Cryovials containing aliquots of serum and plasma will be stored at –80°C.

PBMC collection and culture
Two EDTA or Lithium-Heparin tubes will be layered onto Ficoll and centrifuged at 2000g for 20 minutes to obtain the PBMC fraction, which will be isolated and frozen in FBS + 10% DMSO at –80°C.

Biomarkers
Biomarkers will be analyzed using standardized methodologies. For biomarkers already assessed in clinical practice, values will be retrieved from the patients' medical records. Should new diagnostic biomarkers become available over time, these will also be analyzed according to validated and standardized procedures.

- Sequencing
NGS sequencing will be performed in collaboration with IIT using state-of-the-art sequencing technologies. The laboratory is equipped with both second- and third-generation sequencing platforms, enabling short- and long-read sequencing, respectively, with the possibility of integrating both technologies when advantageous for research purposes.
Short-read Illumina sequencing allows highly accurate analysis of single-nucleotide variants (SNVs) and small insertions/deletions (indels), while long-read sequencing facilitates the analysis and identification of structural variants (SVs). These NGS platforms represent the state of the art for transcriptomic (RNA-Seq), epigenomic, and metagenomic studies.
The **Illumina NovaSeq 6000**, currently available at Fondazione Istituto Italiano di Tecnologia in Genoa and Aosta (CMP3VdA), ensures high throughput, generating up to 6 terabases of sequencing and 20 billion reads per run. Sequencing quality is high, with over 75% of bases reaching a quality score ≥30. Sequencing outputs FASTQ files stored directly in the cluster's analysis storage. A minimum sequencing depth of 30X will be applied, enabling comprehensive variant identification, classification, and association with neurodegenerative diseases.

The **PromethION 24 system** (Oxford Nanopore Technologies, ONT), also available at CMP3VdA, provides flexible, ultra-high throughput third-generation sequencing, with 24 independent flow cells generating terabases of data in real time. This technology enables direct sequencing of native DNA/RNA strands using nanopores embedded in an electro-resistant membrane, through which ionic current alterations induced by nucleic acid passage are converted into genomic sequences via basecalling software.
Third-generation sequencers, due to their ability to process long fragments, offer superior resolution in detecting short repeat units (<1 kb), repeat-rich DNA regions, and large-scale inversions (>10 kb), while also allowing haplotype phasing and improved resolution of compound heterozygosity. Sequencing depth typically ranges between 2–5X, depending on the study design.

7.3 Data Analysis

Bioinformatic analyses
Bioinformatic analyses of the various *omics* datasets will be conducted in collaboration with Fondazione Istituto Italiano di Tecnologia (CMP3VdA, Genoa and Aosta).

- DNA analysis using Illumina technology
For genomics, DNA variants identified will be analyzed jointly by CMP3VdA and the clinical units in Genoa and Milan. The analytical pipeline will generate the following files:

1. **FASTQ** – containing raw sequencing reads and per-base quality scores; the starting point of bioinformatic analyses.
2. **BAM** – containing mapped reads aligned to the human reference genome.
3. **VCF/BCF** – containing identified variants for each individual.

Pipeline steps:
a) Quality assessment, filtering, and trimming of raw reads.
b) Read alignment to the reference genome.
c) Identification of SNVs, indels, structural variations (SVs), rearrangements, and copy number variations (CNVs).
d) Generation of a comprehensive variant database.
e) Compilation of control population variants and allele frequency tables.
f) Comparison of patient-specific variants and allele frequencies with control populations and public cohorts.
g) Prediction of variant effects and association with genomic features.
h) Prioritization of patient-specific variants and their association with neurodegenerative disorders, leading to genomic-based stratification.

The pipeline employs standard tools, notably **GATK (Genome Analysis ToolKit)**[89] from the Broad Institute, widely considered the gold standard for variant analysis. As certain GATK tools are computationally intensive, IIT uses NVIDIA GPUs and **Clara™ Parabricks** software for accelerated processing, achieving a 15–20× reduction in runtime for mapping and variant calling.

- Variant annotation
VCFs will be annotated using tools such as **SnpEff**[66,67], leveraging multiple public databases that catalog variant effects, including:

- ALZFORUM[55] (AD mutations): https://www.alzforum.org/
- PPMI[56] (PD mutations): https://www.ppmi-info.org/
- Alzheimer's Disease Neuroimaging Initiative (ADNI): http://adni.loni.usc.edu/about/
- Gene4PD[69] http://genemed.tech/gene4pd/

- Parkinson's Disease DNA Variant Browser[70] : https://pdgenetics.shinyapps.io/variantBrowser/
- ClinVar[71] (variant–phenotype relationships)
- Databases containing allele frequencies of known variants gnomAD[72], 1000 Genomes, dbSNP [73]
- ALSoD (ALS mutations): https://alsod.ac.uk/

Variants will be classified according to **SIGU** and **ACMG international guidelines**.

- Structural variant analysis
Given the high false-positive rate of short-read SV detection, a consensus approach integrating multiple tools (Manta, Lumpy, CNVnator, BreakDancer)[77] will be applied. Variants will then undergo consensus filtering with **SURVIVOR**, re-genotyping with **Smoove**, and false-positive reduction with **Duphold**, maintaining >99% true positives while removing ~61% false positives. Additionally, **MELT** will be used for transposable element insertions[78]. SV frequency will be estimated using an in-house algorithm, and annotations added with **AnnotSV** and **SVAFotate**. CNVs will also be evaluated using **CNVkit**.
All results will populate a central database for variant prioritization, disease association, and data mining.

- Reproducibility and updates
All analyses will be performed within standardized virtual environments (containers) ensuring reproducibility across samples. Updates to software and databases (e.g., genome builds, annotation tools) will be periodically implemented, reprocessing prior datasets to maintain comparability. Variant annotation will be continuously updated as novel pathogenic variants emerge in the scientific literature.

In summary, the pipelines employed for genomic data analysis utilize a sequence of widely adopted software tools for variant identification (including PARABRICKS, with BWA and GATK adapted for GPU usage, and ANNOVAR and SnpEff for annotation) applied to both coding genes and non-coding regions of the genome. Typically, the most recent human reference genome is used, currently version Hg38. Several databases containing known variants are employed to filter out neutral variants and to identify those associated with the disease of interest. In particular, gnomAD and the 1000 Genomes Project (Table 1) are used, as they are representative of the global population in terms of both neutral and deleterious mutations. This strategy allows us to exclude variants currently lacking functional impact and to identify those likely responsible for disease.
It is important to note that, regardless of the software employed, variant definition and classification are carried out according to internationally recognized standards such as ACMG guidelines.
This approach ensures that all samples are analyzed using the same software and database versions stored within the container (Table 2). As soon as new software updates become available, the virtual machine will be updated accordingly, and all previously processed samples will be reanalyzed. This guarantees reproducibility across different samples. Furthermore, the variant annotation software will be updated frequently, as mutations currently of unknown significance may, in the future, be classified as pathogenic owing to continuous advances in scientific research.

Software or Database	Version
Clara™ Parabrick	4.1.0
SnpEFF	5.0e
ANNOVAR	20221123
Cosmic	v94
GnomAD	v3.1.2
dbSNP	155
PolyPhen	v2
Sift	6 Dec 2019
CADD	v1.6
FATHMM	v2.3
Genoma	Hg38
ADNI	NA
PPMI	NA
ALZFORUM	NA
Gene4PD	NA
Parkinsons's Disease Variant Browser	0.2.1
ClinVar	NA
1000 Genomes Project	NA
MELT	2.2.2
ALSoD	NA
BreakDancer	1.4.5
Manta	1.6.0
CNVKit	0.9.9
Lumpy	0.2.13

CNVnator	0.3.3
Smoove	0.2.6
SVAFotate	0.0.1
AnnotSV	3.2.3

Table 2. List of software and databases employed, with the corresponding version currently in use, where available. NA, Not Applicable.

DNA Analysis with Nanopore Technology

Depending on research outcomes and project requirements, a subset of cases within this study may be sequenced using third-generation Oxford Nanopore Technologies (ONT). This approach may be particularly useful for improved characterization of structural variants, methylation profiling, validation purposes, and other specific analyses.
ONT provides a suite of analysis tools. For DNA sequencing, ONT currently recommends the Guppy basecaller, which is the most performant for initial basecalling. For alignment of sequencing reads to the reference genome, Lra will be employed, while SNV calling will be performed using Clair3.
For structural variant detection, alignment will be performed with Lra, and variant calling with CuteSV.
ONT technology also enables haplotype phasing, allowing determination of the allele on which a variant is located. This will be performed using the WhatsHap tool.

Base Quality Estimation of Sequenced Reads

In sequencing-by-synthesis (SBS) technology (Illumina), the quality of each base in a read is quantified by assigning a Phred Quality Score (PQS), computed by a dedicated algorithm[79,80]. This score represents the probability that a base has been miscalled. A PQS of 10 corresponds to an expected error rate of 1 in 10 bases (90% accuracy), a PQS of 20 to 1 in 100 (99% accuracy), and a PQS of 30 to 1 in 1000 (99.9% accuracy). When sequencing quality reaches Q30—the benchmark for next-generation sequencing (NGS)—virtually all reads are expected to be error-free and unambiguous.

RNA Analysis

RNA sequencing (RNA-Seq) data, generated using commercially available platforms and those installed at CMP3VdA laboratories (currently Illumina and/or ONT), may be used to investigate gene expression and potential correlations between non-coding variants and expression changes. Expression quantitative trait loci (eQTLs) highlight variants that, although not altering protein structure, may affect transcriptional regulation and gene function. Additional analyses of transcript modifications, such as RNA methylation, may also be performed.
Gene expression analyses will be carried out using the DESeq2 package in R[81].
RNA sequencing with ONT technology further enables characterization of transcript isoforms generated by alternative splicing, in addition to quantifying expression. For instance, in Alzheimer's disease, certain TAU isoforms resulting from alternative splicing have been associated with increased disease risk.

Epigenomic Analysis

For the study of molecular modifications regulating gene expression—such as DNA methylation, one of the most common epigenetic modifications—state-of-the-art technologies available both commercially and at CMP3VdA will be employed. Both Illumina and ONT sequencing platforms can be applied for this purpose. In particular, the PromethION 24 platform directly extracts methylation information during DNA sequencing, thereby enabling simultaneous characterization of both the genetic and epigenetic landscape in a single sequencing run. Identification of disease-specific methylation signatures (episignatures) will be performed using the most up-to-date bioinformatic tools available in the scientific community for this class of analyses.

Pathway Analysis
For unresolved cases where the diagnostic solution is not readily identifiable, secondary analyses will be conducted. Variants, including common ones with potential functional effects, will be aggregated and investigated at the pathway level. Specifically, the involvement of patient-specific gene pathways will be compared with the individual's clinical status.
Causative variant identification will be supported by annotation tools such as ANNOVAR and SnpEff, which classify variants as loss-of-function (LoF) or missense. Genes implicated will then be analyzed using available platforms such as STRING and GeneMANIA to assess associations between pathways, gene lists, and variant frequencies.
Gene network and pathway enrichment analyses will be performed to identify biological processes linked to the studied disorders. Commonly used bioinformatic tools such as Gene Set Enrichment Analysis (GSEA)[82] and g:Profiler[83] will be applied for this purpose.

Blood Biomarkers of Neurodegenerative Diseases
Several blood biomarkers currently available and associated with cognitive impairment and other neurodegenerative disorders will be analyzed. These include amyloid-beta 1-42 (Aβ1-42), amyloid-beta 1-40 (Aβ1-40), tau protein phosphorylated at threonine 181 (pTau181), and neurofilament light chain (NFL), a sensitive indicator of neuroaxonal damage in various degenerative conditions. These molecules are key elements in the identification and understanding of pathological mechanisms specifically associated with Alzheimer's disease (Aβ1-42, Aβ1-40, pTau181) and, more broadly, with neurodegenerative diseases (NFL). They also provide information regarding disease progression and patient staging[84,85]. Quantification of these biomarkers will be performed using a chemiluminescent enzyme immunoassay (CLEIA) on the Lumipulse G600 II platform (Fujirebio, Ghent, Belgium). Additional biomarkers, once validated and available, will also be analyzed.

Experimental Validation of Data
From a bioinformatics perspective, NGS-based sequencing methods—particularly Illumina short-read platforms—present a very low error rate (~10^{-5})[86]. However, results obtained through sequencing technologies may be validated using Sanger sequencing, which has a per-base accuracy of 99.999%[87], and/or other conventional approaches, thereby reducing the likelihood of error to a negligible level.

7.4 Statistical Analysis

Sample Size
Given the nature of this study, sample size is not determined by a formal statistical power calculation but rather by projections of i) the number of patients with neurodegenerative diseases expected to be referred to the two experimental centers during the 10-year project period, and ii) an estimated compliance rate of approximately 80%.

We anticipate enrolling at least 1,000 patients with neurodegenerative diseases, distributed approximately as follows:
a) Parkinson's disease (PD): ~30% of cases
b) Parkinsonism: ~5% of cases
c) Dementias:
• Alzheimer's disease (AD): ~30% of cases
• Frontotemporal dementia (FTD): ~10% of cases
• Young Onset Dementia (YOD): ~5% of cases
• Mild Cognitive Impairment (MCI): ~5% of cases
d) Amyotrophic lateral sclerosis (ALS): ~15% of cases.

Additionally, ~200 geriatric controls with negative neurological evaluation and negative family history of neurodegeneration will be enrolled as internal controls.

The projected cohort of ~1,000 patients is based on the epidemiology of neurodegenerative diseases observed at IRCCS Policlinico San Martino (Genoa) and Fondazione IRCCS Policlinico Ca' Granda (Milan). The expected distribution across disease categories should allow reliable characterization of clinical and molecular subgroups.

Variant Analysis and Validation

In genomic studies, initial evaluation focuses on determining whether identified variants are:
i) previously established as causative of neurodegenerative disease,
ii) novel variants potentially acting as disease-driving mutations, or
iii) benign polymorphisms catalogued as non-pathogenic.

Variant assessment will rely on open-access reference databases widely used as benchmarks in clinical and population genomics. To validate associations between identified variants and neurodegenerative diseases, analyses will include:

a) **Control genomes:** Databases such as the 1000 Genomes Project, the International Parkinson's Disease Genomics Consortium (IPDGC), and PPMI will be used to determine the minor allele frequency (MAF) of each variant in European and Italian populations, with a particular focus on the 107 Tuscan genomes sequenced for the 1000 Genomes Project. Considering the prevalence of rare variants in neurodegenerative diseases and the MAF of common variants linked to late-onset AD and PD, the cut-off for rare variants will be set at 0.01. This analysis will allow prioritization of variants absent from control databases.

b) **Disease-specific databases:** Variants will also be analyzed against disease-focused resources such as ALZFORUM, ADNI, PPMI, and Project MinE, which contain genomic data from thousands of patients with AD, PD, and ALS. This will support validation of both individual variant–disease associations and combinations of rare variants enriched in patient genomes. Custom databases will be built from ADNI, PPMI, and MinE to provide allele frequencies and enable systematic comparison between enrolled individuals and these large patient cohorts.

Rare variants will be further assessed for potential disease susceptibility effects in the presence of other variants within the same genome. To this end, the number of variants per sample will be analyzed in both patient and control populations. Odds ratios (ORs) with corresponding 95% confidence intervals (95% CIs) will be calculated. Q-value correction (an extension of the False Discovery Rate) will be applied, and receiver operating characteristic (ROC) curves will be computed to assess test accuracy.

Statistical Approaches

Descriptive statistics will be generated to report the percentage of variants relative to the total enrolled cohort and to the control populations (European and Italian), stratified by disease (PD, parkinsonism, AD, FTD, YOD, ALS, and MCI), individual genes, and specific mutations.
Molecular–genetic data (single variants or variant groups) and clinicopathological data (diagnosis, disease stage, prognostic/predictive parameters, survival) will be analyzed using both numerical and graphical approaches. Frequency and contingency tables will be constructed, and statistical indices (means, percentiles, standard deviation) will be computed. Data will be visualized through histograms, boxplots, and scatterplots. Appropriate statistical tests—including t-tests, Fisher's exact test, and others depending on the formulated hypotheses—will be applied. Enrolled patients will be monitored through active follow-up as part of routine diagnostic care. Data regarding disability progression, disease course, and overall survival will be analyzed by diagnosis, mutation, or mutation combinations, as well as treatment received, using Kaplan–Meier survival analysis.

8. CLINICAL RISK

Blood samples will be collected from these patients (San Martino and Milan) or, if necessary, through an ad hoc collection (IRCCS Foundation, Milan). No additional risks are expected for enrolled patients, since the study does not foresee further procedures but is limited to the collection of clinical data.

Screening Failure

If it is determined that a subject is not eligible for enrollment despite having provided informed consent (e.g., due to an incorrect diagnosis), the patient will be classified as a screening failure. In such cases, all collected data will be deleted and any biological samples destroyed.

Withdrawal from the Study and Early Discontinuation

Participants are free to withdraw from the study at any time without providing any explanation.
Early withdrawal may also occur if the Sponsor decides to terminate the study prematurely, following discussion with the Ethics Committee.
The Sponsor may decide to discontinue the study early under specific circumstances, such as inadequacy of the site or investigators, financial reasons, insufficient recruitment, or intervention by the Ethics Committee.

Management of Adverse Events

As this is not an interventional study, no adverse events are expected. Should an adverse event occur during the collection of biological samples, it will be managed according to standard clinical practice.

9. ETHICAL ASPECTS

The study will be conducted in accordance with the principles of the Declaration of Helsinki and with international standards of Good Clinical Practice. The protocol and the patient information/informed consent form must be approved by the Ethics Committee before enrollment of patients at the Clinical Center.
The investigator is responsible for notifying the Ethics Committee of any protocol amendments, in compliance with current regulations, and for retaining copies in their records.
Projects involving genomic studies raise several challenges related to the increasing use of data. First, technical challenges arise from the need for computational resources for big data analysis, as well as for data storage and management. The growing volume of data produced by research studies requires high-performance computational systems and large storage capacity, which often becomes a limiting factor for small to medium-sized centers.
The most critical challenges, however, are ethical in nature. The accumulation of personal information inevitably makes systems responsible for ensuring privacy vulnerable. Furthermore, the availability of a wide range of information (both clinical and genetic) increases the likelihood of incidental or unexpected findings. For this reason, the scientific community has been working for years on developing regulations for the use of such data.
An example is that of *secondary/incidental findings (SF/IF)* in genomic sequencing. The literature includes numerous publications and guidelines from Scientific Societies (e.g., the American College of Medical Genetics, the European Society of Human Genetics) that define SF/IF as “genetic variants identified through genomic sequencing that are unrelated to the investigated condition” (ACMG, 2013). These findings therefore include many genetic and genomic variants present in every individual.
For the purposes of this project, and according to the relevant literature, secondary findings are defined as “information concerning an individual’s health and/or relevant to their reproductive or existential choices, particularly genetic data emerging in the course of a biomedical research project such as the present one, but not included among its primary or secondary objectives, as a result of practices, analyses, and/or investigative methodologies commonly used in that type of study” (CNR Research Ethics and Bioethics Commission, 2023).

Among the possible secondary findings that may emerge in omics-based research projects, a list of genes has been identified that are associated with highly penetrant diseases, for which early detection is linked to reduced morbidity and mortality (*actionable genes, ACMG 2023*)[91]. This periodically updated list will serve as a reference throughout the present project.

10. PRIVACY AND CONFIDENTIALITY

The Sponsor, Co-Sponsor, and Investigator act as independent data controllers for the processing activities under their respective responsibility, in compliance with current data protection regulations (Regulation (EU) 2016/679, GDPR, as amended), as well as in accordance with the provisions of the Italian Data Protection Authority ("Garante") ruling of July 2019 on the processing of special categories of data pursuant to Article 21, paragraph 1 of Legislative Decree 10 August 2018, no. 101.

Researchers and designated personnel involved in the study and authorized to process data operate in accordance with the "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016."

In particular, each participant is informed, through an appropriate information sheet, of the purposes, conditions, and modalities of data processing. Participants will be assigned an alphanumeric identification code specific to the study, which will enable pseudonymized registration of their data. Throughout the study, participant data will be handled exclusively in association with the assigned study code, including for scientific congress presentations or publications, whether as individual cases or in aggregated form.

Similarly, codes assigned by the healthcare facility to biological samples and patients will not be shared with collaborating centers. Only the Principal Investigator, or a designated member of the study team (clinical staff directly involved), will be able to link such codes to the study ID.

For data verification purposes, authorized representatives of the Sponsor and Competent Authorities may request direct access to parts of the study documentation, including participant-related documents.

Privacy compliance is supervised by the Data Protection Officers (DPOs) of the individual project partners. Their role is to ensure the correct application of privacy regulations, confirm that information systems comply with legal requirements, conduct risk assessments, and recommend any additional technical safeguards, as well as define the data management policies to be followed by operational staff.

The Data Controllers are:

- **IRCCS Ospedale Policlinico San Martino**, Largo Rosanna 10, 16132 Genoa, Italy.
- **Fondazione Istituto Italiano di Tecnologia (IIT)**, Via Morego 30, 16163 Genoa, Italy (Tel. +39 010 28961).
- **Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico**, Via Francesco Sforza 28, 20122 Milan, Italy (Tel. +39 0255033802).

All data processing will follow the provisions outlined in the *Information Sheet and Consent Form for the Processing of Personal Data* attached to this Protocol.

11. DATA SECURITY AND DATA BREACH

The risk of a data breach—i.e., unauthorized access to data with potential consequences for participants providing their health information and for the project's outcomes and financial sustainability—requires that data protection be addressed both technically and organizationally.

At the technical level, the entire IT infrastructure of the research center has been designed to ensure substantial isolation of the research laboratory from the external environment, thereby minimizing the risk of unauthorized access. All data exchange points with external entities are secured by dedicated firewalls and security appliances (Fortinet Fortigate 101s physical appliances and Fortinet Fortigate VM virtual appliances on ESX), which monitor and restrict network traffic strictly to the flows essential for the center's functioning.

Connections between the different nodes of the IT infrastructure are encrypted to prevent interception, and storage systems are equipped with encryption mechanisms to protect data at rest, ensuring confidentiality even in the event of disk repair at external maintenance facilities. Central systems maintain access logs, in compliance with sector regulations, collected centrally in a dedicated system accessible only to the security monitoring team.

The use of administrative "super-accounts" (root access) is restricted to emergency situations requiring system recovery, while routine administrative activities are performed through personal technician accounts, with privilege escalation applied only when necessary. The infrastructure undergoes weekly vulnerability assessments using the Tenable Nessus platform managed by the SOC Engineering team. Issues identified are resolved within scheduled monthly maintenance windows unless urgent intervention is required.

The entire data center infrastructure is ISO 27001 and ISO 27017 certified, and compliant with ISO 27018 best practices for cloud and data center data security. Security events are monitored 24/7 by the SOC Engineering team, with specific countermeasures in place to mitigate denial-of-service (DDoS) and flooding attacks. A dedicated Security Officer is responsible for ensuring the enforcement of security policies and coordinating with public authorities in the event of cybersecurity incidents.

The local network of the Aosta laboratory is composed of laptops and workstations with preconfigured, centrally managed software platforms controlled by the parent company's IT department. Users are not authorized to install unauthorized software, and all administrative and functional system activities are managed by dedicated IT support staff.

All software solutions developed in this context are designed and implemented according to the principles of *Privacy by Design* and *Privacy by Default*, which translate technically into *Security by Design* and *Security by Default*. This includes secure operator credential validation, application-level encryption of data, and authentication mechanisms—also between software modules—based on dual-key systems and digital certificates.

At the organizational level, all staff involved in the project across the partner institutions have signed formal letters of appointment committing to compliance with the GDPR. Staff have also undergone dedicated training on GDPR regulations and the operational procedures applicable within the project, which are published and accessible via the project's document repository.

Particular attention is drawn to the Co-Sponsor's *Data Breach Management Procedure*, which defines the steps for handling suspected personal data breaches:

- Internal reporting of the suspected breach
- Identification of the breach
- Risk assessment and qualification
- Notification to the Supervisory Authority
- Communication to affected data subjects
- Registration of the Data Breach, reporting to Management, and archiving of documentation

12. PUBLICATION AND DISSEMINATION POLICY

The Investigators and the Sponsor commit to making the data generated by the study available to the scientific community. The results of the Study may be published in anonymized form in scientific journals and presented at scientific congresses, events, or seminars, exclusively for scientific, educational, and/or institutional purposes.

Funding, Support, and Activities

This is an investigator-initiated project, the costs of which will, for the time being, be covered partly by IIT and partly through the budgetary resources of the participating units.

IIT will be primarily responsible for the NGS sequencing of the samples, big data analysis, and data storage within the Engineering D.HUB HPC infrastructure.

13. INSURANCE COVERAGE

Not applicable.

BIBLIOGRAPHY

1. Weir S, Samnaliev M, Kuo TC, Tierney TS, Walleser Autiero S, Taylor RS, Schrag A. Short- and long-term cost and utilization of health care resources in Parkinson's disease in the UK. Mov Disord. 2018 Jul;33(6):974-981. doi: 10.1002/mds.27302. Epub 2018 Mar 30. PMID: 29603405.
2. Reese JP, Hessmann P, Seeberg G, Henkel D, Hirzmann P, Rieke J, Baum E, Dannhoff F, Müller MJ, Jessen F, Geldsetzer MB, Dodel R. Cost and care of patients with Alzheimer's disease: clinical predictors in German health care settings. J Alzheimers Dis. 2011;27(4):723-36. doi: 10.3233/JAD-2011-110539. PMID: 21876256.
3. Çoban M, Bilge U, Balseven H, Uysal H, Artut B. The economic evaluation of ALS care: quality and cost. Amyotroph Lateral Scler Frontotemporal Degener. 2023 Aug;24(5-6):502-513. doi: 10.1080/21678421.2023.2176776. Epub 2023 Feb 16. PMID: 36794629.
4. Parnetti L, Chipi E, Salvadori N, D'Andrea K, Eusebi P. Prevalence and risk of progression of preclinical Alzheimer's disease stages: a systematic review and meta-analysis. Alzheimers Res Ther. 2019 Jan 15;11(1):7. doi: 10.1186/s13195-018-0459-7. PMID: 30646955; PMCID: PMC6334406.
5. Menniti-Ippolito F, Spila-Alegiani S, Vanacore N, Bonifati V, Diana G, Meco G, Raschetti R. Estimate of parkinsonism prevalence through drug prescription histories in the Province of Rome, Italy. Acta Neurol Scand. 1995 Jul;92(1):49-54. doi: 10.1111/j.1600-0404.1995.tb00465.x. PMID: 7572060.
6. Chiò A, Logroscino G, Traynor BJ, Collins J, Simeone JC, Goldstein LA, White LA. Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. Neuroepidemiology. 2013;41(2):118-30. doi: 10.1159/000351153. Epub 2013 Jul 11. PMID: 23860588; PMCID: PMC4049265.
7. Bacigalupo I, Mayer F, Lacorte E, Di Pucchio A, Marzolini F, Canevelli M, Di Fiandra T, Vanacore N. A Systematic Review and Meta-Analysis on the Prevalence of Dementia in Europe: Estimates from the Highest-Quality Studies Adopting the DSM IV Diagnostic Criteria. J Alzheimers Dis. 2018;66(4):1471-1481. doi: 10.3233/JAD-180416. PMID: 30412486; PMCID: PMC6294583.
8. Kovacs GG. Concepts and classification of neurodegenerative diseases. Handb Clin Neurol. 2017;145:301-307. doi: 10.1016/B978-0-12-802395-2.00021-3. PMID: 28987178.
9. Brettschneider J, Del Tredici K, Lee VM, Trojanowski JQ. Spreading of pathology in neurodegenerative diseases: a focus on human studies. Nat Rev Neurosci. 2015 Feb;16(2):109-20. doi: 10.1038/nrn3887. Epub 2015 Jan 15. PMID: 25588378; PMCID: PMC4312418.
10. Ciurea AV, Mohan AG, Covache-Busuioc RA, Costin HP, Glavan LA, Corlatescu AD, Saceleanu VM. Unraveling Molecular and Genetic Insights into Neurodegenerative Diseases: Advances in Understanding Alzheimer's, Parkinson's, and Huntington's Diseases and Amyotrophic Lateral Sclerosis. Int J Mol Sci. 2023 Jun 28;24(13):10809. doi: 10.3390/ijms241310809. PMID: 37445986; PMCID: PMC10341997.
11. Riboldi GM, Frattini E, Monfrini E, Frucht SJ, Di Fonzo A. A Practical Approach to Early-Onset Parkinsonism. J Parkinsons Dis. 2022;12(1):1-26. doi: 10.3233/JPD-212815. PMID: 34569973; PMCID: PMC8842790.
12. Wang H, Guan L, Deng M. Recent progress of the genetics of amyotrophic lateral sclerosis and challenges of gene therapy. Front Neurosci. 2023 May 12;17:1170996. doi: 10.3389/fnins.2023.1170996. PMID: 37250416; PMCID: PMC10213321.
13. Ciurea VA, Covache-Busuioc RA, Mohan AG, Costin HP, Voicu V. Alzheimer's disease: 120 years of research and progress. J Med Life. 2023 Feb;16(2):173-177. doi: 10.25122/jml-2022-0111. PMID: 36937482; PMCID: PMC10015576.
14. Karch CM, Cruchaga C, Goate AM. Alzheimer's disease genetics: from the bench to the clinic. Neuron. 2014 Jul 2;83(1):11-26. doi: 10.1016/j.neuron.2014.05.041. PMID: 24991952; PMCID: PMC4120741.
15. Kunkle BW, Grenier-Boley B, Sims R, et al.; Alzheimer Disease Genetics Consortium (ADGC),; European Alzheimer's Disease Initiative (EADI),; Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE),; Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES),. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 2019 Mar;51(3):414-430. doi: 10.1038/s41588-019-0358-2. Epub 2019 Feb 28. Erratum in: Nat Genet. 2019 Sep;51(9):1423-1424. PMID: 30820047; PMCID: PMC6463297.

16. Vieira RT, Caixeta L, Machado S, Silva AC, Nardi AE, Arias-Carrión O, Carta MG. Epidemiology of early-onset dementia: a review of the literature. Clin Pract Epidemiol Ment Health. 2013 Jun 14;9:88-95. doi: 10.2174/1745017901309010088. PMID: 23878613; PMCID: PMC3715758.

17. Borroni B, Benussi A. Recent advances in understanding frontotemporal degeneration. F1000Res. 2019 Dec 13;8:F1000 Faculty Rev-2098. doi: 10.12688/f1000research.20330.1. PMID: 31885860; PMCID: PMC6913220.

18. Mackenzie IR, Neumann M, Bigio EH, Cairns NJ, Alafuzoff I, Kril J, Kovacs GG, Ghetti B, Halliday G, Holm IE, Ince PG, Kamphorst W, Revesz T, Rozemuller AJ, Kumar-Singh S, Akiyama H, Baborie A, Spina S, Dickson DW, Trojanowski JQ, Mann DM. Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update. Acta Neuropathol. 2010 Jan;119(1):1-4. doi: 10.1007/s00401-009-0612-2. Epub 2009 Nov 19. PMID: 19924424; PMCID: PMC2799633.

19. Neumann M, Mackenzie IRA. Review: Neuropathology of non-tau frontotemporal lobar degeneration. Neuropathol Appl Neurobiol. 2019 Feb;45(1):19-40. doi: 10.1111/nan.12526. PMID: 30357887.

20. Greaves CV, Rohrer JD. An update on genetic frontotemporal dementia. J Neurol. 2019 Aug;266(8):2075-2086. doi: 10.1007/s00415-019-09363-4. Epub 2019 May 22. PMID: 31119452; PMCID: PMC6647117.

21. Benussi A, Premi E, Gazzina S, Brattini C, Bonomi E, Alberici A, Jiskoot L, van Swieten JC, Sanchez-Valle R, Moreno F, Laforce R, Graff C, Synofzik M, Galimberti D, Masellis M, Tartaglia C, Rowe JB, Finger E, Vandenberghe R, de Mendonça A, Tagliavini F, Santana I, Ducharme S, Butler CR, Gerhard A, Levin J, Danek A, Otto M, Frisoni G, Ghidoni R, Sorbi S, Le Ber I, Pasquier F, Peakman G, Todd E, Bocchetta M, Rohrer JD, Borroni B; Genetic FTD Initiative (GENFI). Progression of Behavioral Disturbances and Neuropsychiatric Symptoms in Patients With Genetic Frontotemporal Dementia. JAMA Netw Open. 2021 Jan 4;4(1):e2030194. doi: 10.1001/jamanetworkopen.2020.30194. Erratum in: JAMA Netw Open. 2021 Mar 1;4(3):e217664. PMID: 33404617; PMCID: PMC7788468.

22. Benussi A, Libri I, Premi E, Alberici A, Cantoni V, Gadola Y, Rivolta J, Pengo M, Gazzina S, Calhoun VD, Gasparotti R, Zetterberg H, Ashton NJ, Blennow K, Padovani A, Borroni B. Differences and similarities between familial and sporadic frontotemporal dementia: An Italian single-center cohort study. Alzheimers Dement (N Y). 2022 Jul 25;8(1):e12326. doi: 10.1002/trc2.12326. PMID: 35898667; PMCID: PMC9310192.

23. Hendriks S, Peetoom K, Bakker C, van der Flier WM, Papma JM, Koopmans R, Verhey FRJ, de Vugt M, Köhler S; Young-Onset Dementia Epidemiology Study Group; Withall A, Parlevliet JL, Uysal-Bozkir Ö, Gibson RC, Neita SM, Nielsen TR, Salem LC, Nyberg J, Lopes MA, Dominguez JC, De Guzman MF, Egeberg A, Radford K, Broe T, Subramaniam M, Abdin E, Bruni AC, Di Lorenzo R, Smith K, Flicker L, Mol MO, Basta M, Yu D, Masika G, Petersen MS, Ruano L. Global Prevalence of Young-Onset Dementia: A Systematic Review and Meta-analysis. JAMA Neurol. 2021 Sep 1;78(9):1080-1090. doi: 10.1001/jamaneurol.2021.2161. PMID: 34279544; PMCID: PMC8290331.

24. Loi SM, Pijnenburg Y, Velakoulis D. Recent research advances in young-onset dementia. Curr Opin Psychiatry. 2023 Mar 1;36(2):126-133. doi: 10.1097/YCO.0000000000000843. Epub 2022 Dec 28. PMID: 36705011.

25. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med. 2014 Mar;275(3):214-28. doi: 10.1111/joim.12190. PMID: 24605806; PMCID: PMC3967548.

26. Hughes, C.P., Berg, L., Dan-ziger, W.L., Cohen, L.A. and Martin, R.L. (1982) A New Clinical Rating Scale for the Staging of Dementia. The British Journal of Psychiatry, 140, 1225-1230. http://dx.doi.org/10.1192/bjp.140.6.566

27. Raza C, Anjum R, Shakeel NUA. Parkinson's disease: Mechanisms, translational models and management strategies. Life Sci. 2019 Jun 1;226:77-90. doi: 10.1016/j.lfs.2019.03.057. Epub 2019 Apr 10. PMID: 30980848.

28. Bloem BR, Okun MS, Klein C. Parkinson's disease. Lancet. 2021 Jun 12;397(10291):2284-2303. doi: 10.1016/S0140-6736(21)00218-X. Epub 2021 Apr 10. PMID: 33848468.

29. Olgiati S, Quadri M, Bonifati V. Genetics of movement disorders in the next-generation sequencing era. Mov Disord. 2016 Apr;31(4):458-70. doi: 10.1002/mds.26521. Epub 2016 Feb 22. PMID: 26899883.

30. Basavarajappa BS, Subbanna S. Histone Methylation Regulation in Neurodegenerative Disorders. Int J Mol Sci. 2021 Apr 28;22(9):4654. doi: 10.3390/ijms22094654. PMID: 33925016; PMCID: PMC8125694.

31. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A,

Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita VM, Kaivorinne AL, Hölttä-Vuori M, Ikonen E, Sulkava R, Benatar M, Wuu J, Chiò A, Restagno G, Borghero G, Sabatelli M; ITALSGEN Consortium; Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A, Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011 Oct 20;72(2):257-68. doi: 10.1016/j.neuron.2011.09.010. Epub 2011 Sep 21. PMID: 21944779; PMCID: PMC3200438.

32. Woollacott IO, Mead S. The C9ORF72 expansion mutation: gene structure, phenotypic and diagnostic issues. Acta Neuropathol. 2014 Mar;127(3):319-32. doi: 10.1007/s00401-014-1253-7. Epub 2014 Feb 11. PMID: 24515836.

33. Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. Nat Rev Neurol. 2018 Sep;14(9):544-558. doi: 10.1038/s41582-018-0047-2. PMID: 30120348; PMCID: PMC6417666.

34. Masrori P, Van Damme P. Amyotrophic lateral sclerosis: a clinical review. Eur J Neurol. 2020 Oct;27(10):1918-1929. doi: 10.1111/ene.14393. Epub 2020 Jul 7. PMID: 32526057; PMCID: PMC7540334.

35. Churkina Taran AS, Shakhov AS, Kotlobay AA, Alieva IB. Huntingtin and Other Neurodegeneration-Associated Proteins in the Development of Intracellular Pathologies: Potential Target Search for Therapeutic Intervention. Int J Mol Sci. 2022 Dec 8;23(24):15533. doi: 10.3390/ijms232415533. PMID: 36555175; PMCID: PMC9779313.

36. Andrews SJ, Renton AE, Fulton-Howard B, Podlesny-Drabiniok A, Marcora E, Goate AM. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. EBioMedicine. 2023 Apr;90:104511. doi: 10.1016/j.ebiom.2023.104511. Epub 2023 Mar 10. PMID: 36907103; PMCID: PMC10024184.

37. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, Bras J, Young E, von Coelln R, Simón-Sánchez J, Schulte C, Sharma M, Krohn L, Pihlstrøm L, Siitonen A, Iwaki H, Leonard H, Faghri F, Gibbs JR, Hernandez DG, Scholz SW, Botia JA, Martinez M, Corvol JC, Lesage S, Jankovic J, Shulman LM, Sutherland M, Tienari P, Majamaa K, Toft M, Andreassen OA, Bangale T, Brice A, Yang J, Gan-Or Z, Gasser T, Heutink P, Shulman JM, Wood NW, Hinds DA, Hardy JA, Morris HR, Gratten J, Visscher PM, Graham RR, Singleton AB; 23andMe Research Team; System Genomics of Parkinson's Disease Consortium; International Parkinson's Disease Genomics Consortium. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019 Dec;18(12):1091-1102. doi: 10.1016/S1474-4422(19)30320-5. PMID: 31701892; PMCID: PMC8422160.

38. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 2020 May 18;12(1):44. doi: 10.1186/s13073-020-00742-5. PMID: 32423490; PMCID: PMC7236300.

39. Tudini E, Haas MA, Mattiske T, Spurdle AB. Reporting clinically relevant genetic variants unrelated to genomic test requests: a survey of Australian clinical laboratory policies and practices. J Med Genet. 2023 Jun;60(6):609-614. doi: 10.1136/jmg-2022-108808. Epub 2023 Jan 5. PMID: 36604177.

40. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG; American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013 Jul;15(7):565-74. doi: 10.1038/gim.2013.73. Epub 2013 Jun 20. Erratum in: Genet Med. 2017 May;19(5):606. PMID: 23788249; PMCID: PMC3727274.

41. Miller DT, Lee K, Abul-Husn NS, Amendola LM, Brothers K, Chung WK, Gollob MH, Gordon AS, Harrison SM, Hershberger RE, Klein TE, Richards CS, Stewart DR, Martin CL; ACMG Secondary Findings Working Group. Electronic address: documents@acmg.net. ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2022 Jul;24(7):1407-1414. doi: 10.1016/j.gim.2022.04.006. Epub 2022 Jun 17. PMID: 35802134.

42. Miller DT, Lee K, Abul-Husn NS, Amendola LM, Brothers K, Chung WK, Gollob MH, Gordon AS, Harrison SM, Hershberger RE, Klein TE, Richards CS, Stewart DR, Martin CL; ACMG Secondary Findings Working Group. Electronic address: documents@acmg.net. ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2023 Aug;25(8):100866. doi: 10.1016/j.gim.2023.100866. Epub 2023 Jun 22. PMID: 37347242; PMCID: PMC10524344.

43. Andrews SJ, Renton AE, Fulton-Howard B, Podlesny-Drabiniok A, Marcora E, Goate AM. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. EBioMedicine. 2023 Apr;90:104511. doi: 10.1016/j.ebiom.2023.104511. Epub 2023 Mar 10. PMID: 36907103; PMCID: PMC10024184.

44. Levy S, Guttmann-Beck N, Shweiki D. Clustering Alzheimer's Disease Gene Expression Dataset Reveals Underlying Sexually Dimorphic and Disease Status Profiles. J Alzheimers Dis Rep. 2021 Jun 30;5(1):541-547. doi: 10.3233/ADR-210014. PMID: 34514337; PMCID: PMC8385427.

45. Qiao M, Lee AJ, Reyes-Dumeyer D, Tosto G, Faber K, Goate A, Renton A, Chao M, Boeve B, Cruchaga C, Pericak-Vance M, Haines JL, Rosenberg R, Tsuang D, Sweet RA, Bennett DA, Wilson RS, Foroud T, Mayeux R, Vardarajan BN. Polygenic risk score penetrance & recurrence risk in familial Alzheimer disease. Ann Clin Transl Neurol. 2023 May;10(5):744-756. doi: 10.1002/acn3.51757. Epub 2023 Mar 22. PMID: 36946865; PMCID: PMC10187719.

46. Cammann D, Lu Y, Cummings MJ, Zhang ML, Cue JM, Do J, Ebersole J, Chen X, Oh EC, Cummings JL, Chen J. Genetic correlations between Alzheimer's disease and gut microbiome genera. Sci Rep. 2023 Mar 31;13(1):5258. doi: 10.1038/s41598-023-31730-5. PMID: 37002253; PMCID: PMC10066300.

47. Ibnidris A, Fußer F, Kranz TM, Prvulovic D, Reif A, Pantel J, Albanese E, Karakaya T, Matura S. Investigating the Association Between Polygenic Risk Scores for Alzheimer's Disease With Cognitive Performance and Intrinsic Functional Connectivity in Healthy Adults. Front Aging Neurosci. 2022 May 11;14:837284. doi: 10.3389/fnagi.2022.837284. Erratum in: Front Aging Neurosci. 2023 Mar 23;15:1186490. PMID: 35645768; PMCID: PMC9131016.

48. Pan H, Liu Z, Ma J, Li Y, Zhao Y, Zhou X, Xiang Y, Wang Y, Zhou X, He R, Xie Y, Zhou Q, Yuan K, Xu Q, Sun Q, Wang J, Yan X, Zhang H, Wang C, Lei L, Liu W, Wang X, Ding X, Wang T, Xue Z, Zhang Z, Chen L, Wang Q, Liu Y, Tang J, Zhang X, Peng S, Wang C, Ding J, Liu C, Wang L, Chen H, Shen L, Jiang H, Wu X, Tan H, Luo D, Xiao S, Chen X, Tan J, Hu Z, Chen C, Xia K, Zhang Z, Foo JN, Blauwendraat C, Nalls MA, Singleton AB, Liu J, Chan P, Zheng H, Li J, Guo J, Yang J, Tang B; Parkinson's Disease & Movement Disorders Multicenter Database and Collaborative Network in China (PD-MDCNC). Genome-wide association study using whole-genome sequencing identifies risk loci for Parkinson's disease in Chinese population. NPJ Parkinsons Dis. 2023 Feb 9;9(1):22. doi: 10.1038/s41531-023-00456-6. PMID: 36759515; PMCID: PMC9911746.

49. Grenn FP, Kim JJ, Makarious MB, Iwaki H, Illarionova A, Brolin K, Kluss JH, Schumacher-Schuh AF, Leonard H, Faghri F, Billingsley K, Krohn L, Hall A, Diez-Fairen M, Periñán MT, Foo JN, Sandor C, Webber C, Fiske BK, Gibbs JR, Nalls MA, Singleton AB, Bandres-Ciga S, Reed X, Blauwendraat C; International Parkinson's Disease Genomics Consortium (IPDGC). The Parkinson's Disease Genome-Wide Association Study Locus Browser. Mov Disord. 2020 Nov;35(11):2056-2067. doi: 10.1002/mds.28197. Epub 2020 Aug 31. PMID: 32864809; PMCID: PMC7754106.

50. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, Bras J, Young E, von Coelln R, Simón-Sánchez J, Schulte C, Sharma M, Krohn L, Pihlstrøm L, Siitonen A, Iwaki H, Leonard H, Faghri F, Gibbs JR, Hernandez DG, Scholz SW, Botia JA, Martinez M, Corvol JC, Lesage S, Jankovic J, Shulman LM, Sutherland M, Tienari P, Majamaa K, Toft M, Andreassen OA, Bangale T, Brice A, Yang J, Gan-Or Z, Gasser T, Heutink P, Shulman JM, Wood NW, Hinds DA, Hardy JA, Morris HR, Gratten J, Visscher PM, Graham RR, Singleton AB; 23andMe Research Team; System Genomics of Parkinson's Disease Consortium; International Parkinson's Disease Genomics Consortium. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019 Dec;18(12):1091-1102. doi: 10.1016/S1474-4422(19)30320-5. PMID: 31701892; PMCID: PMC8422160.

51. Makarious MB, Leonard HL, Vitale D, Iwaki H, Sargent L, Dadu A, Violich I, Hutchins E, Saffo D, Bandres-Ciga S, Kim JJ, Song Y, Maleknia M, Bookman M, Nojopranoto W, Campbell RH, Hashemi SH, Botia JA, Carter JF, Craig DW, Van Keuren-Jensen K, Morris HR, Hardy JA, Blauwendraat C, Singleton AB, Faghri F, Nalls MA. Multi-modality machine learning predicting Parkinson's disease. NPJ Parkinsons Dis. 2022 Apr 1;8(1):35. doi: 10.1038/s41531-022-00288-w. PMID: 35365675; PMCID: PMC8975993.

52. Vasilopoulou C, McDaid-McCloskey SL, McCluskey G, Duguez S, Morris AP, Duddy W. Genome-Wide Gene-Set Analysis Identifies Molecular Mechanisms Associated with ALS. Int J Mol Sci. 2023 Feb 16;24(4):4021. doi: 10.3390/ijms24044021. PMID: 36835433; PMCID: PMC9966913.

53. Hop PJ, Zwamborn RAJ, Hannon E, Shireby GL, Nabais MF, Walker EM, van Rheenen W, van Vugt JJFA, Dekker AM, Westeneng HJ, Tazelaar GHP, van Eijk KR, Moisse M, Baird D, Al Khleifat A, Iacoangeli A, Ticozzi N, Ratti A, Cooper-Knock J, Morrison KE, Shaw PJ, Basak AN, Chiò A, Calvo A, Moglia C, Canosa A, Brunetti M, Grassano M,

Gotkine M, Lerner Y, Zabari M, Vourc'h P, Corcia P, Couratier P, Mora Pardina JS, Salas T, Dion P, Ross JP, Henderson RD, Mathers S, McCombe PA, Needham M, Nicholson G, Rowe DB, Pamphlett R, Mather KA, Sachdev PS, Furlong S, Garton FC, Henders AK, Lin T, Ngo ST, Steyn FJ, Wallace L, Williams KL; BIOS Consortium; Brain MEND Consortium; Neto MM, Cauchi RJ, Blair IP, Kiernan MC, Drory V, Povedano M, de Carvalho M, Pinto S, Weber M, Rouleau GA, Silani V, Landers JE, Shaw CE, Andersen PM, McRae AF, van Es MA, Pasterkamp RJ, Wray NR, McLaughlin RL, Hardiman O, Kenna KP, Tsai E, Runz H, Al-Chalabi A, van den Berg LH, Van Damme P, Mill J, Veldink JH. Genome-wide study of DNA methylation shows alterations in metabolic, inflammatory, and cholesterol pathways in ALS. Sci Transl Med. 2022 Feb 23;14(633):eabj0264. doi: 10.1126/scitranslmed.abj0264. Epub 2022 Feb 23. PMID: 35196023; PMCID: PMC10040186.

54. 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68-74. doi: 10.1038/nature15393. PMID: 26432245; PMCID: PMC4750478.

55. Kinoshita J, Clark T. Alzforum. Methods Mol Biol. 2007;401:365-81. doi: 10.1007/978-1-59745-520-6_19. PMID: 18368375.

56. Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol. 2011 Dec;95(4):629-35. doi: 10.1016/j.pneurobio.2011.09.005. Epub 2011 Sep 14. PMID: 21930184; PMCID: PMC9014725.

57. Wang Y, Gao JZ, Sakaguchi T, Maretzky T, Gurung P, Short S, Xiong Y, Kang Z. LRRK2 G2019S promotes the development of colon cancer via modulating intestinal inflammation. bioRxiv [Preprint]. 2023 Jun 30:2023.06.28.546897. doi: 10.1101/2023.06.28.546897. PMID: 37425755; PMCID: PMC10326997.

58. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM, Sperling R; Contributors. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018 Apr;14(4):535-562. doi: 10.1016/j.jalz.2018.02.018. PMID: 29653606; PMCID: PMC5958625.v

59. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):270-279. doi:10.1016/j.jalz.2011.03.008

60. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):263-269. doi:10.1016/j.jalz.2011.03.005

61. Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, van Swieten JC, Seelaar H, Dopper EG, Onyike CU, Hillis AE, Josephs KA, Boeve BF, Kertesz A, Seeley WW, Rankin KP, Johnson JK, Gorno-Tempini ML, Rosen H, Prioleau-Latham CE, Lee A, Kipps CM, Lillo P, Piguet O, Rohrer JD, Rossor MN, Warren JD, Fox NC, Galasko D, Salmon DP, Black SE, Mesulam M, Weintraub S, Dickerson BC, Diehl-Schmid J, Pasquier F, Deramecourt V, Lebert F, Pijnenburg Y, Chow TW, Manes F, Grafman J, Cappa SF, Freedman M, Grossman M, Miller BL. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain. 2011 Sep;134(Pt 9):2456-77. doi: 10.1093/brain/awr179. Epub 2011 Aug 2. PMID: 21810890; PMCID: PMC3170532.

62. Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer JD, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, Grossman M. Classification of primary progressive aphasia and its variants. Neurology. 2011 Mar 15;76(11):1006-14. doi: 10.1212/WNL.0b013e31821103e6. Epub 2011 Feb 16. PMID: 21325651; PMCID: PMC3059138.

63. McKeith IG, Boeve BF, Dickson DW, Halliday G, Taylor JP, Weintraub D, Aarsland D, Galvin J, Attems J, Ballard CG, Bayston A, Beach TG, Blanc F, Bohnen N, Bonanni L, Bras J, Brundin P, Burn D, Chen-Plotkin A, Duda JE, El-Agnaf O, Feldman H, Ferman TJ, Ffytche D, Fujishiro H, Galasko D, Goldman JG, Gomperts SN, Graff-Radford NR, Honig LS, Iranzo A, Kantarci K, Kaufer D, Kukull W, Lee VMY, Leverenz JB, Lewis S, Lippa C, Lunde A, Masellis M, Masliah E, McLean P, Mollenhauer B, Montine TJ, Moreno E, Mori E, Murray M, O'Brien JT, Orimo S, Postuma RB, Ramaswamy S, Ross OA, Salmon DP, Singleton A, Taylor A, Thomas A, Tiraboschi P, Toledo JB, Trojanowski JQ, Tsuang D, Walker Z, Yamada M, Kosaka K. Diagnosis and management of dementia with Lewy bodies: Fourth

consensus report of the DLB Consortium. Neurology. 2017 Jul 4;89(1):88-100. doi: 10.1212/WNL.0000000000004058. Epub 2017 Jun 7. PMID: 28592453; PMCID: PMC5496518.

64. McKeith IG, Ferman TJ, Thomas AJ, Blanc F, Boeve BF, Fujishiro H, Kantarci K, Muscio C, O'Brien JT, Postuma RB, Aarsland D, Ballard C, Bonanni L, Donaghy P, Emre M, Galvin JE, Galasko D, Goldman JG, Gomperts SN, Honig LS, Ikeda M, Leverenz JB, Lewis SJG, Marder KS, Masellis M, Salmon DP, Taylor JP, Tsuang DW, Walker Z, Tiraboschi P; prodromal DLB Diagnostic Study Group. Research criteria for the diagnosis of prodromal dementia with Lewy bodies. Neurology. 2020 Apr 28;94(17):743-755. doi: 10.1212/WNL.0000000000009323. Epub 2020 Apr 2. PMID: 32241955; PMCID: PMC7274845.

65. Al-Chalabi A, Hardiman O, Kiernan MC, Chiò A, Rix-Brooks B, van den Berg LH. Amyotrophic lateral sclerosis: moving towards a new classification system. Lancet Neurol. 2016 Oct;15(11):1182-94. doi: 10.1016/S1474-4422(16)30199-5. PMID: 27647646.

66. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, Obeso J, Marek K, Litvan I, Lang AE, Halliday G, Goetz CG, Gasser T, Dubois B, Chan P, Bloem BR, Adler CH, Deuschl G. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord. 2015 Oct;30(12):1591-601. doi: 10.1002/mds.26424. PMID: 26474316.

67. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. Fly. 2012;6(2):80-92.

68. Wang K, Li MY, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16).

69. Li B, Zhao G, Zhou Q, Xie Y, Wang Z, Fang Z, et al. Gene4PD: A Comprehensive Genetic Database of Parkinson's Disease. Front Neurosci. 2021;15:679568.

70. Kim JJ, Makarious MB, Bandres-Ciga S, Gibbs JR, Ding J, Hernandez DG, et al. The Parkinson's Disease DNA Variant Browser. Mov Disord. 2021;36(5):1250-1258.

71. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. Nucleic Acids Res. 2020;48(D1):D835-D844.

72. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.

73. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.

74. Flanagan SE, Patch AM, Ellard S. Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. Genet Test Mol Bioma. 2010;14(4):533-7.

75. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-5.

76. Shihab HA, Gough J, Mort M, Cooper DN, Day INM, Gaunt TR. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. Hum Genomics. 2014;8:11.

77. Ricardo A Vialle, Katia de Paiva Lopes, David A Bennett, John F Crary, Towfique Raj. Integrating whole-genome sequencing with multi-omic data reveals the impact of structural variants on gene regulation in the human brain. Nat Neurosci. 2022 Apr;25(4):504-514. doi: 10.1038/s41593-022-01031-7.

78. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, et al. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res. 2017;27(11):1916-1929.

79. Ewing B, Hillier L, Wendl MC, Green P. (1998): Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8(3):175-185.

80. Ewing B, Green P. (1998): Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8(3):186-194

81. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550.

82. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50. doi: 10.1073/pnas.0506580102. Epub 2005 Sep 30. PMID: 16199517; PMCID: PMC1239896.

83. Reimand J et al. (2007) g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res 35: W193–W200.

84. Altomare D, Stampacchia S, Ribaldi F, Tomczyk S, Chevalier C, Poulain G, Asadi S, Bancila B, Marizzoni M, Martins M, Lathuiliere A, Scheffler M, Ashton NJ, Zetterberg H, Blennow K, Kern I, Frias M, Garibotto V, Frisoni GB. Plasma biomarkers for Alzheimer's disease: a field-test in a memory clinic. J Neurol Neurosurg Psychiatry. 2023 Jun;94(6):420-427. doi: 10.1136/jnnp-2022-330619. Epub 2023 Apr 3. PMID: 37012066; PMCID: PMC10314086.

85. Barro C, Chitnis T, Weiner HL. Blood neurofilament light: a critical review of its application to neurologic disease. Ann Clin Transl Neurol. 2020 Dec;7(12):2508-2523. doi: 10.1002/acn3.51234. Epub 2020 Nov 4. PMID: 33146954; PMCID: PMC7732243.

86. Ma XT, Shao Y, Tian LQ, Flasch DA, Mulder HL, Edmonson MN, et al. Analysis of error profiles in deep next-generation sequencing data. Genome Biol. 2019;20:50.

87. Shendure J, Ji HL. Next-generation DNA sequencing. Nat Biotechnol. 2008;26(10):1135-45.

88. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440-5

89. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FASTQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013. https://doi.org/10.1002/0471250953.bi1110s43.

90. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440-5.

91. Miller DT, Lee K, Abul-Husn NS, Amendola LM, Brothers K, Chung WK, Gollob MH, Gordon AS, Harrison SM, Hershberger RE, Klein TE, Richards CS, Stewart DR, Martin CL; ACMG Secondary Findings Working Group. Electronic address: documents@acmg.net. ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG).Genet Med. 2023 Aug;25(8):100866. doi: 10.1016/j.gim.2023.100866. Epub 2023 Jun 22.PMID: 37347242