ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

IDENTIFICATION AND CHARACTERIZATION OF *CDNF* IN AN EARLY-ONSET NEURODEGENERATION

AYBIKE SEHRIBAN BULUT

M.Sc. THESIS

DEPARTMENT OF GENOME STUDIES

SUPERVISOR
Asst. Prof. Dr. Kaya Bilguvar

ISTANBUL-2024

ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

IDENTIFICATION AND CHARACTERIZATION OF *CDNF* IN AN EARLY-ONSET NEURODEGENERATION

AYBIKE SEHRIBAN BULUT

M.Sc. THESIS

DEPARTMENT OF GENOME STUDIES

SUPERVISOR

Asst. Prof. Dr. Kaya Bilguvar

ISTANBUL-2024

Department:	Institute of Health Sciences
Program:	Genome Studies
Thesis Title:	Identification and Characterization of *CDNF* in an Early-onset Neurodegeneration
Student's name and Surname:	Aybike Sehriban Bulut
Date of Defence:	/ /

This is to certify that I have examined this copy of master thesis. I have found that she/he prepared after fulfilling the specified requirements in the associated legislations before the final examining committee whose signatures are below.

Jury Member (Head of the Defense)	Signature
Jury Member (Thesis Supervisor)	Signature
Jury Member	Signature

DECLARATION

I declare that this thesis work is my own work, I had no unethical behavior at any stages from the planning to the writing of the thesis, I obtained all the information in this thesis in accordance with academic and ethicel rules, I cited all the information and comments that were not obtained with this thesis work, and I provided resources in the list of references. I also declare that there was no violation of any patents and copyrights during the study and writing of this thesis.

Date

Aybike Sehriban Bulut

(Signature)

PREFACE AND ACKNOWLEDGEMENT

To Prof. Dr. Murat Gunel: Dear Hocam, the first time I saw you, I realized why you are "Murat Gunel". I have learned what personality I need to have to be a successful scientist from you. I am grateful for the opportunity to work in your lab and for your support and contributions to the project I am currently involved in. Thank you for your guidance, mentorship, and taking care of me. It is truly an honor to meet someone of your stature and influence.

To Kaya Bilguvar: I am always touched by your kindness and support. I find your way of making sense of things and your approach intriguing and inspiring. From you, I have learned good mottoes to live by and a great deal of science. I would like to express my gratitude for your trust in me. Thank you for being a loving and dedicated mentor in every way.

To Ozlem Akgun Dogan: For me, you are the embodiment of grace, kindness, and happiness. Thank you for always being there to support me.

To Merve Nur and Ayca: Having you as a friend brings me great joy (short but concise). Merve Nur, your support has helped me more than you know.

I would like to express my gratitude to all the professors and colleagues in the Genome Studies.

I express my infinite gratitude to my beloved parents. You mean the world to me. My dear brother and sister, you have been a great support to me in this life. May we have many beautiful days full of laughter ahead. I extend my thanks to those in my gene pool for their support.

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

ACMG	American College of Medical Genetics and Genomics
NDD	Neurodegenerative Diseases
CDNF	Cerebral Dopamine Neurotrophic Factor
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
RNA-seq	RNA sequencing

LIST OF FIGURES

LIST OF TABLES

ÖZET

Erken Başlangıçlı bir Nörodejenerasyonda CDNF'in Tanımlanması ve Karakterizasyonu

Nörodejeneratif hastalıklar (NDH), sinir sistemindeki nöronal ve nöronal olmayan hücrelerin ilerleyici işlev bozukluğu ve kaybı olarak tanımlanır**.** NDH'nin etiyolojisi karmaşıktır ve tam olarak anlaşılamamıştır; özellikle erken başlangıçlı Mendel formlarında gelişimsel defektlere neden olan genetik mutasyonlar dikkat çekmektedir. Bugüne kadar, tüm ekzom dizileme (WES) öncelikle nadir hastalıklarda sporadik formların yanı sıra Mendel kalıtımı ile hastalığa neden olan genetik varyasyonları ortaya çıkarmak için uygulanmıştır. Bu çalışmaya, akraba bir Türk çiftten doğan ve erken nörodejenerasyon belirtileri gösteren iki kız kardeş dahil edilmiştir. WES verileri analiz edilerek erken başlangıçlı nörodejenerasyonu olan kardeşlerde yeni bir gen olan *CDNF* tanımlanmıştır. Tanımlanan homozigot *CDNF*:c.136C>T anlamsız mutasyonu, ER stres yanıt mekanizmalarının işlevsizliği ile ilişkili hastalıklarda görülen nörodejenerasyon fenotiplerinin çeşitli ayırt edici özelliklerini yansıtmakta ve belirtilen mekanizma ile ilişkili hastalık gruplarının altını çizmektedir. Gelecekteki hastalık modelleme hedeflerimize yönelik olarak *CDNF*:c.136C>T mutasyonunu karakterize etmek için, hasta kaynaklı keratinositlerin birincil kaynak olarak yetiştirilmesine yönelik basit, güvenilir ve optimal bir protokol geliştirdik. Son olarak, yabanıl tip ve *CDNF* nakavt farelerin thalamus, striatum ve korteks bölgelerinden elde edilen toplu RNA-seq verileri analiz edilerek kardeşlerde görülen fenotipin patolojisi hakkında bilgi edinilmiştir. Striatum ve kortekse kıyasla en büyük varyasyon kaynağı talamusta tespit edilmiştir; bu da bozulmuş bazal ganglia ağının bir sonucu olarak motor işlev bozukluğuyla ilişkilendirilebilir. Diferansiyel olarak ifade edilen genler özellikle bağışıklık tepkisi, nörotransmitter alımı/salgılanması ve beyindeki taşıyıcı sistem ile ilgili mekanizmalarda zenginleşmiştir.

Anahtar Sözcükler: Erken başlangıçlı nörodejenerasyon, tüm ekzom dizileme, keratinosit kültürleme, RNA sekanslama

ABSTRACT

Identification and Characterization of CDNF in an Early-onset Neurodegeneration

Neurodegenerative diseases (NDD) are defined as progressive dysfunction and loss of neuronal and non-neuronal cells in the nervous system. The etiology of NDD is complex and not fully understood, with genetic mutations causing developmental defects notably drawing attention, especially in early-onset Mendelian forms. To date, whole exome sequencing (WES) has primarily been applied to reveal disease-causal genetic variations with Mendelian inheritance along with sporadic forms in rare diseases. In this work, two sisters with early signs of neurodegeneration born in a consanguineous Turkish couple were recruited. A novel gene, *CDNF*, was identified in the siblings with early-onset neurodegeneration by analyzing WES data. Identified homozygous *CDNF*:c.136C>T stop-gained mutation reflects the several hallmarks of neurodegeneration phenotypes that is seen in the diseases associated with the dysfunction of ER stress response mechanisms, underscoring the indicated mechanism-associated disease groups. To characterize *CDNF*:c.136C>T mutation for our future disease-modeling goals, we developed a simple, reliable, and optimal protocol for the cultivation of patient-derived keratinocytes as a primary source. Lastly, bulk RNA-seq data from the thalamus, striatum, and cortex regions of wild-type and *CDNF* knockout mice was analyzed, leading us to gain insight into the pathology of the phenotype seen in siblings. The major source of variation was detected in the thalamus compared to the striatum and cortex, which may be associated with motor dysfunction as a result of the impaired basal ganglia network. Differentially expressed genes were particularly enriched in the mechanisms related to immune response, neurotransmitter secretion/uptake, and transporter system in the brain.

Keywords: Early-onset neurodegeneration, whole exome sequencing, cultivation of keratinocytes, RNA sequencing

1 INTRODUCTION AND AIM

1.1 An early-onset Neurodegeneration and Gene Discovery

Neurodegenerative diseases (NDD), characterized by progressive dysfunction and loss of neuronal and non-neuronal cells in the nervous system, pose problems in the overlapping s chiefly mediated by the nervous system. (1,2). Anomalies in these three main functions lead to a lack/ excess/loss of stimuli, disrupted nerve impulses, and the inability of muscles and glands to receive or respond to signals, respectively, ultimately causing disease phenotypes (3). Early onset forms commonly describe the condition where neurodegeneration begins in the prenatal period and progresses throughout an individual's life. In recent years, developmental defects have started to be identified in well-known NDDs that are classically characterized by progressive neuronal dysfunction and death (4). These findings highlight the need for investigating developmental defects, especially in early-onset forms of NDD.

Among about 6000/7000 rare disorders, 85% of which are ultra-rare and the majority of them are Mendelian/monogenic. Our ability to sequence the human genome with high-throughput sequencers to manage and analyze data, combined with a substantial decrease in sequencing cost and time, have led to the discovery of novel genes and increased diagnostic yield in patients with rare diseases. Especially, whole-exome sequencing is progressively becoming a key tool for Mendelian disorders with the diagnosis yield ranging between 25-50% and is widely implemented by clinicians (5,6).

Although the etiology of NDD is not fully understood, it is evident that rare genetic variations with Mendelian inheritance play a significant role in the development of degenerative conditions in the nervous system. Our primary aim is to elucidate the genotype-phenotype relationship in the identified family. In accordance with our aim, we firstly reported two sisters with early-onset neurodegeneration, born to a consanguineous Turkish couple, with homozygous rare *CDNF* variant defined as a disease-causing mutation, detected upon WES analysis.

CDNF, like other neurotrophic factors, protects and restores neurons, but it differs from others in terms of its structure and mechanism of action. It has been proposed that it acts as a modulator in the UPR system activated by ER stress and an inhibitor for inflammation and apoptosis (7,8). Additionally, the neuroprotective and neurorestorative effects of CDNF have been extensively observed, particularly as a therapeutic option in animal models of Parkinson's, Huntington's, Alzheimer's diseases, and amyotrophic lateral sclerosis (7,8,9,10,11,12).

1.2 Functional Genomics: Transcriptomics

Today, many projects are being launched both to sequence thousands of genomes and to examine other facets of the genome raising the questions such as how/which genes are differentially expressed, how the genome is organized in three dimensions, which transcription factors are involved, what the epigenetic changes are, how/which proteins interact, etc. These questions are addressed using high-throughput sequencing-based methods, within the realm of functional genomics, such as RNA sequencing, chromatin immunoprecipitation with sequencing (ChIP-seq), Hi-C and its derivatives for 3D genome analysis, bisulfite-seq, ATAC-seq and DNase-seq to assess chromatin accessibility (13). RNA sequencing (RNA-seq) has become a crucial method for full transcriptome analysis (not just a few selected transcripts). RNA-seq allows the identification of changes in gene expression and the characterization of various non-coding RNA forms. The architecture of transcriptome such as gene fusions, single nucleotide variants, transcript isoforms, and other unknown features can be detected by using RNA-seq. Especially for quantifying gene expression levels across the transcriptome, it delivers successful profiling of genes with low expression levels, revealing genes with differentially expressed. Moreover, the advancements of RNA-seq to spatial, single-cell, and single molecular transcriptome have furthered our understanding in single-cell and morphological levels. The data generated by RNA-seq can provide valuable information about gene networks and pathways associated with cellular and disease mechanisms (13,14,15). For instance, the presence of developmental problems in Huntington's disease has been substantiated by profiling differentially expressed genes between HTT mutant glial progenitor cells derived from

embryos and the control group by performing RNA-seq analysis (4). Another recent study showed distinctive dysregulation of long-non-coding RNA (lncRNAs) transcripts in Parkinson's patients and control group by using RNA-seq following pathway analysis, indicating the role of lncRNAs in disease pathology (16). In this context, our secondary aim is to analyze CDNF knockout and wild-type mouse models at the transcriptome level, acquiring information that will help elucidate the role of our candidate gene in the disease pathology of patients.

1.3 Establishment of Primary Keratinocyte Culture for Human Hair Follicle Cells

In recent years, the approach of using patient-specific induced pluripotent stem cells(ipsc) for disease modeling has significantly increased. However, the invasiveness in other methods such as skin biopsy pose a significant obstacle in patients. Using keratinocyte cells derived from hair follicles as initial somatic cells overcomes this challenge, as sample collections can be conducted in simple and non-invasive manner (17,18). Moreover, keratinocytes are originated from the same neuroectodermal embryonic layer similar to neurons. Using patient derived keratinocytes also allows the disease models generated from these initial cells to have the same genetic and epigenetic background with the patients (19). Furthermore, they can be cultured in feeder-cell and serum-free low-calcium media, preventing other cell contaminations. This method also enables rapid proliferation and resistance to apoptosis after viral infection. Compared to fibroblasts, they require more attention, and challenges such as apoptosis at low density become more pronounced, and issues related to differentiation and senescence become more evident upon reaching confluence. However, keratinocytes remain a valuable source of highly proliferating epithelial cells (20). In this study, as our tertiary aim, we developed a reliable, standard, and optimized method for culturing hair-derived keratinocytes and establishing a cell line, to continue our further objectives (generation of organoid models from ipsc).

2 BACKGROUND

2.1 High-throughput Sequencing and Rare Disease Research Development

2.1.1 Progression in DNA Sequencing: From Early-generation to High Throughput Sequencing

Commencing with the discovery of DNA structure, great advancements have unfolded in understanding the intricacies of the genome in diseases (21,22). The idea of revealing the complexity of the human genome has brought multitude of explorations in instruments and reagents for DNA sequencing. Early DNA sequencing methods were limited by efficiency, cost, and time. In 1968, Wu introduced primer extension methods determining 12 bases of bacteriophage lambda (23). Subsequently, Gilbert and Maxam sequenced 24 bases of the lactose-repressor binding site by copying the site into RNA and sequencing fragments, taking two years equivalent to one base per month, in 1973 (24). Two methods developed by Sanger-Coulson, and Maxam-Gilbert, in around 1976, have enabled decode of hundreds of bases by immediately reading the sizes of fragments present on polyacrylamide gel by putting it onto X-ray film (25,26). Gilbert's chemical cleavage method was based on the base-specific cleavage of nucleotides, following radiolabeling of DNA. This was the initial approach widely embraced and could be deemed as the real birth of the first-generation sequencing. However, Sanger's chain termination method came in 1977 and was the breakthrough. Sanger's chain termination method consisted of four reactions each with a trace amount of different radiolabeled chain terminating dideoxide nucleotide, thereby producing fragments in different sizes. While working on the same mechanism based on measuring the sizes of DNA fragments by electrophoresis as in the other techniques, Sanger's chain termination method persisted for years due to its ease of use, accuracy, and robustness. Nevertheless, these methods were insufficient for the sequencing of longer DNA. In 1979, another strategy called Shotgun sequencing - random sequencing of cloning vectors carrying fragmented DNA followed by sequence reads assembled into contigs to establish longer sequence *de novo*- was suggested by Staden (27). By 1985-1990, improvements that the replacement of

radiolabeling with fluorescence-based detection and the usage of capillary-based electrophoresis contributed the development of automated, fluorometric-based Sanger Sequencing machines generating 1,000 bases per day empowering the growth of sequence data (28).

All the accumulated data and innovations in DNA sequencing during this timeframe supported the emergence of the Human Genome Project (HGP) started in 1990 (29). The dominant strategy in the HGP was the hierarchical shotgun method, large DNA fragments cloned into bacterial artificial chromosomes (BACs) followed by sequencing of BAC-derived shotgun library with automated-developed Sanger machines to establish contiguous sequence. However, challenges in wet-lab processes and software, the need for each independent step to work effectively to eliminate doubt -the quality-, and the need for reasonable cost and time have led to noteworthy improvements in this field after 1990. HGP Consortium and Celera led by Craig Venter synergistically exerted efforts to sequence the human genome by tracing shotgun strategy. These two groups differed strategically and in 2001 they published the first human genome draft simultaneously; however, Celera's strategy, while reasonable, did not result in as high-quality reference as in HGP's clone-based approach. In connection with these processes, the current human reference sequence has its origin from the achievement of HGP in 2004 (30, 31).

During the nearly two decades following HGP's completion, the release of high throughput sequencing (HTS) machines has enabled to sequence large human genome in cost effective, time-efficient, and high-quality manner. In the mid 2000s, an innovation in the field named as 'massively parallel sequencing' or 'next generation sequencing (NGS)' replaced Sanger sequencing. The general term for sequencing techniques emerging in the mid 2000s referred as HTS fall under two categories: 'short read sequencing' and 'long read sequencing' (22,32).

Short read sequencing incorporates cycles of biochemical processes prior to imaging, overall, under two approaches called 'sequencing by ligation' (not widely adopted) and 'sequencing by synthesis, SBS'. In SBS, DNA polymerase and modified

dNTPs are used to get sequence readout, eventually giving a signal by incorporated to the strand. The key focus of short read HTS in SBS approach is to multiply fragmented DNA (35-700bp) to generate thousands of identical DNA fragments and to sequence each clonal clustered template in parallel. For amplification, there exist three different strategies, which are not restricted to a sole supplier; 'solid-phase amplification (Illumina)', 'bead-based emulsion PCR (454, Roche; Ion Torrent and SOliD, Thermo Fisher; GeneReader, Qiagen)' and 'in-solution DNA nanoball generation (Complete Genomics, BGI)' (32). For template amplification's first step, fragmented DNAs are ligated to the common adapter to clonally amplify the template on three amplification platforms. Solid-based amplification uses the technique that template DNAs ligate to complementary primer immobilized on the surface of the slide enabling the encoded position for each cluster. For Illumina's solid-based bridge amplification, the ends of hybridized templates bind to nearby primers covalently bound to the solid support to generate 100-200 million clonal clusters. SOLiD's solid-based sequencer, on the other hand, uses the 'template walking' technique instead of using the amplification of DNA templates on the solid support. In the bead-based system, each fragmented template DNA is immobilized on a bead by hybridizing to bead-attached primer. Then, template DNAs on different beads are amplified by using emulsion PCR (32,34,22). The solution-based amplification operates the method that enriches the template in solution. Four adapters iteratively are ligated to a fragmented DNA template, followed by rolling circle amplification in solution to form circular concatemers (DNA nanoballs), thereafter generated nanoballs are distributed and immobilized on flow cells (32, 36, 35, 22).

In addition to the amplification methods of suppliers, the sequencing readout techniques also vary with three main strategies. The pyrosequencing approach, released pyrophosphate starts the cascade reaction generating light upon incorporation of complementary dNTP by DNA polymerase, is based on SBS. Only when one dNTP is added at a time to the platform, in which amplified-immobilized DNAs is located (37, 38, 32). Another strategy is based on the detection of fluorescent signals from hybridized fluorescently labeled dNTPs by DNA polymerase. In the third approach,

the specificity of DNA ligase is employed to bind fluorophore oligonucleotides to the template, overall called sequencing by ligation (39, 40, 32).

The newest approach in the HTS landscape is long-read sequencing. This approach enables to sequence of larger DNA ranging in size from 1,000 and 20,000 bases rather than the short read sequence. It is well known that the genome has complex and highly repetitive regions such as complex structural variants, transposable elements, and centromeric regions (41, 42, 43). Short read sequencing is limited to span and analyze these regions. Long-read sequencing can encapsulate such complex and repetitive regions by generating a single-long continuous read. Moreover, long read sequencing delivers an opportunity to detect base modifications due to remaining in its native state in contrast to short read sequencing. Since sample preparation does not need amplification via PCR, it also prevents PCR bias due to artifacts (32). Two main strategies have been available for long-read sequencing; 'single-molecule real-time sequencing, SMRT seq (PacBio, Oxford Nanopore Tech)' (44) and 'generation of long read construct from existing short reads, (Illumina Synthetic Long Read, 10X Genomics)' (32,44). PacBio immobilizes the DNA polymerase to the zero-mode waveguide wells and allows the DNA polymerase to move along the template DNA (45). As one of the labeled nucleotides is hybridized to DNA template, emitted light upon incorporation of one of four labeled nucleotides is recorded by the camera. Oxford nanopore utilizes the leader sequence interacting with both motor protein and nanopore directing the template DNA into the nanopore (46). When the DNA template is tethered to a nanopore with the help of motor protein, the ion current provides the movement of the template DNA to pass through the nanopore followed by the sequencing of DNA according to a shift in voltage characteristic to the specificity of the nucleotide. Unlike real-long template DNA sequencing, synthetic long-read sequencing generates long reads from existing short fragments. Illumina's synthetic long-read method starts with the distribution of fragmented DNA (8-10kb) into a microliter plate, ensuring approximately 3,000 templates per well. Subsequently, each template is further fragmented into 350 base pairs in the microwells, and each well is barcoded with a single barcode. All templates are then pooled and processed through short-read pipelines. Barcoded pools generate a synthetic long-read assembly.

In 10X Genomics, similar to Illumina's long read method, short fragments of DNAs are generated to construct long DNA. Large fragments up to 100kb, first, are partitioned into micelles including unique barcode per micelle, dNTPs, and enzyme (GEMs). Fragments within the GEMs are further fragmented into 350bp and barcoded. After sequencing, aligned reads are connected to form a sequence of anchored fragments spanning 50kb, in other words, reads with the same barcode are assembled back into larger fragments. This method does not get end-to-end coverage of the full fragment. Instead, overall coverage is obtained from GEMs fragments with dispersed yet interconnected reads (32).

Over the past 15 years, HTS methods have enabled a wide range of applications with varying accuracy, time, and cost. The widely used HTS methods based on SBS, in addition to their advantages, also incorporate their shortcomings. Just like in all fields, to eliminate the shortcomings in this area, sequencing technologies continue to advance incessantly. This has led to the emergence of new methods. Avidity sequencing, the latest among these technologies, was introduced to the market by Element Biosciences in 2023 (47). In SBS, efficient incorporation of labeled nucleotide necessitates a high molar concentration of nucleotide. Allowing a longer time for polymerization requires extended cycles and times. With Avidity sequencing, incorporation of nucleotides is controlled and optimized although the method uses very low concentration of nucleotides by using avidity for nucleotide detection. As in the nearly same with methods previously mentioned, template DNA fragments are circularized and bounded to the surface of the flow cell. Each template in the flow cells then is amplified with rolling circle amplification, generating concatemers, however, unlike the nanoball methods, DNA templates are amplified on the surface of the platform, forming the polonies. It's another differential approach is to use multivalent avidite, a polymer with dye and identical nucleotides attached. DNA polymerase, along with the mixture of four avidities, is added into flow cell and employed for base discrimination. The avidite, additionally, does not incorporate to DNA but form stable complex rather than other labelled nucleotides. A single avidite also interacts with many DNA copies within the polonies. In consequence, the use of avidity sequencing, overcomes the problems in the incorporation of monovalent labelled nucleotide and

increase the base quality(>Q40). Innovation of avidity sequencing have paved the way for the emergence of a new subclass of sub-sequencing; so, called ‘sequencing by chemistry’ (47).

2.1.2 Rare Disease Development and Utilization of High-Throughput Sequencing

2.1.2.1 Rare Disease Development

Rare diseases are heterogeneous and numerous conditions affecting 1 in 10 people in the US and 1 in 2,000 people in Europe, varying countries’ own official definition of a rare disease. There are approximately more than 6,000 in the EU and 7,000 rare diseases in the US affecting up to 30 million people. Of the rare diseases, approximately 80% have genetic origins, among those, 70% manifest during childhood (48, 49). While some diseases have little impact on an individual's life, others cause conditions, impairing normal functioning, and can even result in death. The processes in the diagnosis, treatment, and therapy of individuals with rare disorders pose a distinctive challenge for patients, their families, and the healthcare system. The challenges arising from their low prevalence, including limited knowledge, their chronic, degenerative, and life-threatening characteristics, and a shortage of expertise have positioned rare diseases as a priority in the EU and US. In this regard, numerous national and international institutions and organizations provide support and guidance to patients and their families (50, 51,52,53).

The International HapMap Project, between 2003-2010, was aimed to decipher genetic variations in human genome to relate these variations with diseases. These variations, can be combinations of alleles/segments or single nucleotide polymorphisms, are often inherited together, called haplotype. The hypothesis that limited sets of SNPs would be adequate to detect relevant haplotypes in any population led to HapMap project (54,55,56). In pursuit of this goal, the project was conducted in two main steps: generation of SNP data through SNP arrays and grouping SNPs by linkage disequilibrium (57, 58). The question that comes to mind in this project is why

they dominantly genotype SNPs instead structural variants such as CNVs, insertions, deletions, inversions. The reason for this is 1) the structural variants are not as efficiently detected by genotyping technologies as SNPs, and 2) these variants are not as common as SNPs. Right at this point, the presence of rare variants in these studies comes to mind, which will be explained later.

After the completion of the HapMap Project, genome-wide association studies (GWAS), investigating the impact of common variants on common diseases, become a more feasible approach (59). Since GWA genotyping arrays capture frequent variants, rare variants with larger effect sizes are hard to detect and they do not possess a large effect size enough to be identified by linkage analysis in family studies (Figure 1). In this context, that the identification of rare variants (>0.005 MAF) with HTS technologies, is a more powerful strategy can be concluded (60,61). This information has empowered the use of HTS technologies, for the detection of both SNP and non-SNP variants. The 1000 Genomes Project, The Genome Aggregation Database and Exome Aggregation Consortium have sequenced genomes and exomes from various populations, creating a comprehensive dataset and increasing the accuracy of allele frequency in populations (62,63,64). These projects have accelerated the development and studies in this field.

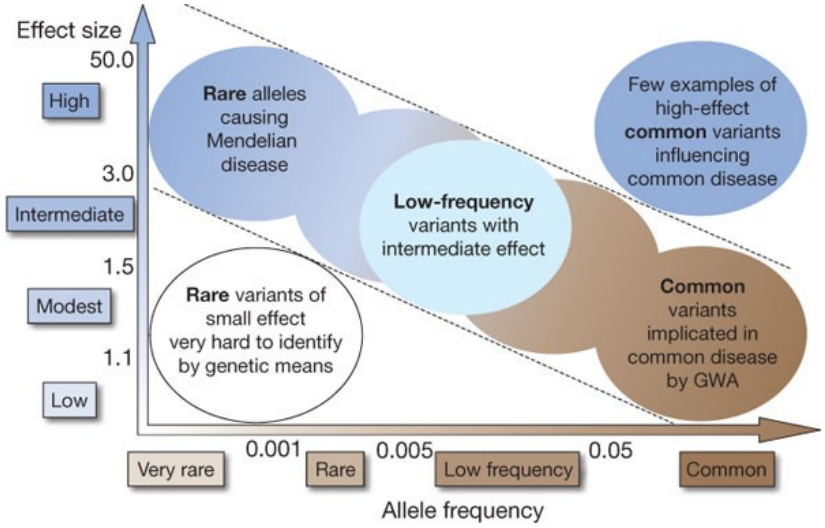


Figure 1. Architecture of diseases. The possibility of detecting genetic variants based on the frequency of risk alleles and the magnitude of the genetic effect (61).

2.1.2.2 Utilization of High-Throughput Sequencing in Rare Diseases

Understanding the genetics and molecular background of rare diseases (orphan diseases) is a unique challenge. With the advent of HTS technologies, they are now a routine part of the clinical system. Tens of thousands of genomes can be sequenced rapidly and at a low cost with HTS technologies. Whole exome and whole genome sequencing have been used to discover new genes and variations to detect and interpret genetic causes in Mendelian and sporadic diseases, increasing the diagnostic yield.

2.1.2.2.1 Whole Exome and Whole Genome Sequencing

Whole exome sequencing (WES) and whole genome sequencing (WGS) are two subtypes of HTS sequencing. While WES captures all the exons of known genes, WGS enables us to analyze entire human DNA including both coding and non-coding regions. WGS has the capability of detecting variants in non-coding regions and structural variations such as CNVs (deletions, duplications, multiallelic CNVs) and balanced structural variants (insertions, inversions) in comparison with WES. To perform WES, specific probes designed to capture and enrich the exomes are used. The targeted sequencing approach in WES stands out from the WGS due to lower cost, simplification of data analysis, and storage (65,66).

WES and WGS also differ in terms of coverage, read depth and accuracy of variant calling. WGS offers lower and uniform coverage rather than WES. Most popular exome kits -SureSelect, MedExome, Nextera Rapid, Truseq Exome- exhibit less overall even coverage compared to WGS, with no significant unevenness, meaning that unevenness highly varies between coding regions in WES (67,68). For the well-regarded WES capture kit (SureSelect), 1180kbp are estimated to have low coverage at 100X, and at 200X, this drops to 970kbp. For 30X WGS, 788kbp coding sequences are predicted to have coverage less than 10X. This profile reflects that both WGS and WES have coverage bias, despite WGS performing better. Additionally, this brings with a different conclusion. WGS delivers more accurate variant calling in comparison with WES due to its uniform coverage and less coverage bias (67,68,69).

2.2 Neurodegenerative Diseases: Maintenance of the Nervous System

2.2.1 An Overview

Neurologic disorders are disease groups affecting neurons throughout the body and non-neuron cells in central&peripheral nervous systems. More precisely, neurological development begins in the embryonic period, continues through the adolescent period, and ends in adulthood. Maintenance of the system continues throughout human life. The developmental and maintenance processes rely on the protection of neural integrity/morphology/function and the establishment of accurate synaptic communication. Monogenic/familial, sporadic/complex, and environmental reasons, affecting extensive range of molecular and cellular mechanisms and pathways, can disrupt the normal functioning of the neurons resulting in variety of common or variable neurological symptoms/phenotypes (70,71).

2.2.2 Neurodegenerative Diseases

Neurodegenerative diseases (NDD) are characterized by the progressive dysfunction or death of neurons. Although it is commonly understood that degeneration of neurons, may ultimately cause death, plays a role in defining NDD

pathology, neuronal cell death is not exclusively mediated by mechanisms inherent to the cells themselves. Cells of the central nervous and peripheral nervous system and even their progenitors contribute the initiation and progression of neurodegeneraiton depending on their roles (72,73). This situation represents a novel awareness in contrast to the common definition of neurodegenerative diseases. NDD includes broad range of conditions such as Alzheimer's disease, Huntington's disease, Parkinson's disease, motor neuron disease, amyotrophic lateral sclerosis, multiple sclerosis, Batten disease frontotemporal dementia, prion disease etc. Overall, typical physiological signs of NDD are demyelination, dendrite and axon abnormality and loss.

2.2.2.1 Delving into the Etiology of Neurodegenerative Diseases

Genetic, epigenetic, and environmental factors lie reasons at the core of neurodegeneration. The genetic factors underlying Mendelian and sporadic forms impair the functioning of complex molecular biological processes, as in the case of other factors. Apoptosis, impaired organelle function, problems in external compartments that ensure neuron health, disrupted protein structures, and cellular stress factors such as dysfunction in protein regulation and homeostasis, oxidative stress, inflammation contribute to the development, onset, and progression of NDD (74,75).

The cause of this group of diseases is not always a single genetic factor. Multiple genes and their interactions may be involved in this process. The presence of rare mutations may tend to increase susceptibility to or be a causative factor in NDD. As many genes have been cataloged in the monogenic form of NDD with Mendelian inheritance, some genetic loci, and variants for sporadic NDD have been associated (Table 1) (76).

Table 1. Disease causing genes associated with certain NDD (76).

Disease Name	Disease Causing Genes (OMIM)
Alzheimer's disease	*PSEN1, PSEN2, APOE, NOS3, PLAU, MPO, APP*

Table 1. Disease causing genes associated with certain NDD (76) (continued).

Parkinson's disease	*PINK1, SNCA, PRKN, DNAJC6, DJI, GIGYF2, EIF4G1, SYNJ1, TMEM230, CHCHD2, VPS13C, RIC3, PTPA*
Multiple Sclerosis	*MS4, MS3, MS2, PDCD1, TNFRSF1A*
Huntington's disease	*PRNP, HTT, JPH3*
Frontotemporal dementia	*PSEN1, MAPT, C9orf72, CHMP2B, SQSTM1, VCP, TBK1, CCNF, FUS, CYLD, CHCHD10*
Amyotrophic lateral sclerosis	*TRPM7, TARDBP, TIA1, DCTN1, ALS2, TUBA4A, ERBB4, MATR3, SQSTM1, FIG4, LRP12, C9orf72, SPTLC1, SETX, OPTN, ANXA11*

** Not all the associated genes are listed. Some genes do not represent direct disease cause, reflecting susceptibility conferred by mutation.*

Genetic findings are related to pathophysiological of NDD. The hallmarks of neurodegeneration are considered in four pathological events (74). 1) Intracellular and extracellular deposition of proteins and sequential series of response to protein aggregates. For instance, in the familial form of Alzheimer's disease, mutation in *PSEN1,* encoding the subunits of γ secretase, disrupts the γ secretase complex, leading to change in the processing of amyloid precursor protein (APP). Upon change in the pattern of processing APP, increased Aβ42/Aβ40 ratio causes beta amyloid plaques (77,78). Mutations in *SCNA* associated with Parkinson's disease, encoding alpha-synuclein, cause the aggregation of monomer alpha-synuclein growing into filaments, driving cell death. This extracellular deposition of beta amyloid plaques and intracellular alpha-synuclein filaments can start inflammation and impair synaptic function by accumulated in synaptic area, thereby damage the neuron itself and surrounded neurons (79,80). Additionally, inhibition of ubiquitin proteosome system, which is proteolytic system that maintain protein homeostasis and regulates DNA repair, stress response interacting in coordinated manner with unfolded protein response and ER-associated degredation, have been implicated in the neurodegeneration. 2) Mitochondrial dysfunction. Mutations or other influences that impact the function of mitochondria, leading to mitochondrial dysfunction, are a fundamental factor in the development of these diseases (81,82,83). Mutations in the *PINK1 and PARK2* -related to early onset parkinsonism- are responsible for disruption PINK1/PARKIN signaling, resulting in uncontrolled mitophagy and mitochondrial

quality control pathways. Impaired mitophagy and disrupted PINK1/PARKIN signaling are also observed in neurodegenerative conditions such as Alzheimer's, Huntington's diseases, and amyotrophic lateral sclerosis (84). 3) Chronic inflammation. Neuroinflammation can be observed concurrently with degeneration or as a response. Commonly, immune response functions as a beneficial mechanism by repairing the tissue, eliminating pathogens, and restoring tissue balance. Factors such as protein aggregates, impaired down-stream pathways and mutations in immune-coding genes contribute to disrupted immune response, resulting in neurotoxicity and promotion to degeneration. *TREM2* is highly expressed in mycroglia and functions as a reducing factor of constitute production of inflammatory cytokines and an activator for immune cells. Homozygous *TREM2* mutations -associated with frontotemporal dementia- both increase the secretion of cytokines and weakens the binding capability of TREM2 which hinders migration and proliferation of microglia (85). Which, in turn, leads to a chronic immune response and the malfunction of this mechanism. In another scenario, the immune system chronically responds to clear protein plaques, thereby increasing inflammation that contributes to degeneration. 4) Oxidative stress. Another contributor of pathophysiology of neurodegeneration is oxidative stress when counterbalance of oxidants/free radicals is insufficient (74). In pathological circumstances, production of free radicals can be generated excessively, leading to detrimental effects. *SOD1* is responsible for coding superoxide dismutase 1, a primary cytoplasmic antioxidant enzyme that transforms superoxide radicals into molecular oxygen and hydrogen peroxide. In doing so, it acts as a defense mechanism against oxygen toxicity. Mutations in *SOD1* are linked to amyotrophic lateral sclerosis cause either loss of function or gain of function of SOD1 (86). There exist two potential mechanisms. Reduced SOD1 activity causing the buildup of harmful superoxide radicals or increased SOD1 activity resulting in elevated levels of hydrogen peroxide and an extremely toxic hydroxyl radical. The latter can be formed when hydrogen peroxide reacts with a transition metal like iron. As schematically illustrated in the Figure 2, oxidative stress not only leads the cell to degeneration on its own but also contributes to it by influencing other mechanisms collectively. Simultaneously, the presence of other mechanisms also generates oxidative stress within the cell, forming a mutually reinforcing cascade of events (87,88).

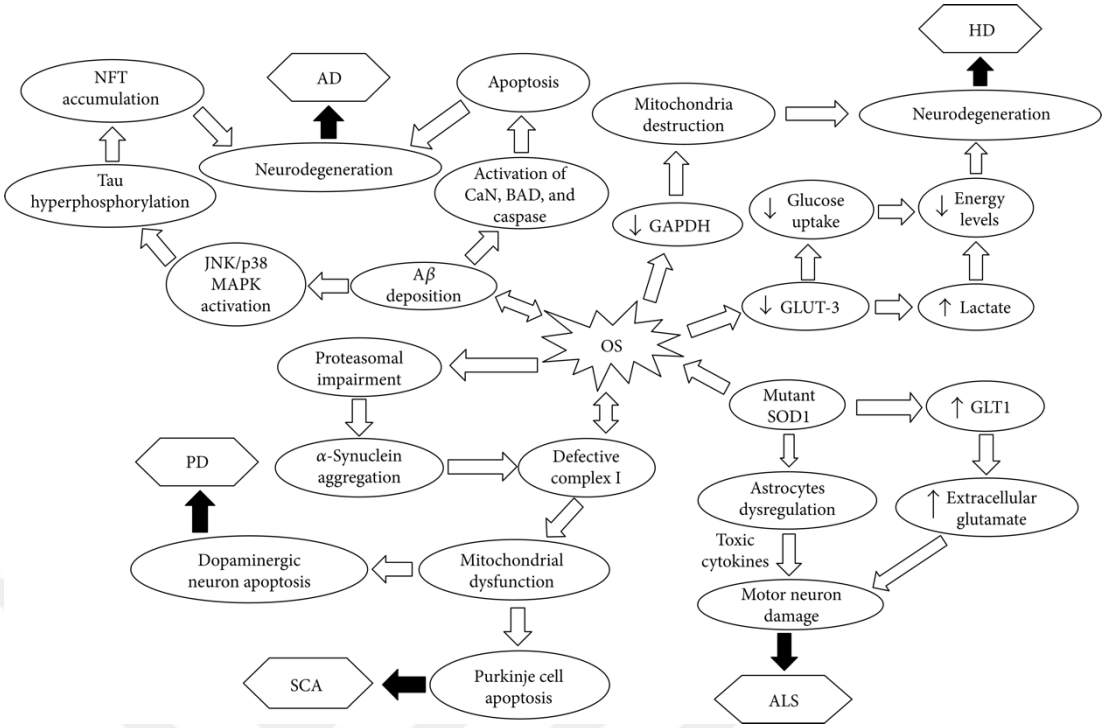


Figure 2. Schematic representation of oxidative stress interplays contributing neurodegeneration (88).

2.3 Cerebral Dopamine Neurotrophic Factor and Unraveling Its Intriguing Mechanism

2.3.1 Neurotrophic Factors

Neurotrophic factors (NF) are small peptides and small proteins playing crucial role in promoting growth, morphology, survival, and differentiation of neurons, both during development and in mature stages, in both the central and peripheral nervous systems. Neurotrophic factors are classified under main families built upon different cellular mechanisms; neurotrophins (nerve growth factor (NGF), brain derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3) and neurotrophin-4 (NT-4)), ciliary neurotrophic factors (CNTF), glial cell line derived neurotrophic factor (GDNF), mesenchymal astrocyte-derive neurotrophic factor(MANF), cerebral dopamine neurotrophic factor (CDNF), transforming growth factor (TGF)/ epidermal growth factor(EGF) family and others (89,90, 91).

2.3.2 Cerebral Dopamine Neurotrophic Factor (CDNF)

2.3.2.1 The Structure of CDNF

CDNF encodes cerebral dopamine neurotrophic factor, a protein differs in other neurotrophic factors in terms of homology and mechanism of action. It shares 59% homology with MANF; however, it is not homologous to other NFs.

CDNF/MANF family is a compact and soluble protein (18-20 kDa and 158/161aa). The structure reveals a globular N-terminal and unstructured C-terminal, linked by a flexible bridge. N terminal domain also consists of three disulfide bridges which is analogous with a saposin-like protein (7). Bai and others showed MANF interacted with sulfatide is found extracellular leaflet of cell membrane, facilitating MANF uptake to cell (92). Sulfatide, a lipid, is abundantly expressed in myelin sheath of oligodendrocytes, Schwann cells while it is expressed to a lesser extent in neurons and astrocytes. Despite homologous to MANF, CDNF cannot interact with sulfatide as it lacks the interacting MANF part. This can be attributed to the substitution of the conserved lysine at 122nd residue in the MANF human protein with leucine in CDNF. Another characteristic of CDNF/MANF is to have eight cystine residues which is important for protein folding and stabilization. Helix-loop-helix domain in C terminal of CDNF/MANF represents similar pattern with SAP (SAF-A/B, Acinus, and PIAS protein superfamily) with neuroprotective activity. Both proteins have also CTXX motifs, located in many proteins with antioxidant properties; however, they do not show oxidoreductase property. This motif provides MANF to function as to enable correct protein folding in endoplasmic reticulum (ER) (7,8). Given ER resident proteins have a lysine-aspartic acid-glutamic acid-leucine (KDEL) sequence, it is seen that CNDF/MANF also carry KDEL-like ER retention signal in the C terminal of CDNF/MANF proteins (Figure 3) (7).

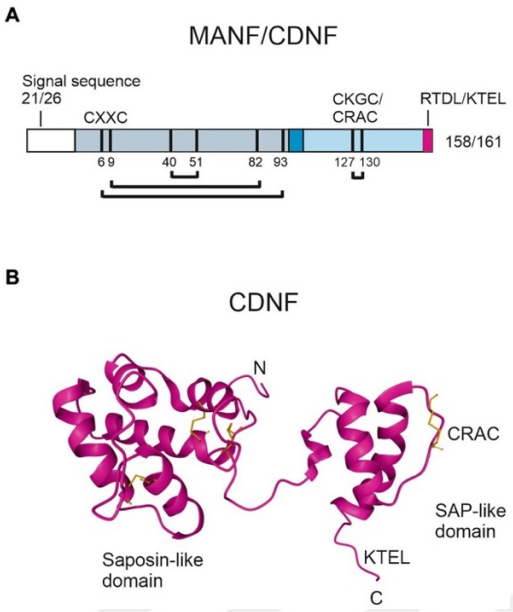


Figure 3. Structure of CDNF&MANF proteins. A) Primary structure of CDNF&MANF proteins. Gray: N terminal sposin like domain, light blue: SAP like domain, black bars: eight cysteine domains, pink: KDEL like ER retention signal, connecting lines: four disulfide bridges between cysteines. B) Nuclear magnetic resonance spectroscopy solution structure of CDNF. Yellow: disulfide bridges (7).

2.3.2.2 ER Stress and Novel Mechanism of CDNF

The endoplasmic reticulum (ER) functions as a comprehensive quality control machine for proteins. It controls protein folding, degradation and homeostasis and is also maintaing calcium homeostasis, protein synthesis, lipid, and carbohydrate metabolism. ER has defense mechanism under stress conditions. Stress conditions, such as aggregation of misfolded/mutant/exogenous viral proteins, imbalance in calcium stores, metabolic dysfunction, presence of environmental toxins, and perturbation of ATP levels, cause saturated protein folding capacity. The ER recognizes this as a stress condition, thereby initiating the UPR defense system, alleviating ER stress. Normally, there is the ERAD (ER-associated degradation) system for the clearance of misfolded proteins. This system transports misfolded proteins to the ER, and they undergo degradation in ubiquitin proteasome with a series of cascades. Persistent misfolded proteins are directed to the UPR. The UPR system protects the cell from ER stress by transiently suppressing mRNA translation decreasing protein synthesis and by increasing expression chaperons mediating protein

folding. It also contributes ERAD in the cytosol. While UPR is homeostatic mechanism, in cases of severe and unresolved ER stress, UPR signaling become persistent/chronic, resulting in the activation of UPR-related apoptotic pathways. Chronic UPR activation can eventually cause death of cells, thereby playing a role in the development of varying pathologies, including neurodegenerative diseases (93, 94).

Three receptors are activated upon ER stress: Ire1α (Inositol-requiring enzyme 1), Perk (protein kinase R-like ER kinase) and ATF6 (activating transcription factor 6). Bip (GRP78), which is normally bound to three UPR receptors, dissociates from receptors and binds to unfolded proteins, leading to activation of UPR. Upon Bip binding to Ire1α, it heterodimerizes, activates kinase domain and trans-autophosphorylates itself, resulting in stimulated cytosolic RNase domain, respectively. XBP1 mRNA is spliced to XBP1s by activated RNase domain. This transcription factor induces of the genes of ER quality control, lipid synthesis and ERAD. Apoptotic signaling and inflammatory response are mediated by IRE1α adapter TRAF2. Upon Perk activation, eukaryotic initiation factor 2 (eIF2) is phosphorylated, leading to transient arrest of translation which in turns reducing protein load. However, there is an exception; ATF4 mRNA and mRNAs with a 5' open reading frame continue to be translated. It triggers the expression of genes related to aminoacidic metabolism, redox control, protein folding, and autophagy. If ER stress persists and the cell needs to give a more severe response, CHOP induces apoptosis-related genes, while GADD34 reverses the inhibition of protein translation. Active Perk also phosphorylates NFR2 transcription factor mediating antioxidant response. In contrast to IRE1α and PERK, activated ATF6 translocates to the Golgi apparatus, where it undergoes specific cleavage, resulting in the release of the transcription factor ATF6(N). ATF6(N) then translocates to the nucleus and triggers the expression of genes related to protein folding and ERAD (Figure 4) (93,94,7,8).

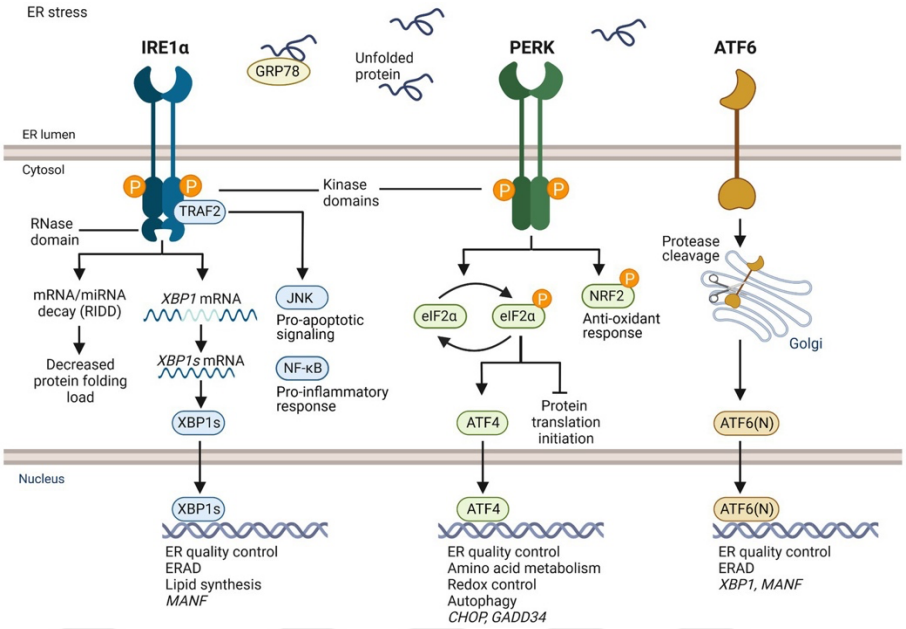


Figure 4. Overview of UPR signaling (7).

The modulatory role of CDNF in UPR has emerged from its similarity to the MANF protein. Both are ER-localized proteins, and it is suggested that CDNF, similar to MANF, regulates the UPR system. MANF blocks the UPR system by both binding to Bip (GRP78) and by inhibiting ATP-Bip and ADP release from Bip. It also binds Ire1α to inhibit Ire1α downstream pathway. It is thought that CDNF blocks UPR as in seen in MANF. Studies showed that CDNF and MANF have common interaction partners (Figure 5). However, some studies create dilemmas about the working mechanism of CDNF. Interestingly, when neurons were treated with CDNF in 2D culture, it did not change the neuronal survival capability in contrast to other NFs (7,8). However, Zhang and colloquies showed that when proteostasis disrupted by a drug, called tunicamycin, there observed increasing expression of CNDF in neurons in vitro and showed neuroprotective activity of CDNF (95). This ER stressor drug has shown to increase CDNF expression in mouse liver and kidneys, proving the CDNF as ER stress regulator. Inverse relation in the expression of CDNF and Bip similar to MANF was also observed. Under ER stress, the expression of XBP*1, Atf6* and *Grp78* genes under the branches of IRE1α and ATF6 UPR are reduced when dopamine neurons were treated with CDNF and MANF. These data suggest that CDNF involves in UPR

system as modulator and requires ER stress condition for its neuroprotective activity (7,8).

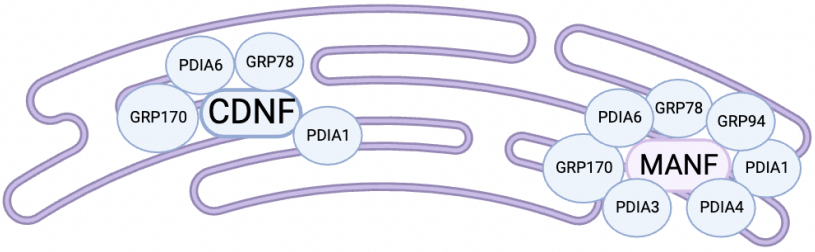


Figure 5: CDNF and MANF interaction partners detected by purification-mass spectrometry (AP-MS) (96,97).
This figure was created in *BioRender (https://www.biorender.com/)*.

Apoptosis, a significant indicator of neurodegenerative diseases, is a consequence that occurs when abnormal metabolic events and increased cellular stress take place. CDNF can protect against neuronal degeneration and inhibit apoptosis by directly regulating pathways associated with apoptosis, in addition to its role in suppressing ER stress. Mai and Niu demonstrated dose-dependent attenuating effect of CDNF on PC12 cells induced by 6-OHDA. Additional investigation reveals that the increase in the Bcl-2/Bax ratio and the decrease in caspase-3 activity are observed in a manner dependent on the dosage, whether CDNF is administered before or after treatment. This implies a regulatory pathway of apoptosis by CDNF (98,7,8).

Several studies have demonstrated the anti-inflammatory role of CDNF. Overexpression of CDNF in primary astrocytes through lentivirus vector delivery was shown to reduce astrocyte damage induced by the ER stressor drug, as well as the secretion of inflammatory cytokines IL-1β, IL-6, and TNFα. Decreased pro-inflammatory signals was observed in LPS induced primary microglia by inhibiting Akt and JNK phosphorylation and by inhibiting FoxO1/mTOR signaling. LPS is an endotoxin, enhancing the expression of pro-inflammatory cytokines. CDNF also reduced 6-OHDAinduced inflammation markers in rat substantia nigra. Overall, CDNF does not only reduce cytokines, but it also involves additional mechanisms to suppress the inflammatory response(99,100, 101).

Due to the similarity between CDNF and MANF, it is believed that their functions are alike; however, the exact mechanism of action is still not fully understood.

2.3.2.3 CDNF Deficient Model Organisms

The study generating and analyzing initial knockout (KO) mouse model for CDNF was published in 2020 by Lindahl and colleagues (102). Despite the recent increase in characterization studies of CDNF, there is limited research with model organisms. However, existing studies have extensively addressed CDNF deficiency and conducted comparisons in an age-dependent and control-dependent manner (Table 2). These studies suggested that CDNF is crucial for the development and sustained maintenance of dopaminergic and various submucosal neuron populations.

Table 2. CDNF -/- model organisms.

Model Organism	Outcomes and phenotypes of CDNF -/-	References
Mouse	Viable, fertile, normal life span, age dependent loss of enteric neurons in the myenteric plexus as a consequence of neurodegeneration and autophagy, presence of neurodegeneration and autophagy markers in submucosal plexus of ileum, duodenum and colon, unaltered number of dopamine neurons and dopamine in substantia nigra, altered dopaminergic axon terminal function in the striatum (high amount of dopamine secretion upon amphetamine administration).	102
Zebrafish	No morphological change in the phenotype, elevated levels of th2 expression in the brain, normal dopamine amount, abnormal clustering of dopaminergic neurons, decreased number of histamine and GABA neurons throughout the brain, increased seizure susceptibility, reduced sociability.	103
Mouse	Deficiency of expression of dopaminergic neuronal markers, dopamine, TH, dopamine transporter in ileum, elevated age-related decline in the number of submucosal dopaminergic neurons, elevated age-related decline in the number of submucosal neurons (NOS, GABA, and CGRP expressing)	104

2.3.2.4 CDNF Therapy for Neurodegenerative Diseases

Models of neurodegenerative diseases are widely used to reveal effects of treatment options. Several studies showed the efficacy of CDNF on disease models (Table 3).

Table 3. Summary of pre-clinical studies in several neurodegenerative diseases.

Disease Model	CDNF Delivery Method	Outcomes of CDNF Treatment	References
6-OHDA induced Parkinson's Model /mouse	Single intrastriatal injection	Contribution to dopaminergic (TH+) neuron restoration, prevention of degeneration of dopaminergic neurons in striatum	105
6-OHDA induced Parkinson Model /rat	Chronic intrastriatal injection	Reduced parkinsonian behavior (rotation), inhibition of loss of TH+ cells in substantia nigra pars compacta, enhanced survival of TH+ fibers in striatum	106
MTPT induced Parkinson Model/ mouse	Striatal injection	Pretreatment before MTPT injection: Improved motor behavior, increased TH+ cell number in substantia nigra pars compacta Post-treatment: Increased motor behaviour, increased dopamine fiber densities in striatum, increased TH+ cell number in substantia nigra	107
6-OHDA induced Parkinson Model /rat	AAV2 vector injection to the striaatum Lentivrial vector injection into the striatum	Reduced parkinsonian behavior (rotation), increased dopamine fiber densities in striatum, increased TH+ cell number in substantia nigra, recovery of dopaminergic neurons	108,109
6-OHDA induced Parkinson Model /rat	Intranigral transfection	Reduced nitrosative stress, reduced inflammation (decrease in the number of glial markers and in the level of IL-6)	110
APP/PS1 Alzheimer Model/ mouse	AAV2 vector intrahippocampal injection	Improved memory in both APP/PS1 and Wild type	111

Table 3. Summary of pre-clinical studies in several neurodegenerative diseases (continued).

6-OHDA induced Parkinson Model /marmoset monkeys	Implanted cannula	Increase in dopamine transporter activity and in the number of TH+cells in lesioned caudate nucleus, tolerable in high doses	112
METH induced Parkinson Model/ rat	AAV8 vector intrastriatal injection	Attenuation of METH induced TH and dopamine loss, not lower METH induced hyperthermia	113
6-OHDA induced Parkinson Model / rat	AAV8 vector intrastriatal injection	Improved motor function and increased TH level, limited effect on rat with severe lesion	114
quinolinic acid induced Huntington Models / Wistar rat	Intrastriatal injection into striatum	Improved motor function, decreased ataxia, stimulated neurogenesis (DCX+ and NeuN+ cells in the striatum)	115,116
ALS Models 1)acetyltransferase-tTA/TRE-hTDP43-M337V rat 2) SOD1-G93A mouse 3) TDP43-M337V mouse	Intracerebroventricular injection	Improved motor function, contribution to motor neuron rescue from ER stress induced apoptosis and survival, attenuation in ER stress response	117
Paraquat induced Parkinson Model/ mouse	Cell-penetrating peptides with CDNF injected to striatum	Protection of neurons and glial cells, thereby preventing motor and cognitive dysfunction	118
Protofibrils of α-synuclein solution induced Parkinson Model /rat	Subcutaneous administration /HER96	Reduced α-synuclein aggregates, increase protection for dopamine neurons, reduced inflammation, reduced ER stress	119

3 MATERIALS AND METHODS

3.1 Materials

3.1.1 Patients

Two sisters, 5 and 13 years old at the time of evaluation, with early onset neurodegeneration and their unaffected biological parents are reported here for the first time. Both sisters were born at term to consanguineous parents of Turkish descent. The clinical presentation of patients was considered severe and progressive. Clinical phenotypes and detailed information are presented in the methods (Table 6). Subjects of this study were included from a large neurogenetics cohort**.** This family was recruited within this thesis by Kaya Bilguvar, MD Ph.D., and Okay Caglayan, MD. Ph.D.

3.1.2 Software, Hardware, Computer-Based Programs, Websites

The software and computer-based programs used in this thesis are listed in Table 4**.**

Table 4. The list of Software, Hardware, Computer-based programs, Websites used in this thesis.

SOFTWARE, HARDWARE, PROGRAM and WEBSITE NAME	COMPANY and WEBSITE	PURPOSE of USE
macOS Ventura 13.3.1	Apple	Bash scripting
Solid State Drive	Western Digital	Data storage and free space usage
NCBI	https://www.ncbi.nlm.nih.gov/	Reference genome, transcript, and protein sequence / Viewing the protein conservation among species.

Table 4. The list of Software, Hardware, Computer-based programs, Websites used in this thesis(continued).

Ensemble Genome Browser	https://useast.ensembl.org/index.html	Reference human and mouse genome, transcript, and protein sequence & gene/transcript/exon annotations
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/	Viewing submitted variations among individuals
Gtex	https://gtexportal.org/home/	Visualizing tissue specific gene expression
Gnomad v3.1.2	https://gnomad.broadinstitute.org/	Viewing and tracing the variations among populations from variety of large-scaled sequencing projects
Annovar	https://annovar.openbioinformatics.org/en/latest/	Annotating genetic variants
VEP	https://useast.ensembl.org/info/docs/tools/vep/script/index.html	Annotating genetic variants
SnpEff	https://pcingola.github.io/SnpEff/se_introduction/	Annotating genetic variants
Conda 23.3.1	https://docs.conda.io/en/latest/	Creating environment and installing packages from the conda channels
Fastqc	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	Checking the quality of high-throughput sequencing data
Star 2.7.9	https://github.com/alexdobin/STAR	Splice transcript alignment to the reference mouse genome (mm39)
Featurecounts	Rsubread package	Quantification of mapped reads

Table 4. The list of Software, Hardware, Computer-based programs, Websites used in this thesis(continued).

R studio	http://www.rstudio.com	Analysis of RNA seq data
Bioconductur packages	bioconductor.org	Analysis of bulk RNA seq in R
IGV	https://igv.org	Visual exploration of genomic data

3.1.3 Chemicals and Media Components

All the chemicals and media components used in this thesis are listed in Appendix 1.

3.1.4 Equipment

All the equipment used in this thesis is listed in Appendix 2.

3.1.5 Solutions

Coating Matrix: TC well plates were coated with collagen type I that is reconstituted with sterile 0.01 N HCl solution.

3.1.6 Growth Media

Storage Medium: Hair plucks from the human donor were stored in filter sterilized DMEM that is supplemented with 1x antibiotic-antimycotic solution.

Washing Medium 1: Hair plucks from the human donor were washed in DMEM/F12 that is supplemented with 1x antibiotic-antimycotic solution.

Washing Medium 2: Hair plucks from the human donor were washed in PBS that is supplemented with 1x antibiotic-antimycotic solution.

DMEM: Hair follicle cells were centrifuged in DMEM that is supplemented with 10% fetal bovine serum.

Growth Medium for Keratinocytes 1: Hair follicle cells were plated in KGM-Gold Keratinocyte Medium that is supplemented with 1x KGM-Gold Supplement, 1x penicillin-streptomycin, and 10µm Rock Inhibitor.

Growth Medium for Keratinocytes 2: Hair follicle cells were plated in EpiLife Keratinocyte Medium that is supplemented with 1x Epidermal Defined Growth Supplement, 1X penicillin-streptomycin, and 10µm Rock Inhibitor.

Keratinocyte Maintenance Medium 1: Hair follicle cells were plated in EpiLife Keratinocyte Medium that is supplemented with 1x Epidermal Defined Growth Supplement, and 1x penicillin-streptomycin.

Keratinocyte Maintenance Medium 2: Hair follicle cells were plated in KGM-Gold Keratinocyte Medium that is supplemented with 1x KGM-Gold Supplement, and 1x penicillin-streptomycin.

Fibroblast Growth Medium: Fibroblast cells were plated in DMEM supplemented with 10%FBS, 1x GlutaMAX and 1x MEM non-essential amino acids.

3.1.7 Enzymes

Enzymes	**Supplier Company**
0.25% Trypsin	Gibco, USA
TrypLE™ Select	Gibco, USA

3.2 Methods

Keratinocyte cell culture protocol was developed and applied according to a study conducted by Kaya Bilguvar, Adjunct Professor, and Ce Zhang, MD Ph.D. at Yale University.

3.2.1 Family Recruitment and Clinical Presentation

In this study, we described a Turkish family with early signs of neurodegeneration, and with two affected children. Detailed family history and clinical features were obtained. The pedigree, detailed family history, and information are shown in the Table 5 and Figure 6.

Table 5: Family information.

Cohort ID	Affection	Age	Sex	Alive/dead
NG1163-1	Affected	13	F	Alive
NG1163-2	Affected	5	F	Alive
NG1163-3	Unaffected father	47	M	Alive
NG1163-4	Unaffected mother	36	F	Alive

F, female; M, male

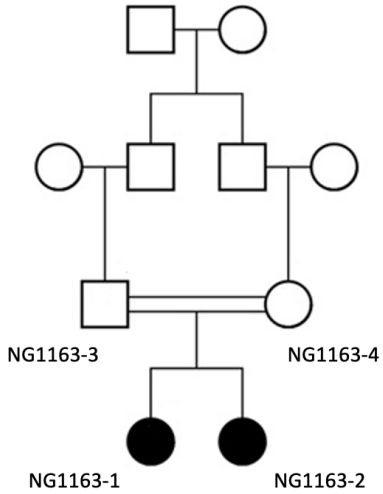


Figure 6. Pedigree chart of the patients.

3.2.2 Whole Exome Sequencing & Variant Calling and Annotation

Genomic DNAs were isolated and evaluated from peripheral blood samples of 4 family members by Kaya Bilguvar and his colleagues at Yale University. Whole exome sequencing of DNAs from 4 subjects was performed at Yale Genome Center (YCGA) (https://medicine.yale.edu/keck/ycga/) with a mean coverage of 62X.

The whole exome analysis (gatkExome.rms) pipeline for variant calling and annotation developed by The Knight Lab at Yale University was executed (Figure 7**).** Raw sequencing reads were aligned to the GRCh38/hg38 assembly using BWA v 0.715. Read duplicates were removed by Picard's MarkDuplicates, (v2.18.29). Systematic errors made by the sequencer for each base were detected and recalibrated by BQSR. Insertion/deletion realignment was performed by HaplotypeCaller. Single nucleotide variations and short insertion/deletions were obtained using HaplotypeCaller. To assign scores to variants based on their likelihood of being real, variant quality score recalibration (VQSR) was applied with the quality of depth, mapping quality, MQRankSum, ReadPosRankSum, SOR, FS annotation parameters. After the joint variant calling, variants were filtered using either VQSR and HardFiltering. Finally, the resulting variants were annotated using SnpEff (v5.1), Annovar, Ensemble Variant Effect Predictor (v107). Annotation databases are indicated in Figure 8.

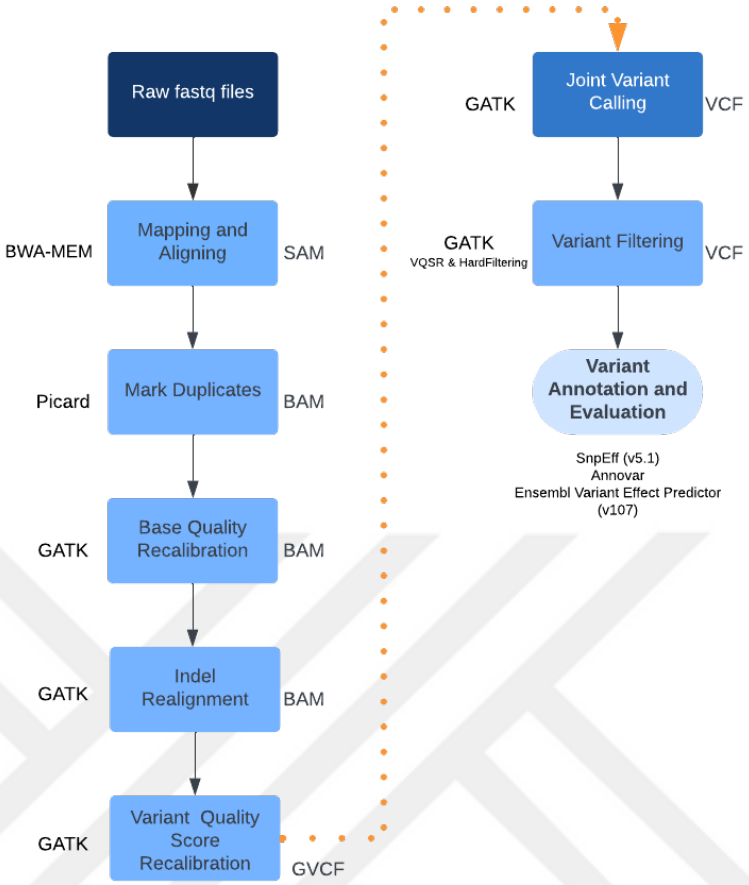


Figure 7. Whole exome sequencing pipeline for the alignment, variant calling, and annotation (https://campuspress.yale.edu/knightlab/ruddle/gatkexome/).
This flowchart was created in *Lucidchart (https://www.lucidchart.com/)*.

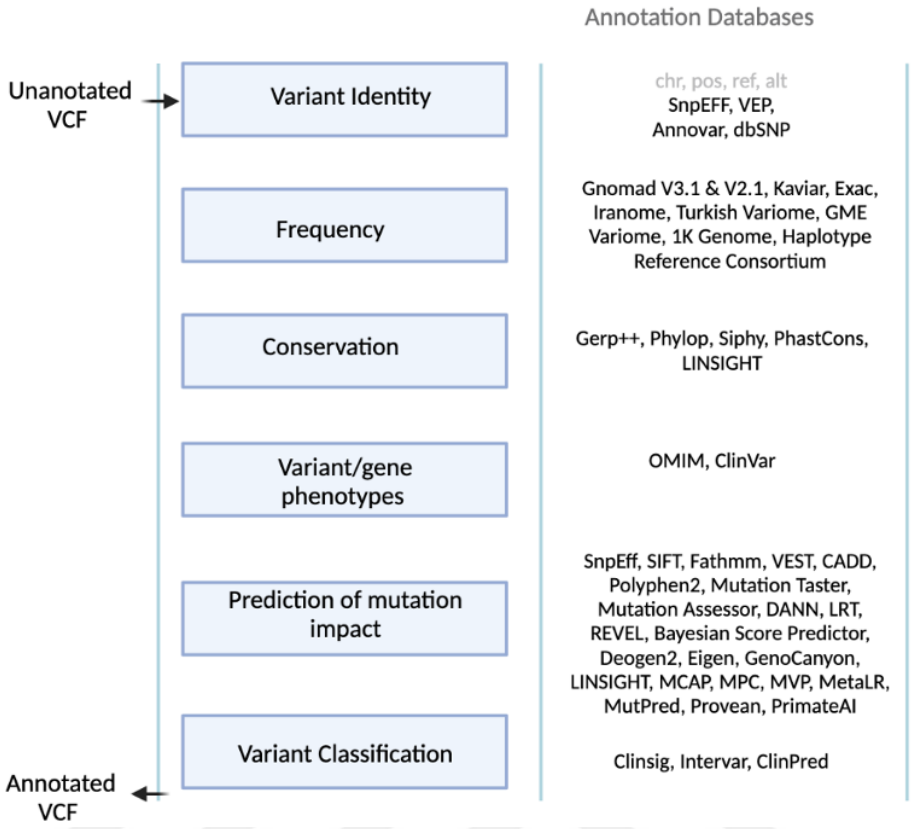


Figure 8. Variant annotation databases.
This figure was created in *BioRender (https://www.biorender.com/)*.

3.2.3 Variant Filtering, Classification and Prioritazition

Variants of NG1163-1, NG1163-2, NG1163-3 and NG1163-4 were filtered, classified, and prioritized with the following criteria. Homozygous shared variants in both NG1163-1 and NG1163-2 were evaluated and assessed as a first step. Then, heterozygous and compound heterozygous variants of all family members were assessed.

Variants with alt dept≤ 4 *and* 8 (respectively for homozygous and heterozygous variants) were removed. Synonymous, intron, upstream and downstream variants were removed. Variants were not considered if homozygous ones had minor allele frequency ≥ 5e-3 and heterozygous and compound heterozygous ones had minor allele frequency ≥ 5e-5 in the population allele frequency databases indicated in Figure 8.

Variants were assessed with specific tools to check the evolutionary conservation and to predict functional impact. Variants also were classified according to American College of Medical Genetics (ACMG) criteria as being benign, likely benign, uncertain significance, likely pathogenic, and pathogenic. Uncertain significance, likely pathogenic and pathogenic variants were retained, and those classified as likely benign and benign were removed.

If variants are known to be involved or not in any disease was checked in OMIM and ClinVar databases. It was examined whether candidate genes coincided with their function in the pathology of patient phenotypes. Variants fulfilling all the criteria were discussed in terms of genotype and phenotype association with clinicians recruited this family in this study.

3.2.4 Variant Validation and Family Segregation Analysis

The candidate disease-causing variant, *CDNF*:c.136C>T, was confirmed by viewing the read alignments in the specific position for both patients and the segregation of our candidate variant for NG1163-1 was within the NG1163 family using the Integrative Genome Browser.

Kaya Bilguvar, MD Ph.D., and his colleagues at Yale University have confirmed our candidate variant for the NG1163-1 subject with Sanger Sequencing.

3.2.5 *CDNF* Knockout Mouse Model

ACTB-Cre/CDNF-/- mouse was generated by crossing the floxed *CDNF* mouse with ACTB-Cre mice by Louvi Lab at Yale University.

3.2.6 RNA-seq and Data Analysis

Generation of mouse models, RNA extraction, library preparation and sequencing steps were completed by Bilguvar Lab and Louvi Lab at Yale University.

3.2.6.1 RNA-seq Data Generation

ACTB-Cre/CDNF-/- mouse and wild-type mouse were used. Animals were sacrificed, thereafter, three samples from the cortex, hypothalamus, and striatum, making up to 54 samples, were collected at 15 days, 3 months, and 21 months of age for RNA extraction. The library was prepared, resulting in an average 101 bp fragments without adaptor size. Hiseq 3000 system was used to sequence the RNA libraries, leading to the generation of single-end reads of 101bp fragments. The quality of reads was controlled by FASTQC v.0.12.1.

3.2.6.2 RNA-seq Analysis

Fastq files belonging to the cortex, hypothalamus, and striatum of ACTB-Cre/CDNF-/- mouse and wild-type mouse at 15 days, 3 months, and 21 months of age were obtained. There existed three different fastq data from three flow cells for every brain region and for every certain age that is mentioned above. Briefly, the experiment was done by three technical replicates from three samples of every brain region of the mouse. The flow of RNA-seq analysis were indicated in Figure 9. The quality of merged fastq files were checked with FASTQC v.0.12.1.

```
# Merging of all corresponding fastq files belonging a timeframe and a brain region
cat *.fastq.gz > <sample_id>.fastq.gz
```

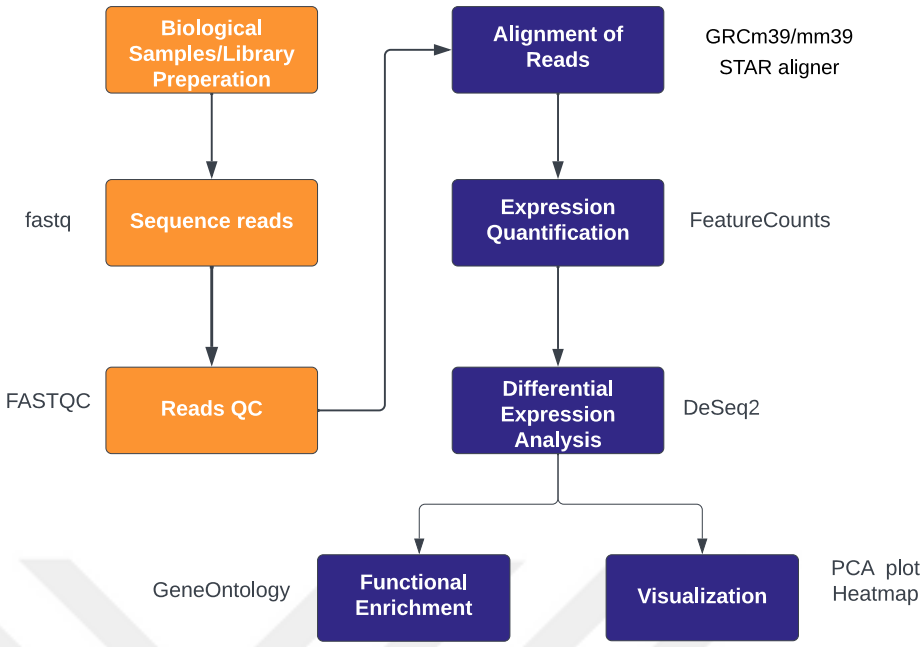


Figure 9. RNA-seq analysis flowchart.

3.2.6.2.1 Aligning Reads and Generating Count Table

RNA sequencing reads from ACTB-Cre/CDNF-/- mouse and wild-type mouse samples were aligned mouse genome GRCm39/mm39 using STAR aligner, which is genome-based alignment tool.

Alignments of the reads using Star aligner v.2.7.10

```
STAR --runMode alignReads --genomeDir genome/ --outSAMtype BAM
SortedByCoordinate --readFilesIn
user<sample_id>/Unaligned/<sample_id>.fastq --runThreadN 12
```

Generating the read counts

```
FeatureCounts -S 1 -a geneannotationfile.gtf -o
nameoftheoutputfile.txt inputbamfile.bam
```

3.2.6.2.2 Differential Gene Expression Analysis

Once data ready for analysis, differentially expressed genes are defined by comparing eighteen variables with the pairwise comparisons and variance across the groups. Firstly, three design factors, mutationally/region/day, with their covariates were assigned. To explore interactions of variations, a complex design was implemented. Then, expression differences were evaluated by DESEq2. The effect of mutationally on region and day was explored by specifying the design formula.

```
# Filtering out the rows
Counts <-read.delim("path/to/counts/file.csv", header = TRUE,
row.names = 1, sep =";")
Counts <- Counts[which(rowSums(Counts) >100), ]

# Assigning conditions to the samples
coldata <- read.csv("path/to/countdata/file.csv.", header=TRUE,
sep=";", row.names= 1)
coldata$mutationality <- factor(coldata2$mutationality, levels =
c("mutant","wildtype"))
coldata$region <- factor(coldata2$region,levels =
c("cortex","striatum", "thalamus"))
coldata$day <- factor(coldata2$day, levels = c("15d","21m", "3m"))

# Construction of DESeq dataset
dds <- DESeqDataSetFromMatrix(countData = Counts,
                        colData = coldata,
                        design = ~ mutationality + region + day +
region:day +day:mutationality + mutationality:region)
dds <- DESeq(dds)
```

To assess similarity among the samples, hierarchical clustering and principal component analysis were applied.

```
# QC of the data
vsdata <- vst(dds, blind = FALSE)
```

```
pcavds <- plotPCA(vsdata, intgroup = c("mutationality", "region",
"day"))

sampleDists <- dist(t(assay(rld)))
sampleDistMatrix <- as.matrix(sampleDists)
colnames(sampleDistMatrix)
colors <- colorRampPalette(rev(brewer.pal(9, "Blues"))) (225)
pheatmap(sampleDistMatrix, clustering_distance_rows = sampleDists,
         clustering_distance_cols = sampleDists,
         col=colors)
```

3.2.6.2.3 Gene Enrichment Analysis and Biological Interpretation

Biological and molecular functions were evaluated according to enriched genes according to experiment's designed factors by using the Gene Ontology Annotator.

```
# Gene ontology
genes_to_test <- rownames(dds.df[dds.df$log2FoldChange > 0,])
GO_results <- enrichGO(gene = genes_to_test, OrgDb = "org.Mm.eg.db",
keyType = "ENSEMBL", ont = "BP")
as.data.frame(GO_results)
fit <- plot(barplot(GO_results, showCategory = 20))
```

3.2.7 Optimization and Establishment of Keratinocyte Cell Line from Human Hair Follicles

Our protocol for keratinocyte cell culture from human hair follicles has been optimized for our further studies with the NG1163 family.

Hair Follicle Collection: 11-16 ml of storage medium was poured to a 100mm sterile petri dish. 50ml tube was filled until it was full. Both including storage medium was equilibrated to room temperature.

Hair Plucking Procedure: Flat tweezer was sterilized under the UV light. Immediately before hair plucking, a flat tweezer was put over a flame for 40sec-1min

to make sure that it is as sterile as much of we can. 10 anagen phase hair follicles were plucked. The follicle was placed in the 100mm petri dish with the storage medium immediately. Only plucked hairs in the anagen phase, with a visible white outer root sheath, were used. 10 hair follicles in the anagen phase were transferred to the 50ml tube with the storage medium. Lastly, the tube with hairs was capped tightly.

Precoating Culture Plate: Two well of 24 well plates was coated with collagen type I (3-10µg/cm^2).

Keratinocyte Plating Protocol: Hairs were transferred to a 100mm petri dish with washing media I using tweezers. Hairs without follicles were discarded. Hairs were transferred to another 100mm petri dish with washing media II. Hairs were cut shorter to obtain hair follicle area. Then, hair follicles with short hair shafts were transferred to 6 well dish or 5ml tube with 1ml of 0.25% Trypsin and incubated for ten minutes at 37^0C. Follicle cells detached from hair were transferred to a 15ml tube with 10% FBS in DMEM. After centrifugation at 200g for 5 minutes, keratinocytes were suspended in the growth media for keratinocytes 1. Keratinocytes were plated in one well of pre-coated plate were incubated at 37^0C for 48 hours. Lastly, the growth medium for keratinocytes was replaced with keratinocyte maintenance medium 1. The medium was changed every other day with keratinocyte maintenance medium 1 until it reaches 80-90% confluency.

Hairs were transferred to a 100mm petri dish with washing media I using tweezers. Hairs without follicles were discarded. Hairs were transferred to another 100mm petri dish with washing media II. Hairs were cut shorter to obtain hair follicle area. Then, hair follicles with short hair shafts were transferred to 6 well dish or 5ml tube with 1ml of 0.25% Trypsin and incubated for ten minutes at 37^0C. Follicle cells detached from hair were transferred to a 15ml tube with 10% FBS in DMEM. After centrifugation at 200g for 5 minutes, keratinocytes were suspended in the growth media for keratinocytes 2. Keratinocytes were plated in one well of the pre-coated plate were incubated at 37^0C for 48 hours. Lastly, the growth medium for keratinocytes was replaced with keratinocyte maintenance medium 2. The medium was changed every

other day with keratinocyte maintenance medium 2 until it reaches 80-90% confluency.

3.2.8 Fibroblast cell line from patient NG1163-1

As of today, Bilguvar Lab members have attempted to establish patient-derived cell cultures for further studies. NG1163 family is one of the samples whose fibroblast cell lines have been established by them; however, cryovials from P0 and P1 were thawed and waited in unstable temperature conditions for a while due to the problem from the nitrogen storage system, resulting in dropped cell viability. Firstly, two of the cryovials were thawed. Thawed cells were transferred to a 15ml tube. Then, 9 ml of prewarmed DMEM was added dropwise on the cells. The cells were centrifugated at 300g for 3 minutes. After fibroblasts were suspended in fibroblast media, the cell viability was calculated by trypan blue exclusion test. Cells were plated to one well of 24 well plates. As a last step, 50μM Rock inhibitor was added onto a well including the fibroblast cells suspended in fibroblast medium.

4 RESULTS

4.1 Clinic and Radiologic Manifestation

The clinical picture was mainly characterized by early-signs of neurodegeneration (Table 6), all of which point to multisystemic neurodegeneration. The result of the MR imaging examination of the brain revealed cerebral and cerebellar atrophy on T2-weighted imaging (Figure 10).

The parents are first cousins and are alive and asymptomatic. This couple suffered 10 spontaneous abortions not exceeding 3-5 months of gestation. Both sisters demonstrate early signs of cerebellar degeneration tremendously impacting their lives. NG1163-1 and NG1163-2 were admitted to the clinic with the diagnosis of developmental delay, positive Babinski sign, head tremor (onset ~3yo), blepharospasm, spastic diplegia, constriction of retinal arteries, photophobia, retinitis pigmentosa sign, and ear abnormality. Infantile nystagmus has been present in both sisters. Body tremors reportedly have existed starting from the age of 3. While NG1163-1 presents with aphasia, abasia, severe delayed intellectual development, and hypersalivation, NG1163-2 also has gait ataxia, mild intellectual disability, and inability to sit up with support. Both have normal metabolism/homeostasis, sphincter function, spinal magnetic resonance imaging, and active deep tendon reflexes.

Table 6. Clinical features.

Cohort ID	Gender and Age	Affected Status	Tremor Age at onset	Clinical Manifestations
NG1163-1	F, 13	Affected Sibling	3 yo	Developmental delay, intellectual disability, head tremor, blepharospasm, spastic diplegia, aphasia, abasia, hypersalivation, photophobia, retinitis pigmentosa-like signs, constriction of retinal arteries, ear anomaly

Table 6. Clinical features (Continued).

NG1163-2	F, 5	Affected Sibling	3 yo	Developmental delay, intellectual disability, head tremor, blepharospasm, spastic diplegia, ataxia, photophobia, retinitis pigmentosa-like signs, constriction of retinal arteries, ear anomaly
NG1163-3	M, 47	Unaffected Father		
NG1163-4	F, 36	Unaffected Mother		

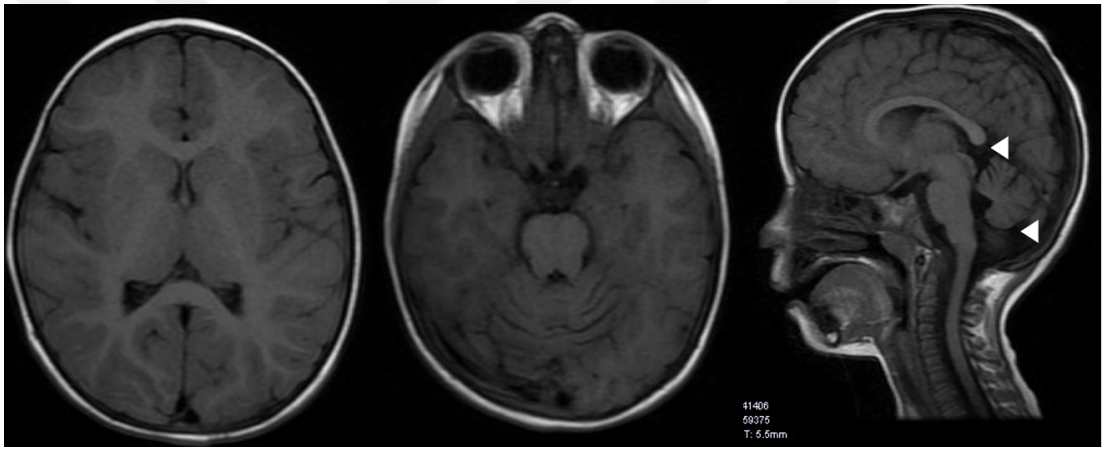


Figure 10. T2 weighted MRI examination of the brain of NG1163-1. Sagittal image showing atrophy in the cerebellar region and cerebellum (*arrowhead*).

4.2 WES Analysis

WES performed for all four members of the NG1163 family. WES data quality metrics were indicated in Tables 7 &8. Counts of consequence types of SNPs and indels predicted by SnpEff and VEP were listed in Table 9.

Table 7. Statistical analysis of VCF file.

# of total variants	# of pass variants	# of SNP	# of Indels	# of Multiallelic Sites	# of Multiallelic SNP sites	TiTv Ratio
130351*	119260	110041	20687	4834	125	2.18

* Multiallelic sites were splitted.

Table 8. Exome Metrics of NG1163 Family.

	NG1163-1	NG1163-2	NG1163-3	NG1163-4
Read Length:	101	101	101	101
# reads (M):	47.4	38.3	69.6	29.2
# bases (G):	4.8	3.9	7.0	2.9
Mean coverage:	64.0	54.1	91.8	40.9
Median coverage:	61	53	89	40
PCR duplicates:	14.01%	8.24%	10.30%	6.36%
Multiply mapped:	3.90%	3.60%	3.68%	3.61%
Unmapped:	0.27%	0.32%	0.25%	0.28%
Reads on-target:	63.92%	66.91%	63.16%	66.80%
Bases on-target:	45.69%	47.71%	44.58%	47.39%
Proper Pairs:	99.77%	99.81%	99.72%	99.74%
Mean error rate:	0.47%	0.47%	0.45%	0.47%
1x target base coverage:	98.3%	98.3%	98.6%	98.3%
4x target base coverage:	98.0%	98.1%	98.3%	98.0%
8x target base coverage:	97.8%	97.8%	98.2%	97.7%
10x target base coverage:	97.7%	97.7%	98.1%	97.4%
20x target base coverage:	96.0%	96.5%	97.7%	92.7%
50x target base coverage:	66.0%	58.5%	87.8%	27.1%
100x target base coverage:	11.8%	1.9%	39.3%	0.4%
Targets:	IDTv2	IDTv2	IDTv2	IDTv2

Table 9. Number of effects by type.

Consequence Type	Count / SnpEff	Count / VEP
Missense	41005	129491
Splice Region	24226	45215
Splice Acceptor	618	602
Splice Donor	222	882
Frameshift	742	2174
Stop Gained	325	931
Start Lost	85	296
Stop Lost	57	189

Table 9. Number of effects by type (continued).

Synonymous	51	157
5_prime_UTR	12165	30525
3_prime_UTR	28849	57002
Non-Coding Transcript Exon	73571	82628
Intron	550977	1318494
Inframe deletion	827	2403
Inframe insertion	617	1821
Upstream Gene	142332	204219
Downstream Gene	176120	259605
Intergenic	1039	8

**Not all the consequences were listed.*

Exome datasets were analyzed to pinpoint variants that are clinically relevant to observed phenotype. Homozygous shared rare variants were listed in Table 10. Following filtering and prioritizing homozygous, heterozygous, and compound heterozygous variants, it was identified that both NG1163-1 and NG1163-2 were found to harbor homozygous *CDNF*:c.136C>T stop-gained variant . This variant was the only one fitting all our filtering and prioritization criteria mentioned in the methods. Variant in silico pathogenicity scores and classification is given in Table 11. This c.136C>T variant causes in premature stop codon that would probably result in a truncated protein. This truncated protein would be also lack of five cysteine domains, SAP like domains and ER-retention signal, thereby causing loss of function of CDNF protein.

Table 10. The list of homozygous shared rare (GnomADV3) variants in the coding region. Synonymous variants were eliminated.

Gene	**Chr Position**	**Status**	**cDNA Change**	**Aminoacid Change**	**GnomADV3**	**Kavir**	**GME**	**1K Genome**	**ExAC**
SNIP1	Chr1: 37540562G>A	missense	c.521C>T	p.Ala174Val	1.1e-3	1.31e-3	2.5e-2	1.1e-3	1.3e-3
SGIP1	Chr1:66695454G>A	missense	c.1591G>A	p.Val531Ile	6,299e-05	7.12e-5	5.054e-4	-	9.07e-5

Table 10. The list of homozygous shared rare (GnomADV3) variants in the coding region. Synonymous variants were eliminated (continued).

CDNF	**Chr10: 14828252G>A**	**Stop-gained**	**c.136C>T**	**p.Arg46Ter**	**1.12e-4**	**3.23e-5**	**2e-3**	**-**	**4.12e-5**
HCRTR	Chr1:31623577C>A	missense	c.793C>A	p.Leu265Met	1.5e-3	1e-3	1.1e-2	3.9e-4	1e-3
ANKRD26	Chr10: 27017749C>T	missense	c.4259G>A	p.Cys1420Try	1.5e-3	3.2e-3	1.8e-2	4e-3	1.5e-3

*NG1163-1: hom, NG1163-2: hom, NG1163-3: het, NG1163-4: het; Hom: homozygous; Het: heterozygous

Table 11. In silico scores and classification of prioritized variant.

Chromosome Position	Chr10:14828252G>A	
Depth and Mapping Quality	133865/60	
Gene	Cerebral Dopamine Neurotrophic Factor, CDNF	
Change	C*DNF*:NM_001029954:exon2:c.C136T:p.R46X	
In silico Pathogenicity Scores	CADD: 36 GERP++:0.144 PhdSNPg: 0.866	Deleterious Conserved Pathogenic
Classification	Likely Pathogenic (ACMG criteria; PM2, PVS1)	

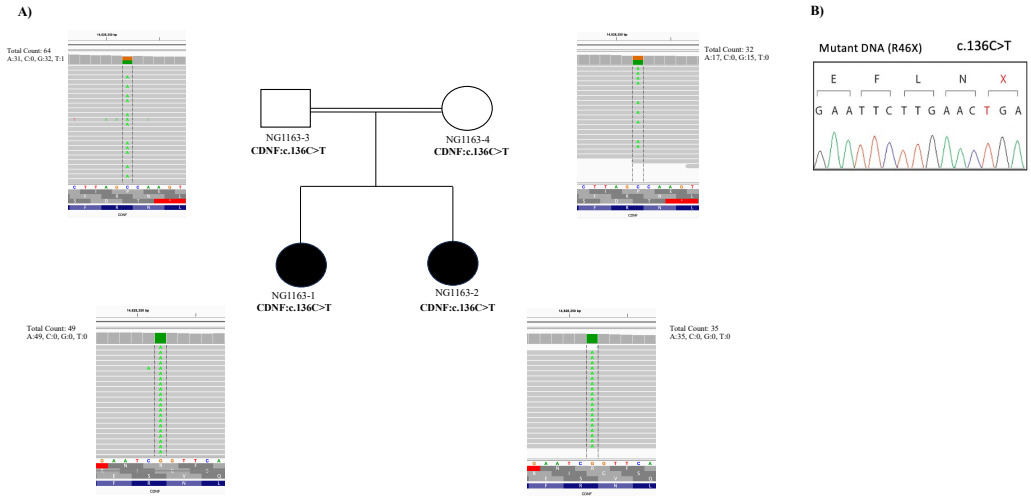


Figure 11. *CDNF* stop-gained mutation in NG1163 family. A) IGV snapshots of c.136C>T variant in *CDNF* gene. B) Sanger Confirmation of NG1163-1.

4.3 RNA-seq Analysis

To understand the impact of CDCNF mutation on the cortex-basal ganglia-cerebellum network, RNA sequencing was performed on samples taken from the cortex, striatum, and thalamus of the CDNF knockout mouse on 15 days, 3 months, and 21 months.

4.3.1 Quality control of the data

Principal component analysis and hierarchical clustering methods were performed (Figure 12 and 13). The data analysis indicated minimal variance in the striatum of the wild-type and mutant mice. The highest variation was observed in the thalamus. Cortex regions in both wild-type and mutant were clustered together, indicating that there observed less number of differentially expressed gene upon mutation (Figure 12).

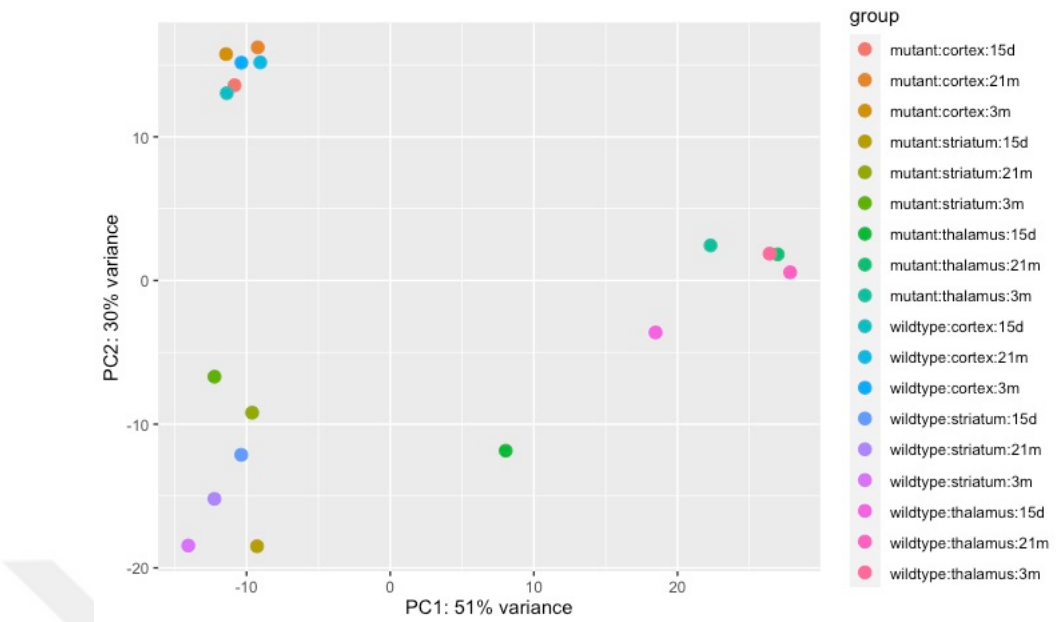


Figure 12. Principal component analysis showing both QC and variation in the data belonging to the striatum, thalamus, and cortex (wild-type and CDNF knockout mice) at 15 days, 3 months, and 21 months.

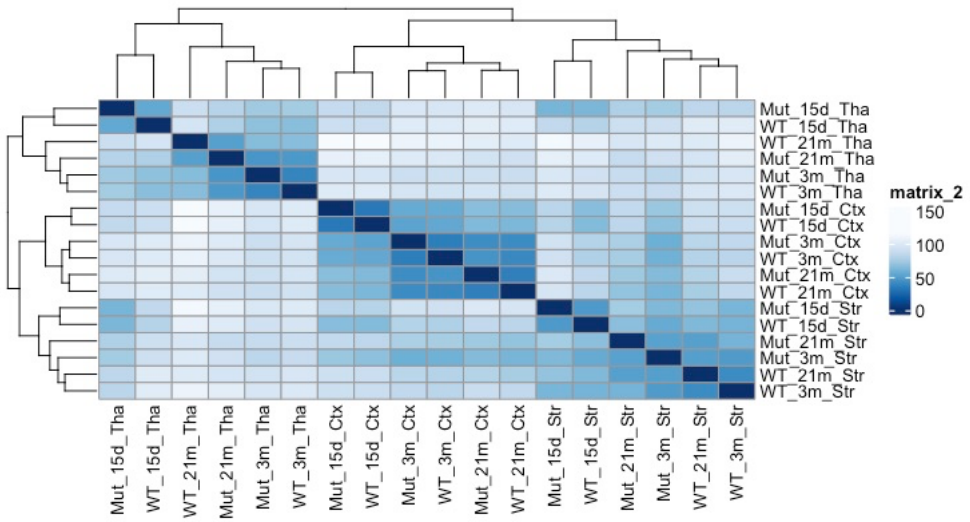


Figure 13. Hierarchical clustering of samples from wild type. Plot shows samples with same mutationally, region and time are more similar to each other, indicating quality of the data. A higher correlation was detected.

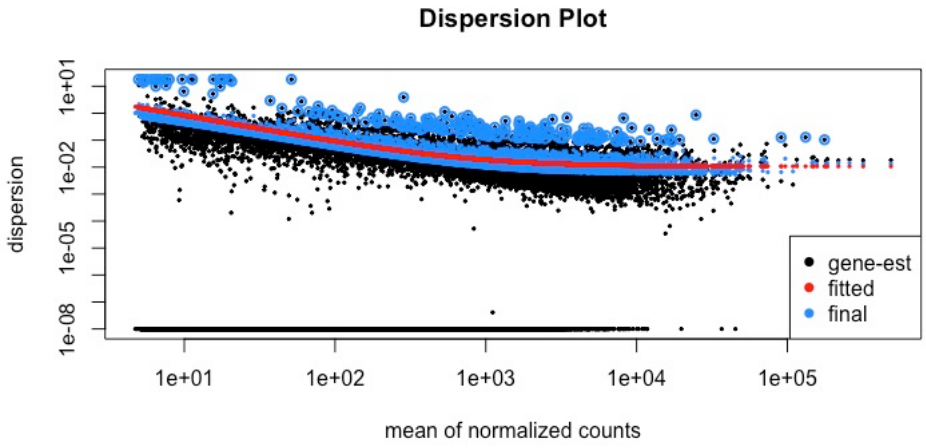


Figure 14. Dispersion of the data. Data indicates variability of data for each gene.

4.3.2 Differential Gene Expression and Enrichment Analysis

Eight differentially expressed genes were detected upon the pairwise comparison of the effect of mutationality on the region and day. These genes were mostly enriched in the mechanisms involved in neurotransmitter uptake and secretion, immune response and cellular traffics (Figure 15, 16).

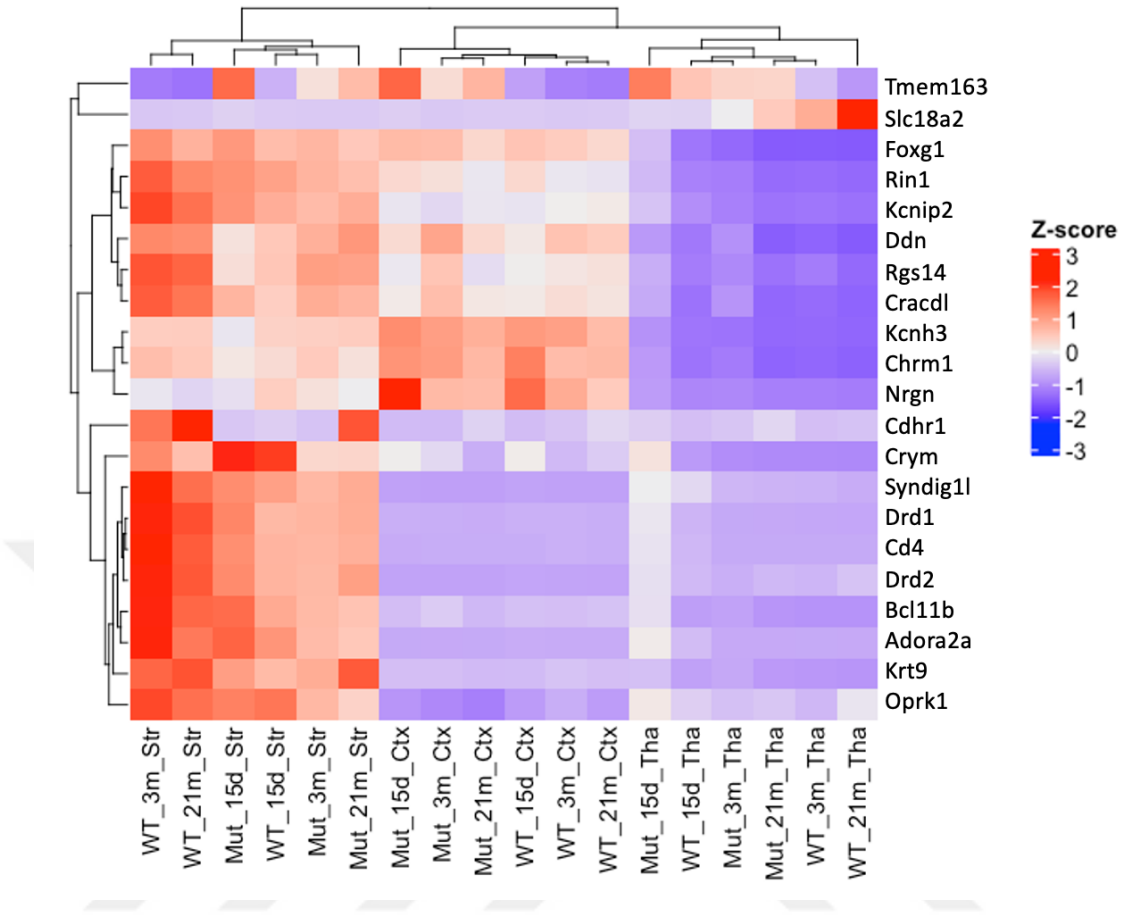


Figure 15. Heatmap indicating differentially expressed genes between wild-type and mutant mice (base mean=100, log2FoldChange (log2FC) >1). Red and blue color gradients represent gene expression levels according to normalized counts per gene. Twenty-one genes showed significance in terms of expression. *Tmem163, Cd4, Crym, Adora2a, Cdhr1, Krt9, Drd1* and Foxg1 genes are the ones with log2FC>1.5. All of the genes were downregulated except for *Tmem163*.

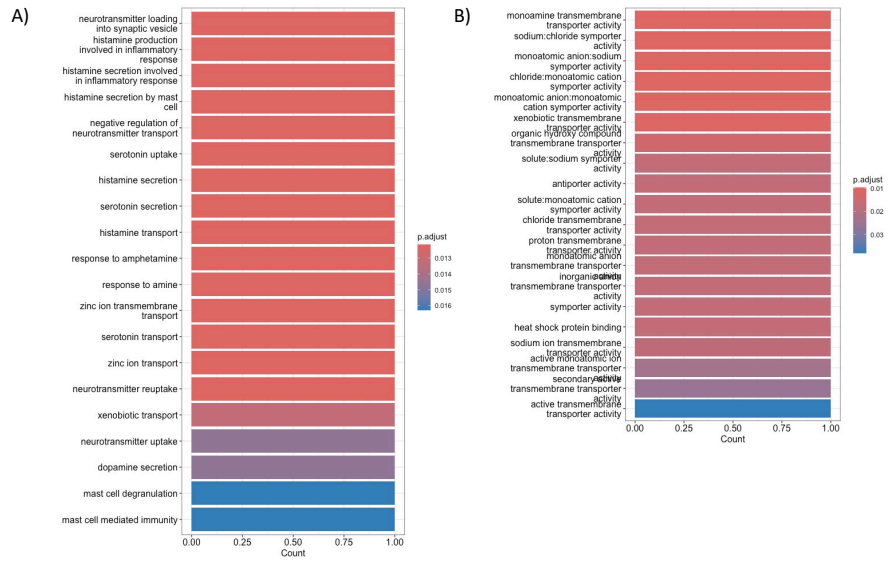
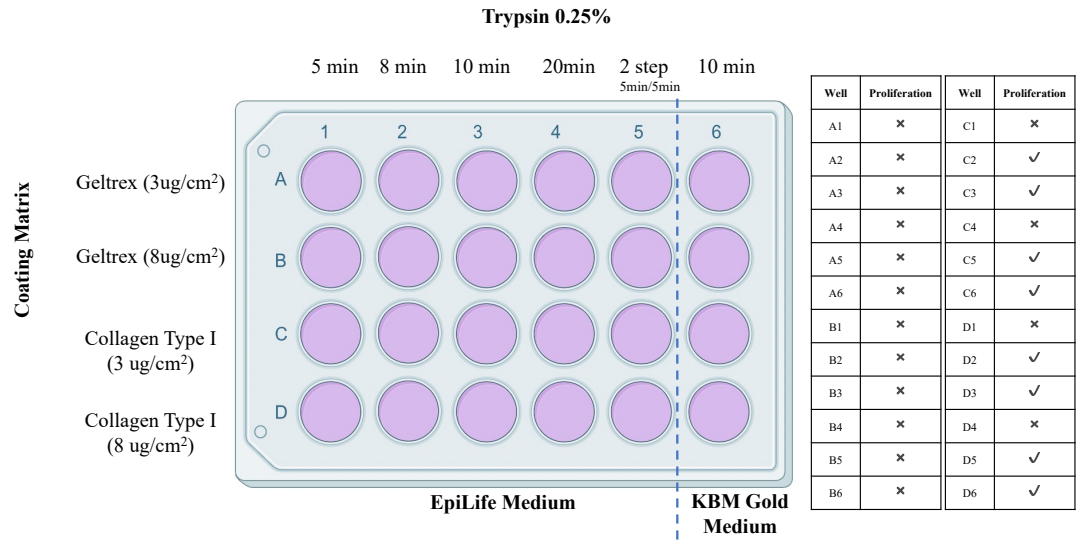


Figure 16. Bar plot of top twenty GO enriched terms. A) Biological functions. B) Molecular functions.

4.4 Keratinocyte Culture

Cultivation of keratinocytes from a low number of hair follicles was managed. Various parameters such as transportation, trypsin incubation time, and different coating matrices and mediums, were employed to reach proliferative keratinocyte culturing. Culturing keratinocytes on collagen type I showed the formation of a confluent cell layer. As the results showed collagen type I coating was the best for the cell proliferation and maintenance contrast to Geltrex including different extracellular proteins (Figure 17 and 18).

One another issue of culturing keratinocytes was their vulnerability to trypsin. As indicated in the Figure 16, as the digestion time of keratinocytes in the trypsin increased, successful cultivation could not be achieved in given incubation time (20 min) more likely due to the loss of proliferative capacity or potential cell death. Both KBM and EpiLife media were considered to be suitable culture media.



Well	Proliferation	Well	Proliferation
A1	×	C1	×
A2	×	C2	✓
A3	×	C3	✓
A4	×	C4	×
A5	×	C5	✓
A6	×	C6	✓
B1	×	D1	×
B2	×	D2	✓
B3	×	D3	✓
B4	×	D4	×
B5	×	D5	✓
B6	×	D6	✓

Figure 17. For optimized keratinocyte cultivation, the combination of different enzyme durations, medium types, and coating matrices were illustrated. Geltrex coating matrix includes collagen IV, entactin, laminin and heparin sulfate proteoglycans.

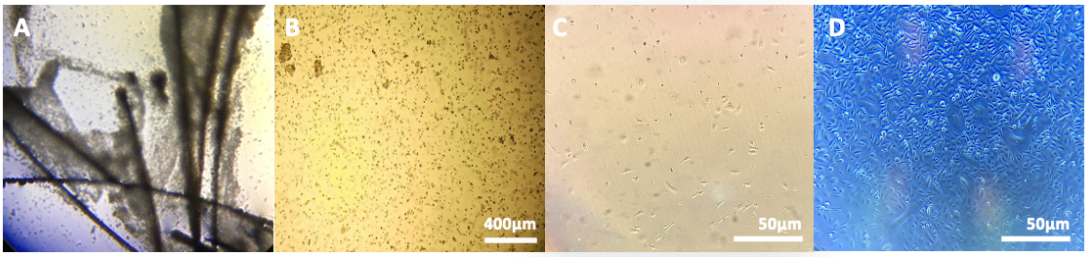


Figure 18. Cultivation of keratinocytes from plucked hairs in anagen phase. A) Five minutes of trypsinization of hair follicles. Note the detachment of cells in the outer root sheath. B) The cells from the outer root sheath are separated from the hair shaft and seeded as single cells (day 1). C) First growing keratinocytes (day 4) D) The confluency of keratinocytes after an average of 6-9 days.

5 DISCUSSION

CDNF encodes cerebral dopamine neurotrophic factor which operates the UPR system which is an ER stress response mechanism. Its molecular mechanism is not fully understood; it is suggested to function in a similar way to the MANF protein, which shares sequence homology with this protein. Its cytoprotective effect has been associated with its role as a modulator in the UPR system. The alteration of ER stress markers has also been shown in 6-OHDA induced Parkinson's models (rat) four weeks after injection of CDNF. Additionally, similar alteration has been observed in also ALS models.

Collectively, different studies showed role of CDNF in brain and especially in dopaminergic system varies among model organisms. While *CDNF* knockout zebra fish models showed dopaminergic phenotype and impaired development of dopamine neurons, there detected no defect in the survival of striatal dopamine neurons in CDNF lack mice. However, another study with CDNF -/- mouse showed degeneration in the enteric dopamine neurons and functional defect in the dopamine neurons of substantia nigra. The differential expression patterns of CDNF across species, coupled with its absence, result in varying phenotypes in model organisms. The alterations in the dopamine transporter system observed in *CDNF*-/- mice resemble early signs of Parkinson's disease. Consequently, the specific roles of CDNF, particularly in dopamine synapses, remain unknown and present a topic for future investigation. The incomplete understanding of the biology and role of MANF and CDNF raises questions about how their absence leads to diverse phenotypes in models and what specific phenotype their deficiency might manifest in the human body (102,103,104).

Given homozygous *CDNF*:c.136C>T mutation causes motor and non-motor symptoms in affected individuals within NG1163 family , it is anticipated that anomalies in the UPR system may give rise to a phenotype similar to these diseases. Synofzik and colleagues' study supports this anticipation. They demonstrated that the loss of function mutation in *DNAJC3*, encoding ER resident Bip co-chaperone DNAJC3 protein, leads to multisystemic neurodegeneration. *EIF2AK3* mutations,

encoding PERK which is one of the member of UPR, also cause Wolcott-Rallison syndrome, thereby representing several hallmarks of DNAJC3 and CDNF deficiency. Likewise, mutations in *SIL1*, upstream protein of DNAJC3 in UPR, cause Marinesco-Sjogren Syndrome, again reflecting the several features of DNAJC3 and CDNF deficiency. Mutations in *WFS1*, encoding negative ER stress regulator wolframin protein, give rise to Wolfram Syndrome, which also shares several features of UPR component deficient phenotype (120, 76).

Keratinocytes offer a variety of advantages compared to other types of somatic cells. Hair collection is non-invasive and can be collected by non-medically trained individuals. Another significant benefit is the option for the specimen to be promptly transported in transportation media at room temperature without the need for time-consuming pre-processing lasting hours or days. However, there are still things that have to be learned and optimized, when working on this cell type.
In the initial stage, reducing the time the hair remains exposed to air after plucking is crucial for the quality of the specimen. Once the hair follicle dried due to air exposure, there is no solution or compensator condition to obtain proliferative cells and cultivate them. Commonly available DMEM is appropriate for transportation medium. The usage of 1X antibiotic-antimycotic solution in DMEM is recommended by us. Serum-based medium causes terminal differentiation. Therefore, we recommend the use of specialized FBS for trypsin neutralization in stem cells. Mediums with varying contents are not suitable for keratinocytes, and serum-free and low calcium media are recommended for keratinocytes, followed by supported with specialized animal-free growth hormones and proteins. Eplife medium seems the right choice when considering proliferation rate and cell morphology according to our expertise in the lab. Another major issue is the choice of coating. The combination of Geltrex matrix+ Epilife medium and Geltrex matrix + KBM2 medium did not present proper culturing condition, after plating of the cells, growth of keratinocytes was not seen in the plate. Either few number of cell or non-proliferative cells were observed in Geltrex coating condition. However, EDGS supplementation in Epilife combined with collagen type 1 coating offers the best proliferation rate and cell morphology.

Considering the partially unknown biology of CDNF, we aimed to generate important findings for the continuation of our research. Since mouse and human shares 99% homology, *CDNF*-/- mice has been generated by Louvi Lab, led to reveal findings enlightening our future studies. Transcriptional changes associated with loss-of function of CDNF in cortex, striatum and thalamus has been evaluated by employing bulk RNA seq. Among three regions, compared to other regions, PCA plot showed more variation between wildtype and *CDNF*-/- thalamic region dataset. In the cortico-basal ganglia network, the thalamus conveys firing to regions of the cortex that govern motor functions. The highest variability in differentially expressed genes within the thalamic region may potentially lead to dysfunction in the thalamus, thereby inhibiting fire onto the cortex in the basal ganglia network. Therefore, we may observe motor function impairments in NG1163-1 and NG1163-2. Significantly differential expressed genes also were enriched in the mechanisms of neurotransmitter uptake/secretion and immune response, which is consistent with the molecular pathology of the phenotypes seen in animal models.

6 CONCLUSION

A novel homozygous mutation of CDNF, which is thought to function as a modulator of UTR under ER stress besides its unknown functions, was detected in two sisters representing early signs of neurodegeneration at the multisystemic level. Mutations in the component of the ER stress mechanism have also mirrored several hallmarks of the CDNF deficiency phenotype of NG1163-1 and NG1163-2. Loss of function in CDNF can be associated with other protein deficiencies involved in ER stress response and especially UPR system. This sheds light on investigating the dysfunction of all proteins involved in the indicated mechanism.

Over the years, many improvements in the collection, transportation, and cultivation of keratinocytes were achieved. An optimized reliable protocol for culturing keratinocytes offers the possibility of generating patient-derived keratinocytes with the same genetic background with patients for further modeling diseases. We describe a simple protocol to generate keratinocytes that increases the quality and reliability of further experiments. The combination of Epilife medium and Collagen type I coating matrix is the best approach to generate proliferative keratinocyte cell culture is recommended according to our expertise.

CDNF knockout mice exhibited differential expression of many genes that underlie the insight into the disease pathology. Moreover, increased variance in gene expression between the thalamic and striatic region of wild type and mutant mice on day 15 can be linked to motor function dysfunction in *CDNF* deficient phenotype seen in siblings of the NG1163 family. Furthermore, our analyses showed that differentially expressed genes are involved in neurotransmitter uptake and release and immune response, providing evidence for biological functions underlying the *CDNF* deficiency phenotype.

7 REFERENCES

1. SEER Training Modules, Module Name. U. S. National Institutes of Health, National Cancer Institute [Internet][cited 2023 December 1].Available from <https://training.seer.cancer.gov/>.
2. Thau L, Reddy V, Singh P. Anatomy, Central Nervous System. StatPearls. Treasure Island (FL). 2023.
3. Bertram L, Tanzi RE. The genetic epidemiology of neurodegenerative disease. J Clin Invest. 2005;115(6):1449-1457.
4. Osipovitch M, Asenjo Martinez A, Mariani JN, et al. Human ESC-Derived Chimeric Mouse Models of Huntington's Disease Reveal Cell-Intrinsic Defects in Glial Progenitor Cell Differentiation. Cell Stem Cell. 2019;24(1):107-122.e7.
5. Nurchis MC, Altamura G, Riccardi MT, et al. Whole genome sequencing diagnostic yield for paediatric patients with suspected genetic disorders: systematic review, meta-analysis, and GRADE assessment. Arch Public Health. 2023;81(1):93.
6. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A. 2009;106(45):19096-19101.
7. Pakarinen E, Lindholm P. CDNF and MANF in the brain dopamine system and their potential as treatment for Parkinson's disease. Front Psychiatry. 2023;14:1188697.
8. Lindholm P, Saarma M. Cerebral dopamine neurotrophic factor protects and repairs dopamine neurons by novel mechanism. Mol Psychiatry. 2022;27(3):1310-1321.
9. Cordero-Llana Ó, Houghton BC, Rinaldi F, et al. Enhanced efficacy of the CDNF/MANF family by combined intranigral overexpression in the 6-OHDA rat model of Parkinson's disease. Mol Ther. 2015;23(2):244-254.
10. Kemppainen S, Lindholm P, Galli E, et al. Cerebral dopamine neurotrophic factor improves long-term memory in APP/PS1 transgenic mice modeling Alzheimer's disease as well as in wild-type mice. Behav Brain Res. 2015;291:1-11.
11. Stepanova P, Srinivasan V, Lindholm D, Voutilainen MH. Cerebral dopamine neurotrophic factor (CDNF) protects against quinolinic acid-induced toxicity in in vitro and in vivo models of Huntington's disease. Sci Rep. 2020;10(1):19045.
12. Stepanova P, Kumar D, Cavonius K, et al. Beneficial behavioral effects of chronic cerebral dopamine neurotrophic factor (CDNF) infusion in the N171-82Q transgenic model of Huntington's disease. Sci Rep. 2023;13(1):2953.
13. De Lorenzo F, Lüningschrör P, Nam J, et al. CDNF rescues motor neurons in models of amyotrophic lateral sclerosis by targeting endoplasmic reticulum stress. Brain. 2023;146(9):3783-3799.
14. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet. 2019;20(11):631-656.

15. Hong M, Tao S, Zhang L, et al. RNA sequencing: new technologies and applications in cancer research. J Hematol Oncol. 2020;13(1):166.
16. Müller-Nedebock AC, Cuttler K, Pfaff AL, Kõks S. Longitudinal dysregulation of long non-coding RNAs in Parkinson's disease. Exp Biol Med (Maywood). 2023;248(20):1780-178
17. Re S, Dogan AA, Ben-Shachar D, et al. Improved Generation of Induced Pluripotent Stem Cells From Hair Derived Keratinocytes - A Tool to Study Neurodevelopmental Disorders as ADHD. Front Cell Neurosci. 2018;12:321.
18. Wüstner LS, Klingenstein M, Frey KG, Nikbin MR, Milazzo A, Kleger A, Liebau S, Klingenstein S. Generating iPSCs with a High-Efficient, Non-Invasive Method—An Improved Way to Cultivate Keratinocytes from Plucked Hair for Reprogramming. Cells. 2022; 11(12):1955.
19. Aasen T, Izpisúa Belmonte JC. Isolation and cultivation of human keratinocytes from skin or plucked hair for the generation of induced pluripotent stem cells. Nat Protoc. 2010;5(2):371-382.
20. Aasen T, Raya A, Barrero MJ, et al. Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol. 2008;26(11):1276-1284.
21. Watson JD, Crick FH. The structure of DNA. Cold Spring Harb Symp Quant Biol. 1953;18:123-131.
22. Shendure, J., Balasubramanian, S., Church, G. et al. DNA sequencing at 40: past, present and future. Nature. 2017:550, 345–353.
23. Wu, R. & Kaiser, A. D. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. J. Mol . Biol.1968; 35, 523–537.
24. Gilbert, W. & Maxam, A. The nucleotide sequence of the lac operator. Proc. Natl Acad. Sci. 1973; 3581–3584.
25. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. Proc. Natl Acad. Sci., 1977;745463–5467.
26. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. Proc. Natl Acad. Sci. 1977;74, 560–564.
27. Staden R. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 1979;6(7):2601-2610
28. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016;107(1):1-8.
29. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
30. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science. 2001;291(5507):1304-1351.
31. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004:431, 931–945.

32. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-351.
33. Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res. 2006;34(3):e22.
34. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci U S A. 2003;100(15):8817-8822.
35. Severins I, Joo C, van Noort J. Exploring molecular biology in sequence space: The road to next-generation single-molecule biophysics. Mol Cell. 2022;82(10):1788-1805.
36. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010;327(5961):78-81.
37. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 1996;242(1):84-89.
38. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057).
39. Tomkinson AE, Vijayakumar S, Pascal JM, Ellenberger T. DNA ligases: structure, reaction mechanism, and function. Chem Rev. 2006;106(2):687-699.
40. Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays [published correction appears in Nat Biotechnol 2000 Oct;18(10):1021]. Nat Biotechnol. 2000;18(6):630-634.
41. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. Nat Genet. 2007;39(7 Suppl):S37-S42.
42. Mirkin SM. Expandable DNA repeats and human disease. Nature. 2007;447(7147):932-940.
43. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med. 2010;61:437-455.
44. Ardui S, Ameur A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res. 2018;46(5):2159-2168.
45. Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science. 2003;299(5607):682-686.
46. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol. 2009;4(4):265-270.
47. Arslan, S., Garcia, F.J., Guo, M. et al. Sequencing by avidity enables high accuracy with low reagent consumption. Nature Biotechnol. 2023.
48. Rare Diseases. [Internet]. [cited 2023 December 10]. Available from: https://health.ec.europa.eu/european-reference-networks/rare-diseases_en

49. Rare Disease at FDA. [Internet]. [cited 2023 December 10]. Available from: https://www.fda.gov/patients/rare-diseases-fda
50. Rare Disease Database. [Internet]. [cited 2023 December 11]. Available from: https://rarediseases.org/rare-diseases/
51. Rare Diseases Europe. [Internet]. [cited 2023 December 11]. Available from: https://www.eurordis.org/
52. Nguengang Wakap, S., Lambert, D.M., Olry, A. et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet . 2020;**28**, 165–173.
53. Undiagnosed Diseases Network International. [Internet]. [cited 2023 December 11]. Available from:https://www.udninternational.org/
54. Manolio TA, Collins FS. The HapMap and genome-wide association studies in diagnosis and therapy. Annu Rev Med. 2009;60:443-456.
55. Haplotype Map. [Internet]. [cited 2023 December 12]. Available from:https://www.broadinstitute.org/international-haplotype-map-project/haplotype-map
56. HapMap3. [Internet]. [cited 2023 December 12]. Available from:https://www.broadinstitute.org/medical-and-population-genetics/hapmap-3
57. Takeuchi F, Yanai K, Morii T, et al. Linkage disequilibrium grouping of single nucleotide polymorphisms (SNPs) reflecting haplotype phylogeny for efficient selection of tag SNPs. Genetics. 2005;170(1):291-304.
58. Negro SS, Millet EJ, Madur D, et al. Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. BMC Plant Biol. 2019;19(1):318.
59. GWAS Catalog. [Internet]. [cited 2023 December 12]. Available from:https://www.ebi.ac.uk/gwas/docs/about
60. Ku CS, Loy EY, Pawitan Y, Chia KS. The pursuit of genome-wide association studies: where are we now?. J Hum Genet. 2010;55(4):195-206.
61. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-753.
62. GnomAD Browser. [Internet]. [cited 2023 December 13]. Available from: https://gnomad.broadinstitute.org/about
63. The International Genome Sample Resource (IGSR). [Internet]. [cited 2023 December 13]. Available from: https://www.internationalgenome.org/sample_collection_principles
64. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-291.
65. Ewans, L.J., Minoche, A.E., Schofield, D. et al. Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis. Eur J Hum Genet. 2022;**30**:1121–1131.
66. Structural Variants. [Internet]. [cited 2023 December 105. Available from: https://gatk.broadinstitute.org/hc/en-us/articles/9022476791323-Structural-Variants

67. Barbitoff, Y.A., Polev, D.E., Glotov, A.S. et al. Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. Sci Rep.2020;257(10).
68. Chilamakuri, C.S.R., Lorenz, S., Madoui, MA. et al. Performance comparison of four exome capture systems for deep sequencing. BMC Genomics. 2014;449(15).
69. Barbitoff, Y.A., Abasov, R., Tvorogova, V.E. et al. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. BMC Genomics. 2022;155(23).
70. Toda, T. Genetics of neurological and psychiatric disorders. J Hum Genet. 2023;113(**68)**. 113
71. Macleod S, Appleton RE. Neurological disorders presenting mainly in adolescence. Arch Dis Child. 2007;92(2):170-175
72. .Han S, Gim Y, Jang EH, Hur EM. Functions and dysfunctions of oligodendrocytes in neurodegenerative diseases. Front Cell Neurosci. 2022;16:1083159.
73. Murakami T, Sango K, Watabe K, et al. Schwann cells contribute to neurodegeneration in transthyretin amyloidosis. J Neurochem. 2015;134(1):66-74.
74. Wilson DM 3rd, Cookson MR, Van Den Bosch L, Zetterberg H, Holtzman DM, Dewachter I. Hallmarks of neurodegenerative diseases. Cell. 2023;186(4):693-714.
75. Dugger BN, Dickson DW. Pathology of Neurodegenerative Diseases. Cold Spring Harb Perspect Biol. 2017;9(7):a028035.
76. An Online Catalog of Human Genes and Genetic Disorders. [Internet]. [cited 2023 December 17]. Available from: https://www.omim.org/
77. Andrade-Guerrero J, Santiago-Balmaseda A, Jeronimo-Aguilar P, et al. Alzheimer's Disease: An Updated Overview of Its Genetics. Int J Mol Sci. 2023;24(4):3754.
78. MacLeod R, Hillert EK, Cameron RT, Baillie GS. The role and therapeutic targeting of α-, β- and γ-secretase in Alzheimer's disease. Future Sci OA. 2015;1(3):FSO11.
79. Bartels T. A traffic jam leads to Lewy bodies. Nat Neurosci. 2019;22(7):1043-1045.
80. Deng H, Wang P, Jankovic J. The genetics of Parkinson disease. Ageing Res Rev. 2018;42:72-85.
81. Qu J, Zou T, Lin Z. The Roles of the Ubiquitin-Proteasome System in the Endoplasmic Reticulum Stress Pathway. Int J Mol Sci. 2021;22(4):1526.
82. Cybulsky AV. The intersecting roles of endoplasmic reticulum stress, ubiquitin- proteasome system, and autophagy in the pathogenesis of proteinuric kidney disease.
83. Zheng Q, Huang T, Zhang L, et al. Dysregulation of Ubiquitin-Proteasome System in Neurodegenerative Diseases. Front Aging Neurosci. 2016;8:303.
84. Quinn, P.M.J., Moreira, P.I., Ambrósio, A.F. et al. PINK1/PARKIN signalling in neurodegeneration and neuroinflammation. acta neuropathol commun . 2020; 189(**8).**
85. Samanci B, Bilgiç B, Gelişin Ö, et al. TREM2 variants as a possible cause of frontotemporal dementia with distinct neuroimaging features. Eur J Neurol. 2021;28(8):2603-2613.

86. Berdyński, M., Miszta, P., Safranow, K. et al. SOD1 mutations associated with amyotrophic lateral sclerosis analysis of variant severity. Sci Rep. 2022;103(12).
87. Rekatsina M, Paladini A, Piroli A, Zis P, Pergolizzi JV, Varrassi G. Pathophysiology and Therapeutic Perspectives of Oxidative Stress and Neurodegenerative Diseases: A Narrative Review. Adv Ther. 2020;37(1):113-139.
88. Liu Z, Zhou T, Ziegler AC, Dimitrion P, Zuo L. Oxidative Stress in Neurodegenerative Diseases: From Molecular Mechanisms to Clinical Applications. Oxid Med Cell Longev. 2017;2017:2525967
89. Skaper SD. Neurotrophic Factors: An Overview. Methods Mol Biol. 2018;1727:1-17.
90. Kim, J., He, M.J., Widmann, A.K. et al. The role of neurotrophic factors in novel, rapid psychiatric treatments. Neuropsychopharmacol. 2024;**49**, 227–245.
91. Wise, B.L., Seidel, M.F. & Lane, N.E. The evolution of nerve growth factor inhibition in clinical medicine. Nat Rev Rheumatol. 2021;**17**, 34–46.
92. Bai, M., Vozdek, R., Hnízda, A. et al. Conserved roles of C. elegans and human MANFs in sulfatide binding and cytoprotection. Nat Commun. 2018;9;897.
93. Chen, X., Shi, C., He, M. et al. Endoplasmic reticulum stress: molecular mechanism and therapeutic targets. Sig Transduct Target Ther **8**, 352 (2023).
94. Adams CJ, Kopp MC, Larburu N, Nowak PR, Ali MMU. Structure and Molecular Mechanism of ER Stress Signaling by the Unfolded Protein Response Signal Activator IRE1. Front Mol Biosci. 2019;6:11.
95. Zhang GL, Wang LH, Liu XY, et al. Cerebral Dopamine Neurotrophic Factor (CDNF) Has Neuroprotective Effects against Cerebral Ischemia That May Occur through the Endoplasmic Reticulum Stress Pathway. Int J Mol Sci. 2018;19(7):1905.
96. Eesmaa A, Yu LY, Göös H, et al. CDNF Interacts with ER Chaperones and Requires UPR Sensors to Promote Neuronal Survival. Int J Mol Sci. 2022;23(16):9489.
97. Eesmaa A, Yu LY, Göös H, et al. The cytoprotective protein MANF promotes neuronal survival independently from its role as a GRP78 cofactor. J Biol Chem. 2021;296:100295.
98. Mei JM, Niu CS. Effects of CDNF on 6-OHDA-induced apoptosis in PC12 cells via modulation of Bcl-2/Bax and caspase-3 activation. Neurol Sci. 2014;35(8):1275-1280.
99. Zhao H, Cheng L, Liu Y, et al. Mechanisms of anti-inflammatory property of conserved dopamine neurotrophic factor: inhibition of JNK signaling in lipopolysaccharide-induced microglia. J Mol Neurosci. 2014;52(2):186-192
100. Zhang Y, Xiang Y, Wang X, et al. Cerebral dopamine neurotrophic factor protects microglia by combining with AKT and by regulating FoxO1/mTOR signaling during neuroinflammation [published correction appears in Biomed Pharmacother. 2019 Apr;112:108548
101. Nadella R, Voutilainen MH, Saarma M, et al. Transient transfection of human CDNF gene reduces the 6-hydroxydopamine-induced neuroinflammation in the rat substantia nigra. J Neuroinflammation. 2014;11:209.

102. Lindahl M, Chalazonitis A, Palm E, et al. Cerebral dopamine neurotrophic factor-deficiency leads to degeneration of enteric neurons and altered brain dopamine neuronal function in mice. Neurobiol Dis. 2020;134:104696.
103. Chen YC, Baronio D, Semenova S, Abdurakhmanova S, Panula P. Cerebral Dopamine Neurotrophic Factor Regulates Multiple Neuronal Subtypes and Behavior. J Neurosci. 2020;40(32):6146-6164.
104. Chalazonitis A, Li Z, Pham TD, et al. Cerebral dopamine neurotrophic factor is essential for enteric neuronal development, maintenance, and regulation of gastrointestinal transit [published correction appears in J Comp Neurol. 2023 Jan;531(1):186].
105. Lindholm P, Voutilainen MH, Laurén J, et al. Novel neurotrophic factor CDNF protects and rescues midbrain dopamine neurons in vivo. Nature. 2007;448(7149):73-77.
106. Voutilainen MH, Bäck S, Peränen J, et al. Chronic infusion of CDNF prevents 6-OHDA-induced deficits in a rat model of Parkinson's disease. Exp Neurol. 2011;228(1):99-108.
107. Airavaara M, Harvey BK, Voutilainen MH, et al. CDNF protects the nigrostriatal dopamine system and promotes recovery after MPTP treatment in mice. Cell Transplant. 2012;21(6):1213-1223.
108. Bäck S, Peränen J, Galli E, et al. Gene therapy with AAV2-CDNF provides functional benefits in a rat model of Parkinson's disease. Brain Behav. 2013;3(2):75-88.
109. Cordero-Llana Ó, Houghton BC, Rinaldi F, et al. Enhanced efficacy of the CDNF/MANF family by combined intranigral overexpression in the 6-OHDA rat model of Parkinson's disease. Mol Ther. 2015;23(2):244-254.
110. Nadella R, Voutilainen MH, Saarma M, et al. Transient transfection of human CDNF gene reduces the 6-hydroxydopamine-induced neuroinflammation in the rat substantia nigra. J Neuroinflammation. 2014;11:209.
111. Kemppainen S, Lindholm P, Galli E, et al. Cerebral dopamine neurotrophic factor improves long-term memory in APP/PS1 transgenic mice modeling Alzheimer's disease as well as in wild-type mice. Behav Brain Res. 2015;291:1-11.
112. Garea-Rodríguez E, Eesmaa A, Lindholm P, et al. Comparative Analysis of the Effects of Neurotrophic Factors CDNF and GDNF in a Nonhuman Primate Model of Parkinson's Disease. PLoS One. 2016;11(2):e0149776.
113. Wang L, Wang Z, Xu X, et al. Recombinant AAV8-mediated intrastriatal gene delivery of CDNF protects rats against methamphetamine neurotoxicity. Int J Med Sci. 2017;14(4):340-347.
114. Wang L, Wang Z, Zhu R, et al. Therapeutic efficacy of AAV8-mediated intrastriatal delivery of human cerebral dopamine neurotrophic factor in 6-OHDA-induced parkinsonian rat models with different disease progression. PLoS One. 2017;12(6):e0179476.
115. Stepanova P, Kumar D, Cavonius K, et al. Beneficial behavioral effects of chronic cerebral dopamine neurotrophic factor (CDNF) infusion in the N171-82Q transgenic model of Huntington's disease. Sci Rep. 2023;13(1):2953.

116. Stepanova P, Srinivasan V, Lindholm D, Voutilainen MH. Cerebral dopamine neurotrophic factor (CDNF) protects against quinolinic acid-induced toxicity in in vitro and in vivo models of Huntington's disease. Sci Rep. 2020;10(1):19045.
117. De Lorenzo F, Lüningschrör P, Nam J, et al. CDNF rescues motor neurons in models of amyotrophic lateral sclerosis by targeting endoplasmic reticulum stress. Brain. 2023;146(9):3783-3799.
118. Villa-Cedillo SA, Matta-Yee-Chig D, Soto-Domínguez A, et al. CDNF overexpression prevents motor-cognitive dysfunction by intrastriatal CPP-based delivery system in a Parkinson's disease animal model. Neuropeptides. 2023;102:102385.
119. Kulesskaya N, Bhattacharjee A, Holmström KM, et al. HER-096 is a CDNF-derived brain-penetrating peptidomimetic that protects dopaminergic neurons in a mouse synucleinopathy model of Parkinson's disease. Cell Chem Biol. Published online November 24, 2023.
120. Synofzik M, Haack TB, Kopajtich R, et al. Absence of BiP co-chaperone DNAJC3 causes diabetes mellitus and multisystemic neurodegeneration [published correction appears in Am J Hum Genet. 2015 Mar 5;96(3):514].

8 APPENDIX

APPENDIX 1: Chemicals and media components used in the study

Chemicals and Media Components	Supplier Company
Ethanol	Merck, Germany
Hydrochloric acid	Merck, Germany
Ultrapure distilled water	Invitrogen, USA
DMEM	Gibco, USA
DMEM/F12	Gibco, USA
FBS	Gibco, USA
MEM NEAA (100X)	Gibco, USA
GlutaMAX (100X)	Gibco, USA
Epilife Medium	Gibco, USA
EpiLife Defined Growth Supplement (EDGS)	Gibco, USA
KGM-Gold Keratinocyte Cell Basal Medium	Lonza, USA
KGM-Gold Supplement (HKGS)	Lonza, USA
Cryopreservation Medium, serum-free	CAPRICORN Sci, Germany
Rock Inhibitor	Adooq Bioscience, Germany
Collagen type I	STEMCELL Tech, USA
Geltrex™ LDEV-Free Reduced Growth Factor Basement Membrane Matrix	Gibco, USA
PBS	EuroClone, Italy
Penicillin-Streptomycin	CAPRICORN Sci, Germany
Antibiotic-Antimycotic	WISENT INC, Canada

APPENDIX 2: Equipment used in this study.

Equipment	**Company**
Autoclave	Nuve, Turkey
Balance	Mettler Toledo, USA
Centrifuge	Eppendorf, Germany
CO_2 Incubator	Esco, USA
DryBath	Thermo Sci, USA
Laminar Flow Hood	Thermo Sci, USA
Microscope	Leica, USA
Microliter Pipettes	Eppendorf, Germany
Serological Pipettes	Sarsted; Germany
Parafilm	Ted Pella, USA
Sterile Gloves	Kimberly Clark, Germany
Tubes (50ml, 15ml)	Corning, USA
Microcentrifuge Tubes	Corning, USA
Cryotubes	Thermo Sci, USA
100mm Petri Dishes	IsoLab, Turkey
TC Treated Plates (24well, 6 well)	Sorfa, China
Non-TC Treated Plates (6 well)	Sorfa, China
Flat tweezers	IsoLab, Turkey
Refrigerator ($+4^0$C)	KIRSCH, Germany
Deepfreeze (-80^0C, -20^0C)	Arctiko, Denmark
Hemocytometer	Hausser Scientific,USA

Cell Counter	Biorad, Germany
Vortex	Biosan, Latvia
Shaker	VWR, USA
Distilled Water	Milipore, France
Ice machine	Scotsman Inc., USA
Nitrogen Storage System	Thermo Sci., USA

9 CURRICULUM VITAE