Alma Mater Studiorum · Università di Bologna

Department of Pharmacy and Biotechnology

First Cycle Degree in Genomics

Phenotype-driven variant prioritization tools: analysis of Whole Exome Sequencing in patients with hereditary optic neuropathy

Candidate:

Giorgia Del Missier

Matriculation number:

0000830635

Supervisor:

Luisa Iommarini

Co-Supervisor:

Leonardo Caporali

A.Y. 2019-2020

6th session

Contents

Abstract

Hereditary optic neuropathies (HONs) are one of the most common causes of vision loss and often present a distinctive mitochondrial dysfunction. Most known causative genes encode mitochondrial proteins; however, half of the patients remain without genetic definition.

Variant discovery is carried out mainly by high-throughput approaches, such as Whole Exome Sequencing (WES). Phenotype-driven variant prioritization can be exploited to speed up the process of NGS data analysis and ease the research for the causative gene, in particular in the context of rare Mendelian diseases like HONs.

In this study, we used four tools (The Exomiser, MutationDistiller, GenIO and OVA) to match HPO (Human Phenotype Ontology) terms to variants from WES, with the aim of finding new disease genes for optic atrophy. Softwares were trained on a set of 11 solved whole-exomes and we kept only the first two tools for variant prioritization in unsolved WES cases.

Eight patients with a recessive pedigree and five dominant families, all with isolated or syndromic forms of optic atrophy, were subjected to gene prioritization. Some interesting new findings were obtained, including possibly pathogenic variants in genes not previously prioritized, including *SNF8*, *TSPOAP1*, *NYX* and *SYNJ1*.

To conclude, The Exomiser and MutationDistiller are effective support tools for rare Mendelian phenotype-based variant prioritization.

Chapter 1

INTRODUCTION

1.1 Mitochondrial disorders

Mitochondrial diseases are inherited disorders that mostly target the oxidative phosphorylation system (OXPHOS) and the process of ATP synthesis. They are caused by mutations either in the mitochondrial DNA (mtDNA) or in genes coding for proteins that make up the mitochondrial proteome but reside in the nuclear DNA (nDNA). As a group, they are one of the most frequent genetic disorders both in adults and children, affecting around 1:5000 of the population (Schon et al., 2020). Accordingly, the field of mitochondrial medicine has been growing exponentially in the last 30 years.

Mitochondrial diseases stand out for their enormous clinical heterogeneity, which might involve one or multiple organ systems, in any combination. The most commonly affected tissues are the ones with high energy demand. In fact, disorders with a mitochondrial etiology often target the brain, causing encephalopathies characterized by strokes, seizures and neurodevelopmental delay, skeletal muscles, including symptoms such as myopathy, exercise intolerance and weakness, the heart and sensory organs, specifically the eyes and ears. Peripheral neuropathy, kidney dysfunction as well as endocrine disorders like diabetes mellitus are also quite frequent (Rahman, 2020). Some of these phenotypes are common in the population and could be mistaken for phenocopies; oligosymptomatic patients and cases of incomplete penetrance make the clinical diagnosis an even more challenging task. As a result, many patients remain trapped in a so-called protracted "diagnostic odyssey" (Schon et al., 2020).

Moreover, from a molecular and genetic point of view, mitochondrial disorders are quite difficult to recognize: no single reliable biomarker exists, all kinds of mode of inheritance are reported and both familial and sporadic cases have been observed.

This genotype-phenotype complexity can be in part explained by their dual genetic deter-

mination and the need for mtDNA and nDNA to cross-talk and be tightly coordinated. Each cell harbours a dynamic network of budding and fusing mitochondria and contain hundreds to thousands copies of mtDNA, depending on the tissue type and developmental stage. The mitochondrial DNA is a double-stranded, circular molecule of ∼16.5 kb that codes for a total of 37 genes: 22 tRNAs and 2 rRNAs, needed for mtDNA translation, and 13 proteins that are essential respiratory chain subunits. During evolution, over 1000 other proteins that are involved in mitochondrial pathways were transferred to the nuclear DNA and hence need to be translated in the cytoplasm and then translocated into the mitochondria.

While it has long been known that these organelles have a role in energy generation through the OXPHOS, last decades studies highlight their implication in other complex cellular activities. Mitochondria, in fact, participate in classic biochemical pathways, for example the Krebs cycle, beta-oxidation and lipid metabolism, but are also involved in the organelle's dynamics, morphology and quality control as well as in cell signalling, including calcium homeostasis, immune response and the intrinsic pathway leading to apoptosis (Rath et al., 2020).

The recently updated MitoCarta3.0 inventory (`www.broadinstitute.org/mitocarta`) describes 1136 human proteins with strong evidence of being part of the mitochondrial proteome, encoded both by the nuclear and mitochondrial genomes. Insertions, deletions or single-nucleotide variants (SNVs) affecting them can then follow either classic Mendelian rules in the case of nDNA or the population genetics of a multicopy genome for mtDNA.

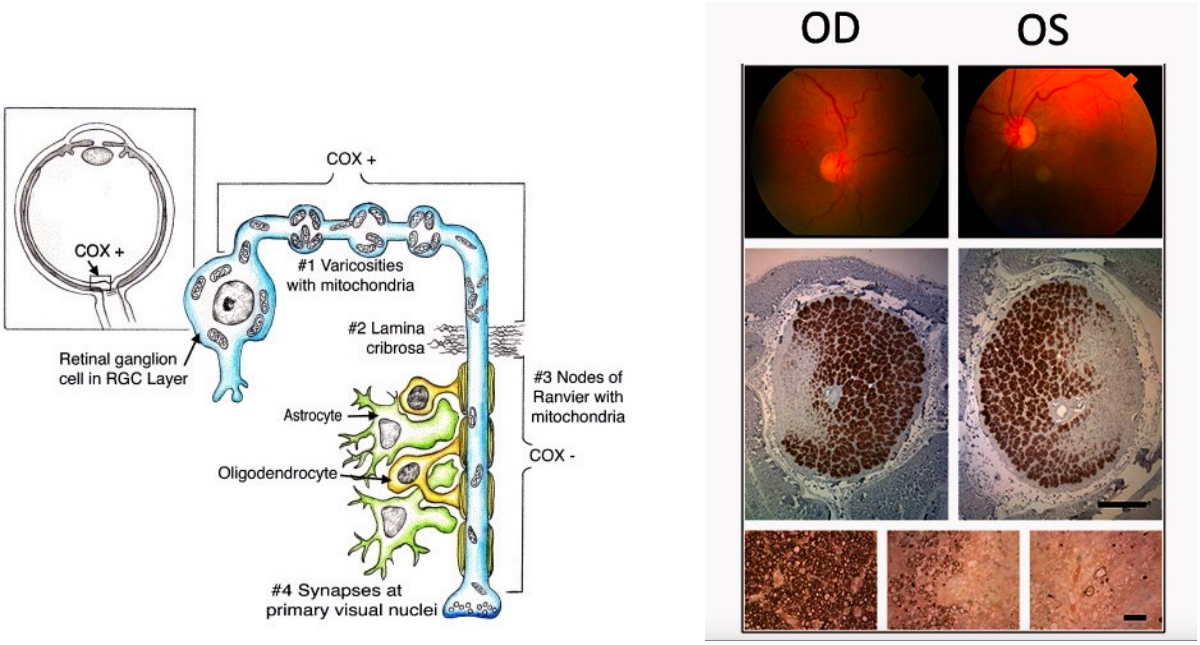
The latter in particular includes key features such as maternal inheritance and a high mutational rate that results in population-specific haplogroups. The heteroplasmy of mtDNA molecules, i.e. the coexistence of different sequences in the same cell, tissue or individual, is another critical aspect from the clinical point of view. In fact, frequently not all mitochondria harbour the pathogenic mutation and disease expression might depend on a "threshold-effect", in which a certain amount of mutated molecules is needed to induce the cellular dysfunction, or a phenotype severity that correlates with the heteroplasmy level (Stenton & Prokisch, 2020). Additionally, the heteroplasmic burden can vary depending on the cell and the tissue type and can also change over time, leading to worsening of the clinical phenotype but also to a loss of the pathogenic mutation, especially in rapidly

dividing tissues such as blood. Some variants instead affect all mtDNA molecules (homoplasmy) and tend to have reduced clinical penetrance (Stenton & Prokisch, 2020).
Thus, mitochondrial diseases may be subdivided into 3 main groups, depending on the underlying type of variant: primary mutations that directly affect mtDNA, secondary mutations of the mitochondrial genome that are a by-product of defects in proteins responsible for mtDNA maintenance and mutations that alter nuclear genes of the mitochondrial proteome. Hence, around 11% of the disease genes are encoded by the mtDNA, with over 1000 rearrangements and 300 possibly pathogenic point mutations reported, while the remaining 89% are the result of variants in the nuclear DNA (La Morgia et al., 2020; Stenton & Prokisch, 2020).

1.2 Optic neuropathies

The wide-phenotypic variability and complexity of these disorders can be exemplified by the group of mitochondrial optic neuropathies, which are characterized by selective dysfunction of the retinal ganglion cells (RGCs), whose axons compose the optic nerve (Figure 1.1a).
RGCs degeneration results clinically in central vision loss, central scotoma and pallor of the optic disc and may eventually lead to complete optic atrophy and legal blindness (Figure 1.1b). The tissue specificity might be explained by the need for mitochondria to be distributed in a skewed fashion: RGCs's axons are not myelinated in the intraretinal tract and, as a consequence, are highly dependent on energy, and thus mitochondria, to drive potentials (La Morgia et al., 2020). Optic neuropathies can be non-syndromic or be one of the phenotypes in a more complex, multisystem disorder. The most frequent are Leber's hereditary optic neuropathy (LHON) and dominant optic atrophy (DOA).

LHON was the first disorder found to be caused by pathogenic point mutations in the mitochondrial genome. Currently, over 90% of LHON patients harbour one of three variants, all affecting genes coding for subunits of Complex I (m.11778G>A/*MT-ND4*, m.3640G>A/*MT-ND1*, m.14484T>C/*MT-ND6*). More recently, other rarer mtDNA mutations have been shown to cause LHON, making up the remaining 10% of cases (Maresca et al., 2014).



(a) (b)

Figure 1.1: **The RCG system and selective cellular degeneration.** (a) The RGC cellular system and the skewed distribution of mitochondria in the unmyelinated portion of the axon. Source: Carelli et al. (2004). (b) Fundus picture and cross-sectional optic nerve histopathology of a LHON case, demonstrating pallor of the temporal sector of the optic disc with selective loss of the axons of the papillomacular bundle. Source: Carelli et al. (2015).

Most of the time, LHON mutations are homoplasmic; however, familial pedigrees often show incomplete penetrance, where some individuals never develop the clinical phenotype. Another peculiar feature of LHON is the male prevalence, with about 50% of men affected and only 10% of women (La Morgia et al., 2020). Despite many attempts to find an involvement of chromosome X, up to now no significant results were found. This difference is now partially explained by the metabolic role played by estrogen and its ability to increase antioxidant defenses and activate mitochondrial biogenesis, a common compensatory mechanism found in mitochondrial diseases (Pisano et al., 2015). Environmental factors such as tobacco and alcohol consumption have been shown to be associated with expression and penetrance of the disease, by means of preventing mitochondrial biogenesis (Kirkman et al., 2009). Finally, European mtDNA background seems to be relevant: functional studies supported the role of specific variants found in haplogroup J, which affect Complex I and III subunit genes and might modulate OXPHOS reactions and reactive oxygen species production (Pello et al., 2008). It should be taken into notice that, in spite of all these findings, penetrance is still incomplete, as shown by the Brazilian

pedigree SOA-BR (Sadun et al., 2013) pointing to the possibility that other non-genetic modifiers might be implicated (Maresca et al., 2014).

Furthermore, a study published this year has given new insights on the enigma of LHON in the absence of pathogenic mtDNA variants: biallelic mutations in the nuclear gene *DNAJC30* can result in a recessive phenocopy of mtLHON, showing all its peculiar hallmarks, including incomplete penetrance and male predominance. Knock-out cellular models demonstrated DNAJC30 to a be a chaperone protein needed for exchange of Complex I subunits after reactive oxygen species exposure and hence essential as repair mechanism (Stenton et al., 2021).

From the clinical point of view, LHON arises early in life, with a median age of onset of 20 years (Rahman, 2020). Patients usually undergo rapid and painless central visual loss, in most cases bilateral or sequential, which stabilizes within the first year after onset. The endpoint is optic atrophy and permanent loss of central vision. Spontaneous recovery has been observed occasionally, with the most favourable prognostic factors being young age of onset and the mutation type, with highest rate in the case of m.14484/*MT-ND6* (Maresca et al., 2013).

LHON is molecularly characterized by a high selectivity for tissue expression, targeting specifically RCGs and the small fibers of the papillomacular bundle involved in central vision. However, LHON-like optic neuropathy has been also described in association with more complex clinical features, overlapping with neurological syndromes such as MELAS (mitochondrial encephalomyopathy, lactic acidosis, stroke-like episodes) and Leigh syndrome. This "plus" phenotype can vary within families, spanning from single ophthalmological symptoms to severe encephalomyopathies, even in cases where the same mtDNA mutation is the causative one; adjunctive private mtDNA variants may be the cause of modulation of clinical expression (Maresca et al., 2014).

DOA clinically presents with a slow, progressive loss of central vision during childhood, even though this may change depending on individuals, even of the same pedigree, showing some phenotype variability. The endpoint is, as in LHON, legal blindness. In this case, however, no spontaneous recovery, gender differences or influence of environmental factors have been observed or fully investigated (Maresca et al., 2014).

In 70% of cases, DOA is associated with heterozygous mutations in the nuclear *OPA1*

gene. This protein is involved in important mitochondrial pathways, such as fusion of the inner membrane, cristae morphology, control of apoptosis and mtDNA stability and maintenance of the respiratory activity and of membrane potential. Pathogenic point mutations in *OPA1* can result either in a truncated protein, leading to haploinsufficiency, or in a dominant negative model of inheritance in the case of missense mutations.

However, disease development has been associated with many other variants, that can follow different inheritance patterns: heterozygous mutations in the *OPA3* and *WFS1* genes result in DOA with cataract and deafness, respectively (Reynier, 2004; Eiberg, 2005). *DNM1L* (*OPA5*) encoding DRP1, a protein involved in mitochondrial fission, has also been associated with dominant optic atrophy (Gerber et al., 2017). More recently, heterozygous mutations in *AFG3L2* and *SPG7* were described: these proteins make up a mitochondrial matrix AAA metalloprotease, an ATP-dependent proteolytic complex that carries out protein quality control in the IMM and takes part in processing of some mitochondrial proteins (Caporali et al., 2020; Charif et al., 2020). Some loci for DOA that await identification of the causative genes include *OPA4* (Kerrison, 1999), *OPA6* (Barbet et al., 2003) and *OPA8* (Carelli et al., 2011). Rarer forms of genetic optic neuropathies are X-linked, as in the case of the *OPA2* locus (Assink et al., 1997; Katz et al., 2006), or recessive, including mutations in the genes *TMEM126A* (Hanein et al., 2009) and *RTN4IP1* (Charif et al., 2018).

Furthermore, the mitochondrial single-strand binding protein 1 (*SSBP1*), a replisome factor, and aconitase 2, a tricarboxylic acid cycle enzyme, have been recently identified as causative genes for optic atrophy, following both dominant and recessive mode of inheritance (Del Dotto et al., 2020, Metodiev et al. 2014, Neumann et al. 2020).

The picture gets even more complex when syndromic cases are taken into consideration: a DOA "plus" phenotype was described for specific *OPA1* missense mutations affecting the GTPase domain and involving more severe features, such as sensorineural deafness, cerebellar ataxia, axonal sensory-motor polyneuropathy, parkinsonism and dementia (Yu-Wai-Man et al., 2010; Carelli et al., 2015). Moreover, all these patients had a mitochondrial myopathy with accumulation of mtDNA deletions, most probably due to OPA1 inability to carry out mtDNA maintenance (Amati-Bonneau et al., 2008).

Both *WSF1* and *OPA3* with recessive mutations lead to syndromic forms of DOA, known as Wolfram and Costeff syndromes (Maresca et al., 2013); Caporali et al. (2020) and Dotto

et al. (2019) also reported patients with additional clinical symptoms in the cases of *AFG3L2* and *SSBP1* mutations, respectively. Many other mitochondrial genes, causative for multisystem disorders that also target the eye can be found in the literature, such as *SLC25A46* (Abrams et al., 2015), *MECR* (Heimer et al., 2016), *NDUFAF2* (Ogilvie, 2005), *MTFMT* (Haack et al., 2012) and *POLG* (Felhi et al., 2019).

1.3 Mitochondrial pathways

This genetic and phenotypic complexity is the effect of mitochondrial biology. In fact, OXPHOS defects represent the hallmark of mitochondrial disorders, but a wide number of proteins with very different functional roles converge and influence the overall OXPHOS reaction (Figure 1.2). That is why it is interesting to better understand the pathways leading to RCGs neurodegeneration (Maresca et al., 2014).

OXPHOS reactions are carried out by means of five large protein complexes (Complexes I to V) and two mobile electron carriers (Coenzyme Q and Cytochrome c). Complex I, the largest, with subunits encoded both by mtDNA and nDNA, transfers electrons from NADH to CoQ while translocating four protons from the matrix to the inner mitochondrial space. During this process, electrons may escape leading to the production of reactive oxygen species (ROS) (Duchen, 2004). CI dysfunction is a common feature of mitochondrial disorders and optic neuropathies: all three LHON mutations affect CI subunits and also OPA1 has been found to interact with the complex assembly process, also helping stabilization (Zanna et al., 2007).

ROS are always present at low concentration and are essential in cell signalling. However, if their production increases and overcomes a threshold, they can damage DNA, proteins and lipids, further provoking genomic instability and mutation accumulation (Maresca et al., 2013). Increased oxidative stress in DOA has been demonstrated in *Drosophila* models, thus providing a link between ROS and optic atrophy (Yarosh et al., 2008; S. Tang et al., 2009).

Mitochondrial biogenesis is a common compensatory mechanism in mitochondrial disorders and requires expression of genes from the nDNA and mtDNA. This process can be induced by estrogen and in some cases also by ROS. Other ways used by cells to make up for energetic defects is increase in mitochondrial mass, incremented Complex III activity and a greater mtDNA copy number (Maresca et al., 2013).

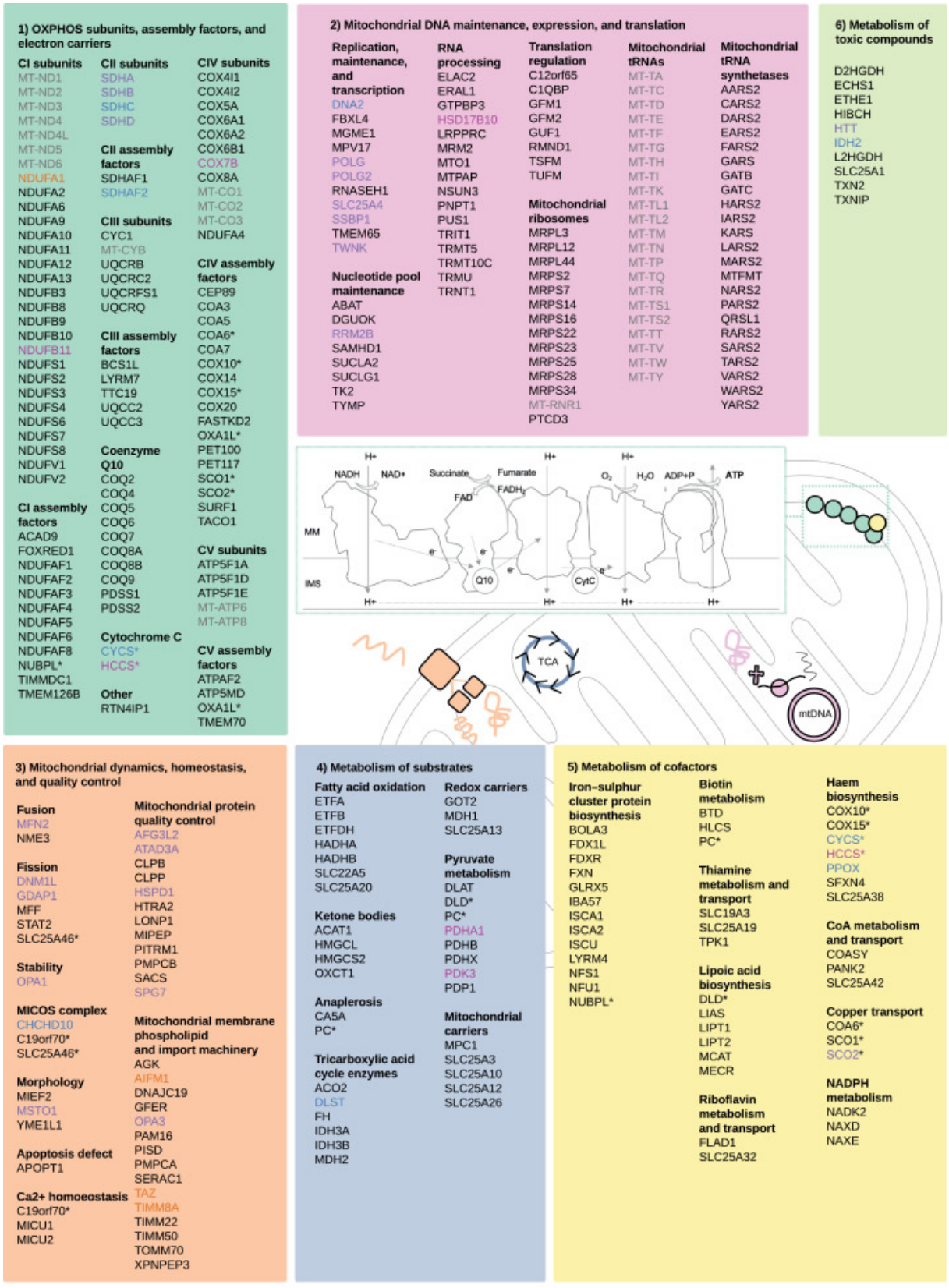


Figure 1.2: **Mitochondrial disease genes.** 338 mitochondrial disease genes, each demonstrated to cause a defect in an enzyme involved in aerobic energy metabolism. The genes are divided into six subsets according to their functional roles; the inheritance mode is also reported.

Source: Stenton and Prokisch (2020).

Mitochondria are also involved in the intrinsic apoptosis pathway, in which caspases are activated giving rise to proteolytic cascades and leading to cell death (Taylor et al., 2008): this might be the last step in the neurodegenerative process leading to optic neuropathies. Lastly, mitochondrial dynamics resulting from a balance between fusion and fission is responsible for the morphology of this organelle and needs to be tightly regulated. In the last 15 years, many proteins involved in this pathway have been identified, including mitofusins MFN1 and MFN2 and OPA1. Better understanding and characterization of the role played by fission-fusion imbalance is thus required to explain the pathogenesis of optic neuropathies (Maresca et al., 2013).

1.4 Next Generation Sequencing for mitochondrial disorders

Candidate gene and mtDNA sequencing combined with linkage analysis were the conventional methods used for disease gene discovery up until the first decade of the 2000s. Over 150 genes in a time period of 25 years were uncovered, but a rapid acceleration was marked by the transition into the Next Generation Sequencing (NGS) era: if pre-NGS five new disease genes were identified per year, this number increased to 15 thanks to the use of high-throughput techniques (Stenton & Prokisch, 2020). This exponential growth has slowed down in the last 3 years, suggesting that the so-called "low-hanging fruit" has already been picked (Figure 1.3). However, only 1/3 of the predicted mitochondrial genes has yet been associated with specific disorders. It might be that some mutations are particularly rare, or that they affect essential genes whose dysfunction is not compatible with life. On the contrary, other predicted mitochondrial genes might be functionally redundant and not pathogenic if mutated. Still, some need yet to be discovered (Schon et al., 2020).

Different genetic analyses exist in order to find the causative variant: candidate gene sequencing is still the method of choice in cases of clear clinical syndromes but, most times, NGS is needed in order to reach a clinical diagnosis.

Multiple approaches can be used to find pathogenic mutations in the context of nuclear DNA: large panels targeted to known genes involved in mitochondrial diseases, ranging from 100 to over 1000, reach a diagnostic rate of 7-31%. The main pros of using panels are the possibility to reach a high coverage (200x) and the easy interpretation of results; however, continual discovery of new genes makes them short-lived and the need for a

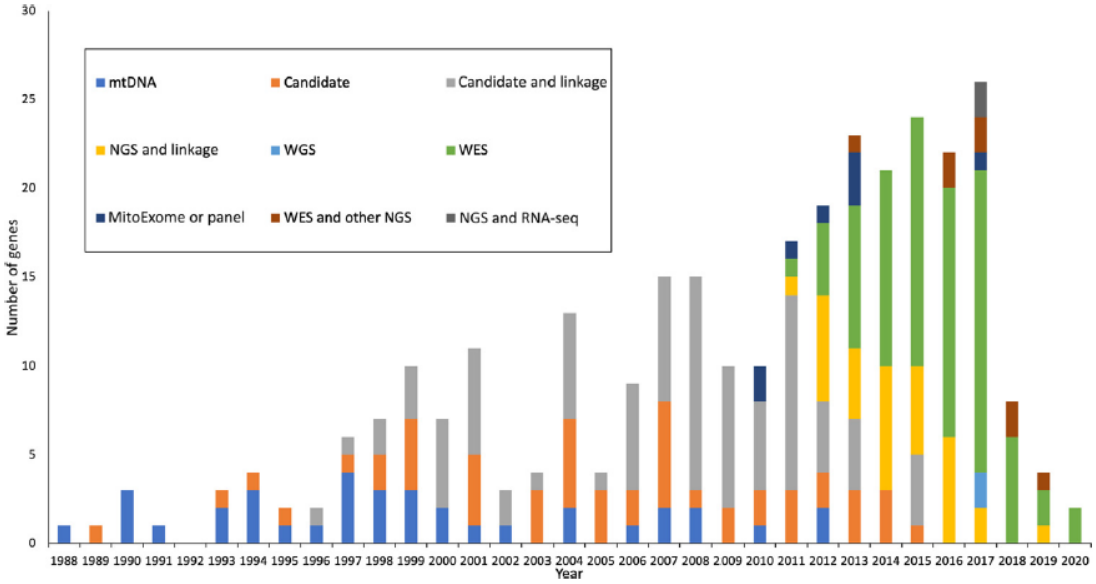


Figure 1.3: **Discovery rate of novel genes underlying mitochondrial disorders.** The graph illustrates the shift from mtDNA and candidate gene sequencing to NGS, which is also reflected by an acceleration in gene discovery.

Source: Schon et al. (2020).

correct *a priori* hypothesis is essential.

Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS) interrogate protein coding and all regions of the genome, respectively. Evaluation of WES reached diagnostic yields of 35-70% and an improvement of 2% with respect to WES is seen when considering WGS. These approaches are unbiased, since they do not require previous hypotheses on the causative gene and can also investigate novel ones, not previously implicated in the disease. Also, mitochondrial disease phenocopies might be identified in this way. Furthermore, re-analysis of negative cases is possible: very often, a clinical diagnosis is reached in 1-3 years after the initial analysis, mostly thanks to the discovery of novel genes. WES and WGS can be performed on trios, by parallel analysis of the proband and parents, in order to identify *de novo* variants.

The main disadvantage of sequencing just the exome is its inability to detect mosaic variants that require deep sequencing, large and small deletions/insertions – less and more than 50bp respectively, and chromosomal rearrangements such as translocations, inversions and duplications; repeats, deep intronic and regulatory variants cannot be examined, either.

Some of these challenges can be overcome using WGS: this technique was recently used to identify three novel mitochondrial disease genes where the causative variants were an intronic deletion, a deep intronic mutation and a 3'UTR duplication that all resulted in aberrant splicing (Kremer et al., 2017; Tamiya et al., 2014; Malicdan et al., 2017). Therefore, Schon et al. (2020) suggest, as soon as costs fall, a transition to a 'WGS-first' approach, in which a single diagnostic pipeline can be followed for all patients.

The pitfall of Next Generation Sequencing is the discovery of many Variants of Uncertain Significance (VUS). The American College of Medical Genetics (ACMG) provided guidelines to classify a variant's impact through different criteria, creating a 5-tier system going from "Pathogenic", in which the variant has a 90-99% chance of being disease-causing, to a "Likely benign" classification; VUS are variants for which there is not enough evidence to determine whether they might be associated with a disorder or not.
Further analyses are then needed to assess their clinical significance: familial segregation remains the most powerful way to ascertain causality, but other recent advances can be exploited. The same applies in the case of negative WES and WGS results.
The first way to improve genome-wide variant interpretation is the use of transcriptomics, which provides information on mRNA transcription and can detect aberrant splicing, anomalous low or high expression of genes, imbalance in allele specific expression or a combination of these events. Transcriptome analysis is often the missing link in a diagnosis and even though it is a still evolving field, diagnostic rates vary from 8 to 35% following negative WES or WGS. High-throughput techniques do not perform as well in such a scenario because splicing algorithms might discredit the real causative variant, due to wrong prediction on their disruptive effect. Current challenges in RNA-sequencing are tissue selection and data analysis.
Proteomics might also be used to quantify all detected proteins in one assay and determine destabilization of the protein structure, in particular in the case of missense variants.
Quantification of metabolites is another "omics" technique that can give clues on genetic defects, even though generation and data analysis is still quite complex and further studies are needed to establish its true value.
Structural rearrangements identification is limited by NGS's short-read length: third-generation techniques such as Nanopore have longer reads, but still suffer a relatively high error-rate and cost.

Data sharing is also very important and can aid diagnosis in rare patients. In this context, phenomics, i.e. the systematic study of measurable physical and biochemical attributes of an individual, can be harnessed. To this aim, the Human Phenotype Ontology (`https://hpo.jax.org/app/`) was developed, in order to provide a standardized vocabulary for phenotypes that can be used to describe rare disorders symptoms.

1.5 Scope of the thesis

Discovery of novel genes associated with mitochondrial disorders is an ever-expanding field of research. Nowadays, the preferred way to identify pathogenic mutations is through the use of high-throughput techniques, in particular Whole Exome Sequencing, and this approach well applies to optic atrophy. This rare disease affects more than 50.000 persons in Europe and mutations in mtDNA genes coding for Complex I subunits and OPA1 are the most frequent genetic cause. Other genes have been associated with non-syndromic or syndromic optic atrophy, but over 50% of patients remain undiagnosed, implying genetic heterogeneity and other causative genes that need to be discovered.

The most difficult task at the moment is interpretation of Next Generation Sequencing data: this step is often carried out manually by researchers and thus subject to human error. As a consequence, in the last decade an increasing number of computational algorithms have been implemented in order to perform data analysis and identify possibly pathogenic mutations in the most unbiased way as possible.

The scope of this study is to analyse WES data in a cohort of patients with hereditary optic neuropathies, negative for known genes, to identify the genetic defect.

To this end, four phenotype-driven variant prioritization tools, The Exomiser, MutationDistiller, GenIO and OVA, will be first evaluated using patients' data for which the causative gene has already been identified to test their performance in our cohort.

Then, the best performing tools will be used in order to look out for variants in genes that might be involved in mitochondrial pathways, even not yet associated with any pathology, or that might lead to phenotypes similar to that of the patient under analysis.

Chapter 2

MATERIALS AND METHODS

2.1 Cohort description

The cohort under analysis consists in 24 families (29 individuals), with isolated or syndromic optic atrophy, clinically and genetically screened at Units of Neurogenetics and Clinical Neurology of IRCSS Institute of Neurological Sciences of Bologna, within the Ministry of Health Project "Italian Project on Hereditary Optic Neuropathies (IPHON): from genetic basis to therapy" (GR-2016-02361449). All patients presented with clinical hallmarks of mitochondrial hereditary optic neuropathies (HON) and their phenotypes were annotated through HPO-encoded terms.
Genomic DNA was extracted using standard procedures and subjects' DNA was analysed by Whole Exome Sequencing (WES) using different exome targets, as shown in Table 2.1.

ID	**Diagnosis**	**HPO terms**	**Library prep**	**Exome target**
NG_1	Optic atrophy and polyneuropathy	HP:0000648 HP:0008587 HP:0003390 HP:0003128 HP:0002514 HP:0002624 HP:0100502	Nextera Rapid Capture	Illumina

ID	Diagnosis	HPO terms	Library prep	Exome target
NG_4	Mitochondrial encephalomyopathy	HP:0000648 HP:0004901 HP:0002075 HP:0002403 HP:0030891 HP:0003390 HP:0002076 HP:0100507	Nextera Rapid Capture	Illumina
NG_5	LHON	HP:0000648 HP:0004901	Nextera Rapid Capture	Illumina
NG_7	LHON	HP:0000648 HP:0004901 HP:0000739	Nextera Rapid Capture	Illumina
NG_12	Optic atrophy	HP:0000648 HP:0007642 HP:0000556	Nextera Rapid Capture	Illumina
NG_16	Optic atrophy	HP:0000648 HP:0001889 HP:0100507 HP:0005264	Nextera Rapid Capture	Illumina
NG_19	Optic atrophy	HP:0000648 HP:0000407	Nextera Rapid Capture	Illumina
NG_39	Optic atrophy	HP:0000648	Nextera Rapid Capture	Illumina
NG_73	Dominant optic atrophy	HP:0000648 HP:0000822	Lotus IDT	xGen Exome v1 IDT
NG_110	Dominant optic atrophy	HP:0000648	NEBNext Ultra II	xGen Exome v1 IDT
NG_127	Dominant optic atrophy	HP:0000648	NEBNext Ultra II	xGen Exome v1 IDT
NG_3	Optic atrophy	HP:0000648	Nextera Rapid Capture	Illumina

ID	Diagnosis	HPO terms	Library prep	Exome target
NG_21	Optic atrophy	HP:0000648 HP:0001249 HP:0000708 HP:0000609 HP:0025163 HP:0002079	Nextera Rapid Capture	Illumina
NG_22	Optic atrophy	HP:0000648 HP:0011342 HP:0001249 HP:0002079 HP:0000639 HP:0002066 HP:0030217	Nextera Rapid Capture	Illumina
NG_23	Optic atrophy	HP:0000648	Nextera Rapid Capture	Illumina
NG_43	Optic atrophy	HP:0000648 HP:0003002 HP:0000563 HP:0011505	Nextera Rapid Capture	Illumina
NG_47	Optic atrophy	HP:0000648 HP:0011342 HP:0001272 HP:0003458 HP:0001260 HP:0012043 HP:0020036 HP:0001348	Nextera Rapid Capture	Illumina

ID	Diagnosis	HPO terms	Library prep	Exome target
NG_51	Optic atrophy	HP:0000648 HP:0000486 HP:0000639 HP:0000407 HP:0003690 HP:0003124 HP:0025379 HP:0004901 HP:0006827 HP:0100561	Nextera Rapid Capture	Illumina
NG_68	Optic atrophy	HP:0000648	Lotus IDT	xGen Exome v1 IDT
NG_2	Dominant optic atrophy	HP:0000648 HP:0003124 HP:0002155 HP:0001875	Nextera Rapid Capture	Illumina
NG_6	Dominant optic atrophy	HP:0000648	Nextera Rapid Capture	Illumina
NG_24	Dominant optic atrophy	HP:0000648 HP:0200026	Nextera Rapid Capture	Illumina
NG_137	Dominant optic atrophy	HP:0000648	NEBNext Ultra II	xGen Exome v1 IDT
NG_53	Dominant optic atrophy	HP:0000648	Nextera Rapid Capture	Illumina
NG_54	Dominant optic atrophy	HP:0000648	Nextera Rapid Capture	Illumina
NG_66	Dominant optic atrophy	HP:0000648 HP:0002077 HP:0030880 HP:0030890 HP:0012683 HP:0011003	Nextera Rapid Capture	Illumina

ID	Diagnosis	HPO terms	Library prep	Exome target
NG_67	Dominant optic atrophy	HP:0000648	Nextera Rapid Capture	Illumina
NG_96	Dominant optic atrophy	HP:0000648	NEBNext Ultra II	xGen Exome v1 IDT
NG_100	Dominant optic atrophy	HP:0000648	NEBNext Ultra II	xGen Exome v1 IDT

Table 2.1: The cohort under analysis.

HPO terms legend:

HP:0000648: Optic atrophy
HP:0008587: Mild neurosensory hearing impairment
HP:0003390: Sensory axonal neuropathy
HP:0003128: Lactic acidosis
HP:0002514: Cerebral calcification
HP:0002624: Abnormal venous morphology
HP:0100502: Vitamin B12 deficiency
HP:0004901: Exercise-induced lactic acidemia
HP:0002075: Dysdiadochokinesis
HP:0002403: Positive Romberg sign
HP:0030891: Periventricular white matter hyperdensities
HP:0002076: Migraine
HP:0100507: Reduced blood folate concentration
HP:0000739: Anxiety
HP:0007642: Congenital stationary night blindness
HP:0000556: Retinal dystrophy
HP:0001889: Megaloblastic anemia
HP:0005264: Abnormality of the gallbladder
HP:0000407: Sensorineural hearing impairment
HP:0000822: Hypertension
HP:0001249: Intellectual disability
HP:0000708: Behavioral abnormality
HP:0000609: Optic nerve hypoplasia
HP:0025163: Abnormality of optic chiasm morphology
HP:0002079: Hypoplasia of the corpus callosum
HP:0011342: Mild global developmental delay
HP:0000639: Nystagmus
HP:0002066: Gait ataxia
HP:0030217: Limb apraxia
HP:0003002: Breast carcinoma
HP:0000563: Keratoconus
HP:0011505: Cystoid macular edema
HP:0001272: Cerebellar atrophy
HP:0003458: EMG: myopathic abnormalities
HP:0001260: Dysarthria
HP:0012043: Pendular nystagmus
HP:0020036: Upper limb dysmetria
HP:0001348: Brisk reflexes
HP:0000486: Strabismus
HP:0003690: Limb muscle weakness
HP:0003124: Hypercholesterolemia
HP:0025379: Anti-thyroid peroxidase antibody positivity
HP:0006827: Atrophy of the spinal cord
HP:0100561: Spinal cord lesion
HP:0002155: Hypertriglyceridemia
HP:0001875: Neutropenia
HP:0200026: Ocular pain
HP:0002077: Migraine with aura
HP:0030880: Raynaud phenomenon
HP:0030890: Hyperintensity of cerebral white matter on MRI
HP:0012683: Pineal cyst
HP:0011003: High myopia

2.2 Bioinformatic analyses

Raw data obtained from sequencing contained the amount of short reads in fastq format. The pipeline for bioinformatic analysis was the following:

1. Raw reads were aligned with the Burrows-Wheeler aligner software BWA (`http://bio-bwa.sourceforge.net`) against the reference genome hg19; BWA-MEM, the latest algorithm, was used. The output was a SAM (Sequence Alignment/Map) file that contained the aligned reads.

2. The software SAMtools (`http://samtools.sourceforge.net`) was used to convert the SAM file into its binary equivalent, the BAM (Binary Alignment/Map) format.

3. GATK (Genome Analysis Toolkit, `https://gatk.broadinstitute.org`) RealignerTargetCreator and IndelRealigner commands were used to perform indel-based realignment in order to transform regions with misalignments caused by indels into clean reads.

4. Duplicate reads were removed by means of the Picardtools MarkDuplicates utility (`https://broadinstitute.github.io/picard`) to avoid sequencing errors propagation and allele overrepresentation.

5. Base quality scores were adjusted through GATK's BaseRecalibrator while coverage metrics were computed by GATK's DepthofCoverage.

6. Finally, variant calling is executed with the GATK tool HaplotypeCaller, that calls SNPs, indels and structural variants simultaneously. The workflow involves producing an intermediate multi-sample GVCF (Genomic VCF) and then performing a variant quality score recalibration, that uses machine learning to identify annotation profiles of variants that are likely to be real.

The final output file is a multi-sample VCF (Variant Call Format) containing the standard columns of CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO and FORMAT.

From this file, the single patients' VCFs were extracted through the GATK SelectVariants option and the VCFtools vcf-subset command (`http://vcftools.sourceforge.net`). VCF files of families with dominant mode of inheritance, for which more than one

member had been sequenced, were merged thanks to the BCFtools isec option (`http://samtools.github.io/bcftools/bcftools.html`): in this way, only mutations occurring in both subjects, and hence possibly causing the phenotype, were retained.

In order to reduce noise and variants resulting from sequencing errors, a filtering step was performed: under the hypothesis that pathogenic mutations need to be rare, all variants found in more than 5% of samples in our in-house database were removed.
Since mtDNA is not adequately covered by WES probes, we decided to eliminate mitochondrial genome variant calls from the VCF files, using again VCFtools (–not-chr option). In families where inheritance was compatible with matrilinear transmission, we separately sequenced the complete mtDNA with a previously reported method (Caporali et al., 2018).

2.3 Phenotype-based variant prioritization tools

A literature research was performed in order to find variant prioritization tools that satisfied the following criteria:

- Required as input a VCF file in combination with the corresponding HPO-encoded clinical diagnosis.
- Were freely available, either as web-tools or with command-line access.
- Were recently updated or fairly maintained.

2.3.1 The Exomiser suite

One of the first bioinformatics tools of this kind is Exomiser, originally published in 2014 as a freely available, command-line Java program (Robinson et al., 2013). The latest release (version 12.0.1) with a number of new features was dropped in 2019 and is available at `https://github.com/exomiser/Exomiser` (Cipriani et al., 2020).
It requires as input: i) the called variants resulting from the WES data, stored in a VCF file; optionally a multi-sample VCF together with a pedigree (PED) file can be supplied if more than one family member has been sequenced, and ii) the set of HPO terms describing the patient's phenotype.

The first step in the analysis is the variants' annotation, handled by the Jannovar software library, relative to the University of California, Santa Cruz (UCSC) hg19 transcript coordinates. Then, a set of user-defined variant filtering criteria can be applied. For the scope of this analysis, the following parameters were chosen:

- Intronic, regulatory and intergenic mutations, together with variants in the untranslated regions and up and downstream the gene were removed.
- The maximum minor allele frequency (MAF) taken from large datasets, including the 1000 Genomes Project, the Exome Aggregation Consortium (ExAc) and the Genome Aggregation Database (gnomAD) was used for comparison with a chosen allele frequency cut-off, under the hypothesis than common variants in the population cannot cause a rare disease.
- The mode of inheritance - autosomal dominant (AD), autosomal recessive (AR) or X-linked recessive (XLR) - was set depending on the pedigree of the patient(s) under analysis; in some cases, multiple modes were used simultaneously. Different allele frequency cut-offs were selected for each inheritance pattern (AD: 0.1%; AR and XLR: 1.0%).
- Analysis mode was set to full, in order to keep also non-passing variants.

After filtering, each variant is assigned a score from 0 (benign) to 1 (pathogenic), depending on how well they fit the criteria described above. Moreover, different pathogenicity scores are used to predict the mutation's consequence on the protein product. In our analysis, PolyPhen, MutationTaster, SIFT and CADD predictors were used; also benign variants were considered by the algorithm.
The Exomiser suite contains different prioritization methods, through which each gene containing any filtered variant(s) is ranked by a phenotype score (from 0 to 1). There exist 4 different methods:

- PHIVE, the original algorithm, that computes the phenotypic similarity between the patient's symptoms and observed phenotypes in mouse mutants associated with each candidate gene in the exome data.
- ExomeWalker, for protein-protein interactions, to identify new causative genes that

are involved in the same pathways as others previously known to be implicated in a specific disease.

- PhenIX, that analyses only genes associated to Mendelian diseases and thus cannot be used for gene discovery;

- hiPHIVE, which integrates the previous algorithms, using also data from zebrafish model organisms: in this way known disease-gene associations and novel candidates can be analysed simultaneously; for genes that have no phenotype data, direct and indirect protein-protein interactions and orthology information from other species are used.

The latter method was the one chosen for the analysis of our cohort.
The final Exomiser score used for ranking is obtained using a logistic regression classifier on a training set of 10.000 disease and 10.000 benign variants using a 10-fold cross-validation.
The previously described options were provided to Exomiser through an yml setting file, one for each patient(s); to optimize running time, the whole analysis was run in batch mode using the following command line:

```
java -Xms2g -Xmx4g -jar exomiser-cli-12.0.1.jar -analysis-batch batchFile.txt
```

The program produces output files in HTML format, easier to interpret, or as a tab-separated file that can be easily incorporated into larger bioinformatic pipelines.

2.3.2 MutationDistiller

The second tool taken into consideration is MutationDistiller (Hombach et al., 2019), which is freely available online at `https://www.mutationdistiller.org/`.
Like in The Exomiser, the basic requirements are the upload of a VCF file, together with the corresponding HPO-encoded phenotypes. The tool then combines the pathogenic prediction scores of MutationTaster and the gene prioritisations of GeneDistiller.
Also in this case, some parameters can be set in order to filter the variants: the minimum coverage was fixed to 5. Polymorphisms were filtered out if they were homozygous in more than 2 samples or heterozygous in more than 500 samples in ExAc and in the 1000 Genomes databases.
Moreover, the mode of inheritance was set according to the patient's pedigree and mu-

tations near splice sites and rare benign non-synonymous variants were included in the analysis.

Other useful features included are the possibility to specify the function of the gene, filter the variants using expression data or via genetic regions, in-house or common virtual panels. However, these options were not exploited during our analysis.

The algorithm is optimized to allow discovery of variants in yet unknown genes.

The output presents the list of prioritized genes, with many hyperlinks to relevant resources and the possibility to update and alter the chosen parameters on-the-fly. Furthermore, MutationDistiller contains the ACMG actionable genes panel, a list of genes of medical relevance published by the American College of Medical Genetics and Genomics.

2.3.3 GenIO

Another webtool that can help in the clinical genomics diagnostic process is GenIO (Koile et al., 2018), available at `https://bioinformatics.ibioba-mpsp-conicet.gov.ar/GenIO/index.php`.

As before, a VCF file is uploaded by the user, together with clinical phenotypes of the patient in the form of HPO terms.

GenIO analyses only mutations that passed quality controls (PASS flag in the FILTER field of the VCF file). Variant annotation is performed by Annovar, Anntools and SnpEff, using pieces of information from ClinVar, OMIM, gnomAD and dbSNP: as a result, a merged, annotated VCF file is produced.

Filtering threshold for population frequency according to the condition of rareness can be modified before analysis: in our case variant frequency was set to less than 0.5% in the case of recessive mutations and not observed as dominant ('rare' option).

Phenotype information are analysed by means of Phenolyzer, allowing disease gene discovery.

GenIO then classifies variants that might have pathogenic effects using information from the ClinVar database annotations, the Mendelian Clinically Applicable Pathogenicity (M-CAP) classifier and the InterVar clinical interpretation.

A list of genes for both dominant and recessive mode of inheritance is produced as output, together with a list of secondary findings including variants found in 59 medically

actionable genes from ACMG.

2.3.4 Ontology Variant Analysis

The last tool used in this study is OVA (Ontology Variant Analysis) (Antanaviciute et al., 2015), freely accessible at `http://dna2.leeds.ac.uk:8080/OVA/index.jsp`.
OVA prioritizes data from WES experiments following 3 main steps:

1. Variants from the uploaded VCF file are passed through user-defined options to reduce candidate search space. Filters include variant type, genotype options, minimum quality score and allele frequency data from the Born in Bradford exome sequencing database. Only the first option was used in order to remove mutations in non-protein coding genes, in introns and in UTRs as well as synonymous substitutions.

2. The query phenotypes provided as HPO terms are used to compute a comparison annotation profile from known phenotype-genotype associations, phenotype similarities and cross-links between ontologies. Both human and model organism phenotypes, functional annotations, pathways and cellular localizations are used to find the most relevant genes for the query phenotype.

3. From the previous data, a supervised learning approach (random forest model) is used in order to calculate the probability that each candidate gene has of harbouring the disease-causing variant.

2.4 Data analysis of unsolved WES cases

The output files obtained from the Exomiser and MutationDistiller were merged in order to retain only the variants found by both algorithms.
Manual curation of the results was performed, firstly by looking up the single positions on the corresponding BAM files on IGV (Integrative Genomics Viewer, `http://software.broadinstitute.org/software/igv/`) and excluding the ones that were the result of errors. The remaining variants were evaluated taking into consideration the ACMG Classification obtained from the VarSome search engine (`https://varsome.com`), the allele

count found on the gnomAD database (`https://gnomad.broadinstitute.org`), if known the gene function annotated on UniProt (`https://www.uniprot.org`) and GeneCards (`https://www.genecards.org`), when present the clinical phenotypes associated to the gene found on OMIM (Online Mendelian Inheritance in Man, `https://omim.org`) and gene expression data from The Human Protein Atlas (`http://www.proteinatlas.org`).

Segregation of some of the candidate genes had already been performed and thus was exploited to further support the obtained results or discredit them.
New genes not previously considered as possible causative ones will be segregated as a further development of this study.

Chapter 3

RESULTS AND DISCUSSION

3.1 Variant filtering and tools testing

Personal genomics is now becoming a clinical reality thanks to the availability of health and genetic data. The advances introduced by the advent of Next Generation Sequencing have revolutionized the field of rare monogenic diseases, enabling researchers to discover novel gene-disease associations and precise molecular diagnoses. In particular, Whole Exome Sequencing is a frequently used technique in routine clinical research, mainly because it is the most cost-effective method (Sun et al., 2015).

Despite that, many challenges still exist, as proved by the limited diagnostic yield. This can be in part ascribed to the intrinsic characteristics of high-throughput techniques, such as incomplete coverage of some genomic regions or limitations in calling copy-number variants.

Another major hurdle in WES data analysis is the wealth of variants found: an individual exome is predicted to harbour, on average, around 30.000 variants when compared to the genomic reference sequence. Of these, around 10.000 lead to non-synonymous amino acid substitutions, alterations of conserved splice sites residues and small insertions and deletions. In order to reduce this number, usually a strict frequency filtering step is applied: common variants found in public or in-house sequencing databases are discarded, since they cannot explain a rare disease. Variants that do not follow the expected mode of inheritance derived from the subject's pedigree can also be removed. Moreover, mutations should be taken into consideration only if found in genes belonging to pathways implicated in the observed phenotypes and this is often hindered by incomplete knowledge of protein functions.

After these filtering steps, additional criteria are used to predict how serious the functional consequence of a specific mutation is: a growing number of in silico computational

tools have been developed to predict variants pathogenicity on the basis of characteristics such as conservation, physicochemical properties of the wild-type and variant amino acids and other features. These include function-prediction methods such as MutationTaster (Schwarz et al., 2010), PolyPhen-2 (Adzhubei et al., 2010) and SIFT (Ng & Henikoff, 2001), conservation methods, like GERP^{++} (Davydov et al., 2010) and phastCons (Siepel, 2005) and ensemble methods that integrate information from multiple algorithms; some of these are CADD (Kircher et al., 2014), M-CAP (Jagadeesh et al., 2016) and REVEL (Ioannidis et al., 2016). More recently, some methods using deep neural networks have been developed - MVP (Qi et al., 2021) and PrimateAI (Sundaram et al., 2018). Nevertheless, Li et al. (2018) demonstrated an overall lack of agreement between these pathogenic predictors in classifying variants as benign or disease-causing. In addition to that, each genome is thought to harbour around 100 genuine loss-of-function variants and 20 genes completely inactivated (MacArthur et al., 2012). Prioritizing genes purely on the basis of rarity and predicted pathogenicity found in them is thus expected to generate many false positive candidates.

Another generation of tools for computational disease-associated gene prioritization were developed, trying to integrate expression data, protein-protein interaction networks, functional annotations and information from medical literature to derive a more reliable list of ranked genes. Between these, are phenotype-driven variant prioritization tools that leverage existing genotype to phenotype information from different databases in order to define a list of possible candidate variants in genes that are relevant to the patient's phenotype.

The starting point for the analysis conducted in this study was to identify and select the best performing tools of this kind. The ones not available without registration were excluded, together with the ones that had not been updated for a long time or were not maintained. Moreover, all the tools needed to require as input a VCF file together with the patient's phenotype encoded into terms from the Human Phenotype Ontology (HPO). The ontology currently contains more than 13.000 terms describing human phenotypic abnormalities and over 156.000 annotations of hereditary diseases, becoming the *de facto* standard for deep phenotyping.

In the end, four tools were retained: The Exomiser suite, MutationDistiller, GenIO and OVA.

In the majority of cases, the tools had only been tested on simulated data, usually obtained by introducing known pathogenic variants in publicly available VCF files of exomes of healthy individuals. Software performance statistics on real patient data is so far very limited.

For this reason, we decided to test the before mentioned tools on a cohort of patients with hereditary optic neuropathy from our own database for which the molecular diagnosis was already known. Familial segregation analysis was used to confirm the diagnostic conclusion. When appropriate, functional validation on yeast models had also been performed. The test set included 11 subjects, all presenting with syndromic or non-syndromic HON and whose pedigree followed different modes of inheritance. Moreover, causative variants affected both genes established in the literature to cause optic atrophy (*OPA1*, *WFS1*) but also ones that had been recently identified (*SSBP1*, *DNAJC30*).

The testing results are shown in Table 3.1.

ID	Causative gene	Causative mutation	Exomiser rank	MutationDistiller rank	GenIO rank	OVA rank
NG_1	*NDUFAF2*	c.95A>G:p.(Tyr32Cys) c.148del:p.(Arg50Glufs*3)	1st	1st	1st	Not found
NG_4	*MTFMT*	c.518C>T:p.(Thr173Ile)	2nd	4th	1st	Not found
NG_5	*DNAJC30*	c.152A>G:p.(Tyr51Cys)	78th	21st (per score: 2nd)	Not found	Not found
NG_7	*DNAJC30*	c.152A>G:p.(Tyr51Cys)	29th	24th (per score: 3rd)	2nd	Not found
NG_12	*CACNA1F*	c.2038C>T:p.(Arg680*)	1st	1st	3rd	6th
NG_16	*NDUFB11*	c.276_278del:p.(Phe93del)	8th	6th	Not found	Not found
NG_19	*MECR*	c.772C>T:p.(Arg258Trp)	1st	7th	Not found	Not found
NG_39	*WFS1*	c.1656_1680dup:p.(Ile561Argfs*53) c.2194C>T:p.(Arg732Cys)	1st	1st	3rd	5th
NG_73	*SSBP1*	c.151A>G:p.(Lys51Glu)	74th	Not found	Not found	Not found
NG_110	*OPA1*	c.2878C>T:p.(Arg960*)	1st	10th (per score: 1st)	7th	Not found
NG_127	*OPA1*	c.1212+5G>A:splice defect	1st	8th	Not found	Not found

Table 3.1: Ranking of each gene from the testing cohort for the four phenotype-based variant prioritization tools.

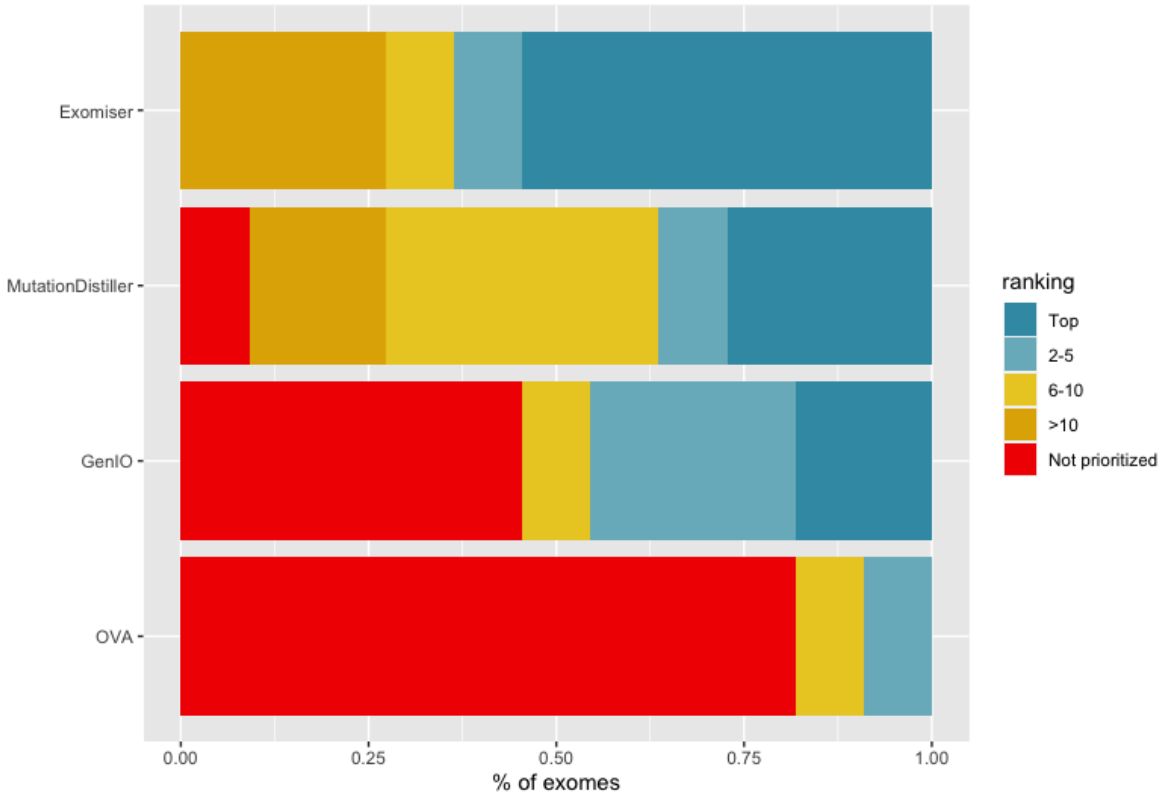


Figure 3.1: **Tools performance.** Categorical percentage distribution of the disease-causing variant ranking according to five mutually exclusive disease-causing ranking bins ("Top", "2–5", "6–10", ">10", and "Not prioritized") per each tool.

The ranking results for the causative variants were categorized into five, mutually exclusive, ranking bins: "Top", "2-5", "6-10", ">10" and "Not prioritized" (Figure 3.1).

In more than 80% of cases (9 out of 11), the OVA algorithm was not able to find the correct causative gene. Such a result might have been expected for subjects carrying mutations in proteins recently identified to have a role in HON development. However, the case of NG_110 is quite emblematic: the tool is not able to prioritize a nonsense mutation (stop gain) in *OPA1*, a gene responsible for optic atrophy in 70% of patients with this disorder.

The tool did identify the correct mutations in samples NG_12 (*CACNA1F*) and NG_39 (*WFS1*). In both cases, the gene was associated in the literature with the phenotypes carried by the two patients; still, the causative variant was not found as top candidate in the list of prioritized genes.

In order to reduce noise, HPO-encoded phenotypes that did not strictly correlate with optic neuropathies were removed from the analysis. However, no improvement was ob-

served.

Since GenIO only accepted a maximum of 3 HPO terms per patient, only the most relevant ones for the phenotype were kept and the analysis was repeated with different combinations of them. Nevertheless, the final ranked list of genes did not change.
GenIO performed better than OVA, finding the causative mutation in half of the patients and always within the first 10 positions.
A few remarks can be made on the results: subjects NG_5 and NG_7 both carried the same mutation (c.A152G:p.Y51C) in *DNAJC30*, a nuclear gene found to be implicated in a recessive phenocopy of mitochondrial LHON. However, GenIO prioritized the variant only in NG_7 but not in NG_5, where it returned as result for the recessive inheritance mode only two genes associated with retinitis pigmentosa (*OFD1* and *RP1L1*).
The algorithm filtered out the correct variant also in the case of NG_127: this time, the pathogenic mutation was found on a +5 site in the splice region. However, GenIO does not give the possibility to include or define the number of bases in the intron/exon boundary to be treated as splice sites, a feature that could be added to improve the tool in future updates.

The most valid tools to prioritize the pathogenic mutation are MutationDistiller and The Exomiser.
In particular, the latter was always able to prioritize the causative gene, half of the times (54%) as the top candidate. The worst results were obtained in the cases of *SSBP1* and *DNAJC30* (">10" ranking bin).
More or less the same results were obtained with MutationDistiller, but with a lower percentage of causative genes identified as top candidates (27%). Still, in 72% of cases the disease-associated gene was found within the first 10 positions.
As in The Exomiser, the tool did not perform very well with newly discovered genes: in particular, MutationDistiller filtered out the mutation in *SSBP1* and did not prioritize it.
In some occasions, the ranking was higher if equally scoring gene-variants were taken into account: for example, in subject NG_110, *OPA1* was found at the 10^{th} place in the list, but had the same score as the ones that preceded it. The same occurred with *DNAJC30* in NG_5 and NG_7, where the 2^{nd} and 3^{rd} positions respectively were achieved when considering the score.

3.2 Unsolved WES cases

The successive step of this study consisted in using these phenotype-driven tools to analyse WES data of patients with hereditary optic neuropathy whose causative variant had not been identified by virtual gene panels and other classic approaches.
According to the software performances obtained for the testing cohort, we decided to retain only The Exomiser and MutationDistiller.

Under the intuition that variants that are too common cannot cause a rare disease phenotype, we also filtered out mutations that were found in more than 5% of samples of our in-house database (N=250): in this way, file size was reduced up to 90%.
The ranked list of genes obtained from the two tools was merged into a single file, so that only variants prioritized by both algorithms were retained. This choice was made mainly to identify genes that were associated to optic atrophy development only recently, and hence not reported in the literature, following the results obtained for *DNAJC30*. In fact, both tools prioritized it: this double corroboration gave us reason to think that, if a new gene was the causative one, it would have been identified both by The Exomiser and MutationDistiller.
The remaining variants' positions were first looked up on IGV, a browser for visual exploration of genomic data obtained during NGS analysis. In this way, mutations caused by sequencing errors (strand bias) or that were incorrectly called and not present in the corresponding BAM file were excluded. Finally, possible candidates were evaluated by means of different clinical databases.

Eight subjects coming from pedigrees following recessive modes of inheritance were analysed and prioritized variants are reported in Table 3.2.

ID	CHR	POS	REF	ALT	Functional class	Gene	Variant	Exomiser rank[1]	Mutation Distiller rank
NG_3	4	6302903	A	C	missense	*WFS1*	c.1381A>C:p.(Thr461Pro)	1	1
	4	6303119	C	T	missense	*WFS1*	c.1597C>T:p.(Pro533Ser)	1	1
	15	98513125	A	C	missense	*ARRDC4*	c.895A>C:p.(Ile299Leu)	14	11
	15	98513932	A	G	missense	*ARRDC4*	c.1159A>G:p.(Ile387Val)	14	11
	17	56389516	C	T	missense	*TSPOAP1*	c.2486G>A:p.(Arg829Gln)	22	5
	17	56389823	C	T	missense	*TSPOAP1*	c.2179G>A:p.(Glu727Lys)	22	5
	17	56402170	A	C	splice donor	*TSPOAP1*	c.762+2T>G:splice defect	22	5
NG_43	1	22168033	G	A	splice region	*HSPG2*	c.9327C>T:p.(=)	27	2
	1	22175460	C	T	missense	*HSPG2*	c.7511G>A:p.(Arg2504His)	27	2
	2	196726465	T	C	missense	*DNAH7*	c.7712A>G:p.(Lys2571Arg)	61	8
	2	196740494	C	T	missense	*DNAH7*	c.6191G>A:p.(Arg2064Lys)	61	8
NG_51	2	236949476	G	A	missense	*AGAP1*	c.1882G>A:p.(Glu628Lys)	4	17
	2	237028850	G	A	missense	*AGAP1*	c.2129G>A:p.(Arg710Gln)	4	17
	5	78326663	A	C	missense	*DMGDH*	c.1676T>G:p.(Ile559Ser)	6	4
	5	78338269	G	C	missense	*DMGDH*	c.1030C>G:p.(Arg344Gly)	6	4
	3	196509522	C	G	missense	*PAK2*	c.5C>G:p.(Ser2Cys)	9	8
	3	196529902	G	C	missense	*PAK2*	c.303G>C:p.(Gln101His)	9	8
	9	4662486	G	C	missense	*PLPP6*	c.111G>C:p.(Arg37Ser)	13	21
	9	4663124	G	C	missense	*PLPP6*	c.749G>C:p.(Arg250Thr)	13	21
NG_23	4	123855300	G	A	missense	*SPATA5*	c.554G>A:p.(Gly185Glu)	11	27
	4	123857284	G	A	missense	*SPATA5*	c.1307G>A:p.(Arg436His)	11	27
	17	10210260	G	A	missense	*MYH13*	c.5291C>T:p.(Thr1764Met)	25	14
	17	10243471	A	G	synonymous	*MYH13*	c.2052T>C:p.(=)	25	14
	17	4540514	C	T	missense	*ALOX15*	c.847G>A:p.(Gly283Arg)	37	7
	17	4542931	G	A	splice region	*ALOX15*	c.136-5C>T:p.?	37	7
	20	62658384	C	T	missense	*PRPF6*	c.1958C>T:p.(Ala653Val)	40	1
	20	62659826	C	T	splice region	*PRPF6*	c.2220-5C>T:p.?	40	1
	X	102832076	T	C	missense	*TCEAL4*	c.118T>C:p.(Trp40Arg)	4	23
	8	100844794	C	G	missense	*VPS13B*	c.9528C>G:p.(Ser3176Arg)	12	3

ID	CHR	POS	REF	ALT	Functional class	Gene	Variant	Exomiser rank[1]	Mutation Distiller rank
	8	100865673	C	T	splice_region	*VPS13B*	c.10062-6C>T:p.?	12	3
	14	103410692	TCCG	T	disruptive inframe deletion	*CDC42BPB*	c.3941_3943del:p.(Ala1314del)	24	14
	14	103442018	T	C	splice region	*CDC42BPB*	c.1507+3A>G:p.?	24	14
	11	76886504	T	C	synonymous	*MYO7A*	c.2148T>C:p.(=)	26	4
	11	76890919	C	T	missense	*MYO7A*	c.2473C>T:p.(Arg825Cys)	26	4
	14	64587777	C	G	missense	*SYNE2*	c.13156C>G:p.(Gln4386Glu)	34	29
	14	64604650	A	G	missense	*SYNE2*	c.14792A>G:p.(Lys4931Arg)	34	29
NG_22	19	40370528	C	T	synonymous	*FCGBP*	c.12138G>A:p.(=)	46	13
	19	40421316	C	T	missense	*FCGBP*	c.2605G>A:p.(Asp869Asn)	46	13
	X	67273643	T	C	missense	*OPHN1*	c.2168A>G:p.(Asp723Gly)	1	2
	X	48935371	T	C	missense	*WDR45*	c.166A>G:p.(Met56Val)	10	6
	X	48649629	C	T	missense	*GATA1*	c.113C>T:p.(Pro38Leu)	7	5
	X	9863702	C	T	missense	*SHROOM2*	c.1754C>T:p.(Pro585Leu)	4	10
	17	47010709	C	G	splice acceptor	*SNF8*	c.423-1G>C:splice defect	3	11
	17	47014427	C	T	missense	*SNF8*	c.304G>A:p.(Val102Ile)	3	11
	11	65347852	C	T	missense	*EHBP1L1*	c.536C>T:p.(Pro179Leu)	14	23
	11	65351863	A	G	missense	*EHBP1L1*	c.3245A>G:p.(Asp1082Gly)	14	23
	11	14486502	C	T	missense	*COPB1*	c.2365G>A:p.(Val789Ile)	24	8
NG_47	11	14502647	A	C	splice region	*COPB1*	c.958-4T>G:p.?	24	8
	X	152815588	G	A	missense	*ATP2B3*	c.1667G>A:p.(Arg556Gln)	1	1
	X	152027402	G	A	missense	*NSDHL*	c.356G>A:p.(Arg119Lys)	5	4
	X	10093084	G	A	missense	*WWC3*	c.1847G>A:p.(Cys616Tyr)	6	19
	X	3229159	T	A	missense	*MXRA5*	c.7085A>T:p.(Asp2362Val)	2	12
	X	89177667	G	T	missense	*TGIF2LX*	c.583G>T:p.(Val195Phe)	10	21
	X	31089928	G	A	missense	*FTHL17*	c.143C>T:p.(Ala48Val)	11	17
NG_21	3	142272098	A	G	missense	*ATR*	c.2776T>C:p.(Phe926Leu)	4	1
	3	142272237	G	A	synonymous	*ATR*	c.2637C>T:p.(=)	4	1
	17	47007930	GGCCTGTAA	G	frameshift truncation	*SNF8*	c.673_680del:p.(Leu225Profs*14)	9	13

ID	CHR	POS	REF	ALT	Functional class	Gene	Variant	Exomiser rank[1]	Mutation Distiller rank
	17	47007943	A	AC	frameshift elongation	*SNF8*	c.667_668insG:p.(Leu223Argfs*19)	9	13
	17	47014427	C	T	missense	*SNF8*	c.304G>A:p.(Val102Ile)	9	13
	X	9707600	C	T	missense	*GPR143*	c.1105G>A:p.(Glu369Lys)	1	4
	X	30322890	T	C	missense	*NR0B1*	c.1219A>G:p.(Thr407Ala)	4	2
	X	138698628	C	T	missense	*MCF2*	c.1004G>A:p.(Arg335His)	5	7
	X	100646816	T	G	splice region	*RPL36A*	c.285+6T>G:p.?	3	11
NG_68	21	34025674	A	G	missense	*SYNJ1*	c.2798T>C:p.(Phe933Ser)	1	1
	21	34072407	T	C	missense	*SYNJ1*	c.220A>G:p.(Met74Val)	1	1
	7	6049010	G	C	splice donor	*AIMP2*	c.-251+1G>C:p.(=)	6	5
	7	6054834	C	T	missense	*AIMP2*	c.193C>T:p.(Arg65Cys)	6	5
	7	6063209	G	A	missense	*AIMP2*	c.850G>A:p.(Val284Ile)	6	5
	17	2227654	T	C	missense	*TSR1*	c.2251A>G:p.(Met751Val)	22	6
	17	2234244	A	G	synonymous	*TSR1*	c.1656T>C:p.(=)	22	6
	X	41332746	G	A	missense	*NYX*	c.40G>A:p.(Val14Met)	1	2
	X	16859560	C	T	missense	*TXLNG*	c.1258C>T:p.(Arg420Cys)	2	7
	X	153033712	C	T	synonymous	*PLXNB3*	c.1095C>T:p.(=)	4	3
	X	106844675	C	T	missense	*FRMPD3*	c.3505C>T:p.(Pro1169Ser)	3	9
	X	152089287	G	A	synonymous	*ZNF185*	c.648G>A:p.(=)	11	4

Table 3.2: Prioritized variants for the 8 subjects with a recessive of X-linked recessive inheritance.
[1] For The Exomiser, ranking positions are reported separately for the AR and XLR modes.

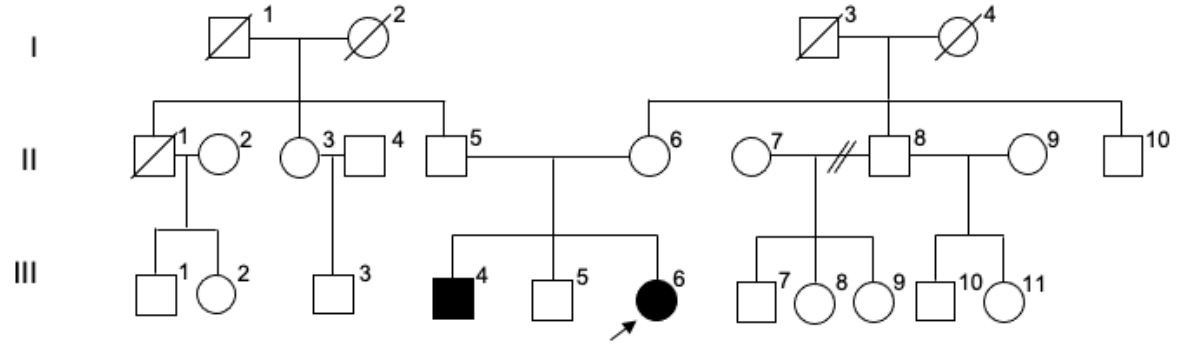


Figure 3.2: **NG_3.** Familial pedigree.

The first subject under analysis, NG_3, follows a recessive inheritance pattern and presents with a non-syndromic optic atrophy (Figure 3.2).
The initial filtered VCF file contained more than 6600 variants. After prioritization, only 59 were retained. All of these were verified on the IGV viewer and after this step only three genes were kept as potential candidates.
Both tools identified as top candidate *WFS1*, a gene causative for Wolfram syndrome, also called DIDMOAD (diabetes insipidus, diabetes mellitus, optic atrophy and deafness). However, previous familial segregation carried out in this pedigree discredited this hypothesis.
The second candidate, *ARRDC4*, is a gene involved in the process of endocytosis. Nevertheless, also these possibly pathogenic mutations had already been segregated and did not follow the expected pattern.
The third gene retained by our analysis is *TSPOAP1*, recently linked to autosomal recessive dystonia and coding for a pre-synaptic active zone protein (Mencacci et al., 2021). Three variants were identified in this protein, two missense and a splice donor. VarSome, a human genomic variant search engine, following the ACMG Classification, predicts the first two variants to be 'Likely benign'; moreover, they are both quite common in the population as heterozygous and both present in one individual as homozygotes. However, the splice donor mutation affects the canonical +2 splice site and it is classified as 'Pathogenic'. Further analyses should hence be performed in order to understand if these mutations found together could lead to phenomena like haploinsufficiency.

No significant results were obtained for subject NG_43, a sporadic case presenting

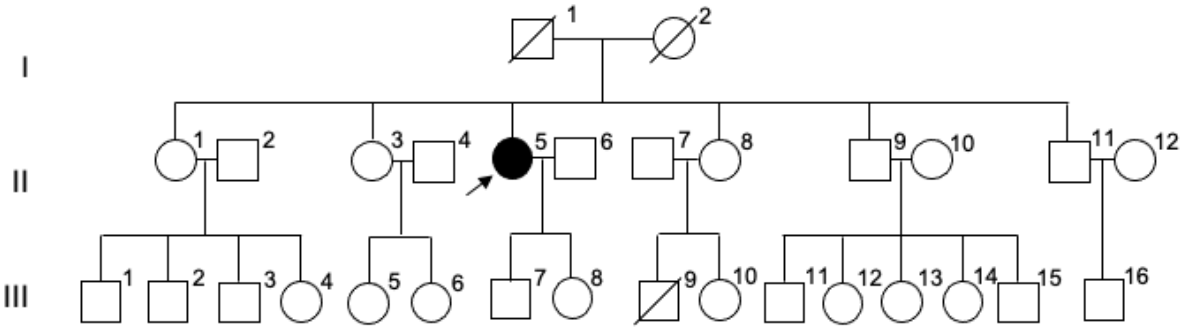


Figure 3.3: **NG_43.** Familial pedigree.

with optic atrophy, keratoconus, cystoid macular edema and breast carcinoma (Figure 3.3). 44 variants were retained from the initial 7513 of the filtered VCF. Only two possible gene candidates were kept after analysis on IGV: the two variants in *HSPG2* are both predicted 'Benign' and are common in the population. In addition, the gene is not associated to the phenotypes observed in the patient. The same conclusions were obtained for mutations found on *DNAH7*: one variant is predicted as 'Benign', while the second is a VUS that might have a role in cancer.

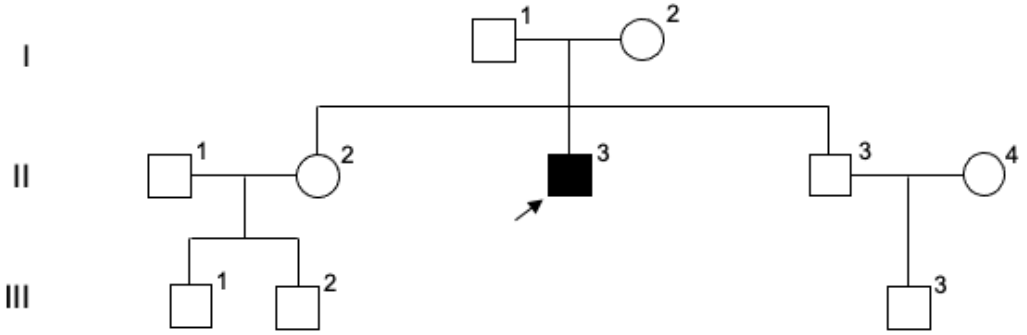


Figure 3.4: **NG_51.** Familial pedigree.

Starting from a total of 6714, 55 variants were retained for subject NG_51. This patient reported a quite complex clinical picture, including optic atrophy as well as sensorineural hearing impairment and limb muscle weakness (Figure 3.4).
Some of the variants that were not excluded after analysis through IGV were later discarded because of gene function not correlated with the phenotype (*PAK2*, *PLPP6*). *AGAP1*, a protein that is part of the endosomal-lysosomal system, had already been segregated and did not follow the expected pattern.

The remaining candidate, *DMGDH*, is a dehydrogenase found in the mitochondrial matrix and associated in the OMIM database with muscle fatigue. Both missense mutations are classified as of uncertain significance and are not common in the population (gnomAD database). The possible pathogenicity of the mutations was also confirmed by familial segregation, but they should be validated at least in other independent cases and by functional studies.

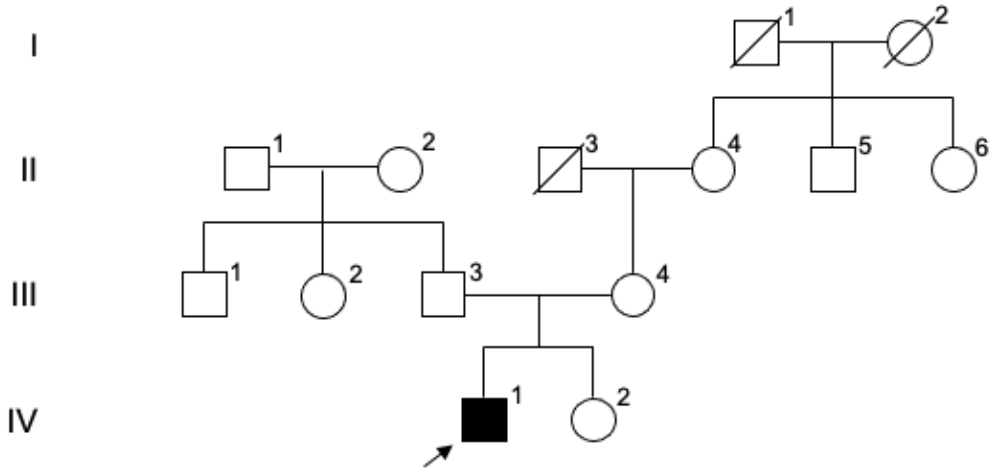


Figure 3.5: **NG_23.** Familial pedigree.

NG_23 is a sporadic case of optic atrophy (Figure 3.5); recessive and X-linked inheritance patterns were investigated.
106 likely pathogenic variants were kept by the prioritization algorithm. Of these, two genes identified as possible candidates by The Exomiser and MutationDistiller (*SPATA5*, encoding a protein that might be involved in mitochondrial transformations during spermatogenesis, and *PRPF6*, a spliceosome component) had already been discredited by familial segregation analysis. *MYH13* and *TCEAL4* were discarded after evaluation of gene function.
The two variants found in gene *ALOX15*, a lipoxygenase found mainly in the cytosol and in the plasma membrane, will need to be further evaluated.

The pedigree of subject NG_22 seems to follow an X-linked mode of inheritance (Figure 3.6). The patient presents with optic atrophy but also intellectual disability, a mild developmental delay, gait ataxia and limb apraxia.
The initial filtered VCF contained around 6500 variants and 79 were kept after prioriti-

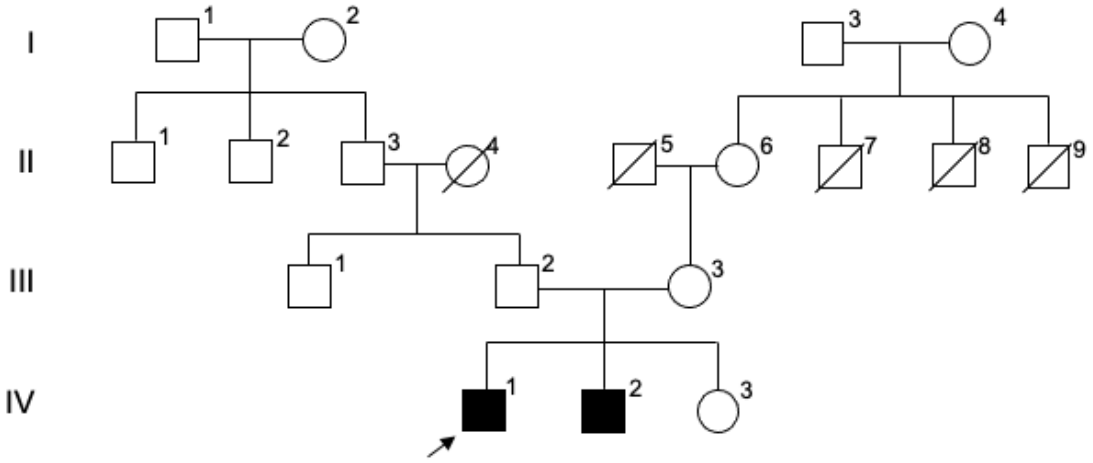


Figure 3.6: **NG_22.** Familial pedigree.

zation.

Four are the most interesting genes emerging from the analysis: *VPS13B* is a gene associated with Cohen syndrome, a multisystem disorder that causes, between other neurological symptoms, also optic atrophy, and this gene encodes a transmembrane protein playing a crucial role in preserving the integrity of the Golgi complex (Momtazmanesh et al., 2020); *CDC42BPB* codes for a serine/threonine kinase and might play a role in cytoskeletal reorganization; *SHROOM2* is a gene highly expressed in the retina and involved in endothelial cell morphology; moreover, the variant segregates correctly in the pedigree.

However, the most significant candidate is *WDR45*: this gene, found on chromosome X, is a component of the autophagy machinery and is associated in the literature with neurodegeneration with brain iron accumulation. Over one hundred unrelated patients were reported with this variable disorder, mostly affecting females carrying truncating mutations (Adang et al., 2020). Since the variant found in the subject under analysis is a missense, this might explain the milder phenotype observed. This hypothesis is corroborated also by familial segregation analysis.

NG_47 is a sporadic case of optic atrophy presenting also with a mild developmental delay, dysarthria, myopathic abnormalities and cerebellar atrophy (Figure 3.7).

Some of the variants prioritized by the algorithm were excluded because of higher frequency than expected in the healthy population (*COPB1*, *NSDHL*, *WWC3*, *FTHL17*) or due to gene function unrelated to mitochondria or neurodegeneration (*EHBP1L1*,

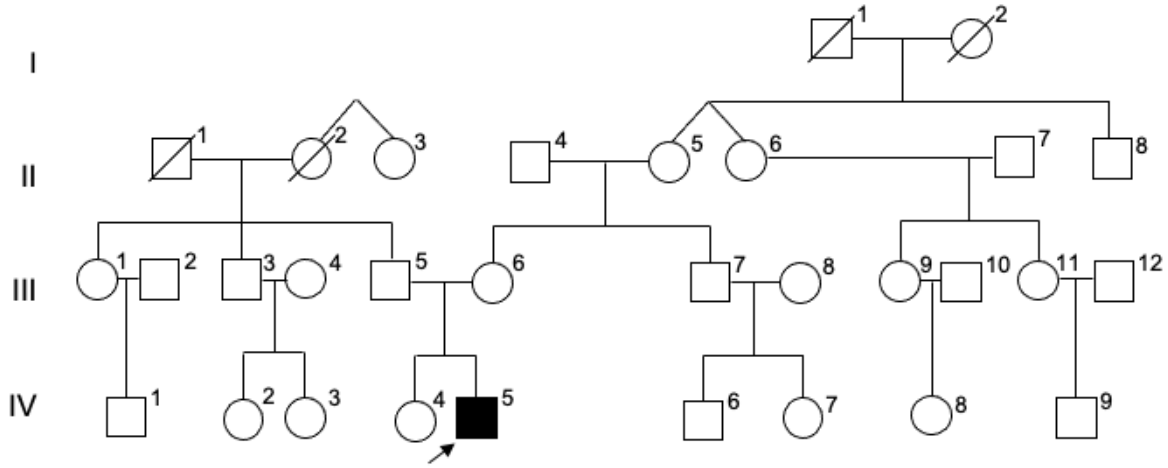


Figure 3.7: **NG_47.** Familial pedigree.

MXRA5, *TGIF2LX*).

The most promising candidate, placed at first position by both tools, was *ATP2B3*, a gene encoding an enzyme found in the plasma membrane and playing a role in calcium homeostasis. Mutations in this protein are associated with an X-linked form of spinocerebellar ataxia. Nevertheless, familial segregation analysis discredited this hypothesis: the variant was found in the unaffected maternal grandfather.

The second most interesting gene is *SNF8*, a component of the endosomal sorting complex. Two mutations were identified: one is a missense, while the other affects a -1 splice site and hence is predicted pathogenic by the ACMG Classification. Familial segregation on these variants was performed, giving more credit to this hypothesis.

Furthermore, the same missense variant found in patient NG_47 was also present in subject NG_21, whose phenotype included optic atrophy, intellectual disability and hypoplasia of the corpus callosum (Figure 3.8). Two other mutations were identified next to one another. Further evaluation of these mutations will be needed in order to prove the possible involvement of this gene in optic atrophy development.

Two other possible candidates in patient NG_21 had already been segregated (the nuclear receptor NR0B1 and the ribosomal protein RPL36A), while variants in *ATR*, *MCF2* and *GPR143* were excluded because of ACMG predictions ("Benign" verdict).

The last subject is NG_68, a sporadic case presenting with optic atrophy (Figure 3.9). Only 20 variants were kept after merging the results of the two tools.

The top candidate for both algorithms is *SYNJ1*, a phosphatase associated with epileptic

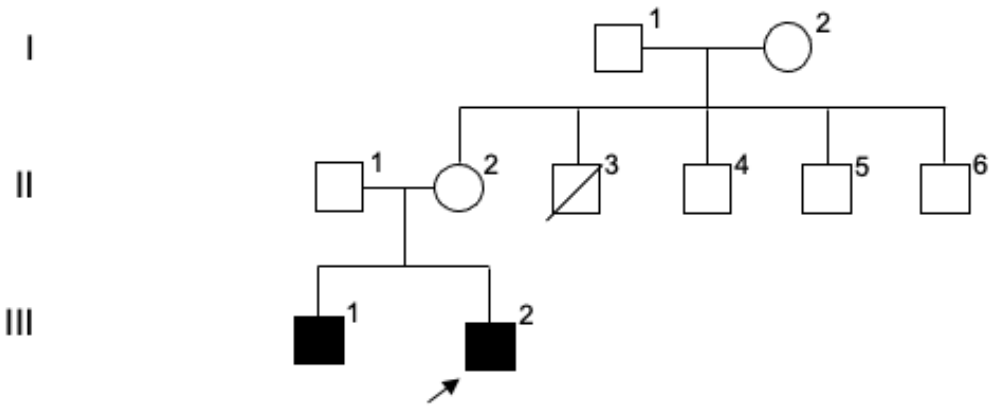


Figure 3.8: **NG_21.** Familial pedigree.

encephalopathy and Parkinson disease with early onset. Both missense mutations are predicted of uncertain significance by VarSome and thus should be further analysed in order to make a conclusion.

For the X-linked inheritance pattern, another interesting gene emerges: mutations in *NYX* are known to cause congenital stationary night blindness. The protein is a member of the subfamily of small leucine-rich proteoglycans and it is mainly localized in the extracellular space. Also in this case, the missense variant is a VUS and needs to be examined by familial segregation.

The mutation in *FRMPD3*, encoding a protein involved in signal transduction, is another VUS and thus may be another possible candidate for the X-linked mode of inheritance.

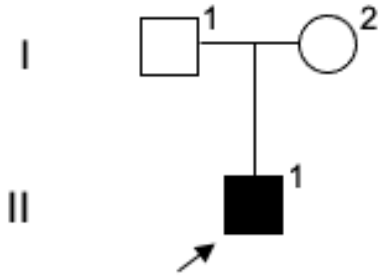


Figure 3.9: **NG_68.** Familial pedigree.

Five families with dominant optic atrophy were analysed. In this case, more than one individual per pedigree had been subjected to WES analysis and only heterozygous variants present in both patients were retained in the VCF file. Moreover, only variants that had a frequency of zero in public databases were taken into consideration (Table 3.3).

ID	CHR	POS	REF	ALT	Functional class	Gene	Variant	Exomiser rank	Mutation Distiller rank
NG_2, NG_6	4	147560457	T	TGGCGGC	disruptive inframe insertion	*POU4F2*	c.195_200dup:p.(Gly67_Gly68dup)	1	15
	11	61121433	C	T	stop gained	*CYB561A3*	c.216G>A:p.(Trp72*)	7	50
	11	64669744	C	T	missense	*ATG2A*	c.3892G>A:p.(Ala1298Thr)	9	25
	15	41195469	T	A	missense	*VPS18*	c.2852T>A:p.(Met951Lys)	12	36
	2	71297970	G	A	splice donor	*NAGK*	c.213+1G>A:splice defect	15	33
	5	16463689	G	C	missense	*ZNF622*	c.788C>G:p.(Ser263Trp)	16	42
	11	1016961	G	C	missense	*MUC6*	c.5840C>G:p.(Thr1947Ser)	21	13
	15	30854417	A	T	missense	*GOLGA8Q*	c.1697A>T:p.(Glu566Val)	26	58
NG_24, NG_137	1	180243596	T	G	missense	*LHX4*	c.1055T>G:p.(Leu352Arg)	3	1
	17	38976536	G	A	missense	*KRT10*	c.1007C>T:p.(Ala336Val)	17	8
	1	203769003	G	C	missense	*ZBED6*	c.2353G>C:p.(Gly785Arg)	18	36
	1	41541020	C	G	missense	*SCMH1*	c.819G>C:p.(Arg273Ser)	32	23
	13	47279253	G	T	missense	*LRCH1*	c.1451G>T:p.(Gly484Val)	42	57
	15	44964599	T	C	missense	*PATL2*	c.502A>G:p.(Ser168Gly)	72	68
	15	72195287	T	C	missense	*MYO9A*	c.2995A>G:p.(Thr999Ala)	73	70
	17	62859060	G	T	missense	*LRRC37A3*	c.3130C>A:p.(Leu1044Met)	89	91
	19	1395480	G	A	missense	*NDUFS7*	c.635G>A:p.(Arg212His)	\ \ [1]	5
NG_53, NG_54	11	62295585	C	T	missense	*AHNAK*	c.6304G>A:p.(Gly2102Ser)	15	46
	16	15847262	AG	A	frameshift truncation	*MYH11*	c.1852del:p.(Leu618Cysfs*17)	17	14
	15	44150960	G	T	missense	*WDR76*	c.1501G>T:p.(Ala501Ser)	24	67
	12	104102217	A	T	splice acceptor	*STAB2*	c.4193-2A>T:splice defect	25	55
	19	18377865	T	A	missense	*IQCN*	c.485A>T:p.(Asp162Val)	30	94
	4	104035714	C	T	splice region	*CENPE*	c.7443-5G>A:p.?	78	6
NG_66, NG_67	8	10470021	C	A	missense	*RP1L1*	c.1587G>T:p.(Glu529Asp)	5	20
	2	228168862	T	A	splice donor	*COL4A3*	c.4153+2T>A:splice defect	8	16
	12	82752043	C	T	missense	*CCDC59*	c.113G>A:p.(Arg38Gln)	12	55
	3	48607157	C	A	splice donor	*COL7A1*	c.7557+1G>T:splice defect	17	17
	2	9514943	A	G	missense	*ASAP2*	c.1616A>G:p.(Lys539Arg)	22	95

ID	CHR	POS	REF	ALT	Functional class	Gene	Variant	Exomiser rank	Mutation Distiller rank
	1	38158372	AG	A	frameshift truncation	*CDCA8*	c.12del:p.(Arg4Argfs*22)	24	22
	13	46542118	T	G	missense	*ZC3H13*	c.3842A>C:p.(Glu1281Ala)	26	56
	19	38834967	A	G	missense	*CATSPERG*	c.583A>G:p.(Arg195Gly)	27	85
	11	66247937	G	T	splice donor	*DPP3*	c.-9+1G>T:p.(=)	30	44
	12	50048748	C	G	missense	*FMNL3*	c.905G>C:p.(Arg302Pro)	31	54
	2	179432530	G	C	missense	*TTN*	c.51710C>G:p.(Thr17237Ser)	52	15
	11	61675656	C	T	missense	RAB3IL1	c.275G>A:p.(Gly92Glu)	55	42
	16	3274530	C	G	missense	*ZNF200*	c.547G>C:p.(Asp183His)	57	66
	10	99393206	A	C	missense	*MORN4*	c.139T>G:p.(Trp47Gly)	61	36
	14	21552102	G	T	stop gained	*ARHGEF40*	c.3682G>T:p.(Glu1228*)	93	57
	19	58132173	A	G	missense	*ZNF134*	c.686A>G:p.(Glu229Gly)	98	93
	17	37369297	G	C	missense	*STAC2*	c.1082C>G:p.(Pro361Arg)	105	72
	3	193361239	T	G	splice region	*OPA1*	c.1215+6T>G:p.?	1	6
	12	52163736	G	A	missense	*SCN8A*	c.3457G>A:p.(Glu1153Lys)	3	3
	17	7097296	C	T	missense	*DLG4*	c.1493G>A:p.(Arg498Lys)	8	79
	11	62397344	G	A	missense	*GANAB*	c.1745C>T:p.(Ala582Val)	10	52
NG_96	13	60667817	T	G	missense	*DIAPH3*	c.440A>C:p.(Asp147Ala)	14	13
NG_100	4	47655645	C	T	missense	*CORIN*	c.1768G>A:p.(Gly590Arg)	18	20
	1	198222290	G	A	missense	*NEK7*	c.178G>A:p.(Val60Ile)	19	39
	10	104247901	C	T	splice region	*ACTR1A*	c.315+6G>A:p.?	59	46
	15	59399518	C	A	splice region	*CCNB2*	c.25-3C>A:p.?	63	70
	12	100904621	T	C	missense	*NR1H4*	c.175T>C:p.(Tyr59His)	69	12

Table 3.3: Prioritized variants for the 5 families with a dominant mode of inheritance.
[1] The gene *NDUFS7* was reported as possible candidate but not ranked by The Exomiser since it did not follow the expected inheritance mode (genotype: 1/1).

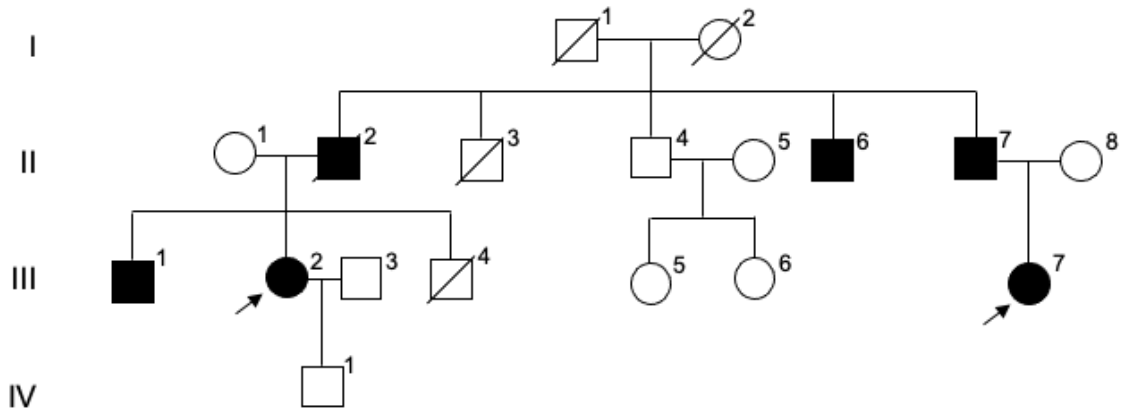


Figure 3.10: **NG_2, NG_6.** Familial pedigree.

For the first family under analysis (III:2, NG_2 and III:7, NG_6), with a clear dominant inheritance (Figure 3.10), the two affected subjects sequenced were cousins and thus the final VCF contained only 855 variants that both cases shared, of which 18 were retained after prioritization.

Some of the variants were frequently found in non-pathogenic genes, such as *MUC6* and *GOLGA8Q*, and were excluded from analysis. The variant found in gene *POU4F2* was located within a glycine stretch and thus not predicted to be disease-causing.

Familial segregation was performed for the remaining mutations: in this way, variants in *VPS18*, *NAGK* and *ZNF622* were discarded, while the ones in *CYB561A3* and *ATG2A* were segregated with disease within the family. The first protein belongs to the cytochrome b561 family and it is a lysosomal ferric-chelate reductase that reduces Fe^{3+} to Fe^{2+} before its transport from the endosome to the cytoplasm, while the second is involved in autophagy pathways, transferring lipids between membranes (Otomo & Maeda, 2019), and interacting with TOM40 at the mitochondria-associated ER membrane (MAM) (Z. Tang et al., 2019). Both genes will thus need further evaluation.

For the second family (IV:1, NG_24 and III:6, NG_137, Figure 3.11), no significant results were obtained: most of the variants were found in genes deemed not pathogenic (*ZBED6*, *SCMH1*, *LRCH1*, *LRRC37A3*); other prioritized genes were associated with different phenotypes from the ones observed in the patients (*LHX4*, *KRT10*, *PATL2*, *MYO9A*). Finally, the variant in *NDUFS7*, a core subunit of the mitochondrial respiratory chain NADH dehydrogenase, had already been discarded by segregation analysis:

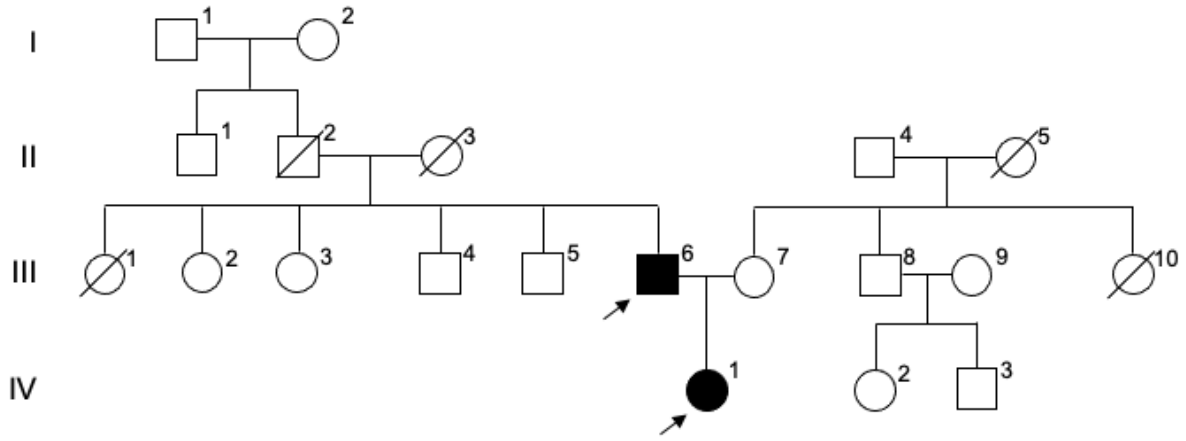


Figure 3.11: **NG_24, NG_137.** Familial pedigree.

this variant was found homozygous in the proband and heterozygous in both parents.

Patients III:2, NG_53 and II:2, NG_54 (Figure 3.12) shared around 2300 variants, but only nine were retained after filtering passages. Mutations in *AHNAK* and *IQCN* were predicted to be benign; genes associated with different phenotypes included *CENPE*, a gene recessive for microcephaly. Three other genes were discredited after segregation (*MYH11*, *WDR76*, *STAB2*).

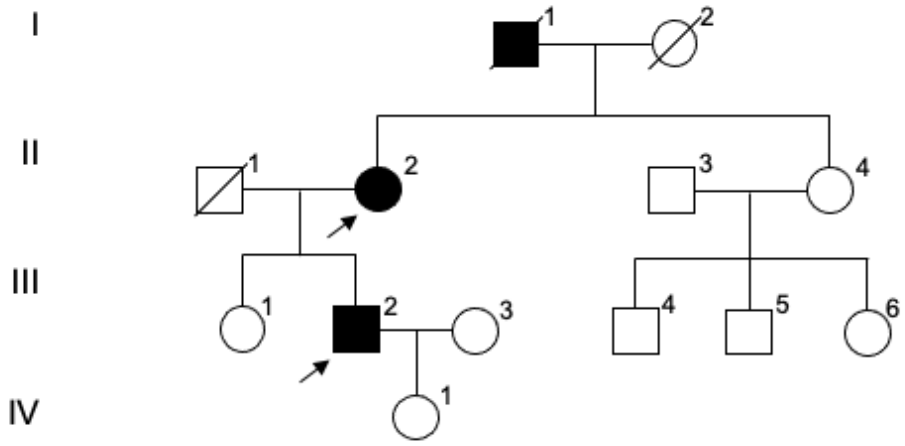


Figure 3.12: **NG_53, NG_54.** Familial pedigree.

21 variants were prioritized for patients III:2, NG_66 and II:2, NG_67 (Figure 3.13), but no significant result was obtained. Many non-pathogenic genes were prioritized by the tool, including *CCDC59*, *ASAP2*, *ZC3H13*, *DPP3*, *FMNL3*, *TTN*, *RAB3IL1*, *ZNF200*

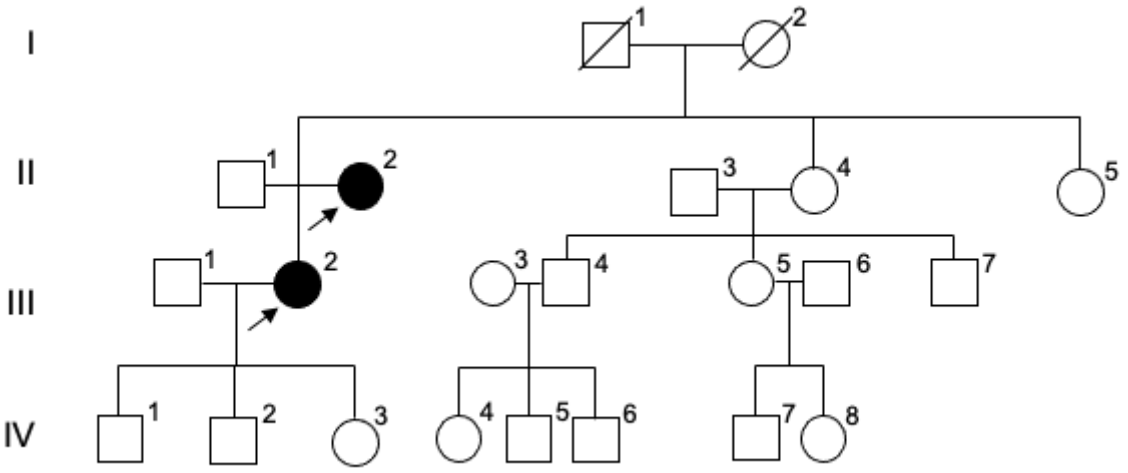


Figure 3.13: **NG_66, NG_67.** Familial pedigree.

and *ZNF134*. *RP1L1* is a dominant gene associated with retinitis pigmentosa and that plays a role in the organization of the outer segment of rod and cone photoreceptors. *COL4A3*, a type IV collagen, is known to cause Alport syndrome; *MORN4* function is not clear but the encoded protein binds *MYO3A*, a gene recessive for deafness. *CATSPERG* is a sperm-specific gene.

Two variants were predicted to be pathogenic: a frameshift truncation in *CDCA8* and a stop gain in *ARHGEF40*. However, the function of both genes does not correlate with a possible disease of mitochondrial etiology: the first is a key regulator of mitosis, while the second is probably a GEF (guanine nucleotide exchange factor) found in the nucleus and in the cytosol.

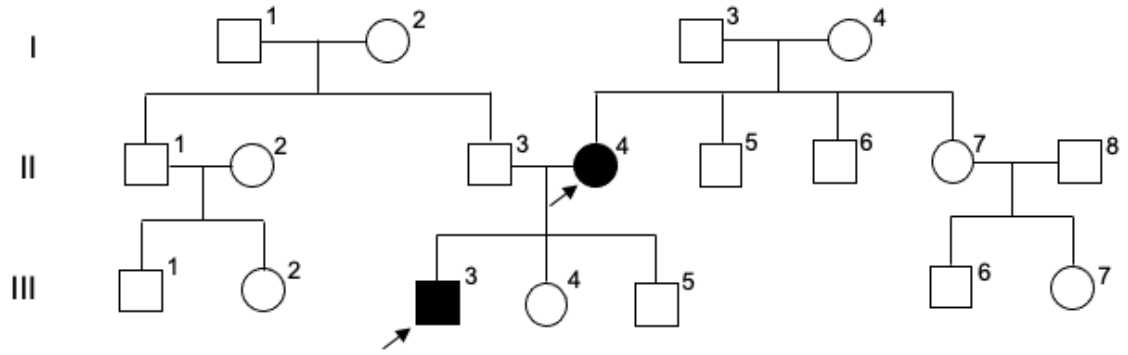


Figure 3.14: **NG_96, NG_100.** Familial pedigree.

The causative variant for optic atrophy in the last family under analysis, III:3, NG_96

and II:4, NG_100 (Figure 3.14), is probably a *de novo* mutation in a +6 site of the splice region of *OPA1*. This is a non-canonical site in splicing; however, mutations in the +5 site of this gene have been shown to be able to cause DOA (Ferré et al., 2009). In addition, familial segregation analysis has been performed, confirming this hypothesis.

A second possible candidate is *SCN8A*, that has also been segregated. This gene encodes a member of the sodium channel alpha subunit family and is known to cause dominant epilepsy. This variant was found in affected cases, but also in one unaffected individual, the maternal grandfather's proband.

All the other genes prioritized had either already been segregated (*DLG4*, *DIAPH3*, *CORIN*, *NEK7*), were non pathogenic (*ACRT1A*, *CCNB2*) or were associated with different phenotypes (*GANAB*, *NR1H4*).

Chapter 4

CONCLUSIONS

In this study, we tested the efficacy of phenotype-based variant prioritization tools in finding causative genes for Mendelian disorders, particularly in the case of mitochondrial optic neuropathies.

Between the four tools taken under analysis, during the testing phase, two did not reach as high statistical performances as expected: GenIO and OVA, in fact, were often not able to prioritize known genes associated in literature with the optic atrophy phenotype. Because of this inaccuracy, they were excluded in the successive step, that consisted in finding novel genes associated with the patients' phenotypes.

It should be pointed out, however, that these two tools were published some years ago and did not undergo any recent update. New features as well as new algorithms implementation might hence improve, in the future, their performance.

The Exomiser and MutationDistiller, instead, proved to be very powerful tools for variant prioritization, in particular in the context of known genes associated to the disease phenotype. Such an approach should hence be integrated into bioinformatic pipelines and exploited in order to support and speed up the whole-exome sequencing data analysis process for molecular diagnosis.

Some interesting results were obtained also for previously unsolved WES cases, highlighting the usefulness of phenotype-driven tools also in the field of new gene discovery, especially in recessive cases.

In particular, two genes emerge as new possible candidates for a role in optic neuropathies development. The first is *SNF8*: not much is known about this protein, a part from it being a component of the endosomal sorting complex that regulates the movement of ubiquitinylated transmembrane proteins to the lysosome for degradation (Kumthip et al., 2017). However, the presence of mutations in this gene in two unrelated probands with a quite similar phenotype gives us reason to investigate further this protein. The second

relevant finding is the one obtained for subject NG_3 and *TSPOAP1*. This gene's expression was specifically enhanced in the brain and the protein localizes in the mitochondria (Galiègue et al., 1999). Moreover, loss-of-function variants for this gene were recently linked to juvenile-onset progressive generalized dystonia, intellectual disability and cerebellar atrophy, a set of neurological phenotypes not specifically observed in our patient but that correlate with a disorder of mitochondrial etiology (Mencacci et al., 2021).
Further developments of this study will include firstly familial segregation analysis for all the previously unconsidered mutations. In cases in which the hypotheses are met, the following step will be to validate these findings by functional experiments. Furthermore, the GeneMatcher database (`https://genematcher.org/`) can be used to find and connect individuals sharing the same mutations and thus help confirm the possible involvement of the gene of interest in the observed clinical symptoms.
In conclusion, we have shown how The Exomiser and MutationDistiller can be used as effective variant prioritization tools for analysis of rare Mendelian HPO-encoded clinical phenotypes and whole-exomes. These approaches, however, should not be regarded as tools for perfect automation of the gene prioritization process, but rather as an aid to support the lengthy process usually carried out manually by researchers and clinicians.

Bibliography

Abrams, A. J., Hufnagel, R. B., Rebelo, A., Zanna, C., Patel, N., Gonzalez, M. A., ... Dallman, J. E. (2015, July). Mutations in SLC25a46, encoding a UGO1-like protein, cause an optic atrophy spectrum disorder. *Nature Genetics*, *47*(8), 926–932. Retrieved from https://doi.org/10.1038/ng.3354 doi: 10.1038/ng.3354

Adang, L. A., Pizzino, A., Malhotra, A., Dubbs, H., Williams, C., Sherbini, O., ... Vanderver, A. (2020, August). Phenotypic and imaging spectrum associated with WDR45. *Pediatric Neurology*, *109*, 56–62. Retrieved from https://doi.org/10.1016/j.pediatrneurol.2020.03.005 doi: 10.1016/j.pediatrneurol.2020.03.005

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... Sunyaev, S. R. (2010, April). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. Retrieved from https://doi.org/10.1038/nmeth0410-248 doi: 10.1038/nmeth0410-248

Amati-Bonneau, P., Valentino, M. L., Reynier, P., Gallardo, M. E., Bornstein, B., Boissiere, A., ... Carelli, V. (2008, February). OPA1 mutations induce mitochondrial DNA instability and optic atrophy 'plus' phenotypes. *Brain*, *131*(2), 338–351. Retrieved from https://doi.org/10.1093/brain/awm298 doi: 10.1093/brain/awm298

Antanaviciute, A., Watson, C. M., Harrison, S. M., Lascelles, C., Crinnion, L., Markham, A. F., ... Carr, I. M. (2015, August). OVA: Integrating molecular and physical phenotype data from multiple biomedical domain ontologies with variant filtering for enhanced variant prioritization. *Bioinformatics*, btv473. Retrieved from https://doi.org/10.1093/bioinformatics/btv473 doi: 10.1093/bioinformatics/btv473

Assink, J. J., Tijmes, N. T., ten Brink, J. B., Oostra, R.-J., Riemslag, F. C., de Jong, P. T., & Bergen, A. A. (1997, October). A gene for x-linked optic atrophy is closely linked to the xp11.4-xp11.2 region of the x chromosome. *The American Journal of Human Genetics*, *61*(4), 934–939. Retrieved from https://doi.org/10.1086/514884 doi: 10.1086/514884

Barbet, F., Gerber, S., Hakiki, S., Perrault, I., Hanein, S., Ducroq, D., ... Kaplan, J. (2003, September). A first locus for isolated autosomal recessive optic atrophy (ROA1) maps to chromosome 8q. *European Journal of Human Genetics*, *11*(12), 966–971. Retrieved from https://doi.org/10.1038/sj.ejhg.5201070 doi: 10.1038/sj.ejhg.5201070

Caporali, L., Iommarini, L., Morgia, C. L., Olivieri, A., Achilli, A., Maresca, A., ... Carelli, V. (2018, February). Peculiar combinations of individually non-pathogenic missense mitochondrial DNA variants cause low penetrance leber's hereditary optic neuropathy. *PLOS Genetics*, *14*(2), e1007210. Retrieved from https://doi.org/10.1371/journal.pgen.1007210 doi: 10.1371/journal.pgen.1007210

Caporali, L., Magri, S., Legati, A., Dotto, V. D., Tagliavini, F., Balistreri, F., ... Taroni, F. (2020, April). ATPase domain AFG3L2 mutations alter OPA1 processing and cause optic neuropathy. *Annals of Neurology*, *88*(1), 18–32. Retrieved from https://doi.org/10.1002/ana.25723 doi: 10.1002/ana.25723

Carelli, V., Musumeci, O., Caporali, L., Zanna, C., Morgia, C. L., Dotto, V. D., ... Zeviani, M. (2015, June). Syndromic parkinsonism and dementia associated with OPA1 missense mutations. *Annals of Neurology*, *78*(1), 21–38. Retrieved from https://doi.org/10.1002/ana.24410 doi: 10.1002/ana.24410

Carelli, V., Ross-Cisneros, F. N., & Sadun, A. A. (2004, January). Mitochondrial dysfunction as a cause of optic neuropathies. *Progress in Retinal and Eye Research*, *23*(1), 53–89. Retrieved from https://doi.org/10.1016/j.preteyeres.2003.10.003 doi: 10.1016/j.preteyeres.2003.10.003

Carelli, V., Schimpf, S., Fuhrmann, N., Valentino, M. L., Zanna, C., Iommarini, L., ... Wissinger, B. (2011, February). A clinically complex form of dominant optic atrophy (OPA8) maps on chromosome 16. *Human Molecular Genetics*, *20*(10), 1893–1905. Retrieved from https://doi.org/10.1093/hmg/ddr071 doi: 10.1093/hmg/ddr071

Charif, M., Chevrollier, A., Gueguen, N., Bris, C., Goudenège, D., Desquiret-Dumas, V., ... Lenaers, G. (2020, May). Mutations in the m-AAA proteases AFG3l2 and SPG7 are causing isolated dominant optic atrophy. *Neurology Genetics*, *6*(3), e428. Retrieved from https://doi.org/10.1212/nxg.0000000000000428 doi: 10.1212/nxg.0000000000000428

Charif, M., Nasca, A., Thompson, K., Gerber, S., Makowski, C., Mazaheri, N., ... Lenaers, G. (2018, January). Neurologic phenotypes associated with mutations in RTN4ip1 (OPA10) in children and young adults. *JAMA Neurology*, *75*(1), 105. Retrieved from https://doi.org/10.1001/jamaneurol.2017.2065 doi: 10.1001/jamaneurol.2017.2065

Cipriani, V., Pontikos, N., Arno, G., Sergouniotis, P. I., Lenassi, E., Thawong, P., ... Smedley, D. (2020, April). An improved phenotype-driven tool for rare mendelian variant prioritization: Benchmarking exomiser on real patient whole-exome data. *Genes*, *11*(4), 460. Retrieved from https://doi.org/10.3390/genes11040460 doi: 10.3390/genes11040460

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010, December). Identifying a high fraction of the human genome to be under selective constraint using GERP++, journal = PLoS Computational Biology. , *6*(12), e1001025. Retrieved from https://doi.org/10.1371/journal.pcbi.1001025 doi: 10.1371/journal.pcbi.1001025

Dotto, V. D., Ullah, F., Meo, I. D., Magini, P., Gusic, M., Maresca, A., ... Carelli, V. (2019, November). SSBP1 mutations cause mtDNA depletion underlying a complex optic atrophy disorder. *Journal of Clinical Investigation*, *130*(1), 108–125. Retrieved from https://doi.org/10.1172/jci128514 doi: 10.1172/jci128514

Duchen, M. R. (2004, August). Mitochondria in health and disease: perspectives on a new mitochondrial biology. *Molecular Aspects of Medicine*, *25*(4), 365–451. Retrieved from https://doi.org/10.1016/j.mam.2004.03.001 doi: 10.1016/j.mam.2004.03.001

Eiberg, H. (2005, September). Autosomal dominant optic atrophy associated with hearing impairment and impaired glucose regulation caused by a missense mutation in the WFS1 gene. *Journal of Medical Genetics*, *43*(5), 435–440. Retrieved from https://doi.org/10.1136/jmg.2005.034892 doi: 10.1136/jmg.2005.034892

Felhi, R., Sfaihi, L., Charif, M., Desquiret-Dumas, V., Bris, C., Goudenège, D., ... Fakhfakh, F. (2019, January). Next generation sequencing in family with MNGIE syndrome associated to optic atrophy: Novel homozygous POLG mutation in the c-terminal sub-domain leading to mtDNA depletion. *Clinica Chimica Acta*, *488*, 104–110. Retrieved from https://doi.org/10.1016/j.cca.2018.11.003 doi: 10.1016/j.cca.2018.11.003

Ferré, M., Bonneau, D., Milea, D., Chevrollier, A., Verny, C., Dollfus, H., ... Amati-Bonneau, P. (2009, July). Molecular screening of 980 cases of suspected hereditary optic neuropathy with a report on 77 novelOPA1mutations. *Human Mutation*, *30*(7), E692–E705. Retrieved from https://doi.org/10.1002/humu.21025 doi: 10.1002/humu.21025

Galiègue, S., Jbilo, O., Combes, T., Bribes, E., Carayon, P., Fur, G. L., & Casellas, P. (1999, January). Cloning and characterization of PRAX-1. *Journal of Biological Chemistry*, *274*(5), 2938–2952. Retrieved from https://doi.org/10.1074/jbc.274.5.2938 doi: 10.1074/jbc.274.5.2938

Gerber, S., Charif, M., Chevrollier, A., Chaumette, T., Angebault, C., Kane, M. S., ... Lenaers, G. (2017, September). Mutations in DNM1l, as in OPA1, result in dominant optic atrophy despite opposite effects on mitochondrial fusion and fission. *Brain*, *140*(10), 2586–2596. Retrieved from https://doi.org/10.1093/brain/awx219 doi: 10.1093/brain/awx219

Haack, T. B., Haberberger, B., Frisch, E.-M., Wieland, T., Iuso, A., Gorza, M., ... Prokisch, H. (2012, April). Molecular diagnosis in mitochondrial complex i deficiency using exome sequencing. *Journal of Medical Genetics*, *49*(4), 277–283. Retrieved from https://doi.org/10.1136/jmedgenet-2012-100846 doi: 10.1136/jmedgenet-2012-100846

Hanein, S., Perrault, I., Roche, O., Gerber, S., Khadom, N., Rio, M., ... Rozet, J.-M. (2009, April). TMEM126a, encoding a mitochondrial protein, is mutated in autosomal-recessive nonsyndromic optic atrophy. *The American Journal of Human Genetics*, *84*(4), 493–498. Retrieved from https://doi.org/10.1016/j.ajhg.2009.03.003 doi: 10.1016/j.ajhg.2009.03.003

Heimer, G., Kerätär, J. M., Riley, L. G., Balasubramaniam, S., Eyal, E., Pietikäinen, L. P., ... Yi, Q. (2016, December). MECR mutations cause childhood-onset dystonia and optic atrophy, a mitochondrial fatty acid synthesis disorder. *The American Journal of Human Genetics*, *99*(6), 1229–1244. Retrieved from https://doi.org/10.1016/j.ajhg.2016.09.021 doi: 10.1016/j.ajhg.2016.09.021

Hombach, D., Schuelke, M., Knierim, E., Ehmke, N., Schwarz, J. M., Fischer-Zirnsak, B., & Seelow, D. (2019, May). MutationDistiller: user-driven identification of pathogenic DNA variants. *Nucleic Acids Research*, *47*(W1), W114–W120. Retrieved from https://doi.org/10.1093/nar/gkz330 doi: 10.1093/nar/gkz330

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., ... Sieh, W. (2016, October). REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*, *99*(4), 877–885. Retrieved from https://doi.org/10.1016/j.ajhg.2016.08.016 doi: 10.1016/j.ajhg.2016.08.016

Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., ... Bejerano, G. (2016, October). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nature Genetics*, *48*(12), 1581–1586. Retrieved from https://doi.org/10.1038/ng.3703 doi: 10.1038/ng.3703

Katz, B. J., Zhao, Y., Warner, J. E., Tong, Z., Yang, Z., & Zhang, K. (2006). A family with x-linked optic atrophy linked to the OPA2 locus xp11.4-xp11.2. *American Journal of Medical Genetics Part A*, *140A*(20), 2207–2211. Retrieved from https://doi.org/10.1002/ajmg.a.31455 doi: 10.1002/ajmg.a.31455

Kerrison, J. B. (1999, June). Genetic heterogeneity of dominant optic atrophy, kjer type. *Archives of Ophthalmology*, *117*(6), 805. Retrieved from https://doi.org/10.1001/archopht.117.6.805 doi: 10.1001/archopht.117.6.805

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014, February). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. Retrieved from https://doi.org/10.1038/ng.2892 doi: 10.1038/ng.2892

Kirkman, M. A., Yu-Wai-Man, P., Korsten, A., Leonhardt, M., Dimitriadis, K., Coo, I. F. D., ... Chinnery, P. F. (2009, June). Gene–environment interactions in leber hereditary optic neuropathy. *Brain*, *132*(9), 2317–2326. Retrieved from https://doi.org/10.1093/brain/awp158 doi: 10.1093/brain/awp158

Koile, D., Cordoba, M., de Sousa Serro, M., Kauffman, M. A., & Yankilevich, P. (2018, January). GenIO: a phenotype-genotype analysis web server for clinical genomics of rare diseases. *BMC Bioinformatics*, *19*(1). Retrieved from https://doi.org/10.1186/s12859-018-2027-3 doi: 10.1186/s12859-018-2027-3

Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., ... Prokisch, H. (2017, June). Genetic diagnosis of mendelian disorders via RNA sequencing. *Nature Communications*, *8*(1). Retrieved from https://doi.org/10.1038/ncomms15824 doi: 10.1038/ncomms15824

Kumthip, K., Yang, D., Li, N. L., Zhang, Y., Fan, M., Sethuraman, A., & Li, K. (2017, October). Pivotal role for the ESCRT-II complex subunit EAP30/SNF8 in IRF3-dependent innate antiviral defense. *PLOS Pathogens*, *13*(10), e1006713. Retrieved from https://doi.org/10.1371/journal.ppat.1006713 doi: 10.1371/journal.ppat.1006713

La Morgia, C., Maresca, A., Caporali, L., Valentino, M. L., & Carelli, V. (2020, May). Mitochondrial diseases in adults. *Journal of Internal Medicine*, *287*(6), 592–608. Retrieved from https://doi.org/10.1111/joim.13064 doi: 10.1111/joim.13064

Li, J., Zhao, T., Zhang, Y., Zhang, K., Shi, L., Chen, Y., ... Sun, Z. (2018, July). Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Research*, *46*(15), 7793–7804. Retrieved from https://doi.org/10.1093/nar/gky678 doi: 10.1093/nar/gky678

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., ... and, C. T.-S. (2012, February). A systematic survey of loss-of-function variants in human protein-coding genes. *Science*, *335*(6070), 823–828. Retrieved from https://doi.org/10.1126/science.1215040 doi: 10.1126/science.1215040

Malicdan, M. C. V., Vilboux, T., Ben-Zeev, B., Guo, J., Eliyahu, A., Pode-Shakked, B., ... Anikster, Y. (2017, November). A novel inborn error of the coenzyme q10 biosynthesis pathway: cerebellar ataxia and static encephalomyopathy due to COQ5 c-methyltransferase deficiency. *Human Mutation*, *39*(1), 69–79. Retrieved from https://doi.org/10.1002/humu.23345 doi: 10.1002/humu.23345

Maresca, A., Caporali, L., Strobbe, D., Zanna, C., Malavolta, D., Morgia, C., ... Carelli, V. (2014, October). Genetic basis of mitochondrial optic neuropathies. *Current Molecular Medicine*, *14*(8), 985–992. Retrieved from https://doi.org/10.2174/1566524014666141010132627 doi: 10.2174/1566524014666141010132627

Maresca, A., la Morgia, C., Caporali, L., Valentino, M. L., & Carelli, V. (2013, July). The optic nerve: A "mito-window" on mitochondrial neurodegeneration. *Molecular and Cellular Neuroscience*, *55*, 62–76. Retrieved from https://doi.org/10.1016/j.mcn.2012.08.004 doi: 10.1016/j.mcn.2012.08.004

Mencacci, N. E., Brockmann, M. M., Dai, J., Pajusalu, S., Atasu, B., Campos, J., ... Acuna, C. (2021, February). Bi-allelic variants in TSPOAP1, encoding the active zone protein RIMBP1, cause autosomal recessive dystonia. *Journal of Clinical Investigation*. Retrieved from https://doi.org/10.1172/jci140625 doi: 10.1172/jci140625

Momtazmanesh, S., Rayzan, E., Shahkarami, S., Rohlfs, M., Klein, C., & Rezaei, N. (2020, June). A novel VPS13b mutation in cohen syndrome: a case report and review of literature. *BMC Medical Genetics*, *21*(1). Retrieved from https://doi.org/10.1186/s12881-020-01075-1 doi: 10.1186/s12881-020-01075-1

Ng, P. C., & Henikoff, S. (2001, May). Predicting deleterious amino acid substitutions. *Genome Research*, *11*(5), 863–874. Retrieved from https://doi.org/10.1101/gr.176601 doi: 10.1101/gr.176601

Ogilvie, I. (2005, October). A molecular chaperone for mitochondrial complex i assembly is mutated in a progressive encephalopathy. *Journal of Clinical Investigation*, *115*(10), 2784–2792. Retrieved from https://doi.org/10.1172/jci26020 doi: 10.1172/jci26020

Otomo, T., & Maeda, S. (2019, August). ATG2a transfers lipids between membranes in vitro. *Autophagy*, *15*(11), 2031–2032. Retrieved from https://doi.org/10.1080/15548627.2019.1659622 doi: 10.1080/15548627.2019.1659622

Pello, R., Martín, M. A., Carelli, V., Nijtmans, L. G., Achilli, A., Pala, M., ... Ugalde, C. (2008, September). Mitochondrial DNA background modulates the assembly kinetics of OXPHOS complexes in a cellular model of mitochondrial disease. *Human Molecular Genetics*, *17*(24), 4001–4011. Retrieved from https://doi.org/10.1093/hmg/ddn303 doi: 10.1093/hmg/ddn303

Pisano, A., Preziuso, C., Iommarini, L., Perli, E., Grazioli, P., Campese, A. F., ... Giordano, C. (2015, September). Targeting estrogen receptor as preventive therapeutic strategy for leber's hereditary optic neuropathy. *Human Molecular Genetics*, ddv396. Retrieved from https://doi.org/10.1093/hmg/ddv396 doi: 10.1093/hmg/ddv396

Qi, H., Zhang, H., Zhao, Y., Chen, C., Long, J. J., Chung, W. K., ... Shen, Y. (2021, January). MVP predicts the pathogenicity of missense variants by deep learning. *Nature Communications*, *12*(1). Retrieved from https://doi.org/10.1038/s41467-020-20847-0 doi: 10.1038/s41467-020-20847-0

Rahman, S. (2020, April). Mitochondrial disease in children. *Journal of Internal Medicine*, *287*(6), 609–633. Retrieved from https://doi.org/10.1111/joim.13054 doi: 10.1111/joim.13054

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T. J., ... Mootha, V. K. (2020, November). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Research*, *49*(D1), D1541–D1547. Retrieved from https://doi.org/10.1093/nar/gkaa1011 doi: 10.1093/nar/gkaa1011

Reynier, P. (2004, September). OPA3 gene mutations responsible for autosomal dominant optic atrophy and cataract. *Journal of Medical Genetics*, *41*(9), e110–e110. Retrieved from https://doi.org/10.1136/jmg.2003.016576 doi: 10.1136/jmg.2003.016576

Robinson, P. N., Kohler, S., Oellrich, A., Wang, K., Mungall, C. J., Lewis, S. E., ... and, D. S. (2013, October). Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Research*, *24*(2), 340–348. Retrieved from https://doi.org/10.1101/gr.160325.113 doi: 10.1101/gr.160325.113

Sadun, A. A., Morgia, C. L., & Carelli, V. (2013, April). Mitochondrial optic neuropathies: our travels from bench to bedside and back again. *Clinical & Experimental Ophthalmology*, n/a–n/a. Retrieved from https://doi.org/10.1111/ceo.12086 doi: 10.1111/ceo.12086

Schon, K. R., Ratnaike, T., van den Ameele, J., Horvath, R., & Chinnery, P. F. (2020, September). Mitochondrial diseases: A diagnostic revolution. *Trends in Genetics*, *36*(9), 702–717. Retrieved from https://doi.org/10.1016/j.tig.2020.06.009 doi: 10.1016/j.tig.2020.06.009

Schwarz, J. M., Rödelsperger, C., Schuelke, M., & Seelow, D. (2010, August). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*, *7*(8), 575–576. Retrieved from https://doi.org/10.1038/nmeth0810-575 doi: 10.1038/nmeth0810-575

Siepel, A. (2005, August). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*(8), 1034–1050. Retrieved from https://doi.org/10.1101/gr.3715005 doi: 10.1101/gr.3715005

Stenton, S. L., & Prokisch, H. (2020, June). Genetics of mitochondrial diseases: Identifying mutations to help diagnosis. *EBioMedicine*, *56*, 102784. Retrieved from https://doi.org/10.1016/j.ebiom.2020.102784 doi: 10.1016/j.ebiom.2020.102784

Stenton, S. L., Sheremet, N. L., Catarino, C. B., Andreeva, N., Assouline, Z., Barboni, P., ... Prokisch, H. (2021, January). Impaired complex i repair causes recessive leber's hereditary optic neuropathy. *Journal of Clinical Investigation*. Retrieved from https://doi.org/10.1172/jci138267 doi: 10.1172/jci138267

Sun, Y., Ruivenkamp, C. A., Hoffer, M. J., Vrijenhoek, T., Kriek, M., van Asperen, C. J., ... Santen, G. W. (2015, April). Next-generation diagnostics: Gene panel, exome, or whole genome? *Human Mutation*, *36*(6), 648–655. Retrieved from https://doi.org/10.1002/humu.22783 doi: 10.1002/humu.22783

Sundaram, L., Gao, H., Padigepati, S. R., McRae, J. F., Li, Y., Kosmicki, J. A., ... Farh, K. K.-H. (2018, July). Predicting the clinical impact of human mutation with deep neural networks. *Nature Genetics*, *50*(8), 1161–1170. Retrieved from https://doi.org/10.1038/s41588-018-0167-z doi: 10.1038/s41588-018-0167-z

Tamiya, G., Makino, S., Hayashi, M., Abe, A., Numakura, C., Ueki, M., ... Hayasaka, K. (2014, September). A mutation of COX6a1 causes a recessive axonal or mixed form of charcot-marie-tooth disease. *The American Journal of Human Genetics*, *95*(3), 294–300. Retrieved from https://doi.org/10.1016/j.ajhg.2014.07.013 doi: 10.1016/j.ajhg.2014.07.013

Tang, S., Le, P. K., Tse, S., Wallace, D. C., & Huang, T. (2009, February). Heterozygous mutation of opa1 in drosophila shortens lifespan mediated through increased reactive oxygen species production. *PLoS ONE*, *4*(2), e4492. Retrieved from https://doi.org/10.1371/journal.pone.0004492 doi: 10.1371/journal.pone.0004492

Tang, Z., Takahashi, Y., He, H., Hattori, T., Chen, C., Liang, X., ... Wang, H.-G. (2019, August). TOM40 targets atg2 to mitochondria-associated ER membranes for phagophore expansion. *Cell Reports*, *28*(7), 1744–1757.e5. Retrieved from https://doi.org/10.1016/j.celrep.2019.07.036 doi: 10.1016/j.celrep.2019.07.036

Taylor, R. C., Cullen, S. P., & Martin, S. J. (2008, March). Apoptosis: controlled demolition at the cellular level. *Nature Reviews Molecular Cell Biology*, *9*(3), 231–241. Retrieved from https://doi.org/10.1038/nrm2312 doi: 10.1038/nrm2312

Yarosh, W., Monserrate, J., Tong, J. J., Tse, S., Le, P. K., Nguyen, K., ... Huang, T. (2008, January). The molecular mechanisms of OPA1-mediated optic atrophy in drosophila model and prospects for antioxidant treatment. *PLoS Genetics*, *4*(1), e6. Retrieved from https://doi.org/10.1371/journal.pgen.0040006 doi: 10.1371/journal.pgen.0040006

Yu-Wai-Man, P., Griffiths, P., Gorman, G., Lourenco, C., Wright, A., Auer-Grumbach, M., ... Chinnery, P. (2010, February). Multi-system neurological disease is common in patients with OPA1 mutations. *Brain*, *133*(3), 771–786. Retrieved from https://doi.org/10.1093/brain/awq007 doi: 10.1093/brain/awq007

Zanna, C., Ghelli, A., Porcelli, A. M., Karbowski, M., Youle, R. J., Schimpf, S., ... Carelli, V. (2007, December). OPA1 mutations associated with dominant optic atrophy impair oxidative phosphorylation and mitochondrial fusion. *Brain*, *131*(2), 352–367. Retrieved from https://doi.org/10.1093/brain/awm335 doi: 10.1093/brain/awm335