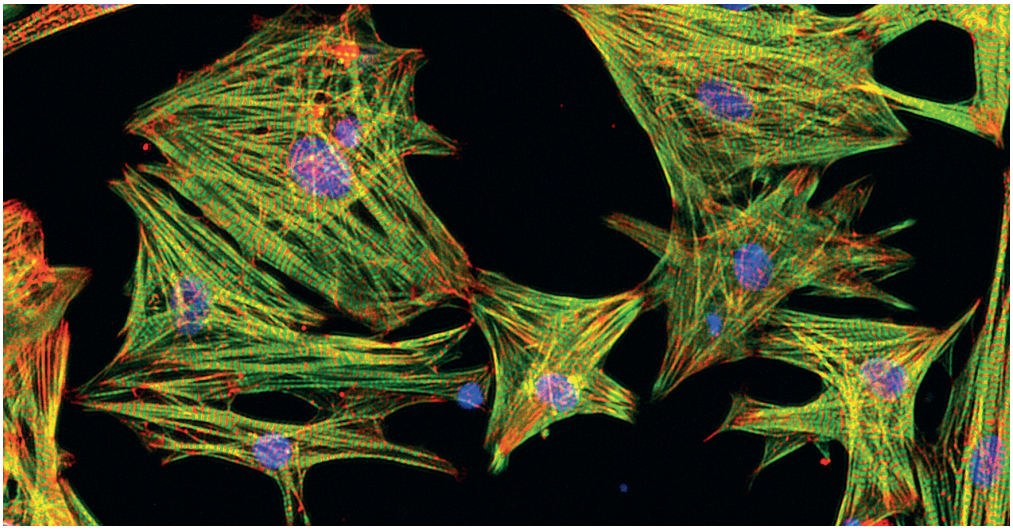
DISSERTATIONES SCHOLAE DOCTORALIS AD SANITATEM INVESTIGANDAM
UNIVERSITATIS HELSINKIENSIS

CATALINA VASILESCU

GENETICS OF RARE CHILDHOOD DISORDERS
WITH SPECIAL FOCUS ON CARDIOMYOPATHIES



STEM CELLS AND METABOLISM RESEARCH PROGRAM
FACULTY OF MEDICINE
DOCTORAL PROGRAMME IN BIOMEDICINE
UNIVERSITY OF HELSINKI

Stem Cells and Metabolism Research Program
&
Doctoral Programme in Biomedicine
University of Helsinki, Finland

GENETICS OF RARE CHILDHOOD DISORDERS

WITH SPECIAL FOCUS ON CARDIOMYOPATHIES

Catalina Vasilescu

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine,
University of Helsinki, for public examination remotely by Zoom, from lecture hall 2,
Metsätalo, on the 4th of December 2020, at 12 o'clock.

Helsinki 2020

Supervised by

Professor Anu Suomalainen Wartiovaara
Stem Cells and Metabolism Research Program
University of Helsinki, Finland

&

PhD Christopher Carroll
Molecular and Clinical Sciences
St George's University of London, United Kingdom

Reviewed by

Professor Katriina Aalto-Setälä
Faculty of Medicine and Health Technology
Tampere University, Finland

&

Professor Johanna Kuusisto
Centre For Medicine and Clinical Research
University of Eastern Finland

Discussed by

Professor Holger Prokisch
Institute of Neurogenomics
Helmholtz Zentrum München, Germany

ISBN 978-951-51-6873-3 (PRINT)
ISBN 978-951-51-6874-0 (ONLINE)
ISSN 2342-3161 (PRINT)
ISSN 2342-317X (ONLINE)

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis

Hansaprint
Helsinki 2020

ABSTRACT

Next-Generation Sequencing (NGS) technologies enabled the characterization of the human genome and its variation in great detail within large cohorts. The current medical research aims towards personalized medicine, whereby identifying the causal disease mechanisms in each individual will promote more tailored forms of treatment. Rare diseases have been in the front line of the individual-level molecular diagnoses, resulting from their typical monogenic inheritance. NGS technologies made it possible to solve cases that were a challenge to traditional methods, either imposed by their extreme rarity or by the genetic heterogeneity of the disorders. In genetic studies, NGS technologies involve targeted sequencing panels of known disease genes, whole-exome sequencing (WES) covering the genome's protein-coding part, and whole-genome sequencing (WGS). Our study applied targeted panels and WES for the genetic diagnosis of severe childhood disorders, starting with a progressive neurological syndrome and continuing with a cohort of 66 children diagnosed with cardiomyopathy, a leading cause of pediatric heart transplants.

First, we focused on genetic leukoencephalopathies, a heterogeneous group of central nervous system disorders with white matter involvement. We studied a Finnish patient presenting genetic hypomyelinating leukoencephalopathy with atrophy of the cerebellar vermis. WES identified a novel homozygous mutation in *HIKESHI*. Hikeshi binds and carries the molecular chaperones HSP70s into the nucleus in response to heat shock. A founder Ashkenazi-Jewish mutation previously linked Hikeshi and its heat shock protective function to leukoencephalopathy. Structural modeling of the amino acid changes and protein analyses revealed that modified interactions inside Hikeshi's hydrophobic pockets induce protein instability. The patient's cells showed impaired nuclear translocation of HSP70s during heat shock and decreased ERO1-Lα, an endoplasmic reticulum (ER) oxidoreductase. Our study was the second report of *HIKESHI* mutations, extended the phenotypic spectrum associated with Hikeshi disease, and showed that the cellular pathogenic process involves nuclear chaperone and ER functions.

Second, we studied the genetics of childhood cardiomyopathies (CMPs), the most common cause of heart failure and the leading indication for a heart transplant in children. Cardiomyopathies can manifest as isolated heart disorders or as components of developmental and metabolic disorders. Despite severe courses, often leading to heart transplantation, childhood cardiomyopathies have been relatively understudied from a genetic perspective. To cover this knowledge gap, we collected a unique cohort of 66 childhood-onset CMP patients (KidCMP) from the single center in Finland performing cardiac transplantations. Overall in the cohort, including six previously published findings, we uncovered the pathogenic variants in 39%

of patients: 46% *de novo*, 34% recessive, and 20% dominantly inherited. The disease genes converge on metabolic causes (*PRKAG2*, *MRPL44*, *AARS2*, *HADHB*, *DNAJC19*, *PPA2*, *TAZ*, *BAG3*), MAPK pathways (*HRAS*, *PTPN11*, *RAF1*, *TAB2*), development (*NEK8*, *TBX20*), calcium handling (*JPH2*, *CALM1*, *CACNA1C*), the sarcomeric function (*TNNC1*, *TNNI3*, *ACTC1*, *MYH7*), and the cytoskeleton (*NRAP*). We found that severe childhood cardiomyopathies are typically caused by rare pathogenic variants, commonly *de novo*, indicating that NGS of trios is the best approach in their diagnosis. Our study was the first countrywide genetic characterization of severe early-onset cardiomyopathies and indicated the importance of early diagnosis in treatment optimization.

Third, we identified a recessive pathogenic variant in *TMOD1* as a novel genetic cause for cardiomyopathy in three patients from two independent families. Tmod1 binds filamentous actin at its pointed, slow-growing end, controlling thin filament length in cardiomyocytes. The overexpression of mutant and wild type protein fused with GFP in U2OS (human osteosarcoma) cells suggested that the mutant is prone to form aggregates. In the patient muscle, however, the protein was at normal levels. We employed induced pluripotent stem cells derived cardiomyocytes (iPSC-CMs) for investigating the distribution of Tmod1 and related proteins in the patient and controls. We found only minor changes in the sarcomeric structure, challenging to interpret in the context of differentiation heterogeneity. Molecular modeling of patients' mutation showed that the variant determines a hydrophobic amino acid exposure at the protein's surface, which may influence the local folding and affinity for actin. Taken together with the cell-based studies, we suggest that the patient mutation likely alters thin filament's pointed end dynamics in cardiomyocytes.

We made substantial progress in understanding the molecular basis of the studied disorders. First, we strengthen the evidence that heat shock response is a novel mechanism underlying leukoencephalopathy. Second, we characterized genetically a cohort of early-onset severe cardiomyopathies, representing the whole of Finland for patients evaluated for cardiac transplantation or receiving inotropic support. Third, we characterized a novel disease gene causing childhood cardiomyopathies, an important step in further deciphering the genetic landscape of these severe heart disorders.

Altogether, this thesis highlights the power of novel technologies in identifying causal genetic variants and characterizing novel disease genes. Our findings enhance knowledge of the underlying molecular mechanisms and potentially aid in developing new therapeutic interventions. For families, the genetic diagnosis enables a causative recognition of the disease and identifying individuals at risk. Our cardiomyopathy project also contributed to establishing a protocol for systematic genetic testing of patients and families at the Pediatric Cardiology Department of the University of Helsinki Central Hospital.

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis comprises three original publications:

I **Vasilescu C.**, Isohanni P., Palomäki M., Pihko H., Suomalainen A., Carroll C.J. Absence of Hikeshi, a nuclear transporter for heat-shock protein HSP70, causes infantile hypomyelinating leukoencephalopathy. European Journal of Human Genetics (2017). 25:366-370.

II **Vasilescu C.**, Ojala T.H., Brilhante V., Ojanen S., Hinterding H., Palin E., Alastalo T.P., Koskenvuo J., Hiippala A., Jokinen E., Jahnukainen T., Lohi J., Pihkala J., Tyni T.A., Carroll C.J., Suomalainen A. Genetic basis of severe childhood-onset cardiomyopathies. Journal of American College of Cardiology (2018). 72(19):2324-38.

III **Vasilescu C.**, Ojala T., Lohi J., Ylanen K., Rahkonen O., Poutanen T., Martelius L., Kumari R., Hinterding H., Brilhante V., Ojanen S, Lappalainen P., Koskenvuo J., Carroll C.J., Suomalainen A. Recessive *TMOD1* mutation causes childhood cardiomyopathy. Manuscript.

The publications are referred to within the text by their roman numerals.

ABBREVIATIONS

3D	Three-dimensional
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
AJ	Ashkenazi-Jewish
ALCAPA	Anomalous left coronary artery from the pulmonary artery
AR	Autosomal recessive
ATP	Adenosine triphosphate
AVC	Arrhythmogenic ventricular cardiomyopathy
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
C. elegans	*Caenorhabditis elegans*
CI-V	Mitochondrial complexes I-V
CADD	Combined annotation dependent depletion
CDM3	Cardiomyocyte medium v3
CMP	Cardiomyopathy
cMRI	Cardiac magnetic resonance imaging
CMs	Cardiomyocytes
CNS	Central nervous system
CNV	Copy number variant
CRISPR	Clustered regularly interspaced short palindromic repeats
CTD	Carboxy-terminal domain
CSF	Cerebrospinal fluid
DAPI	4', 6-Diamidino-2-phenylindole
DCM	Dilated cardiomyopathy
DIDA	Database of digenic disorders
D-loop	Displacement loop
DMEM	Dulbecco's modified Eagle medium
DNA	Deoxyribonucleic acid
ECG	Electrocardiogram
EDTA	Ethylenediaminetetraacetic acid
EF	Ejection fraction
EGTA	Ethylene glycol tetraacetic acid
E-loop	Extended loop
ENCODE	Encyclopedia of DNA elements
ER	Endoplasmic reticulum
ER-UPR	Endoplasmic reticulum - unfolded protein response
ESCs	Embryonic stem cells
ExAC	The Exome Aggregation Consortium
F-actin	Filamentous actin
FBS	Fetal bovine serum

FG-nups	Phenylalanine-glycine repeat-containing nucleoporins
FIMM	Finnish Institute for Molecular Medicine
FLAIR	Fluid attenuated inversion recovery
G	Guanine
GC	Guanine cytosine
GFP	Green fluorescent protein
gnomAD	The Genome Aggregation Database
GpC	Guanine cytosine dinucleotides
GRCh38	Genome Reference Consortium human build 38
GWAS	Genome-wide association studies
HCM	Hypertrophic cardiomyopathy
HGP	Human Genome Project
HICM	Histiocytoid cardiomyopathy
HRP	Horseradish peroxidase
IPA	Integrated Pathway Analysis
iPSCs	Induced pluripotent stem cells
iPSC-CMs	Induced pluripotent stem cell-derived cardiomyocytes
KCl	Potassium chloride
KidCMP	Our Finnish cohort of childhood-onset cardiomyopathy patients
lncRNA	Long noncoding RNA
LRR	Leucine-rich repeat
LVNC	Left ventricular noncompaction cardiomyopathy
MAF	Minor allele frequency
Mb	Megabase
MEFs	Mouse embryonic fibroblasts
MgCl2	Magnesium chloride
NaCl	Sodium chloride
MNV	Multi nucleotide variant
MOPS	3-(N-morpholino)propanesulfonic acid
MRI	Magnetic resonance imaging
mtDNA	Mitochondrial DNA
NCBI	National Center for Biotechnology Information
nDNA	Nuclear DNA
NGS	Next-Generation Sequencing
NTD	Amino-terminal domain
NYHA	New York Heart Association classification
O	Atomic oxygen
OSKM	Transcription factors Oct3/4, Sox2, Klf4, and c-Myc
OXPHOS	Oxidative phosphorylation
PAGE	Polyacrylamide gel electrophoresis
PB	Phosphate-buffered
PBS	Phosphate-buffered saline
PDB	Protein Data Bank
PFA	Paraformaldehyde
PolyPhen-2	Polymorphism Phenotyping v2

PVDF	Polyvinylidene fluoride
qPCR	Quantitative polymerase chain reaction
QT	Interval seen in an electrocardiogram
RCM	Restrictive cardiomyopathy
RIPA	Radioimmunoprecipitation assay
RNA	Ribonucleic acid
RPMI 1640	Roswell Park Memorial Institute 1640 medium
mRNA	Messenger RNA
miRNA	Micro RNA
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
sncRNA	Small noncoding RNA
SIFT	Sorting intolerant from tolerant
SISu	Sequencing Initiative Suomi
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
STRs	Short tandem repeats
SV	Structural variant
T	Thymine
T1, T2	In MRI, the time between magnetic pulses and image acquisition
TADs	Topologically-associated domains
TIM-TOM	Translocases of the inner-outer mitochondrial membrane
tRNA	Transfer RNA
U2OS	Human osteosarcoma cell line
UPR	Unfolded protein response
UTR	Untranslated region
VNTR	Variable number tandem repeats
VUS	Variant of unknown significance
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

*This list does not include abbreviations of genes, proteins, and amino acids.

1. INTRODUCTION

The human genome is built on a four-letter nucleotide language, which ultimately governs all aspects of cell functioning. With the aims of providing a genetic, physical, and sequence map of our genome, the Human Genome Project (HGP) was initiated in 1990, the first 'draft' published in 2001 (Lander et al, 2001; Venter et al, 2001), and continued few more years afterward to finalize the 'essentially complete genome'. This 3.5-billion-dollar, 13-year program was a substantial multicenter effort, which promoted significant technological and bioinformatic advancements. It also inoculated a culture of large-scale cooperation in research and introduced the concept of 'big data' in biology (Hood and Rowen, 2013). The 'catalogs' of genes and proteins assembled by the HGP have been essential for the development of 'systems biology' applications in biomedical sciences. HGP's societal echoes are enormously increasing, as they gradually include novel aspects, such as genetic privacy, potential discrimination, genetic determinism, or non-determinism, and identity politics.

The HGP inspired other large-scale projects in multiple biology disciplines, such as the Cancer Genome Atlas (The Cancer Genome Atlas Research Network et al, 2013), the Human Brain Project (www.humanbrainproject.eu), and the Human Proteome Project (Legrain et al, 2011). Following HGP, the description of human genetic variation was pioneered by the HapMap Project (International HapMap Consortium, 2005; International HapMap Consortium et al, 2007), which aimed to identify haplotype blocks of common single nucleotide polymorphisms (SNPs) in different populations. The 1000 Genomes Project followed (The 1000 Genomes Project Consortium, 2010) and cataloged common and rare single-nucleotide and structural variation in multiple populations.

The data obtained in these large projects supported early clinical genome-wide association studies (GWAS), which correlate specific genetic variants with disease risk based on case-control comparisons. On the other hand, to identify rare, highly penetrant disease-causing variants, pedigree-based whole-exome sequencing offers a more robust approach. The dichotomy of common variants-common disorders *versus* rare variants-rare disorders has been the basis for the technological split between the two genetic areas. Therefore, microchips *versus* exome-sequencing gradually hand over to whole-genome sequencing, which fundamentally aims to capture all genetic variation in a genome, common and rare, coding and noncoding. Presently, hundreds of thousands of individual genomes are sequenced worldwide to identify disease-causing variants or, more broadly, to establish associations between sequence variation and phenotypes (Claussnitzer et al, 2020).

It is foreseeable that, soon, individual genomes will play a larger role in medical practice. Individuals will use the genomic information to improve their health through prevention or therapeutic strategies appropriate for a condition indicated by their genome (Özdemir et al, 2017). The application of systems approaches to diseases shapes our understanding of how metabolic or cell signaling networks become perturbed in different diseased tissues and how they change as the disease progresses. Having a clear picture of how these molecular events unfold will provide novel approaches to diagnosis and treatments. Adding this layer of information to the knowledge of the individual genetic makeup, with its own risk and protective variants, will shape a predictive, preventive, personalized, and participatory medicine: P4 medicine.

Given the simpler genetic architecture of rare diseases, typically of monogenic origin, they are one of the immediate applications of P4 medicine. Moreover, it is easier to study and catalog one-gene perturbation effects in different models and tissues. Rare disorders have also been the front line of genetic investigation using Next-Generation Sequencing (NGS) approaches, with tremendous success in disease gene discovery over the last decade (Frésard and Montgomery, 2018).

Despite this, it is estimated that about half of the causative genes involved in rare disorders are still to be identified (Bamshad et al, 2019; Chong et al, 2015). Many factors can contribute to this 'missing heritability', among them the rarity of different disorders, their genetic and phenotypic variability, the need for technological and bioinformatic advancements to capture all types of genome variation, along with overcoming current challenges in interpreting noncoding functional variation of the genome (Boycott et al, 2019).

In this thesis, we used NGS technologies coupled with functional studies to advance the molecular characterization of rare childhood disorders in Finland, focusing on cardiomyopathy. Induced pluripotent stem cell disease modeling allowed us to explore the possibilities that are at the horizon for personalized approaches in rare disorders.

2. REVIEW OF THE LITERATURE

2.1. STRUCTURE AND EXPRESSION OF THE HUMAN GENOME

2.1.1. CHROMOSOMES AND 3D ORGANIZATION

The DNA represents the genetic material containing all the information essential for life and the basis for heredity. The human genome comprises 46 nuclear chromosomal DNA molecules (22 pairs of autosomes and two sex chromosomes XY) and a multicopy small circular mitochondrial genome (Strachan and Read, 2018). The DNA in the nucleus (nDNA) hosts most genes and contains >6 billion nucleotides that would measure ~2 m in length (Fraser et al, 2015). One circular copy of the mitochondrial genome (mtDNA) contains 16,569 base pairs and harbors only 37 intronless genes. MtDNA exists in multiple copies inside mitochondria, between 100 and 10,000 molecules per cell, on a cell-type-specific basis (Gammage and Frezza, 2019). Exceptionally, oocytes present over 100,000 mtDNA copies that are distributed between daughter cells during initial embryonic cell divisions.

The three-dimensional organization of the genome is crucial for both integrity maintenance and gene regulation. Because of extensive length and functional constraints, the nDNA is hierarchically folded to reach a compaction level estimated at ~10,000 fold for metaphase chromosomes (Grigoryev, 2018).

During interphase, chromosomes exist as territories, defined by the nuclear space occupied by each chromosome, with a roughly ovoidal shape and 2-3 μm in diameter. Interphase chromosomes have irregular folding that implies a less constrained chromatin state, which allows the access of gene regulatory factors to the DNA. Adjacent territories may overlap considerably, and chromatin loops can reach out into neighboring territories (Fraser et al, 2015).

Within chromosome territories, there are two types of chromosome compartments: open A compartments associated with indices of active and poised chromatin (such as specific histone markers, high GC content, increase in DNaseI hypersensitivity sites); closed B compartments of silenced chromatin (such as silencing histone markers, tighter self-association, late replication timing) (Fraser et al, 2015; Lieberman-Aiden et al, 2009).

A smaller-scale organization gives rise to topologically-associated domains (TADs). TADs form distinct globular, self-interacting units that are well delimitated spatially and contain genes that tend to be coexpressed (Dixon et al, 2012). Even finer structures emerge from the ability of the chromatin to fold into loops, promoting long-distance interactions.

Chromatin looping has been the preferred model to explain the interaction between promoters and regulatory elements (Rao et al, 2014a).

Nucleosomes represent the basic packaging level and contain octamers of core histone proteins and a DNA segment wrapping around them. Nucleosomes are separated by stretches of DNA, often occupied by linker histones. This structure is generally compared with beads-on-a-string and forms the classical ~10 nm chromatin fiber that provides 7-fold compaction of the DNA (Krumm and Duan, 2019). Recent studies suggested that, in interphase chromosomes, chromatin fibers are irregularly folded from both nucleosome-containing and nucleosome-depleted DNA stretches, resulting in fibers with diverse diameters (5 nm - 24 nm) and poorly described geometry *in vivo* (Ou et al, 2017). Metaphase chromosomes are at least 100 fold more compacted than interphase chromosomes. This level of compaction in metaphase chromosomes is achieved by tightly packed consecutive chromatin loops of ~100 kb in length (Goloborodko et al, 2016).

At whole nucleus resolution, the genome organization depends on contacts with different nuclear substructures. For example, laminin-associated domains and nucleolus-associated domains frequently contain transcriptionally-inactive heterochromatin and are gene-poor. On the contrary, nuclear pores present in their vicinity 'heterochromatin exclusion zones', suggesting a possible coupling of transcription with nuclear export (Fraser et al, 2015). The 3D organization is largely similar in specific cell types. During development or cell differentiation, chromatin compartments undergo dynamic switches as cells transition between cell types (Zheng and Xie, 2019). Disruption of the 3D genome architecture can lead to disease, often by a pathogenic promoter-enhancer interaction (Lupiáñez et al, 2015, 2015), which disrupts the normal gene expression.

2.1.2. GENES, REGULATORY ELEMENTS, AND REPETITIVE SEQUENCES

From human genetics point of view, much emphasis has been on identifying and characterizing the genome's genes because a large proportion of the variation involved in disease resides within genes. The genes are functional elements of the genome, consisting of transcribed DNA sequences that ultimately generate functional products: proteins, functional RNAs, or both.

Analyses of the human reference genome (GRCh38) indicate that there are approximately 19,950-20,200 protein-coding genes. Nevertheless, the actual number of unique proteins increases by magnitudes using alternative start codons, alternative splicing of messenger RNA (mRNA), and post-translational modifications of proteins. In many instances, these isoforms have different tissue specificities and functions (Boeckmann et al, 2005). For protein-coding genes, the mean exon size is 268 nucleotides (median 129), with an average of 9 exons per gene. If we exclude the first and last exons, typically long and mostly untranslated, the mean length of internal exons is

147 nucleotides (median 121). Introns vary enormously in size, from a few to over 100,000 nucleotides, with a small minority of genes lacking introns (Strachan and Read, 2018). Because transcription of long introns is costly (16 hours to transcribe the 2.3 Mb dystrophin gene) (Tennyson et al, 1995), highly expressed genes tend to have short introns or are intronless (Castillo-Davis et al, 2002).

The human genome also contains RNA genes, which are much more diverse and far less characterized than protein-coding genes. RNA genes raise particular challenges related to their definition. Their identification is also challenging, often because they have been poorly conserved during evolution and they do not have open reading frames to facilitate searching (Kashi et al, 2016). By convention, RNA genes belong to two classes: short noncoding RNAs (sncRNAs), when their mature products are ≤200 nucleotides, or long noncoding RNAs (lncRNAs) if >200 nucleotides. The 200 nucleotide threshold promotes an arbitrary classification, which artificially separates the ribosomal RNAs into two classes (cytoplasmic 28S and 18S, mitochondrial 16S and 12S belong to lncRNAs, while cytoplasmic 5.8S and 5S to sncRNA). Genes encoding lncRNAs can be tens of kilobases long and often contain introns, with an average of 1.8 transcripts per gene. The average length of exons of lncRNA genes is 683 nucleotides, with a median of 129 (Strachan and Read, 2018). On the other hand, the small RNA genes, like those that specify micro RNAs (miRNAs), are often intronless. However, small functional noncoding RNAs can arise even from processed introns after splicing, complicating the RNA gene definition. As a consequence of their diversity, the functional units of RNA genes are the mature noncoding transcripts and not the DNA sequences that encode them (Palazzo and Lee, 2015; Strachan and Read, 2018).

There are at least three reasons why there is variation in the number of genes observed in the human genome: first reason, as mentioned above, stems from some difficulties in defining the genes because their classical view as discrete entities has been challenged by atypical examples (such as overlapping genes, genes-within-genes, 'conjoined genes'); second, genes can vary in number between haplotypes due to structural variation; and third, updates of the reference genome and new releases of patches may refine the number of detected genes (Table 1).

About 75% of the genome seems to be transcribed at least in some cells or at specific time points during development (Djebali et al, 2012). Noncoding transcripts are abundant, with largely uncharacterized functions. Also, the traditional division of genes into protein-coding and RNA genes has been challenged. The average human protein-coding gene produces about six to seven transcripts, with some being noncoding transcripts that may have functional roles. The reverse is also true when genes originally classified as encoding lncRNAs have been found to contain short sequences that encode micropeptides (Li and Liu, 2019). These examples suggest a substantial

revision of the traditional view and definition of genes (Portin and Wilkins, 2017).

Table 1. Statistics of human genes.
Data from NCBI, GRCh38.p12 (released in December 2017) at www.ncbi.nlm.nih.gov/genome/annotation_euk/Homo_sapiens/109/
Data from GENECODE, GRCh38.p13 (released in February 2019), at www.gencodegenes.org/human/stats.html

Feature	NCBI GRCh38.p12	GENCODE GRCh38.p13
Genes and pseudogenes	54,644	60,669
Protein-coding genes	20,203	19,959
mRNAs	113,620	84,068
RNA genes	17,871	17,960 (lncRNA) 7,578 (sncRNA)
Noncoding RNAs	46,854	48,479 (lncRNAs)
Pseudogenes	16,152	14,763
Pseudo transcripts	1,346	nd

Families of genes emerged through duplication events that led to increases in the human genome's functional diversity. These events involved duplications of single genes (Conrad and Antonarakis, 2007) or of larger chromosomal regions (Bailey et al, 2002), in addition to the evolutionarily ancient whole-genome duplication.

Pseudogenes are copies of a functional gene that have acquired inactivating mutations and can no longer produce the gene product. They can arise by copying genomic DNA sequences, as in the case of nonprocessed pseudogenes. Another mechanism is retrotransposition, when a processed RNA transcript is reverse-transcribed and reintegrated into the genome, as in the case of processed pseudogenes (Zheng and Gerstein, 2007). Retrogenes are intronless genes that arose by retrotransposition, acquired the ability to be expressed in the new genomic context, and retained their function under selective pressure (Kaessmann et al, 2009).

The ENCODE project began in 2003 as a pilot study to define all functional elements in the human genome. The functional elements include protein- and RNA-genes, as well as regulatory elements (promoters, enhancers, repressors/silencers, insulators) (Kellis et al, 2014). In general, functional elements are considered DNA stretches that encode a defined product or present a reproducible biochemical signature, such as transcription or a specific chromatin structure (ENCODE Project Consortium, 2011). The ENCODE reports showed initial evidence for ~400,000 regions with features of enhancers and >70,000 promoter-like

regions (ENCODE Project Consortium, 2012). Their finding that over 80% of the human genome is either transcribed or participates in an RNA or chromatin-associated event was broadly interpreted that the 'junk DNA' concept should be abandoned. This idea, however, met strong resistance from evolutionary geneticists (Germain et al, 2014; Palazzo and Gregory, 2014).

Over two-thirds of the human genome consists of repetitive sequences (de Koning et al, 2011). They encompass high copy number tandem repeats (satellite, minisatellite, and microsatellite DNA) and transposon repeats (DNA transposons and retrotransposons).

Among tandem repeats, satellite DNA primarily associates with highly condensed, constitutive heterochromatin. Nevertheless, these sequences can be transcribed, as it is the case for DNA in telomere heterochromatin, with the resulting transcripts regulating different aspects of telomere biology [reviewed in (Biscotti et al, 2015)].

The majority of the human genome sequence originates from transposons, DNA elements that can move from one genomic location to another. The great majority of transposons in the human genome moved by a 'copy-and-paste' mechanism (retrotransposons) and a minority by 'cut-and-paste' (DNA transposons). Most human transposon sequences are now inactive, as they acquired inactivating mutations or lost functional elements [reviewed by (Goodier, 2016)], but sometimes they can cause human disease (Hancks and Kazazian, 2012).

2.1.3. MITOCHONDRIAL DNA

Human mtDNA is typically maternally inherited, with rare observed instances of biparental inheritance (Luo et al, 2018; McWilliams and Suomalainen, 2019; Schwartz and Vissing, 2002). In most cases, paternal mtDNA is degraded after fertilization by a poorly understood mechanism that in *Caenorhabditis elegans* involves endonuclease G, autophagy, and proteasome degradation pathways (Zhou et al, 2016).

As already mentioned, mitochondria have a multicopy genome, but this does not necessarily imply that all mtDNA molecules are identical. Variation in the mtDNA sequence gives rise to heteroplasmy, a phenomenon where some of the mtDNA molecules in the cell vary in sequence from others (Figure 1). The sequence variation can be (nearly) neutral, or pathogenic. A high-level heteroplasmy of an mtDNA mutation is associated with disease.

In the development of a female, the process of oocyte formation may lead to very different mtDNA heteroplasmy levels that can be transmitted to offsprings. Therefore, the fate of future offsprings is determined in the early life of a mother-to-be. The phenomenon has been termed 'mitochondrial bottleneck', a concept borrowed from population genetics because a small number of mitochondrial DNA molecules allows the drift of heteroplasmy in the oocytes. There is a debate on whether the bottleneck occurs in the early

or later stages of oocyte development. Although mtDNA has ~100 fold higher mutation rare than nDNA, evidence exists for mechanisms of purifying selection against the transmission of certain mtDNA mutations [reviewed by (Stewart and Larsson, 2014)].

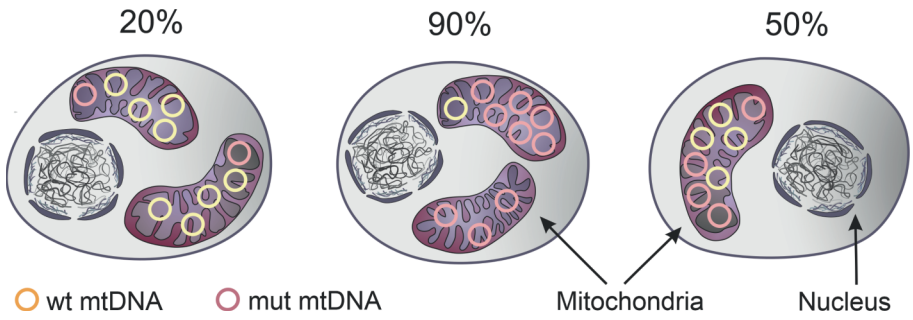


Figure 1. Mitochondrial DNA heteroplasmy. A high heteroplasmy level for a mutant mtDNA is associated with disease.

MtDNA is gene-dense, without introns, encoding 13 proteins of the oxidative phosphorylation system (OXPHOS), as well as two rRNAs and 22 tRNAs that enable the translation of the 13 proteins on mitochondrial ribosomes. A small stretch of noncoding DNA forms a structure called displacement loop (D-loop), which contains major regulatory elements for replication and transcription initiation (Falkenberg et al, 2007). Unlike the DNA in the nucleus, mtDNA does not associate with histones but is complexed with other proteins of the matrix to form nucleoids, connected to the inner mitochondrial membrane. Each nucleoid typically contains a single mtDNA molecule, protected and bent by a high amount of transcription factor A (TFAM) (Ngo et al, 2014).

2.1.4. COMPLEX CONTROL OF GENOME EXPRESSION

The 100 Mb genome of *C. elegans*, a 1 mm long nematode, consists of six chromosomal pairs but contains about the same number of protein-coding genes as the human genome (Hillier et al, 2005). Where does the difference in complexity come from? The answer may not be straightforward. However, the observation that humans have many more RNA genes (~17,800 in human *versus* ~1,300 in worm, www.wormbook.org), with roles in regulating gene expression, suggests that the differences may rely mostly on the complex control mechanisms of human genome expression. The control of gene expression happens at multiple levels: DNA accessibility, transcription, translation, mRNA turnover, mRNA or protein localization, and protein stability.

The regulation of gene expression at the cellular level takes place continuously, both in short-term as response to external signals and in long-

term as part of specific cellular and tissue identity. The transcriptome and proteome vary considerably in different cell types, and the mechanisms that govern the stable tissue-specific identity are epigenetic. Therefore, we can say that the characteristic form and behavior of different cell types result from different epigenetically determined readouts of the same genome (Strachan and Read, 2018).

First, to transcribe a gene, the DNA must be accessible, not buried in densely packed chromatin (Sivakumar et al, 2019). As discussed before, chromatin conformation plays a crucial role in the regulation of gene expression. Chemical modifications of DNA and histones shape the local chromatin organization. Participants in this process are enzymes that methylate cytosines in DNA, enzymes that modify histones in nucleosomes, proteins that bind methylated DNA or modified histones, and ATP-driven chromatin remodeling complexes that can change the positions of nucleosomes (Flaus and Owen-Hughes, 2011). DNA methylation modifies the functional state of regulatory regions but preserves the Watson-Crick base pairing of methylated cytosine. DNA methylation represents the classical epigenetic mark. It is involved in many forms of stable epigenetic repression: imprinting, X-chromosome inactivation, silencing of repetitive DNA, and tissue-specific gene expression. In vertebrates, heritable methylation occurs only at CpG dinucleotides (Schübeler, 2015).

Transcription of a gene requires the transcription initiation complex to be assembled on the promoter. DNA looping brings distant sites close to the promoter, influencing its activity. The promoter activity is affected by sequence-specific DNA-binding proteins that bind either directly to the promoter or at more distant places, such as enhancers and repressors (Venkatesh and Workman, 2015). Co-activators and co-repressors do not bind DNA but are recruited by protein-protein interactions (Schoenfelder and Fraser, 2019).

Protein translation is another control level for gene expression. This control often depends on miRNAs that bind to sequences in the 3' untranslated region of mRNAs (Bartel, 2009). Moreover, translation has a complex regulation at the level of ribosome-mRNA interaction, with specific factors promoting the preferential translation of mRNA subsets. A few examples include regulation of cap-dependant (Richter and Sonenberg, 2005) and cap-independent translation (Walters and Thompson, 2016), or the synthesis of selenoproteins (Kossinova et al, 2013).

Mitochondria have specific ribosomes, which translate the 13 OXPHOS subunits encoded by mtDNA. However, >99% of the proteins residing inside mitochondria are specified by nuclear genes and translated on cytoplasmic ribosomes. They are imported into mitochondria through transmembrane translocase complexes (TIM-TOM), often co-translationally. This process also applies to the remaining 79 nuclear-encoded OXPHOS components. Further, different proteases, chaperones, and assembly factors assist in forming the inner membrane complexes from the dual origin subunits

(Pearce et al, 2017) (Figure 2). Mitochondrial and cytosolic translation seem to be dynamically regulated for the synthesis of OXPHOS components. Studies in yeast have shown that cytosolic translation controls mitochondrial translation unidirectionally (Couvillion et al, 2016).

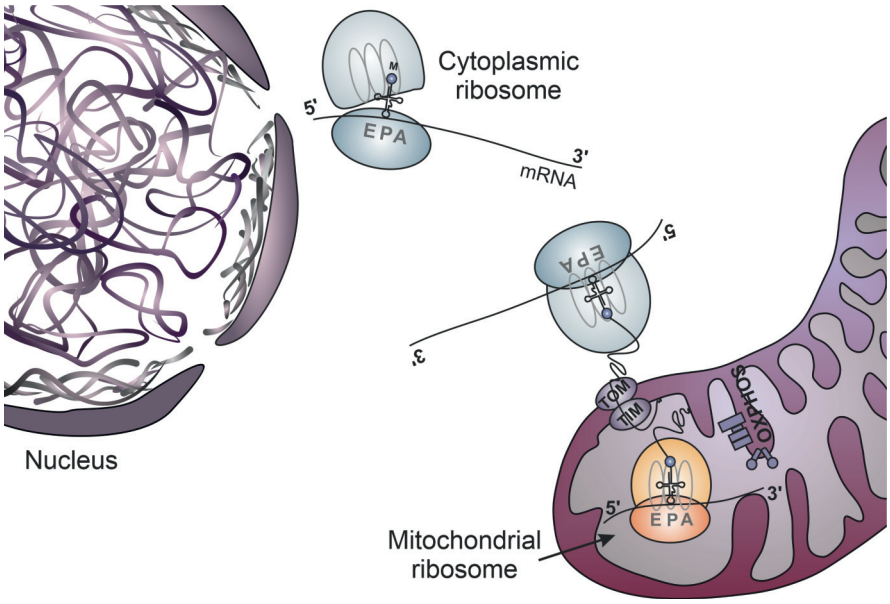


Figure 2. Coordination of cytoplasmic and mitochondrial translation in the synthesis of OXPHOS components.

Besides the epigenetic mechanisms involved in cell identity and tissue-specific regulation, two other significant examples of epigenetic processes are X-inactivation and imprinting. In female cells, a randomly-selected X chromosome is inactivated at the blastocyst stage of embryogenesis (Deng et al, 2014). All daughter cells will inactivate the same X as the mother cell. This process depends on *Xist* lncRNA, expressed only from the inactive X, physically coating it, and repressing gene expression (Chang and Brown, 2010; Moindrot and Brockdorff, 2016). About 15% of X-linked genes escape inactivation, at least in some tissues and in some individuals. There are two pseudoautosomal regions that escape inactivation, but in general, X inactivation is patchy (Carrel and Willard, 2005). Epigenetic marks are usually erased at meiosis, but various instances exist when they are transmitted from parent to offspring. Such an example is imprinting, whereby alleles at a locus carry marks of their parental origin, which ultimately specify their expression. This effect has been described for a few dozen human genes (Soellner et al, 2017). Apart from these well-characterized imprinted genes, other effects of transgenerational epigenetic

memory seem to exist, but their conceptual framework is poorly defined (Miska and Ferguson-Smith, 2016).

2.2. VARIATION OF THE HUMAN GENOME IN HEALTH AND DISEASE

2.2.1. DEFINITION AND TYPES OF VARIANTS

The DNA sequence of the genome determines, together with environmental factors, the characteristics of each individual. Sequence variation underlies traits, such as height and eye color, but also disorders.

The genetic variation encompasses all the permanent changes in the DNA sequence of the genome. Therefore, variants are any differences between newly generated genetic data and the reference genome. Variants range from single nucleotide changes to gains and losses of whole chromosomes, with the following main types [reviewed by (Lappalainen et al, 2019)]:

Single-nucleotide variants (SNVs) are changes of one DNA base to another, estimated to be around 3-4 million per genome. Multi-nucleotide variants (MNVs) are clusters of two or more nearby variants, with functional consequences that may differ from individual ones when they reside within the same codon (Wang et al, 2020).

Small insertions and deletions (indels) are gains or losses of genetic material <50bp, estimated at 0.4-0.5 million in a typical genome. SNVs and indels represent the vast majority of variation in the human population. While most of this variation has no or little functional impact, some can cause severe disorders.

Structural variation (SV) is a larger scale type of variation, quite diverse, defined by ≥50 bp in size. It includes copy-number variants (CNVs), rearrangements, and mobile element insertions. SVs are fewer in number in an average genome comparing with SNVs and indels, accounting for approximately 0.2% of total variants. Typically they have a larger functional impact owing to their size (Sudmant et al, 2015). SVs can disrupt genes and alter gene dosage, rearrange regulatory elements and genomic context. Identifying different SV classes may require distinct technological and algorithmic approaches (Audano et al, 2019).

Tandem repeat variation comprises three classes: (1) short tandem repeats (STRs or microsatellites, which have the repeat unit of 1-6 bp), (2) variable number tandem repeats (VNTRs or minisatellites, with 7-49 bp repeat unit), (3) centromeric and heterochromatin repeats (satellite DNA, with various size repeat unit). Tandem repeats are the most challenging forms of variation to detect with current sequencing technologies.

2.2.2. CAUSES OF GENETIC VARIATION

Genetic variants result from DNA changes that have not been corrected by the cellular repair systems (Shendure and Akey, 2015). No cellular process (e.g., DNA repair system) or enzyme (e.g., DNA polymerase) can function with 100% efficiency (Barnes and Lindahl, 2004). As a consequence, it is thought that endogenous factors determine the majority of genetic variants. Occasionally, external mutagens cause DNA changes. Such mutagens are radiation and chemicals in our environment. The molecular sources of DNA changes are errors in chromosome segregation and recombination, DNA replication and repair, and chemical damage to DNA.

In early development, chromosome segregation errors can give rise to embryos with an altered number of DNA molecules (such as Down syndrome). In somatic cells, the altered chromosome number is a feature of many cancers. CNVs are typically caused by mispaired recombination during meiosis or recombination-like mechanisms in somatic cells (Conrad et al, 2010). CNV formation also has a complex and poorly understood connection with early/late replication timing of the DNA (Chen et al, 2014). SNVs, MNVs, and indels typically arise during DNA replication (random polymerase errors and polymerase slippage at tandem repeats) or as a consequence of chemical damage to DNA (strand breakage, base deletion, base modification, and cross-linking).

There are two categories of molecular pathways dedicated to DNA repair [reviewed by (Chatterjee and Walker, 2017)]: 1) pathways that repair single-stranded damage and 2) the repair of double-stranded damage. Failures in these processes lead to permanent sequence alterations. The alterations are transmitted to all daughter cells when they occur in dividing somatic cells or to offspring if they occur in germline cells.

However, variants do not exist evenly across the genome. There are so-called 'mutation hotspots'. For example, CpG deamination causes more C to T transitions than other dinucleotide pairs. Repeat junctions are another example of sites prone to variation due to polymerase slippage during replication. Moreover, DNA regions with late replication timing associate with higher SNPs in the population, higher newly formed point mutations, and CNVs [reviewed by (Chen et al, 2014)]. The estimated rate of germline SNVs is $1.2x10^{-8}$ per base per generation, about 63 new mutations per person, but in CpG sites, the rate is $1.12x10^{-7}$ (Kong et al, 2012). Germline MNV rate is estimated to be only 1–3% of the SNV rate (Wang et al, 2020). Moreover, germline mutations are mostly of paternal origin, and they increase in number with paternal age, although a maternal age effect was also reported (Goldmann et al, 2016).

2.2.3. THE SPECTRUM OF GENETIC VARIATION AND POPULATION HISTORY

Human population history, with ancient bottlenecks and recent expansion, is reflected in variant allele frequencies. A criterion to classify the variants is the minor allele frequency (MAF). Based on MAF, the variants are: common (>5%), low-frequency (1-5%), rare (<1%), and ultra-rare (<0.01%) (Lappalainen et al, 2019). If we look at the individual level, most variants are common because they arose early in human history and are now present in all major ancestry groups. However, at the population level, most variants are ultra-rare, resulting from recent generations of population growth, with 50-100 new mutations occurring in each individual (Ralph and Coop, 2013). Allele frequency reflects both allele's age and the selective forces acting on it. A low allele frequency may indicate a recent variant in population or a deleterious variant that is inefficiently passed on to new generations, or both. In the light of purifying selection, we can understand that most deleterious variants are rare because they affect reproductive fitness. For this reason, a small allele frequency is the primary variant prioritization criterion in severe Mendelian disorders with strong effects on reproductive fitness. Presently, the most accurate allele frequency estimates come from the databases gnomAD (https://gnomad.broadinstitute.org/) and SISu for the Finnish population (http://www.sisuproject.fi/). However, these databases do not contain structural and repetitive variants, and larger indels are also limited.

Isolated populations resulting from recent bottlenecks typically display a lower level of rare neutral variation and an enrichment in functional and even deleterious variants. Some of these variants can be present at relatively high frequencies, a consequence of increased drift and reduced selective pressure. Finland is an example of such an isolated population with multiple historical bottlenecks (Chheda et al, 2017). The population bottlenecks appeared when small groups of settlers inhabited remote, isolated areas, with the resulting populations further subjected to famine periods. The dual theory of Finland's inhabitation postulates an early migration wave into Finland from the east ~4,000 years ago involving eastern Uralic speakers, followed by a later migration ~2,000 years ago from the south, involving Indo-European speakers (Peltonen et al, 1999). Additional significant migratory movements were within Finland in the 16th century in the eastern and northern parts (Norio, 2003a).

These historical events caused the enrichment of some disease-causing variants and others' losses, a phenomenon known as the founder effect. The concept of 'Finnish disease heritage' was introduced in 1973 (Norio et al, 1973) and describes presently 36 monogenic diseases more frequent in Finland than in any other population (Norio, 2003b). It is a consequence of how founder effects and genetic isolation have shaped the Finns' gene pool. Most Finnish heritage diseases have an autosomal recessive mode of inheritance, while two are autosomal dominant, and two X-linked. The genetic causes for 35 of these diseases have been characterized, with the

particularity that one founder mutation accounts for 70-100% of disease alleles (http://www.findis.org/index.php).

2.2.4. GENETIC VARIATION IN HUMAN DISEASE

Each human genome contains ~4 million SNVs and indels (Pelak et al, 2010), of which ~25,000 occur in protein-coding regions. Already about five decades ago, it was postulated that the vast majority of genetic variation is neutral (Kimura, 1968) or nearly neutral (Ohta, 1973) and does not have a dramatic evolutionary effect. Supporting this view, data from genome sequencing projects have shown that the human genome can tolerate a large degree of variation without any apparent pathogenic effect. For example, a typical human genome contains ~100 rare loss-of-function variants, which are predicted to disrupt the gene function, with ~20 genes being completely inactivated (MacArthur et al, 2012).

However, not all of the genetic variation is neutral or without phenotypic effect. Some variants increase susceptibility to certain disorders or even directly cause them. Moreover, genetic variants can also be advantageous and positively selected during evolution (Hurst, 2009). Developmental genes typically tolerate little variation in the amino acid sequence, whereas nonessential genes can sometimes be completely knocked out without phenotypic effect (Sulem et al, 2015). Given that proteins represent essential functional elements of the cell, it is no surprise that much of the deleterious variation eliminated by negative selection resides within the exome. Along this line, the 1000 Genomes Project showed that exons have 50% less genetic diversity compared to introns (The 1000 Genomes Project Consortium, 2010). However, awareness has been raised on the importance of noncoding variation for human disease. For example, many variants that predispose to multifactorial diseases are located in regulatory regions that affect gene expression levels. Noncoding variation can also have a dramatic effect on gene function, for example, by disrupting splicing.

Genetic variants can be classified based on their consequence on the gene product (Strachan and Read, 2018). Loss-of-function variants result in abolished function. A decreased function characterizes hypomorphic variants. Dominant-negative variants can disrupt the function of the wild-type allele. Gain-of-function variants cause an enhanced or altered function, for example, a kinase that is always in an active conformation.

Another classification of variants specifies their effect on the protein sequence. Silent or synonymous variants do not change the amino acid sequence, whereas missense variants cause a substitution of an amino acid residue to another. Typical loss-of-function DNA changes include splice site (disrupt splicing), nonsense (premature stop codon), and frameshift indel (disrupt reading frame) variants. However, these types of variants do not always cause loss of function. For example, when a nonsense or frameshift variant occurs at the end of the gene's coding sequence, and to a smaller

extent for those at the very beginning (MacArthur et al, 2012). On the other hand, noncoding, missense, and synonymous variants can have complete loss-of-function consequences. An example constitutes recently-described variants in UTR regions (Whiffin et al, 2020) that prevent the ribosome from correctly translating the mRNA. Some additional examples are missense changes that disrupt catalytically active residues or induce protein instability, along with synonymous and deep intronic variants that disrupt splicing.

2.3. GENETICS OF RARE DISORDERS AND NEXT-GENERATION SEQUENCING

2.3.1. DEFINITION AND GENETIC ARCHITECTURE OF RARE DISORDERS

Rare disorders are defined as life-threatening or chronically debilitating disorders with a prevalence of less than 1:2,000 in the population and are mostly genetic in origin. Around 7,000 rare disorders have been reported, with new syndromes continuously being described. With ~350 million people suffering from rare disorders worldwide, among which 50-75% being children, they are common overall (Claussnitzer et al, 2020). Rare disorders are not only severe and chronic but also often difficult to diagnose and without adequate treatment. Understanding their genetic etiology is pivotal for choosing the best available treatment, as well as for promoting research towards the development of improved, targeted treatments.

The heritability of rare genetic diseases typically follows the simple rules of inheritance proposed by Gregor Mendel in the 19th century; therefore, they are also called Mendelian. This type of inheritance is a consequence of genetic defects that are fully penetrant and affect a single gene. For this reason, they are also considered monogenic (Strachan and Read, 2018). An autosomal recessive disease develops if deleterious variants affect both copies of the gene. X-linked recessive disorders affect primarily males when their unique copy of the X chromosome carries the defect. Females can be mildly and variably affected, depending on their X chromosome inactivation. Loss-of-function or hypomorphic variants typically underlie recessive diseases. A dominant disease is caused by heterozygous variation that can be gain-of-function, loss-of-function (haploinsufficient diseases), or dominant-negative.

De novo variants are new sequence changes that occur either in parental germline cells (sometimes mosaic in a parent) or at some point after conception. They contribute to human diseases and, when they affect the individual's reproductive fitness, are not transmitted to the next generations (Veltman and Brunner, 2012). At healthy population level, this severe variation does not exist.

Somatic mutations are not inherited; they affect a subset of cells in the body and are *de novo* genetic variants that occur after conception in any cell type except germ cells. Although they have been studied broadly in cancer, their role in rare and common disorders has also been increasingly recognized (Li and Williams, 2013).

Discovering the disease-causing variants in rare disorders is very important for adequate treatment of patients, for assessing disease prognosis, and for genetic counseling of the family. A genetic diagnosis can also reduce the number of tests, medical appointments, and the emotional burden associated with the illness. Moreover, there is a scientific purpose beyond patient care, as the association of a phenotype to a gene typically brings knowledge on the gene function and related cellular pathways.

2.3.2. TRADITIONAL AND MODERN GENETIC APPROACHES IN RARE DISORDERS

Five years ago, the estimated discovery of causative genes in all known Mendelian disorders was 50% (Chong et al, 2015). Consequently, there are still many novel disease genes to uncover, with the additional consideration that many new disorders are still uncharacterized (Bamshad et al, 2019). This view agrees with the ExAC findings, where over 3,000 genes are nearly depleted of protein-truncating variants, 72% of them without any present disease association (Lek et al, 2016).

Compared to multifactorial, complex conditions, Mendelian disorders' simpler genetic architecture positioned them on the front line of gene discovery for a long time. Disease gene mapping (localization of disease genes using proxies) started in the 1980s, marked by mapping of Huntington's disease gene to chromosome 4 (Gusella et al, 1983). During the next two decades, gene discovery studies employed mainly genome-wide linkage analysis, using polymorphic genetic markers that segregate together with the disease locus in large pedigrees. This method was able to identify the chromosomal region of a disease gene, sometimes further combined with positional cloning to isolate and characterize the gene. Other successfully employed methods were karyotyping and homozygosity mapping (Strachan and Read, 2018). These methods allowed the identification of numerous disease loci, but the genetic findings were limited to large family pedigrees or diseases with a particular clinical presentation.

Moreover, disease gene mapping in isolated populations, such as Finland and Iceland, was incredibly successful because of reduced genetic heterogeneity originating from population bottlenecks in the early phases of population history (Nelis et al, 2009). Traditional methods were underpowered to detect genetic components of complex diseases because the effect sizes of the underlying variants are small compared to Mendelian disorders. The traditional methods also do not tolerate high genetic or phenotypic heterogeneity (Strachan and Read, 2018).

The 'first-generation' DNA sequencing method was the dideoxy chain-termination developed by Frederick Sanger in the 1970s, the reason for which it is also called 'Sanger sequencing' (Sanger and Coulson, 1975; Sanger et al, 1977). The initial gel-based method was subsequently multiplexed and automatized for the detection of products through capillary electrophoresis. Sanger sequencing played a crucial role in the Human Genome Project, but today applies to small-scale applications such as confirmation of variants or rarely to a candidate gene screening (Shendure et al, 2017).

The 'Next-Generation Sequencing' (NGS) era started in 2005 by introducing the first commercial sequencing machine, generating low-cost, high-throughput sequence data. A few chemistries are underlying different NGS technologies. One of the most widely used is sequencing-by-synthesis, initially developed at Solexa and later acquired by Illumina. This method produces short sequence reads, 50-300 bp in length, making their alignment to the reference genome more challenging and limiting the method's capacity to detect repeat variation and complex genomic rearrangements (DePristo et al, 2011).

The 'third-generation sequencing' started with introducing single-molecule technologies (Pacific Biosciences and Oxford Nanopore Technologies), capable of producing long-read data. However, they are not yet widely used because of high error rates, high costs, and fewer bioinformatic tools developed (Heather and Chain, 2016)

The automation and standardization of short-read NGS technologies determined to be widely adopted and promoted a shift in costs and challenges from the sequencing process itself on the storage capacity, bioinformatic analysis, and interpretation of the data.

2.3.3. NEXT-GENERATION SEQUENCING IN GENE DISCOVERY AND DIAGNOSIS OF RARE DISORDERS

NGS technologies have revolutionized gene discovery in human diseases and made it possible to decipher the genetic basis of phenotypically and genetically heterogeneous disorders, a challenge for traditional methods (Wright et al, 2018).

Figure 3 summarizes the gene discovery pace in rare disorders, marking the advancements in research resources and key scientific findings by NGS, with focus on cadiomyopathy.

The collection of large cohorts of patients often comes with the cost of a shallow phenotypic description. NGS-based methods enable using the 'genotype-first' or 'reverse phenotyping' approach, when the precise diagnosis is established primarily on the genetic data. NGS technologies allow the evaluation of all annotated variants in a hypothesis-free manner, important especially in phenotypically heterogeneous diseases (Bacchelli and Williams, 2016).

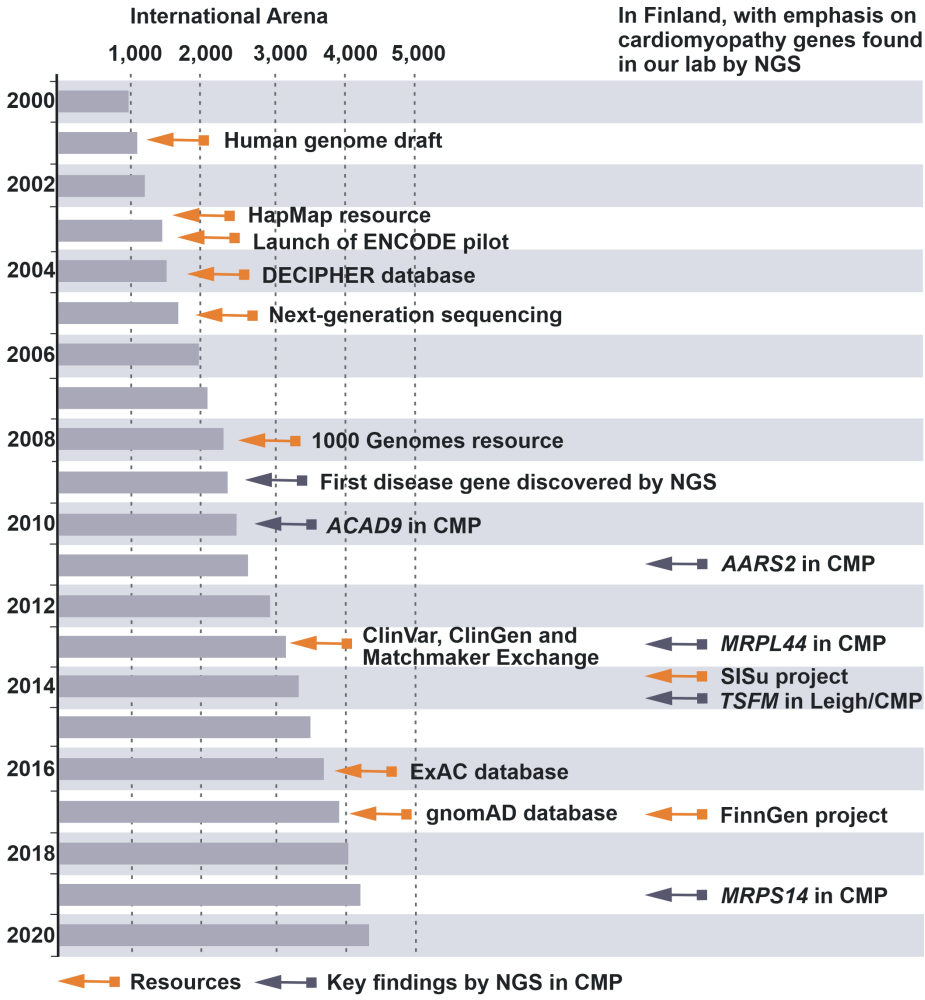


Figure 3. Gene discovery in rare disorders. Significant resources and key findings by NGS in cardiomyopathies. Adapted after Claussnitzer et al, 2019. The publication of the human genome draft in 2001 was a significant milestone for biology and medicine. The human genome, together with the introduction of the first commercially available NGS machines in 2005, sparkled the field of human genetics. The first disease gene uncovered by NGS was *MYH3* in distal arthrogryposis type 2A (Ng et al, 2009). Among the first disease genes that cause a mitochondrial disorder uncovered by NGS were *ACAD9* (Haack et al, 2010) and *AARS2* (Götz et al, 2011), both presenting with hypertrophic cardiomyopathy.

Classical genetic methods relied on delimitating subgroups of patients with a specific and consistent clinical presentation, as well as on large family pedigrees to assert the shared genetic background. The variability of

symptoms in many disorders was limiting for clinical recognition of etiological subgroups.

Moreover, the hypothesis-free approaches allow expansion of the phenotypic spectrum of already known disease genes to completely different disorders without shared symptoms. For example, *AARS2* is a nuclear gene that encodes a mitochondrial protein and was initially described in neonatal fatal hypertrophic cardiomyopathy (Götz et al, 2011). However, it has also been later reported in leukodystrophy with ovarian failure (Dallabona et al, 2014).

There are three main NGS-based applications in human genetics: gene panels, whole-exome sequencing (WES), and whole-genome sequencing (WGS) [reviewed in (Fernandez-Marmiesse et al, 2018; Maroilley and Tarailo-Graovac, 2019)].

Gene panel sequencing typically targets known disease genes, most often only the coding parts and intron-exon boundaries. Panels can range from one to thousands of genes. The most comprehensive panel targets all exons of known Mendelian genes (e.g., TruSight One Expanded Sequencing Panel from Illumina, covering ~6700 genes). By focused analysis, gene panels are timesaving, cost-effective, and preferred when the clinical phenotype is suggestive for a specific disorder with well-characterized genetic background. They are also useful to screen a replication cohort for candidate genes identified in a discovery cohort or when high coverage is essential, for example, to uncover somatic mutations. Another benefit of panels is that they can reliably exclude known disease genes due to adequate coverage and depth (Rehm, 2013). Their limitations include missing the genetic diagnosis as a consequence of the phenotypic and genetic heterogeneity of the disease, multiple overlapping diagnoses not all represented in the panel, novel disease genes, as well as types of genetic variation not detected by panels (such as noncoding variants and SVs).

Whole-exome sequencing (WES) enables a more comprehensive molecular diagnostic approach, targeting virtually all protein-coding genes in the genome. Given that the exome constitutes only ~1.1% of the genome and contains a large proportion of the variation involved in monogenic disorders, WES is a cost-effective method (<500€/sample) that facilitated unprecedented discoveries in the last decade. WES typically has been the next step when gene panel testing was negative. Further, it is preferred when the phenotype is unspecific for a panel or in disorders with high genetic heterogeneity, with hundreds of disease genes.

The first gene discovered by WES was published in 2009 (Ng et al, 2009). Since then, many genetic discoveries have involved novel disease genes, novel genotype-phenotype correlations, somatic mosaicism, or novel causes of phenotypic variability (double mutations, genetic modifiers, multiple diagnoses). WES' success rate varies widely with disease type, patient selection criteria, and type of test applied (singleton *versus* trio-WES, the latter including the patient and two unaffected family members, most

often parents). The overall diagnosis rate of trio-WES in rare disorders ranges between 30 and 50% (Chung et al, 2020; Fitzgerald et al, 2015; Sawyer et al, 2016), a significant improvement comparing to 10% diagnosis with traditional methods (Wright et al, 2018). Many of these patients, however, undergo different types of genetic investigations beforehand. When WES was used as a first-line genetic test in infants with a suspected monogenic disorder, the diagnosis yield amounted to 58%. In contrast, the standard care approach, which included panel sequencing when clinically indicated, led to a 14% diagnosis in the same individuals (Stark et al, 2016).

Nevertheless, there are still many limitations to the method, including an unequal coverage depth between different genomic regions, with GC-rich areas particularly challenging to cover. An additional limitation is the inability of WES to detect all genomic variation like functional noncoding variants (deep intronic variants affecting splicing, regulatory elements, functional noncoding RNAs), SVs, and repeats. Recent capture kits enable better target coverage and also capture some functional RNA genes. For example, the capture kits from different companies can isolate 40-60 Mb of DNA, representing 1.5-2% of the genome. These DNA regions consist of ~180,000 protein-coding exons plus a variable amount of intronic and regulatory sequences, as well as different RNA genes. WES has lower sensitivity and resolution to detect CNVs than other approaches, with only a minority of CNVs being identifiable from short read depth analysis.

Whole-genome sequencing (WGS) is the most comprehensive approach, with the potential to identify every genetic variation that plays a role in human biology. Over the next few years, once prices come down sufficiently, WGS will become the predominant technology for genetic analysis. The main advantage of WGS is the avoidance of exon capture because this process often leads to uncovered coding sequences in WES (Lelieveld et al, 2015). There are examples when WGS found pathogenic coding variants after a negative WES analysis on the same samples (Alfares et al, 2018). Most variants in noncoding sequences are uninterpretable in the present state of knowledge and are often not expected to be high-penetrance causes of disease. An additional benefit of WGS is the ability to detect and characterize SVs, which are relatively rare causes of Mendelian conditions but are crucial for tumor analysis.

Currently, there are three general WGS strategies (Lappalainen et al, 2019): (1) short-read WGS (Illumina mainly) with paired-end reads of ~150 bp and low error rates in the range of 0.1-0.5%; (2) long-read WGS using single-molecule technologies (Pacific Biosciences and Oxford Nanopore Technologies) with 10-100 kb reads and high error rates in the range of 10-15%; (3) linked-read WGS (10xGenomics) using Illumina bar-coded short-reads derived from a single long molecule (of ~50kb) and assembled *in silico*.

Given the accuracy, ease of use, and costs, most studies employ Illumina short-read technology. However, the short reads are challenging for accurate mapping of pseudogenes and tandem repeats, as well as for repeat expansion

and SV detection. All these types of variants are better detected by long-read sequencing, which has its own limitations related to high cost per sample, high error rates, lower throughput, and less bioinformatic tools developed.

Holistic approaches may integrate short and long-read technologies, together with multiple bioinformatic approaches to detect variants and predict their functionality. Although the integration of data sources and tools is superior, its use is limited currently by time, costs, and expertise.

The higher costs (~double WES), additional computational burden, less bioinformatic tools, and limited functional interpretation of noncoding variants have restricted the use of WGS.

As laboratories transition from gene panels to WES and WGS, genetic information on additional disease susceptibility loci can be detected in the patient data. Ethical considerations have been raised related to their disclosure (Christenhusz et al, 2013). These are incidental and secondary findings, defined as genetic test results that provide information about known or expected pathogenic variants in a gene unrelated to the test's primary purpose. The second condition might not be manifesting yet (Roche and Berg, 2015).

The American College of Medical Genetics and Genomics (ACMG) developed a set of recommendations on reporting the incidental and secondary findings, first in 2013 (Green et al, 2013), with an update in 2017 (Kalia et al, 2017). They defined secondary findings as those in genes intentionally analyzed from a defined list, while incidental findings are discovered non-intentionally. The aim is to identify and manage the risk for highly penetrant genetic disorders with established interventions that prevent or significantly reduce morbidity and mortality (Saelaert et al, 2018). The updated list includes 59 medically actionable genes associated with cancers, cardiomyopathies, arrhythmogenic disorders, and other life-threatening conditions (Kalia et al, 2017). The recommendation is to report only pathogenic or likely pathogenic variants in these genes (class 5 and 4 on ACMG criteria of variant pathogenicity), but not variants of unknown significance (VUS), whose involvement in disease at the time of analysis is unclear.

The secondary findings may help prevent diseases or guide their management, but they can also cause anxiety for individuals and families. For this reason, the patient or family can choose to 'opt-out' the analysis of the secondary findings (Saelaert et al, 2019). Soon, it is expected the incorporation of selected pharmacogenomics variants to the minimum secondary finding screening list because of known variants and haplotypes that dramatically modify the metabolism of specific drugs, with potentially severe side effects (Caudle et al, 2014; Hicks et al, 2018).

2.4. LEUKOENCEPHALOPATHIES

2.4.1. MYELIN IN THE CNS

The central nervous system (CNS) broadly consists of gray matter structures, defined by the densely packed cell bodies, dendrites, and synapses, connected by extensive white matter tracts, which are defined by the heavily myelinated axons. Besides the myelinated axons, the white matter contains numerous glial cells (oligodendrocytes, astrocytes, microglia) and nourishing blood vessels. On a dry weight basis, 40-50% of white matter is myelin. Comparing with other plasma membranes, myelin presents a unique composition with low water content (40%) and very high lipid content. Dry myelin is 70% lipids and 30% proteins, in contrast to most biological membranes with a higher ratio of proteins to lipids (Morell and Quarles, 1999a).

In the CNS, the myelin sheath is a modified oligodendrocyte plasma membrane that extends and surrounds in a spiral a segment of an axon (Stassart et al, 2018). One oligodendrocyte may form myelin segments on up to 40 axons. However, on the same axon, adjacent myelin segments belong to different oligodendrocytes. The nodes of Ranvier are small unmyelinated axonal regions between two segments of myelin and contain clusters of voltage-gated sodium channels that allow membrane depolarization (Salzer, 1997). The high resistance and low capacitance of myelin determine the electrical impulse to jump from node to node, which represents the basis of saltatory, fast nerve conduction, crucial for functional integration of the CNS (Cohen et al, 2020).

Myelination starts during the fifth month of fetal life in the spinal cord. It peaks during the first year of life and progresses slowly until around 20 years of age in some cortical fibers, especially in associative areas (Morell and Quarles, 1999b). Active myelination is a tightly regulated cellular and molecular process, but also more vulnerable to insults. In contrast, mature myelin is relatively stable and resistant to alterations (Williamson and Lyons, 2018).

2.4.2. CLINICAL ASPECTS OF LEUKOENCEPHALOPATHIES

Failure of white matter formation or maintenance in the CNS appears in heritable conditions termed leukoencephalopathies and leukodystrophies, which may or may not present peripheral nervous system involvement.

The incidence of pediatric heritable leukoencephalopathies and leukodystrophies is not clearly established due to diagnosis challenges. Their estimated combined incidence ranges from 1:50,000 (Heim et al, 1997) to 1:8,000 (Bonkowsky et al, 2010).

Most leukoencephalopathies and leukodystrophies present with motor symptoms. In an infant or young child, the manifestation can be delayed

acquisition or regression of motor skills. In older children, the first signs can be frequent falls or a clumsy gate, and in adolescents, deterioration of sports activities. In addition to neurological symptoms, several other red flags are suggestive of a white matter disorder: endocrine disturbances, dysmorphic facial features, ophthalmologic abnormalities, cortical visual impairment, hearing impairment, dental, skin, tendinous, or skeletal abnormalities, hepatosplenomegaly, ovarian dysgenesis, and gastrointestinal symptoms (Parikh et al, 2015).

Genetic leukoencephalopathies and leukodystrophies often have a pediatric onset and are caused by defects in any of the white matter structural components, including oligodendrocytes, astrocytes, microglia, axons, and blood vessels (van der Knaap and Bugiani, 2017). Leukodystrophies refer to those disorders in which glial cells are primarily affected. By definition, the pathology in leukoencephalopathies is primarily neuronal, vascular, or systemic, and the white matter changes occur secondarily or are not the predominant clinical picture (Vanderver et al, 2015).

White matter disorders may present different types of underlying pathology, most often in combination, and can be determined by both genetic and acquired factors (Stadelmann et al, 2019). Examples of encountered pathologies are: hypomyelination, when the myelin never reaches the normal developmental levels; dysmyelination, a disturbance that leads to irregular, patchy myelin; axonal degeneration is another pathology accompanied by myelin disturbance as a consequence of the interdependence between the axon and the myelin sheath; demyelination of previously normal myelin sheath; intramyelin vacuole formation, which may result in myelin splitting, with or without demyelination; and oligodendrocyte death, which affects both the formation and maintenance of myelin.

Brain magnetic resonance imaging (MRI) is the most important clinical investigation in patients suspected of leukoencephalopathy or leukodystrophy. There are several types of MRI techniques, but fundamental to diagnosis are T1-weighted and T2- weighted images (Parikh et al, 2015).

T1-weighted images are useful for revealing the anatomic structures in detail and assessing myelination in infants until 9-10 months old because small amounts of myelin are better visualized on these types of images. On T1-weighted images, normal white matter appears bright and cerebrospinal fluid (CSF) black. Pathological white matter gives a decreased signal and appears dark.

T2-weighted images are superior in evaluating the pathological changes of white matter and the stage of myelination in subjects older than 9-10 months. On T2-weighted images, normal white matter appears dark and CSF bright. Pathological white matter leads to abnormally high signal and appears bright.

Other MRI investigations may be employed during diagnostic workup (such as FLAIR and diffusion-weighting imaging). Magnetic resonance

spectroscopy, which reveals biochemical changes in the brain, is useful in mitochondrial disorders and Canavan disease to assess lactate and N-acetyl aspartate, respectively.

Following the MRI pattern analysis, the diagnostic approach aims to resolve the etiology as quickly as possible because some white matter disorders present established therapeutic interventions. For example, acquired white matter abnormalities that result from B12 (cobalamin) deficiency are reversible with vitamin supplementation. Some of the genetic leukoencephalopathies that present an established intervention include the amino acidemias (e.g., phenylketonuria), Wilson disease, and others. Biochemical testing is essential for a reliable diagnostic. A minimum set of tests includes plasma very-long-chain fatty acids, leukocyte lysosomal enzymes, blood lactate, pyruvate, and amino acids, urine sulfatides, and organic acids (Parikh et al, 2015). In parallel to biochemical tests, the diagnostic approach continues with genetic testing.

2.4.3. GENETICS OF LEUKOENCEPHALOPATHIES

The genetic defects leading to myelin disorders are heterogeneous (Parikh et al, 2015), with novel disease genes continuously being identified. Here we highlight several pathways commonly recognized in myelin disorders: a) maintenance of a balanced myelin lipid and protein composition b) synthesis and degradation of lipids c) secretory pathway and endoplasmic reticulum (ER) stress d) adherence of myelin sheath to axolemma and of multilamellar myelin membranes to each other e) axonal integrity and axon-myelin communication f) immunity activation g) astrocyte function and astrocyte-oligodendrocyte communication h) cytoplasmic and mitochondrial translation i) amino acid metabolism j) mitochondrial metabolism. Table 2 contains most of the genes described so far to cause white matter disorders.

The genetic heterogeneity in white matter disorders is best addressed in genetic testing by NGS technology. Gene panel approaches, encompassing a minimum set of genes defining both disease groups, can be used but often are not sufficient. The number of novel genes is continuously increasing, and the phenotypic spectrum of disorders with secondary white matter involvement continues to broaden. A study suggested that detailed MRI pattern analysis coupled with NGS, may lead to a higher molecular diagnosis rate (Vanderver et al, 2016). A recent randomized trial of immediate WGS in pediatric white matter disorders showed quicker and greater diagnosis efficacy than standard care (Vanderver et al, 2020).

Table 2. Genetics of leukodystrophies and leukoencephalopathies. Cellular pathways affected by disease genes. Adapted after Parikh et al, 2015.

Affected compartment or process	Genes	Disorders
Leukodystrophies		
Gene transcription	*POLR3A, POL3B* *SOX10*	Pol-III related disorders (4H leukodystrophy) SOX10-associated disorders
Oligodedrocyte plasma membrane	*FAM126A* *GJC2* *PLP1*	Hypomyelination and congenital cataract Hypomyelinating leukodystrophy 2 Pelizaeus-Merzbacher disease
Cytoskeleton	*GFAP* *TUBB4A* *LMNB1*	Alexander disease Hypomyelinating leukodystrophy Adult onset autosomal dominant leukodystrophy
Nucleic acids metabolism	*ADAR1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, TREX1*	Aicardi-Goutières Syndrome
Breakdown of N-acetyl-L-aspartate	*ASPA*	Canavan disease
Cytoplasmic translation	*EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5* *DARS*	eIF2B related disorders Hypomyelination with brainstem and spinal cord involvement and leg spasticity
Mitochondrial translation	*DARS2, EARS2* *AARS2*	Leukoencephalopathies with high lactate Leukoencephalopathy with ovarian failure
Glycogen biosynthesis	*GBE1*	Polyglucosan body disease
Lysosomal degradation	*FUCA1* *ARSA*	Fucosidosis Metachromatic leukodystrophy
RNA catabolism	*RNASET2*	RNAse T2 deficient leukoencephalopathy
Lipid metabolism	*CYP27A1* *GALC* *PSAP* *ALDH3A2*	Cerebrotendinosus Xanthomatosis Krabbe disease Krabbe disease/ Metachromatic leukodystrophy Sjögren Larsson syndrome
Peroxisome biogenesis	*PEX1, PEX2, PEX3, PEX6, PEX10, PEX12, PEX26*	Peroxisome biogenesis disorders

Table 2. Genetics of leukodystrophies and leukoencephalopathies. (Continuation)

Affected compartment or process	Genes	Disorders
Peroxisomal transport of fatty acids and beta oxidation	*ABCD1* *HSD17B4* *SCP2* *ACOX1*	X-linked adrenoleukodystrophy D-bifunctional protein deficiency Sterol carrier protein 2 deficiency Peroxisomal acyl-coA-oxidase deficiency
Sialic acid transport	*SLC17A5*	Sialic acid storage disease
Ion transport and gap junction	*MLC1* *CLCN2* *GJA1*	Megalencephalitic leukoencephalopathy with subcortical cysts Chloride ion channel-related leukoencephalopathy with intramyelinic edema Oculodentodigital dysplasia
Protein post translational modification	*SUMF1*	Multiple sulfatase deficiency
Cell adhesion	*HEPACAM*	Megalencephalitic leukoencephalopathy with subcortical cysts
Leukoencephalopathies		
Mitochondrial complex 1	*NDUFS1, NDUFS4, NDUFS7, NDUFS8, NDUFV1, NDUFS2, NDUFAF1*	OXPHOS disorders
Mitochondrial complex 2	*SDHA, SDHB, SDHAF1*	
Mitochondrial complex 3	*BCS1L*	
Mitochondrial complex 4	*SURF1, SCO1, SCO2, COX10, COX15, TACO1*	
Mitochondrial complex 5	*ATPAF2*	
Mitochondrial metabolism	*D2HGDH, L2HGDH, SLC25A1* *ETFDH*	L2-Hydroxyglutaric aciduria Glutaric acidemia
Mitochondrial Translation	*MRPS16, GFM1, TUFM*	Combined OXPHOS disorders
MtDNA metabolism	*POLG, POLG2, TYMP, SUCLA2, TWINKLE, RRM2B, SLC25A4, DGUOK*	Mitochondrial DNA depletion syndromes
Mitochondrial protein quality control	*SPG7* *HSP60*	Spastic paraplegia Hypomyelinating leukodystrophy
Glutamate/ aspartate	*SLC25A12*	Global cerebral hypomyelination

Table 2. Genetics of leukodystrophies and leukoencephalopathies. (Continuation)

Affected compartment or process	Genes	Disorders
Transcription-coupled excision repair	*ERCC6, ERCC8* *ERCC2, ERCC3, GTF2H5*	Cockayne syndrome Trichothiodystrophy with hypersensitivity to sunlight
Aminoacid synthesis and interconversion	*PHGDH* *PSAT*	Phosphoglycerate dehydrogenase deficiency Phosphoserine aminotransferase deficiency
Cholesterol transport	*NPC1* *NPC2*	Niemann Pick type C1 Niemann Pick type C2
Lipid metabolism	*FA2H* *CYP2U1* *CYP7B1*	Spastic paraplegia
Coenzyme Q biosynthesis	*COQ2, COQ9, ADCK3*	Coenzyme Q deficiency
Microtubule-associated	*SPAST, SPG20*	Spastic paraplegia
Immunity	*TYROBP* *STXBP2, UNC13D, STX11, PRF1, RAB27A, SH2D1A, BIRC4* *ACP33*	Dementia with bone cysts Hemophagocytic lymphohistiocytosis Spastic paraplegia
Vascular signaling	*NOTCH3* *HTRA1* *COL4A1*	Cerebral arteriopathy with subcortical infarcts and leukoencephalopathy Hereditary angiopathy with neuropathy, aneurysms, and muscle cramps
Tight junction signaling	*OCLN*	Band-like intracranial calcification with simplified gyration and polymicrogyria
ER chaperone	*BCAP31*	Deafness, dystonia, and cerebral hypomyelination
Cell cycle/Cell division	*MPLKIP* *ZFYVE26*	Trichothiodystrophy, nonphotosensitive Spastic paraplegia
Thyroid hormone transport	*SLC16A2*	Allan-Herndon-Dudley syndrome
Regulation of membrane trafficking	*OCRL*	Lowe syndrome
Lysosomal degradation	*GLB1* *HEXA* *PPT1* *MCLON1* *GLA*	GM1-gangliosidosis GM2-gangliosidosis (Tay-Sach disease) Infantile neuronal ceroid lipofuscinosis Mucolipidosis IV Fabry disease
Copper homeostasis	*ATP7A* *ATP7B*	Menkes disease Wilson disease

OXPHOS – Oxidative phosphorylation

2.5. CARDIOMYOPATHIES

2.5.1. DEFINITION AND TYPES OF CARDIOMYOPATHIES

Childhood cardiomyopathies (CMPs) are rare but severe heart muscle disorders characterized by pathological changes of the myocardium, unexplained by abnormal loading conditions, or congenital heart disease (Lee Teresa M. et al, 2017). They present an annual incidence of about 1:100,000 children worldwide, with the highest in infants under one year of age (Lipshultz et al, 2013). An epidemiology study from Finland estimated an annual incidence of 0.65:100,000 individuals up to 20 years of age, but this study included only idiopathic CMPs (Arola et al, 1997). The highest incidence emerged in an Australian study, with 1.24:100,000 children younger than ten years of age, a period when CMP is most typically diagnosed (Nugent et al, 2003). In a study from the United States, the annual incidence was 1.1:100,000 children younger than 18 years (Lipshultz et al, 2003).

CMPs present some of the least favorable prognoses in pediatric cardiology. In children, CMPs are more severe comparing with adults, with higher morbidity and mortality rates. Pediatric CMPs are also the most common cause of heart transplant in children older than one year of age (Lee Teresa M. et al, 2017). The cardiomyopathy in children is more often part of a metabolic, multi-organ disorder than in adults (El-Hattab and Scaglia, 2016).

In infants and young children, the presenting symptoms are usually poor feeding, growth failure, tachypnea (rapid breathing), hepatomegaly, extreme sleepiness, and excessive sweating. In older children, CMPs may manifest with abdominal symptoms caused by hepatomegaly and low cardiac output, tachypnea, dizziness and fainting, chest pain, and extreme fatigue. Some affected children may have no symptoms (Hollander et al, 2013; Kantor et al, 2013).

Although there are different classification systems for CMPs [reviewed by (McKenna William J. et al, 2017)], the morpho-functional status of the heart is a central consideration (Elliott et al, 2008; Rapezzi et al, 2013). Based on this criterion, the following main types of CMPs exist (Lipshultz et al, 2013):

Dilated cardiomyopathy (DCM) is characterized by left ventricular dilation, systolic dysfunction, and myocardial fibrosis. The heart muscle becomes thin, the left ventricle enlarged, and the heart pumps blood insufficiently. At the cardiomyocyte level, impaired generation or transmission of force ultimately leads to protein and organelle degradation and apoptosis. Approximately 50% of childhood CMPs are dilated, with 10-25% of them attributed to acute myocarditis (Lipshultz et al, 2003; Towbin et al, 2006). In DCM, the five-year transplant-free survival rate ranges from 60-75%, with most events occurring within the first two years of presentation. About 20-45% of patients regain normal cardiac function five years after diagnosis [reviewed by (Lee Teresa M. et al, 2017)].

Hypertrophic cardiomyopathy (HCM) manifests with left ventricle hypertrophy and impaired diastolic function. The hypertrophy can occur at the interventricular septum, at the inferior wall, at the apex, or throughout the entire left ventricle. When hypertrophy is severe, the muscle can block blood flow from the left ventricle to the aorta, condition termed outflow tract obstruction. In HCM, the cardiomyocytes appear disorganized and irregular. This appearance is called cardiomyocyte disarray and may cause changes in the electrical signals traveling through the ventricles leading to ventricular arrhythmias (abnormal heart rhythm). Fibrosis can also be observed at the tissue level. The hypertrophic form represents 35-50% of total childhood CMPs (Arola et al, 1997; Lipshultz et al, 2003; Nugent et al, 2003). The survival for pediatric HCM is 97% at five years and 94% at ten years after diagnosis. The mortality for HCM peaks before one year of age and again between eight and 17 [reviewed by (Lee Teresa M. et al, 2017)].

Restrictive cardiomyopathy (RCM) manifests with diastolic dysfunction. The heart muscle becomes rigid and unable to relax and fill with blood, causing abnormal heart rhythms and heart failure symptoms. RCMs are diagnosed in <5% of childhood cardiomyopathies and have notably poor outcomes. The five-year transplant-free survival rate is 22% for children with pure RCM and 68% for children with an RCM/HCM phenotype (Webber Steven A. et al, 2012).

Hallmarks of left ventricular noncompaction cardiomyopathy (LVNC) are trabeculations (networks of muscle fibers) of the left ventricular myocardium and left ventricular dysfunction. During development, the myocardium resembles a network of muscle fibers. As normal development progresses, the fibers become compacted, transforming the heart muscle from network-like to smooth and solid. In LVNC, these trabeculations typically occur at the apex but can be seen anywhere in the left ventricle. LVNC may co-occur with other types of heart muscle disorders (HCM, DCM, or RCM). It is estimated to account for ~5% of childhood CMPs, and the prognosis differs on the clinical subtype (isolated or in combination). LVNC/DCM has the least favorable prognosis, with a 43% five-year rate for combined death or transplantation (Jefferies et al, 2015).

Ventricular arrhythmias and impaired systolic function clinically characterize arrhythmogenic ventricular cardiomyopathy (AVC). At the tissue level, a fibro-fatty replacement of the myocardium takes place. Although AVC is very rare in children, it raises many concerns because 55% of AVC pediatric patients experience life-threatening arrhythmias, indication for a heart transplant, or (sudden) cardiac death (te Riele et al, 2015).

Histiocytoid cardiomyopathy (HICM) is another rare cardiac disorder of infancy and childhood. It predominantly affects girls younger than two years of age (female: male ratio 3:1). Clinically, it manifests as severe cardiac arrhythmias or dilated cardiomyopathy. Sudden death is a common presentation (Gilbert-Barness, 2004). At the tissue level, HICM is characterized by the development of abnormal Purkinje fibers, which

interfere with normal cardiac conduction, and by subendocardial, epicardial, or valvular yellow nodules. The nodules contain histiocyte-like cardiomyocytes, filled with an increased number of normal and abnormal mitochondria (Finsterer, 2008). HICM is considered mitochondrial cardiomyopathy and is typically fatal (Bates et al, 2012).

The morpho-functional evaluation of the heart is required for screening, diagnosis, risk stratification, prognosis, and treatment. For example, in DCM, the degree of ventricular dysfunction and dilation is a predictor of death or transplantation (Alvarez Jorge A. et al, 2011). In HCM, septal wall thickness is associated with sudden death (Bharucha Tara et al, 2015).

The morpho-functional assessment is typically accomplished by echocardiography, repeated over time to monitor the heart status. cMRI (cardiac magnetic resonance imaging) can be further used to refine the diagnosis, evaluate the cause (e.g., identify inflammatory processes), assist in disease management and risk stratification. cMRI is a very sensitive technique for assessing the morphological heart phenotype, functional performance, and for tissue characterization (such as fibrotic scar, interstitial fibrosis, edema) (Lipshultz Steven E. et al, 2019). There are several cMRI techniques. Among them, the most frequently used are T1- and T2-mapping and late gadolinium enhancement imaging.

Late gadolinium enhancement imaging evaluates the increased extracellular space in the myocardium, which can be a consequence of necrosis, fibrosis, edema, or infiltration. In healthy myocardium, the extracellular space is limited, but typically it expands in disease. This phenomenon can be visualized with intravenous administered gadolinium, which distributes rapidly to the extracellular space, but does not cross the healthy plasma membranes.

In addition to the heart morpho-functional assessment, the clinical evaluation of a child with CMP includes searching for an underlying metabolic, congenital, or acquired cause. About one-third of children diagnosed with CMP presents a metabolic or syndromic disorder (Albakri, 2019; Kindel et al, 2012).

First, the medical history is reviewed in detail, including growth and development, chronic medical issues, and a three-generation pedigree. This evaluation is followed by a comprehensive physical examination, paying attention to features that may suggest a genetic syndrome. Further, biochemical laboratory tests and genetic testing are essential for an accurate diagnosis. This multi-step evaluation is extensive and involves a multidisciplinary team, including geneticists and neurologists.

2.5.2. GENETICS OF CARDIOMYOPATHIES

Cardiomyopathies (CMPs) are mainly genetic in origin and typically monogenic, although an oligogenic model of inheritance has been sometimes suggested to explain diseases challenging to diagnose. In children, CMPs may manifest as isolated muscle disorders or as components of developmental or metabolic disorders, making their etiology more diverse comparing with adults.

More than 100 disease genes with all modes of inheritance have been found to underlie different types of CMPs. Disease-causing variants affecting sarcomeric and cytoskeletal proteins, desmosomes, calcium signaling, Ras/MAPK pathway, and mitochondrial proteins have been reported in pediatric CMPs (Table 3) (Lee Teresa M. et al, 2017). However, the overall spectrum of genetic causes in children presents still many unknowns.

Autosomal dominant inheritance is the most typical in familial isolated CMP types, many caused by mutations in sarcomeric proteins (Morita et al, 2008). Autosomal recessive inheritance often appears in metabolic disorders and others, such as disorders caused by cytoskeletal defects. X-linked and mtDNA mutations are also involved in CMPs. Noonan syndrome, caused typically by *de novo* pathogenic variants in the Ras/MAPK pathway, constitutes an important fraction of the infantile HCMs, with an ~22% mortality rate before one year of age (Wilkinson et al, 2012).

Neuromuscular, metabolic, mitochondrial, and syndromic causes contribute to ~1/3 of childhood CMPs (Cox et al, 2006). In a large cohort containing 196 HCM children, 40-50% presented a metabolic or syndromic disorder (Lipshultz et al, 2003). HCM caused by inborn errors of metabolism has an ~50% mortality rate within two years of diagnosis. In a large study of 1,731 children with DCM, 8% presented a neuromuscular disorder (Alvarez Jorge A. et al, 2011).

Genetic screening in childhood CMP has become part of the comprehensive approach to diagnosis and treatment. However, negative family history and a negative genetic test result cannot rule out a genetic cause. Small pedigrees, unknown causative genes, *de novo* genetic variants, multiple mutations contributing to disease make the genetic characterization difficult. Moreover, locus and allelic heterogeneity are known concepts in CMPs, also called variable expression. It refers to the instance when the same mutation may lead to different CMP types, different severities, and different ages of onset, sometimes even within the same family. Variable penetrance has also been described in the dominant CMP, and it refers to the individual risk for developing CMP when carrying a known pathogenic variant. Modifier genes, multiple pathogenic variants, and environmental factors can contribute to the clinical picture.

Table 3. Disease genes involved in cardiomyopathies.
Cellular processes affected by disease genes. Modified after Lee et al, 2017 and El-Hattab and Scaglia, 2016.

Affected Structure or Process	Genes	Additional Presentation	CMP Types
Muscle-specific disorders			
Sarcomeric thin filament	*ACTC1*	Atrial septal defect (AD)	Different types of CMPs (typically AD, rarely AR)
	TNNC1, TNNI3, TNNT2, TPM1		
Sarcomeric thick filament	*MYH7*	Myopathies (AD)	Different types of CMPs (typically AD, rarely AR)
	MYBPC3, MYL2, MYL3		
Z-disc	*ACTN2, MYOZ2, CSRP3*		Different types of CMPs (typically AD, rarely AR)
	LDB3	Myofibrillar myopathy (AD)	
	TCAP	Limb-girdle muscular dystrophy (AR)	
	TTN	Hereditary myopathy with early respiratory failure (AD)	
Desmosome	*DSC2*	Palmoplantar keratoderma and wooly hair (AR)	Mainly AVC, sometimes DCM (AD and AR)
	DSG2, PKP2		
	DSP	Carvajal syndrome (AR)	
	JUP	Naxos disease (AR)	
Intermediate filament	*DES*	Limb-girdle muscular dystrophy (AR) Myofibrillar myopathy (AD, AR)	DCM (AD, AR)
Other cytoskeleton components	*VCL*		HCM, DCM (AD)
Nuclear membrane	*EMD, SYNE1, SYNE2*	Emery-Dreifuss muscular dystrophy (X-linked for *EMD*, AD for the other genes)	DCM (X-linked for *EMD*, AD for the other genes)
Plasma membrane	*LMNA*	Congenital muscular dystrophy (AD) Emery-Dreifuss muscular dystrophy (AD, AR)	DCM, HCM, RCM (AD, AR)
	CAV3	Limb-girdle muscular dystrophy (AD, AR) Long QT (AD)	HCM (AD, AR)
	SGCD	Limb-girdle muscular dystrophy (AR)	DCM (AD, AR)

Table 3. Disease genes involved in cardiomyopathies.
(Continuation)

Affected Structure or Process	Genes	Additional Presentation	CMP Types
Protein quality control	*CRYAB*	Myofibrillar myopathy	DCM (AD, AR)
Calcium signaling	*JPH2*		HCM, DCM (AD)
Syndromic cardiomyopathies			
Ras/MAPK pathway	*BRAF, HRAS, KRAS, PTPN11, SOS1, SPRED1, RAF1*	Noonan/Costello/CFC syndrome	HCM, DCM (*de novo*, rarely AD)
Metabolic cardiomyopathies			
Carnitine transport across membranes	*CPT2*	Carnitine palmitoyltransferase II deficiency	DCM (AR)
	SLC22A5	Primary carnitine deficiency	HCM, DCM (AR)
Glycogen metabolism	*GAA*	Pompe disease (glycogen storage disease type II)	HCM (AR)
	LAMP2	Danon disease (glycogen storage disease type IIb)	HCM, DCM (X-linked)
	PRKAG2	Cardiac glycogen storage disease, Wolff-Parkinson-White syndrome	HCM (AD)
Mitochondrial beta-oxidation	*HADHA HADHB*	Systemic disorders	HCM, DCM (AR)
Fatty acid metabolism	*AGK*	Sengers syndrome	HCM (AR)
Mitochondrial cardiolipin synthesis	*TAZ*	Barth syndrome	DCM, LVNC (X-linked)
	DNAJC19	Ataxia	DCM, LVNC (AR)
Mitochondrial translation	*MT-TL1*	MELAS (mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes)	DCM (Mitochondrial)
	MT-TK	MERF (myoclonic epilepsy with ragged red fibers)	DCM, HICM (Mitochondrial)
	AARS2, MTO1, ELAC2 TSFM	Systemic disorders	HCM (AR)

Table 3. Disease genes involved in cardiomyopathies.
(Continuation)

Affected Structure or Process	Genes	Additional Presentation	CMP Types
Mitochondrial complex I	mtDNA: *MT-ND1, MT-ND2, MT-ND4, MT-ND5, MT-ND6* nDNA: *NDUFV1, NDUFV2, NDUFS1, NDUFS2, NDUFS3, NDUFS4, NDUFS6, NDUFS7, NDUFS8, NDUFA2, NDUFA11, NDUFAF3, NDUFA10, NDUFB3, NDUFB9, NDUFA1* Assembly: *NDUFAF2, NDUFAF4, NDUFAF5, NUBPL, NDUFAF1, FOXRED1, ACAD9*	Systemic disorders/ OXPHOS disorders	Different types of CMPs (Mitochondrial, AR)
Mitochondrial complex II	*SDHA, SDHD*		
Mitochondrial complex III	*MT-CYB*		
Mitochondrial complex IV	mtDNA: *MT-CO1, MT-CO2, MT-CO3* nDNA: *COX6B1* Assembly: *COX10, COX14, COX15, COX20, SCO1, SCO2, COA3, COA5*		
Mitochondrial complex V	*TMEM70*		
Coenzyme Q10 biosynthesis	*COQ2, COQ4, COQ6, COQ7, COQ9, ADCK3, PDSS1, PDSS2*	Coenzyme Q10 deficiency/Systemic disorders	HCM (AR)
MtDNA maintenance	*TYMP*	Mitochondrial neurogastrointestinal encephalopathy syndrome	HCM (AR)
Neuromuscular/ neurodegenerative disorders			
Cytoskeleton	*DMD*	Duchenne/Becker muscular dystrophy	DCM (X-linked)
Biosynthesis of heme and iron-sulfur clusters	*FXN*	Friedreich ataxia	HCM (AR)

mtDNA – mitochondrial DNA, nDNA – nuclear DNA
AD – autosomal dominant, AD – autosomal recessive, CMP – Cardiomyopathy, HCM – Hypertrophic cardiomyopathy, DCM – Dilated cardiomyopathy, LVNC – Left ventricular noncompaction cardiomyopathy, ARV – Arrhythmogenic ventricular cardiomyopathy, HICM – Histiocytoid cardiomyopathy, OXPHOS – Oxidative phosphorylation

Finding the genetic cause of disease enables the cascade screening of family members to identify mutation-positive individuals that require cardiac surveillance. However, medical therapy and activity restriction are not indicated for children positive for a genetic mutation but without CMP morpho-functional features. An exception to this rule is endurance and high-intensity athletics in desmosomal mutation carriers, as they predispose to the progression of AVC (James et al, 2013; Sawant et al, 2016). The high genetic heterogeneity of CMPs has been an obstacle for the comprehensive characterization of disease genes. NGS, allowing coverage of broad genomic regions, opened the possibility to overcome this barrier and has been invaluable for discovering new causative genes and variants in recent years (Andreasen et al, 2013; Rosenbaum et al, 2020).

2.5.3. STEM CELLS IN MODELING CARDIAC DISORDERS

Two intrinsic properties define stem cells: long-term self-renewal and the ability to differentiate into other cell types. Long-term self-renewal refers to their capacity to replicate themselves and produce other similarly potent cells over extended periods or through multiple passages while maintaining an undifferentiated state. This property is essential for stem cell pool expansion and maintenance. The differentiation refers to their capacity to produce different cell types, like cardiomyocytes, neurons, or endothelial cells (Chagastelles and Nardi, 2011).

Based on the developmental stage, we can distinguish two main types: embryonic stem cells and adult or somatic stem cells. Embryonic stem cells have a role in development [reviewd by (Martello and Smith, 2014)]. In contrast, adult stem cells maintain the formed tissues' proper function by secreting molecules that support tissue function and by replacing damaged cells [reviewd by (Simons and Clevers, 2011)].

Embryonic and adult stem cells also differ for their capacity to give rise to different lineages, known as potency. From this perspective, embryonic stem cells are totipotent, pluripotent, or multipotent, while adult stem cells are multipotent, oligopotent, or unipotent (Burgess, 2016).

Totipotent cells (sometimes called omnipotent) possess the ability to differentiate into both embryonic and extra-embryonic tissues. In human embryos, all cells from zygote to morula (16-cell stage) are totipotent (Balakier and Pedersen, 1982; Tarkowski, 1959). Pluripotent cells have the capacity to differentiate into all three primary germ layers (endoderm, mesoderm, ectoderm) but cannot contribute to extra-embryonic tissues (e.g., placenta) (Gardner, 1968). In development, after the morula stage, the embryonic stem cells lose totipotency, and the inner mass cells of the blastocyst are pluripotent. These cells propagate indefinitely.

Further down the potency line, multipotent cells are more limited in their differentiation capabilities and ability to divide. They produce multiple, but not all, cell lineages in developing embryos or adults. Few examples

include hematopoietic stem cells, mesenchymal stem cells, and adult neural stem cells. Mesenchymal cardiac-specific stem cells give rise to cardiomyocytes, smooth muscle, and endothelial cells [reviewed by (Díez Villanueva et al, 2012)].

Oligopotent stem cells are found in adult organisms and can differentiate into a small number of cell types, typically two. An example is represented by the oligopotent stem cells on the eye surface that generate new corneal and conjunctival cells (Majo et al, 2008).

Unipotent stem cells, also known as precursor cells, typically arise from multipotent stem cells and differentiate into only one cell type. Stem cells in the epithelium that regenerate skin can exemplify them [reviewed by (Blanpain et al, 2007)].

The accumulation of damage in stem cells causes their functional decline and induces age-associated degeneration of tissues (Sharpless and DePinho, 2007). The potential of stem cell-based therapies for regenerative medicine fuelled the stem cell field (De Luca et al, 2019; Trounson and DeWitt, 2016). However, the use of human embryonic stem cells (ESCs) in research has been challenging due to ethical and legislative considerations regarding their human embryonic origin. A brilliant solution to this issue was the development of reprogramming strategies to convert somatic cells, like fibroblasts, to induced pluripotent stem cells (iPSCs).

In 2006, Yamanaka and Takahashi used the overexpression of transcription factors specific to ESCs to reprogramme mouse fibroblasts to a pluripotent state. They determined that a combination of OCT3/4, SOX2, KLF4, and c-MYC (OSKM), was enough for robust reprogramming of mouse fibroblasts to pluripotency (Takahashi and Yamanaka, 2006). The Japanese researchers and their colleagues soon reported the generation of human iPSCs using the same strategy, highlighting that OSKM is a *bona fide* set of pluripotency transcription factors (Takahashi et al, 2007). A slightly different set of factors containing OCT3/4, SOX2, NANOG, and LIN28 was used by Thomson and colleagues to generate iPSCs, with similar efficiencies (Yu et al, 2007). Nowadays, both sets and their combinations are popular for reprogramming somatic cells to pluripotency. iPSCs can be defined as a type of pluripotent stem cells generated directly from somatic cells by overexpression of a set of early pluripotency transcription factors.

Human iPSCs-derived cardiomyocytes (iPSC-CMs) have been used to model inherited cardiac disorders for investigating pathogenic mechanisms (Chen et al, 2016) and to test novel therapeutic agents (Paik et al, 2020). Moreover, iPSC-CMs offer a personalized platform to test the cardiotoxicity of drugs (e.g., chemotherapy drugs doxorubicin and tyrosine kinase inhibitors) (Gintant Gary et al, 2019; Kussauer et al, 2019; Pang, 2020). Success in these areas, however, requires iPSC-CMs that recapitulate the physiology and pathology of adult cells. The iPSC-CMs generated with current methods are typically embryonic-like, displaying structural and functional immaturity.

Adult cardiomyocytes and iPSC-CMs differ in several important ways (Musunuru Kiran et al, 2018). One of them is gene expression, adult CMs displaying high levels of sarcomeric, calcium handling, and OXPHOS genes. In contrast, iPSC-CMs display high levels of MYH6 as opposed to MYH7 and other embryonic isoforms (Xu et al, 2009).

Structurally, adult CMs exhibit a high length-to-width ratio and are more often bi-nucleated. Moreover, they present T-tubules at the sarcomere's Z-lines, ensuring efficient calcium handling and propagation of contraction. The sarcomeres are highly organized, with a length of 2.2 μm, and display Z-discs, I- and A-bands, H-zones, and M-lines. Adult cardiomyocytes are rich in mitochondria, which account for ~25% of the cell volume. In contrast, immature iPSC-CMs tend to be round, often mono-nucleated, with disorganized and shorter sarcomeres of 1.6 μm length. Typically, they do not form T-tubules and only have Z-discs and I-bands. Their mitochondrial volume is also lower [reviewed in (Yang et al, 2014)].

The metabolism of adult cardiomyocytes relies primarily on fatty acid oxidation for efficient energy production, with high levels of OXPHOS. Although iPSC-CMs possess some capacity to metabolize fatty acids, they rely mostly on glucose and lactate (Correia et al, 2017).

Contractile function in adult cardiomyocytes also varies from iPSC-CMs, with adult cardiomyocytes generating greater force and conduction velocities when stimulated [reviewed by (Yang et al, 2014)].

The immaturity of iPSC-CMs may obscure the patient phenotype in disease modeling and may alter drug responses in screenings and toxicity tests. Therefore, efforts endeavored to develop methods for rapid and consistent maturation of iPSC-CMs, with variable improvements in maturity. These methods include (1) biochemical stimulations (fatty acids and galactose medium, T3 hormone treatment, microRNA overexpression, inhibition of HIF1 alpha and of LDHA, treatment with soluble factors from mesenchymal stem cells), (2) physical stimulations (electrical stimulation, substrate stiffness modulation), (3) co-culturing (with fibroblasts or endothelial cells), (4) as well as more advanced tissue-engineering approaches (organoids, heart-on-chip, engineered cardiac tissues) [reviewed by (Guo Yuxuan and Pu William T., 2020; Nugraha et al, 2019; Tu Chengyi et al, 2018)]. Among the most promising results were obtained by Ronaldson-Bouchard et al (2018, 2019) with their approach of mechanical and high-intensity electrical stimulation of human cardiac tissues grown from early-stage iPSC-CMs. The engineered tissues displayed adult-like expression profiles and remarkably organized ultrastructure (Ronaldson-Bouchard et al, 2018, 2019). These outcomes are promising for late-onset disease modeling and small molecule screening in drug discovery and toxicity assays, as maturity is necessary for achieving physiological responses.

Despite the limitations mentioned above, the use of iPSC-CMs enabled successful disease modeling of several cardiomyopathy-associated genetic defects. A few examples are modeling of Barth syndrome caused by *TAZ*

mutations (Wang et al, 2014), modeling of the Finnish founder mutations in *MYBPC* and *TPM1* leading to adult CMPs (Ojala et al, 2015a), and many others [reviewed in (Brodehl et al, 2019; Buikema and Wu, 2017)]. Interestingly, a recent study also revealed a non-cardiomyocyte contribution to heart disease by engineering 3D cardiac microtissues containing healthy iPSC-CMs and AVC patient iPSC-cardiac fibroblasts (Giacomelli et al, 2020). This approach highlights the role of microenvironmental cues in disease development and the importance of adopting more advanced modeling systems such as engineered microtissues and heart-on-chip technologies.

3. AIMS

The general aim of this work was to investigate the molecular background of rare childhood-onset disorders, with patients belonging to two larger projects in our lab: Exomit study of mitochondrial-suspected disorders and Cardiomyopathy project with severe early-onset patients from the single center in Finland performing cardiac transplants. Our specific aims were:

1. To clarify the underlying genetic cause of an uncharacterized, progressive leukeoncephalopathy syndrome with sudden deterioration of symptoms following a febrile illness and an ovarian malignant tumor (Study I).

2. To define the genetic spectrum of severe childhood cardiomyopathies in Finland, using a unique cohort representative for the whole country with respect to patients considered for cardiac transplant or who required inotropic support (Study II).

3. To characterize a novel disease gene for cardiomyopathy (Study III).

4. PATIENTS AND METHODS

4.1. ETHICAL CONSIDERATIONS

An important ethical consideration in studies of childhood disorders is that most patients are too young to allow their own informed consent to participate in the study. All the samples were taken for diagnostic purposes with informed consent from parents and from patients when they were older than 10 years of age.

The Coordinating Ethical Board of Helsinki University Hospital approved the study plan for the Exomit project, with the ethical permit number 325/13/03/00/2015. The Childhood Cardiomyopathy project's study plan was approved by the Child and Adolescent Psychiatry Ethical Board of Helsinki University Hospital and received the ethical permit number 291/13/03/03/2008.

Patients and family members donated blood for DNA extraction. Besides, for functional studies, cells and tissue samples from selected patients and a few controls were also obtained from Helsinki University Hospital.

4.2. STUDY SUBJECTS

4.2.1. INFANTILE HYPOMYELINATING LEUKOENCEPHALOPATHY

The patient presented a progressive leukeoncephalopathy syndrome of unknown genetic origin. She was recruited to our lab's Exomit project on the genetics of mitochondrial-suspected disorders because of decreased OXPHOS CI + CIII activity in muscle biopsy (Table 4).

Table 4. Mitochondrial respiratory chain activities in skeletal muscle relative to citrate synthase activity (nmol O/min/mg mitochondrial protein).

Measurements	Patient	Controls
C I+III	50.1	295±91
C I+III/CS	**0.14**	**0.34**
C II+III	218.5	276±97
C II+III/CS	**0.60**	**0.32**

Table 4. Mitochondrial respiratory chain activities in skeletal muscle (Continuation)

Measurements	Patient	Controls
C II	77.8	181±54
C II/CS	**0.21**	**0.21**
C IV	1119.4	2620±858
C IV/CS	**3.09**	**3.05**
Citrate synthase	362.0	858 ± 268

O = atomic oxygen, C I+III (NADH:cyt C reductase), C II+III (succinate:cyt C reductase), C II (SDH), C IV (COX), CS= citrate synthase, NADH = reduced nicotinamide adenine dinucleotide, Cyt c = Cytochrome c, SDH = succinate dehydrogenase, COX o cytochrome c oxidase.

At the beginning of this project, only samples from the patient were available, but not from any other family member, and the latter were not obtained either retrospectively.

4.2.2. CHILDHOOD-ONSET CARDIOMYOPATHY (KIDCMP COHORT)

The KidCMP was collected by the effort of pediatric cardiologists and consisted of 66 unrelated childhood CMP patients who visited between 1993 and 2014 the Pediatric Cardiology Department of the Helsinki University Central Hospital. The cohort included patients younger than 15 years at presentation, the majority of Finnish ancestry.

In Finland, all the invasive and pretransplantation investigations are performed in Helsinki for the whole country. The patient inclusion criteria were childhood-onset cardiomyopathy with or without other organ systems involvement, who came for inotropic support, invasive hemodynamic examinations, or pretransplantation evaluation. Exclusion criteria were congenital heart disease, arrhythmia-induced cardiac dysfunction, Kawasaki disease, ALCAPA (anomalous origin of left coronary arising from pulmonary artery), secondary cardiac dysfunction resulting from abnormalities in other organs, and maternal diabetes for infants who presented with septal hypertrophy before four weeks of age. According to the Declaration of Helsinki, informed consent was received from patients when they were older than 10 years of age and from parents. Cardiac examinations, including echocardiography, were performed for all first-degree relatives and other family members.

4.3. NEXT-GENERATION SEQUENCING

4.3.1. TARGETED PANELS

The first goal of any NGS study is to obtain a high-quality map of genetic variation for the investigated samples. This represents the foundation for all downstream analyses because variants that are not accurately genotyped cannot be directly assessed. Panels offer the advantage of obtaining a very accurate map across known disease genes, making them an attractive first step in the genetic investigation of many rare disorders.

We utilized two targeted sequencing panels. CMP-Custom HaloPlex Panel (Agilent Technologies), of 117 cardiac genes, was applied to 32 patients. Pan Cardiomyopathy v1.0 Panel (Blueprint Genetics), of 101 cardiomyopathy-related genes, was applied to 29 patients. The data obtained with the CMP-Custom Haloplex panel were processed using the bioinformatic pipeline of the Finnish Institute for Molecular Medicine FIMM (Sulonen et al, 2011). The data obtained with the Pan Cardiomyopathy panel were processed at Blueprint Genetics with their customized bioinformatic pipeline.

4.3.2. WHOLE-EXOME SEQUENCING

We sequenced the whole exome of >40 patients, using two service providers: the Finnish Institute for Molecular Medicine FIMM (with NimbleGen SeqCap EZ Exome v2.0 kit) and the BGI company (with Agilent SureSelect Human All Exon v6 kit). Ten patients have been screened with multiple NGS methods. Typically, we analyzed only the index patients, except for 3 families, where additional family members have been included. The whole-exome sequencing data were processed using FIMM's bioinformatic pipeline (Sulonen et al, 2011), with a few recent updates.

4.4. IDENTIFICATION OF PUTATIVE DISEASE-CAUSING VARIANTS

4.4.1. *IN SILICO* APPROACHES: VARIANT FILTERING, POPULATION DATABASES, PATHOGENICITY PREDICTIONS, CONSERVATION IN SPECIES, MOLECULAR MODELING

The filtering steps and prioritization of SNVs and indels to identify the likely pathogenic variants included (Figure 4): 1) selection of rare variants with a MAF<0.01 in public databases: ExAC http://exac.broadinstitute.org/ gnomAD http://gnomad.broadinstitute.org/, and the Finnish SISu http://www.sisuproject.fi/; 2) for a dominant inheritance, the occurrence

was not allowed in public databases, while for recessive inheritance only heterozygotes were permitted; 3) prediction of variants' deleteriousness based on CADD C-score (Kircher et al, 2014), SIFT (Sim et al, 2012), and PolyPhen-2 (Adzhubei et al, 2010); 4) further prioritization considered the protein function related to patient's phenotype and the amino acid conservation in species.

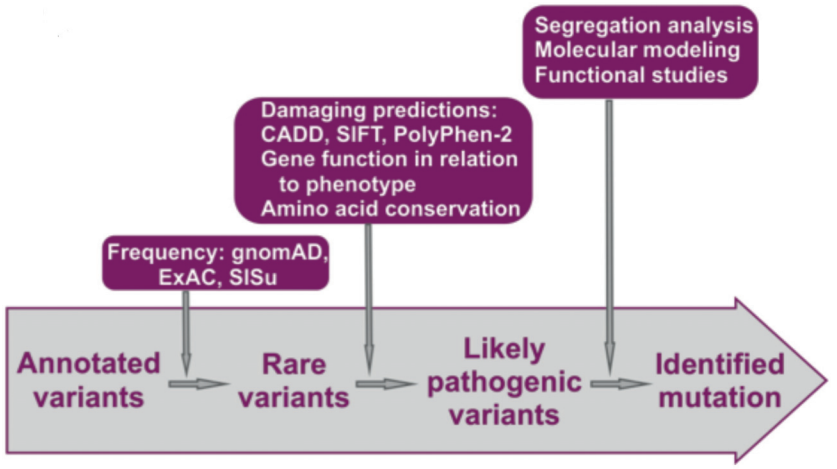


Figure 4. Variant prioritization strategy. Steps in NGS data filtering for SNVs (single nucleotide variants) and indels (insertion-deletions) to identify the likely causative variants.

For interpretation of sequence variants, we followed the guidelines of the American College of Medical Genetics and Genomics (ACMG) (Richards et al, 2015). Variants were considered disease-causing if they met criteria for 'pathogenic' or 'likely pathogenic' in ACMG classification. In short, the causative variants were: 1) known to cause CMP, 2) new variants in genes that were relevant for cardiac function, segregated with disease in families, were absent in control databases if dominant, or present only as heterozygotes if recessive, and altered an evolutionarily conserved amino acid in the protein, 3) *de novo* variants in known CMP genes showing multiple bioinformatic evidence of deleteriousness.

We modeled the consequences of selected variants using the software Discovery Studio 4.5 (Biovia) and as templates either the protein crystal structures from PDB (Protein Data Bank, https://www.rcsb.org/) or homology models deposited at SwissProt (https://www.expasy.org/). The program is set to generate 5 mutant models, which typically are very similar, and the model with the lowest energy is chosen as representative.

4.4.2. FAMILY SCREENING: SEGREGATION, FINGERPRINT ANALYSIS

We verified the putative pathogenic variants by Sanger sequencing and their co-segregation with the disease in families. *De novo* variants, determined by being absent in the parents' samples, were further confirmed by DNA fingerprinting of patients and parents with seven microsatellite markers. Both Sanger dideoxy reactions and the fragment amplification in microsatellite analysis were prepared in our lab and submitted to the FIMM Sequencing Lab for analysis through capillary electrophoresis.

4.5. FUNCTIONAL VALIDATION OF DISEASE-CAUSING VARIANTS

4.5.1. PROTEIN ANALYSES: SDS-PAGE, NATIVE-PAGE, WESTERN BLOTTING, UNTARGETED PROTEOMICS

In our projects, we used proteins extracted from cultured fibroblasts and skeletal muscle biopsies.

To extract proteins for SDS-PAGE, the cultured fibroblasts were collected by trypsinization, total cellular proteins extracted in RIPA buffer (Radioimmunoprecipitation assay buffer: 50 mM Tris-Cl pH8, 150 mM NaCl, 1% TritonX 100, 0.5% sodium deoxycholate, 0.1% SDS) containing protease and phosphatase inhibitors, and quantified by Bradford method. From tissues stored at -80°C, a small piece was cut in a jar kept on dry ice and quickly homogenized in cold PBS containing protease and phosphatase inhibitors. 5x RIPA buffer was quickly added to a final concentration of 1x. The samples were further extracted by incubating for 20 minutes on ice, followed by 20 minutes centrifugation at 20.000xg to pellet the debris. Protein concentration was determined with BCA Pierce Kit (Thermo Fisher).

The proteins were separated on 4-20% Mini-Protean TGX Stain-Free™ gels (Bio-Rad) and transferred to 0.2 μm PVDF (polyvinylidene fluoride) membranes for Western blotting using the Trans-Blot Turbo™ Transfer System (Bio-Rad). The blots were blocked in 5% milk in TBS-Tween 20 (0.1%), incubated overnight with a primary antibody, and followed by one-hour incubation the next day with an HRP-conjugated secondary antibody. The chemiluminescent reaction was used to develop the signal that was finally captured with Bio-Rad ChemiDoc™ XRS+ imager.

Proteomics analysis used the same tissue samples extracted in RIPA buffer as described above. For accurate quantification, the lysates were measured three independent times with BCA Pierce Kit (each time with two technical repeats). The protein concentration was determined as the average value of all measurements. The extracted proteins were further purified to remove detergents, and a small amount of each sample was used for

untargeted mass spectrometry. The data were analyzed using GraphPad Prism 8.

Native-PAGE enabled the isolation of intact mitochondrial complexes. We followed a protocol described approximately 30 years ago (Schägger and von Jagow, 1991), with a few adjustments. The purpose was to evaluate the integrity of mitochondrial complexes in primary fibroblasts from our patients with *PPA2* genetic defect compared to control fibroblasts. The fibroblasts were grown in DMEM medium without glucose (Thermo Fisher) supplemented with 10% FBS, 1x GlutaMAX (Gibco), antibiotics 1x Pen Strep (Gibco), 10 mM D-galactose (Sigma-Aldrich), and 50 μg/ml uridine. When the plate was 90% confluent, the cells were pelleted by centrifugation and resuspended in cold PBS containing protease inhibitors. Further, all the steps were carried out on ice or at +4°C. First, the cells received a digitonin solution, which disrupted the plasma membrane and the outer mitochondrial membrane. Cell lysates were centrifuged at 10,000xg for 10 minutes, and the supernatant was removed. The pellets, containing mitochondria, received a further treatment with lauryl maltoside solution, which released mitochondrial complexes from the inner membrane. After centrifugation at 20,000xg for 20 minutes, the collected supernatant contained the desired protein complexes. After quantification, 10 μg of proteins were separated on a native-PAGE gel (NuPAGE Bis-Tris, Invitrogen). The proteins were further transferred from the gel to a PVDF membrane, which was used in Western blotting, as described above.

4.5.2. EXPRESSION ANALYSIS

Total RNA was isolated from fibroblasts, heart biopsies, and iPSCs using the RNeasy Mini kit (Qiagen). The cDNA was generated from 0.3-1 μg RNA using Maxima First Strand cDNA Synthesis Kit with dsDNase (Thermo Scientific). All qPCR reactions were performed in triplicate, using either IQ™ SYBR Green Supermix (Bio-Rad) or SensiFAST™ SYBR (Bioline).

4.5.3. VISUALIZATION OF CELLULAR STRUCTURES BY IMMUNOFLUORESCENCE

Immunofluorescence is a very versatile technique, and we employed it with different types of cells: fibroblasts, U2OS cells, iPSCs, iPSC-CMs.

The fibroblasts and U2OS cells were fixed in 4% PFA in PBS for 15 minutes, followed by three washes with PBS. iPSCs and cardiomyocytes were prefixed for 10 minutes by adding an equal volume of 4% PFA to their medium, followed by 10 minutes fixation with 4% PFA, and three washes with PBS. A subset of cardiomyocytes was fixed after preincubation with a 'relaxing buffer' to distend the sarcomeres. The relaxing buffer consisted of 150 mM KCl, 5 mM MgCl2, 10 mM MOPS, 1 mM EGTA, adjusted to pH 7.4, and 4mM ATP freshly added.

The fixed cells were kept at +4°C in PBS until were stained (or in PBS with 0.02% sodium azide if they were stored longer). For staining, the cells were first permeabilized for 15 minutes using 0.2% TritonX-100 in PBS, washed with PBS, followed by a blocking step of 1 hour with 10% horse serum, 1% BSA, and 0.1% TritonX-100 in PBS. The cells were washed once with 1% horse serum, 1% BSA, and 0.1% TritonX-100 in PBS and incubated with primary antibodies overnight at +4°C. The primary antibodies were diluted in 1% horse serum, 1% BSA, and 0.1% TritonX-100 in PBS. The next day, the cells were washed three times with 1% BSA in PBS and incubated with secondary antibody, or with phalloidin (to stain F-actin) plus secondary antibody diluted in 1% BSA in PBS. The final washes were three times with PBS and one time with PB. We mounted the coverslips on glass using Vectashield Mounting Medium with DAPI (Vector Laboratories).

4.5.4. DISEASE MODELING USING INDUCED PLURIPOTENT STEM CELL (IPSC)-DERIVED CARDIOMYOCYTES

4.5.4.1. Generation and characterization of iPSCs

The reprogramming process unfolds in two broad phases (Polo et al, 2012; Theunissen and Jaenisch, 2014). (1) The early reprogramming phase, characterized by stochastic binding of OSKM (Oct3/4, Sox2, Klf4, and c-Myc) to their target promoters in the genome. OSKM components possess pioneer transcription factor activity, binding to inaccessible chromatin areas where they remodel the epigenetic markers, thus modifying gene expression. The consequence of epigenetic remodeling is the induction of early pluripotency genes and the repression of somatic cell transcription programs. (2) The reprogramming late phase is more deterministic, distinguished by a gradual activation of the late pluripotency genes and the establishment of full pluripotency by endogenous gene regulatory networks.

We used low-passaged primary fibroblasts to obtain iPSCs from patient and control lines. Neon electroporation system was used to transduce 700,000 cells with three plasmids containing Yamanaka factors OSKM: pCXLE-hOCT3/4-shp53-F, pCXLE-hSK, and pCXLE-hMLN (1 μg each). We used 1,650V, 10 ms, and 3 pulses for electroporation, after which the cells were transferred into 6-cm freshly gelatinized dishes (0.15% gelatin) containing warm fibroblast medium. We considered this day 1. The medium was changed to cells on days 2 and 4. On day 5, we prepared feeder plates for passaging the cells by seeding ~20,000 mitotically inactivated MEFs (mouse embryonic fibroblasts) per freshly gelatinized 6-cm dishes and cultured them overnight in fibroblast medium. On day 6, the electroporated cells were collected with TrypLE™ Express (Gibco), pelleted by centrifugation, resuspended in fibroblast medium, and seeded at 150,000-300,000 cells/6-cm feeder plates. The feeder layers of mitotically inactivated MEFs secrete

growth factors and provide a good surface for culture. More recent techniques grow the electroporated cells directly on matrigel to avoid feeder contamination and mechanical cutting. On day 7, we changed to hES medium, which was refreshed every second day afterward. hES medium promotes pluripotency and contains DMEM/F12 with Glutamine (Life Technologies), 20% KnockOut Serum Replacement (Gibco), 0.0915 mM 2-βMercaptoethanol (Sigma-Aldrich), 1x Non-essential amino-acids (Gibco), 6ng/ml recombinant human FGF basic (Peprotech), and 0.25 mM Sodium butyrate (MilliporeSigma™). Around day 20, iPSC foci may become ready for cutting. We cut the colonies mechanically using a scalpel and plated them on freshly gelatinized 4-well dishes with feeder cells and hES medium. A few initial passages were done by mechanical cutting with a scalpel, on feeder cells with hES medium. Afterward, the cells were transferred on matrigel-coated plates and cultured further in E8 medium (Essential 8™ Basal Medium plus 1x Essential 8™ Supplement, Gibco) until they displayed the characteristic morphology of tightly packed colonies with defined borders. iPS cells are small (~43.5 μm in size), with a small cytoplasm around the nucleus (~0.87 nucleus to cytoplasm ratio). The colonies grow in a single layer and present a cell density of ~6,000/mm^2 (Wakao et al, 2012). The E8 medium contains only 8 elements, including the growth factors TGFB1 and FGF2. The matrigel is a substitute for extracellular matrix, providing cell attachment and promoting survival. iPSCs are particularly sensitive to single-cell dissociation, quickly triggering apoptosis when cells lose any contact, a process termed anoikis (Gilmore, 2005). The addition of Rock inhibitor (Y27632) promotes quick attachment to a coated surface and substantially improves survival.

For passage, the cells were washed once with PBS, after which they were incubated with 0.5 mM EDTA for about 5 minutes at room temperature. The EDTA was aspirated gently, and fresh E8 medium was added to the dish to resuspend the cells, which were distributed to new plates. To freeze the cells, we followed the same procedure as above but resuspended the cells in ice-cold FBS with 10% dimethyl sulfoxide and added them to a cryovial cooled on ice. The patient and control iPSCs were expanded over passage 30, with cells frozen as stocks at different passages.

Several markers of pluripotency were tested at the protein and mRNA levels. At the protein level, we tested the surface markers SSEA4 and TRA-1-60, and the nuclear transcription factor NANOG.

4.5.4.2. Cardiomyocyte differentiation, maturation, and characterization

For cardiomyocyte differentiation, we followed the small molecule protocol (Burridge et al, 2015), with a few modifications. The iPSCs were grown on matrigel (Corning) in E8 medium for 3-4 days until they became 65-85%

confluent. At this stage, they were dissociated to single cells by incubating them for 7 minutes with 0.5 mM EDTA in PBS. Further, we seeded 120,000 cells/well of a 6-well dish coated with growth factor reduced matrigel (Corning) in the presence of 10 μM Rock inhibitor (Y27632). The E8 medium was changed regularly until cells reached confluence, and the differentiation process was started. We used for induction CDM3 glucose medium with 3-4.5 μM CHIR99021 (Sigma-Aldrich), a GSK3B (Glycogen synthase kinase 3 beta) inhibitor that mediates epithelial to mesenchymal transition. CDM3 glucose medium consisted of RPMI 1640 with glucose and L-glutamine (Gibco) supplemented with 75 mg/ml human recombinant albumin (Sigma-Aldrich) and 64 mg/ml L-ascorbic acid 2-phosphate (Sigma-Aldrich). After 48 h, the cells were washed with PBS, and the medium was changed with CDM3 glucose containing 2 μM Wnt-C59, an inhibitor of the WNT pathway. After another 48 h, the cells were washed, and the medium was changed with normal CDM3 glucose and refreshed every other day from this point. The cells typically started to beat between days 7 and 10.

We replated the cells on day 20 at a higher density of 1-1.5 million cells/well of a 24-well plate coated with matrigel. We used the Multi Tissue Dissociation Kit 3 (Miltenyi Biotech) for cardiomyocyte dissociation, and for purification, we used PSC-Derived Cardiomyocyte Isolation Kit (Miltenyi Biotech). Finally, the enriched cells were collected in 15 ml tubes topped with CDM3 medium and 20% FBS. The cells were pelleted at 500xg for 5 minutes, resuspended in 1 ml CDM3 containing 20% FBS, and 10 μM Rock inhibitor (Y27632), counted, and seeded. The cells started to contract after 2-5 days.

At day 30 of culturing in CDM3 glucose medium, we began the maturation of cells by switching to CDM3 galactose and fatty acids. The CDM3 galactose and fatty acids medium consisted of RPMI 1640 without glucose but with L-glutamine (Gibco), supplemented with 75 mg/ml human recombinant albumin (Sigma-Aldrich), 64 mg/ml L-ascorbic acid 2-phosphate (Sigma-Aldrich), 0.025% glucose (Gibco), 7.5 mM D-galactose (Sigma-Aldrich), 100 μM oleic acid-albumin (Sigma-Aldrich), 50 μM sodium palmitate (Sigma-Aldrich), and 100 μM L-carnitine (Sigma-Aldrich). The maturation continued for 15 days. Around day 45, the cardiomyocytes were dissociated and replated on matrigel-coated coverslips for immunofluorescence staining. The seeding density was about 125,000 cells/well of a 24-well plate for immunofluorescence. The cells were fixed with PFA 4%, 3-4 days after replating (detailed above at immunofluorescence method). The purity of cardiomyocyte cultures was assessed by immunofluorescence using cardiac troponin T (cTnT) as marker. iPSC-CM cultures from Patient 1 and Control 1 presented >80% purity.

5. RESULTS AND DISCUSSION

5.1. HEAT SHOCK RESPONSE IS A NOVEL DISEASE MECHANISM UNDERLYING INFANTILE LEUKOENCEPHALOPATHY (STUDY I)

5.1.1. CLINICAL PRESENTATION

The patient was a girl, the first child of a healthy Finnish couple with two younger healthy sons. The parents were distantly related. From birth, she was irritable and presented an increased muscle tone. In the next two years, she was diagnosed with delayed motor development and other symptoms common to white matter disorders. The remarkable feature of her disease was a sudden and dramatic deterioration following a febrile illness at two years and three months, with an encephalitis-like episode, prolonged seizures, and unconsciousness. She was treated in an intensive care unit for four days with antiepileptic, antiviral, and antibiotic medication, but no bacteria, virus, or antibodies could be detected. Another particularity of her disease was a malignant ovarian tumor diagnosed at six years of age and treated with chemotherapy. Her disease was progressive and the diagnostic based on MRI at ten years of age was hypomyelinated leukoencephalopathy with periventricular cysts and atrophy of the cerebellar vermis. At 15 years of age, she died of pneumonia.

5.1.2. IDENTIFICATION OF *HIKESHI* MUTATION

The WES data of our patient was filtered first to detect a probable homozygous mutation. After applying the filtering criteria (consisting of homozygosity, allele frequency <0.01 in Finnish cohorts and ExAC, CADD C-score >10, call depth >10, and absence of homozygotes in ExAC), one homozygous variant was found, which seemed the most likely cause of the patient's disease: *HIKESHI*, c.11G>C, p.Cys4Ser.

The variant was not present in the population database ExAC, which included exome data from more than 3,000 Finns, and was predicted to be deleterious by SIFT and CADD (the C-score was 20.7). The variant causes a missense mutation, exchanging a conserved cysteine to serine (p.Cys4Ser) in the protein's N-terminus.

The function of the protein Hikeshi was characterized in 2012 as the only transporter of HSP70s into the nucleus during heat stress (Kose et al, 2012). A few years later, a founder mutation in the *HIKESHI* gene (p.Val54Leu) was described to cause leukoencephalopathy with acquired microcephaly in Ashkenazi-Jewish (AJ) families (Edvardson et al, 2016). The name of the

gene and protein, Hikeshi, is a Japanese word for firefighter, reflecting its role in protecting cells from heat shock stress (Rahman et al, 2017). This function is achieved through the binding and nuclear transport of HSP70s, a major class of heat shock proteins (HSPs) (Figure 5).

The function of HSP70s relates to protection against stress and maintenance of homeostasis by promoting protein folding and refolding, proteasome-dependent degradation, chaperone-mediated autophagy, and protection against inclusion formation. The HSP70s' numerous functions affect almost all aspects of protein life cycle, from synthesis to degradation (Faust et al, 2020).

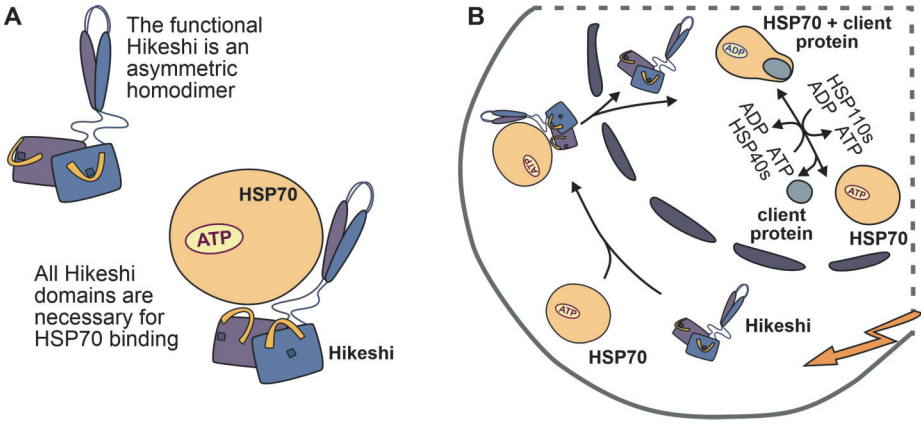


Figure 5. Cellular function of Hikeshi during heat stress.
(**A**) Functional Hikeshi is an asymmetric homodimer. All domains of Hiheshi seem to be essential for HSP70 binding. (**B**) During heat shock, Hikeshi binds HSP70s and transports them into the nucleus, where HSP70s assist in client protein folding.

Our study was the second report of *HIKESHI* mutations, expanded the disorder's clinical phenotype, and added knowledge to cellular pathways contributing to disease.

5.1.3. REMODELING OF INTERACTIONS INSIDE THE HYDROPHOBIC POCKETS OF HIKESHI LEADS TO PROTEIN INSTABILITY

The functional Hikeshi is an asymmetric homodimer, and its asymmetry is important for HSP70 binding (Song et al, 2015). Each monomer consists of an N-terminal domain (NTD) folded into a jelly-roll/β-sandwich structure and a C-terminal domain (CTD) with a coiled-coil-like conformation, which are connected by a flexible linker region (Figure 6A).

The NTDs contain the hydrophobic pockets involved in FG-Nups (phenylalanine-glycine repeat-containing nucleoporins) recognition. The

interaction between Hikeshi and FG-nups is proposed to be regulated by a flexible extended loop (E-loop) in the NTD. The E-loop covers the hydrophobic pocket like a lid. One amino acid in the E-loop, Phe97, protrudes into the hydrophobic pocket, playing a critical role in closing and opening the lid (Figure 6B). The open conformation enables binding to FG-nups, allowing further the translocation into the nucleus.

The amino acids altered in the Finnish and AJ patients, Cys4 and Val54, respectively, are localized within the NTDs and are involved in hydrophobic interactions inside the pockets.

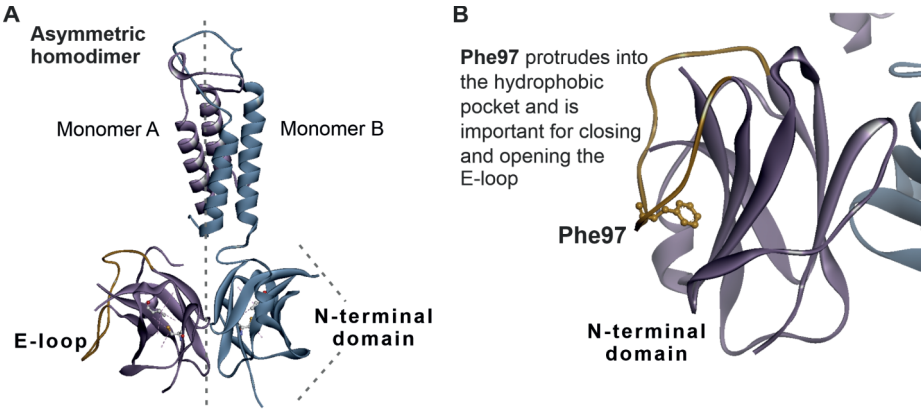


Figure 6. Structure of Hikeshi after Song et al, 2015.
(**A**) The N-terminal domain of Hikeshi contains the hydrophobic pocket, which is covered by the E-loop. Upon binding, the E-loop modifies its conformation to expose the hydrophobic pocket involved in nuclear pore recognition. (**B**) Phe97 protrudes into the N-terminal domain's hydrophobic pocket, playing an important role in covering or exposing the surface involved in FG-Nups recognition.

The wild type Cys4 establishes two hydrophobic interactions in both monomers, with Phe25 and Ile111. Computational modeling of p.Cys4Ser mutation indicated that the missense change would determine the loss of the hydrophobic bonds in both monomers. In monomer A, the side chain of Ser4 would form a new hydrogen interaction with the aromatic ring of Phe25, while in monomer B, Ser4 establishes no interactions. Therefore, the hydrophobic interactions inside the pockets are altered and weaker for p.Cys4Ser mutation. Western blot showed undetectable levels of Hikeshi in patient fibroblasts, consistent with the idea that altered interactions inside hydrophobic pockets induce protein instability.

Edvardson and colleagues (2016) also showed an undetectable level of protein in Western blot for the AJ founder mutation p.Val54Leu; therefore, we also analyzed the AJ missense change by computational modeling. An asymmetry in hydrophobic interactions inside the two monomers' pockets is observed for the wild type Val54, which establishes three hydrophobic

interactions in monomer A, with Val38, Phe82, and Ile113, but only two in monomer B, with Val38 and Phe82. In the model of p.Val54Leu mutation, Leu54 forms four hydrophobic bonds in both monomers, with Val38, Val 73, Phe82, and Ile113. We observe, thus, the loss of binding asymmetry between the two monomers characteristic for the wild type Val54 and remodeling of hydrophobic interactions inside the pockets, which become stronger with Leu54 mutation.

The dramatic conformational changes undergone by Hikeshi upon HSP70 binding require a structure with a high degree of flexibility, which maintains at the same time the ability to bind to FG-nups and translocate into the nucleus. The observation that both decreased and enhanced hydrophobic interactions inside the pockets would lead to protein instability suggests that the pockets' 3D conformation is under functional constraints and cannot tolerate significant alterations without affecting stability.

5.1.4. HIKESHI DISEASE MECHANISM INVOLVES BOTH NUCLEAR CHAPERONE AND ER FUNCTIONS

During heat shock, the expression of HSP70s increases, and they translocate into the nucleus to protect nuclear structures from heat damage. The damage refers to the loss of native conformation of proteins (associated with DNA, RNA, or carrying out other tasks) inside the nucleus (Imamoto and Kose, 2012). After one-hour heat shock at 42°C, control fibroblasts exhibit a substantial punctate accumulation of HSP70 inside the nucleus, while the patient's fibroblasts carrying the p.Cys4Ser mutation show weaker accumulation. However, variability characterized the repeats of this experiment. Therefore, we suggest that other factors might modulate HSP70s transport into the nucleus during heat shock, possibly on a cell-type-specific basis. Both patient and control fibroblasts responded to heat shock by phosphorylating HSF1-S326.

Given that heat shock and endoplasmic reticulum (ER) stress responses have common mediators, we screened the ER stress markers in the patient's cells. Hikeshi^{Cys4Ser} fibroblasts showed normal levels of unfolded protein response (UPR) markers IRE1α and BiP. These markers also show a typical upregulation pattern when patient cells were treated with the ER stressor tunicamycin. Interestingly, we found that the level of ERO1-Lα, an ER oxidoreductase, was markedly reduced in the patient's fibroblasts. Therefore, we can state that ERO1-Lα reduction was independent of a generalized activation of ER-UPR or defective ER-UPR expression.

5.1.5. HEAT SHOCK RESPONSE IN MYELIN DISORDERS

As discussed in the introduction, leukoencephalopathies can result from genetic defects in many cellular pathways. Among them, we encounter pathways related to nucleic acid metabolism, immunity to nucleic acids,

protein quality control, and ER stress, which are all vulnerable pathways in Hikeshi disorder.

The inefficient nuclear import of HSP70s in patients' cells will likely render the nucleus vulnerable to protein and nucleic acid damage and supports the idea that an inadequate response to heat stress is an essential element of the Hikeshi disorder mechanism. The nuclear transport of HSP70s by Hikeshi has been mostly studied in the context of heat stress (Imamoto, 2018). It would be interesting to address the question if other types of cellular stress would affect the Hikeshi pathway.

At the age of two years and three months, our patient manifested pronounced deterioration following a febrile illness with suspected encephalitis. In this context, it is noteworthy that three out of six AJ patients died suddenly during short febrile illnesses, implying abnormal responses to heat stress. Moreover, our patient was diagnosed with a malignant ovarian tumor. Although, as a single finding, this can be coincidental, cellular stress responses mitigated by heat shock proteins have been described in both neurodegenerative disorders and cancer (Lindberg et al, 2015; Sherman and Gabai, 2015).

A rising question concerns the downstream effects of heat shock damage and how the cells respond to nuclear stress. One proposed hypothesis is that the excess of extra-nuclear HSP70s might initiate cytokine release or provide a target for natural killer cells (Edvardson et al, 2016). An alternative to this explanation would be an activation of the innate immune system by altered DNA and RNA metabolism during heat shock, or other types of cellular stress, in patients with Hikeshi mutations. Innate immunity activation induced by disrupted nucleic acid metabolism has been illustrated by another demyelinating syndrome, Aicardi-Goutières (Blumen et al, 2012; Rice et al, 2007).

The synthesis and turnover of plasma membrane components are accomplished through the secretory pathway. This pathway becomes particularly active in myelinating cells, which have to synthesize and deliver considerable amounts of membrane proteins, cholesterol, and other membrane lipids to specific membrane sites. Endoplasmic reticulum (ER) stress disrupts the secretory pathway, and ER unfolded protein response (UPR) appeared as a recurrent theme in myelin disorders as in Pelizaeus-Merzbacher disease, vanishing white matter disease, Charcot-Marie-Tooth, and multiple sclerosis (Lin and Popko, 2009). We found normal levels of ER-UPR markers in the patient's cells, markers that were also normally upregulated during tunicamycin treatment. However, we uncovered a clear reduction of ERO1-Lα, an ER oxidoreductase involved in disulfide bond formation and ER redox state, indicating that ER likely contributes to the cellular pathology of Hikeshi disorders.

The major function of HSPs is to ensure proper folding and prevent protein aggregates, with HSP70s being involved in the active folding and refolding of proteins. Only a few examples of other genes encoding

chaperones and co-chaperones causing diseases with nervous system involvement are *HSPD1* in hypomyelinating leukodystrophy (Magen et al, 2008) and hereditary spastic paraplegia (Hansen et al, 2002); *TBCE* causes Sanjad-Sakati and Kenny-Caffey syndromes (Parvari et al, 2002); *SIL1* leads to Marinesco-Sjögren syndrome (Anttonen et al, 2005; Senderek et al, 2005); *DNAJC6* (Edvardson et al, 2012) and *DNAJC13* (Vilariño-Güell et al, 2014) are involved in Parkinson's disease; *DNAJC19* in dilated cardiomyopathy with ataxia (Davey et al, 2006; Ojala et al, 2012a). The addition of *HIKESHI* to this list highlights the crucial role of adequate stress responses in the nervous system's normal development and maintenance.

5.2. GENETIC LANDSCAPE OF SEVERE CHILDHOOD CARDIOMYOPATHIES IN FINLAND (STUDY II)

5.2.1. THE CHARACTERISTICS OF KIDCMP COHORT REPRESENTATIVE FOR FINLAND FOR SEVERE CHILDHOOD CARDIOMYOPATHIES

The childhood cardiomyopathy (CMP) genetics project started in our lab many years back by dedicated clinicians who have gathered clinical information and biological samples from patients and family members. With centralized data and support for research, the Finnish medical system has played a key role in our project, facilitating the cohort collection. In Finland, all severe pediatric cardiomyopathy patients that require inotropic support or are evaluated for cardiac transplant are investigated at University Hospital in Helsinki. This provided us the unique opportunity to collect the KidCMP cohort, representing the whole country for this subgroup of severe childhood cardiomyopathy patients. As most of the patients did not have a genetic diagnosis at the study initiation, the cohort was considered naïve from a genetic perspective.

The KidCMP cohort consists of 66 patients and presents three characteristics that are making it unique: a) mostly based on Finnish ancestry, with a history of geopolitical isolation and genetic homogeneity; b) early-onset, with a median diagnosis age of 0.33 years (mean 2.05 years); c) representative of severe childhood cardiomyopathy diversity of a country, as it was collected from the single center in Finland performing cardiac transplantations.

There were five CMP types in the cohort: dilated cardiomyopathy (DCM, n=37), hypertrophic cardiomyopathy (HCM, n=20), left ventricular noncompaction (LVNC, n=6), restrictive cardiomyopathy (RCM, n=2), and histiocytoid cardiomyopathy (HICM, n=1). Seventeen patients underwent cardiac transplantation, 18 have died of their disorder, the cardiac findings in 14 patients were stabilized with treatment, and 17 present a progressive disorder. For the 17 transplanted and 18 deceased patients, the average

transplant time was 1.97 years (median 1 year), and the average death time occurred 2.24 years (median 0.315 years) after the initial diagnosis.

Forty-three patients were females, 23 males (a trend towards female preponderance). Of all, 49 had no familial history, suggesting an autosomal recessive inheritance or *de novo* occurrence. Seventeen patients showed a familial disease: in 7 families with multiple affected siblings and healthy parents, we considered recessive inheritance, while 10 families showed dominant inheritance.

5.2.2. HETEROGENEOUS GENETIC BACKGROUND ENRICHED IN *DE NOVO* MUTATIONS

Our lab has been a pioneer in Finland for using NGS for the genetic diagnosis of cardiomyopathies, with the publication of *AARS2* as a novel disease gene underlying fatal neonatal CMP (Götz et al, 2011). Other publications followed, including *MRPL44* as a novel disease gene (Carroll et al, 2013) discovered by WES, and a few clinical reports on *PRKAG2* and *HRAS* (Hiippala et al, 2016; Pöyhönen et al, 2015), uncovered by panels. In the beginning, selected patients have also undergone candidate gene screening based on their clinical presentation, with findings in *HADHB* and *DNAJC19* genes published as clinical reports (Ojala et al, 2012a, 2015b). The above six genetic findings from patients belonging to the KidCMP cohort have been published separately. In the current screening, we identified disease-causing variants in the following genes (Figure 7): *PPA2*, *TAZ*, *BAG3*, *NEK8*, *TBX20*, *TAB2*, *PTPN11*, *RAF1*, *JPH2*, *CALM1*, *CACNA1C*, *TNNC1*, *TNNI3*, *ACTC1*, *MYH7,* and *NRAP*. Each disease-causing variant was family-specific in our screening. Among all 26 molecular diagnoses in the cohort, 12 (46%) were *de novo*, 9 (34%) recessive, and 5 (20%) dominantly inherited.

Therefore, *de novo* variants emerged as a common cause of early-onset CMPs, accounting for almost half of genetic diagnoses, with implications for family counseling and genetic diagnosis approaches. Our results indicate a high genetic heterogeneity in severe childhood cardiomyopathies, with family-specific and frequent *de novo* mutations. We have not identified major recessive ancestral mutations enriched in the country, as it is typical for other diseases described as Finnish disease heritage (Norio, 2003b).

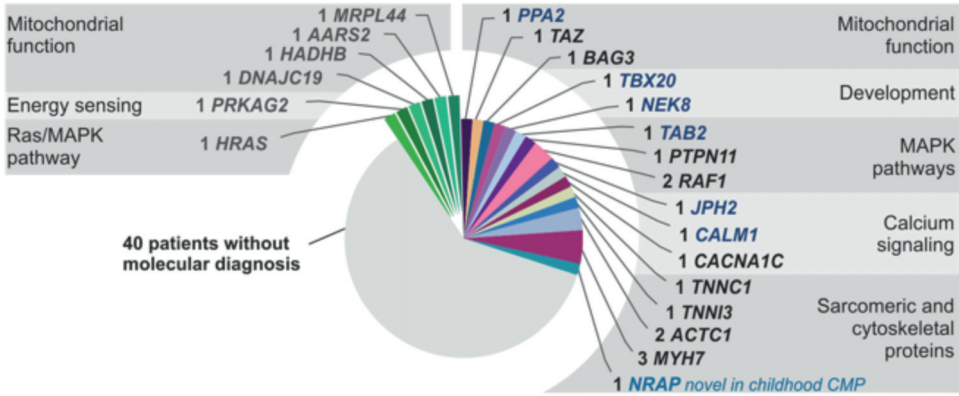


Figure 7. Heterogeneity of genetic causes in the KidCMP cohort.
We encountered a high genetic heterogeneity, each family presenting a different genetic cause. *De novo* pathogenic variants were common in the cohort, accounting for 46% of diagnoses. The right panel depicts genes uncovered in the current screening. Dark blue: a novel clinical presentation. Light blue: *NRAP* novel in childhood CMP. Left panel: genes characterized previously from the cohort.

5.2.3. CONSEQUENCES OF MUTATIONS AT PROTEIN LEVEL

For an etiological understanding of disease pathogenesis and genotype-phenotype correlations, it is essential to study the effects of mutations on protein structure and function.

Table 5 summarizes the predicted structural and functional consequences of the pathogenic variants identified in our screening, based on the current knowledge from literature and our functional assays.

Table 5. Consequences of variants at the protein level.
Current knowledge of biochemical and predicted functional consequences of the variants based on literature and our protein analyses.

Gene (Protein)	Variant	Effects at protein level	Evidence
PPA2 Inorganic pyrophosphatase 2, mitochondrial	p.E172K p.V186M	Loss of function by protein instability	Molecular modeling Western blotting
TAZ Tafazzin	p.G240R	May affect protein stability	*In silico* modeling (Hijikata et al, 2015)
BAG3 BAG3 family molecular chaperone regulator 3	p.P209L	The variant occurs in IVP motif 2, described to mediate the interaction with HSPB8 and HSPB6	(Fuchs et al, 2009)

Table 5. Consequences of variants at the protein level.
(Continuation)

Gene (Protein)	Variant	Effects at protein level	Evidence
TBX20 T-box transcription factor 20	p.M224V	Likely affects binding in the major groove of DNA	Molecular modeling
NEK8 Serine/threonine-protein kinase Nek8	p.I215T p.R352L	Unknown. Other pathogenic variants altered Hippo signaling	(Frank et al, 2013; Grampa et al, 2016)
TAB2 TGF-beta-activated kinase 1 and MAP3K7-binding protein 2	p.S390Qfs*37	Haploinsufficiency	Early stop codon
PTPN11 Tyrosine-protein phosphatase non-receptor type 11	p.Q510E	Gain of function. Decreased phosphatase activity but also decreased autoinhibition, the protein adopting easily an open, active conformation	Activity measurements and crystal structures (Yu et al, 2014)
RAF1 RAF proto-oncogene serine/threonine-protein kinase	p.S257L	Increased kinase activity and enhanced ERK activation	(Pandit et al, 2007; Razzaque et al, 2007)
	p.P261R	The variants, p.P261A, p.P261L, and p.P261S, show increased kinase activity and enhanced MAPK1 activation	(Pandit et al, 2007; Razzaque et al, 2007)
JPH2 Junctophilin-2	Hom p.Q428*	It likely causes loss of function. The transmembrane anchor of the protein is lost	Position of the premature stop codon. The stability of the protein fragment was not investigated
CALM1 Calmodulin	p.F142I	F142 is adjacent to the fourth EF-hand (calcium-binding) domain. p.F142L reduces the affinity for calcium in the C-terminal domain and alters the ability of the protein to transduce calcium signals	Experimental evidence for p.F142L (Crotti et al, 2013; Rocchetti et al, 2017)
CACNA1C Voltage-dependent L-type calcium channel subunit alpha-1	p.G406R	Reduced channel inactivation resulting in maintained depolarizing L-type calcium currents	(Splawski et al, 2005)
TNNC1 Troponin C, slow skeletal and cardiac muscles	p.A31S	Partial activation of a silenced calcium-binding domain, EF-hand 1. Increased affinity for calcium	(Parvatiyar et al, 2012)
TNNI3 Troponin I, cardiac muscle	p.V176M	Not clear. The change occurs in the α-helix 4 of the C-terminal mobile domain. More data is needed regarding the interaction with other sarcomeric/cytoskeletal proteins and about the dynamics of the complex during the contraction cycle	Crystal structure of cardiac troponin (Takeda et al, 2003)

Table 5. Consequences of variants at the protein level. (Continuation)

Gene (Protein)	Variant	Effects at protein level	Evidence
ACTC1 Actin, alpha cardiac muscle 1	p.H175R	May affect the interaction between actin monomers	Cryo-EM structure of a human cytoplasmic actomyosin complex (von der Ecken et al, 2016)
	p.Y220H	The affected amino acid is localized at the edge of the ATP-binding pocket and in the vicinity of the tropomyosin filament	Cryo-EM structure of a human cytoplasmic actomyosin complex (von der Ecken et al, 2016)
MYH7 Myosin-7	p.S842R	Substitution between amino acids with different properties at the beginning of the rod domain: a large positive arginine replaces a small neutral serine Removal of a predicted phosphorylation site	MyoMAPR http://bmf.colorado.edu/myomapr www.phosphosite.org
	p.R925G	Amino acids with different properties: large arginine replaced with small glycine; charge transition from + to 0 in position g of the rod domain repeat three By analogy with p.E924K, may influence the binding affinity of MyBP-C	MyoMAPR http://bmf.colorado.edu/myomapr Experimental evidence for p.E924K (Gruen and Gautel, 1999)
	p.K1879E	Amino acid charge transition from + to - in position e of the rod domain repeat 37	MyoMAPR http://bmf.colorado.edu/myomapr
NRAP Nebulin-related-anchoring protein	Hom p.Y448*	It likely causes loss of function. NRAP mRNA harboring the premature stop codon is not degraded in the patient's heart biopsy	Position of the premature stop codon. qPCR measurement of transcript level The stability of the protein fragment was not determined

5.2.4. MOLECULAR PATHWAYS AFFECTED BY DISEASE GENES

To establish an accurate prognosis and develop future therapies, it is vital to understand the mechanisms involved in disease, how molecular pathways are disturbed by mutations, and how they can be modulated with drugs. In our screening, the identified genes are involved in mitochondrial metabolism (*PPA2*), cardiolipin maturation (*TAZ*), protein quality control (*BAG3*), intracellular calcium regulation (*JPH2*, *CALM1*, *CACNA1C*), Ras/MAPK signaling (*PTPN11*, *RAF1*, *TAB2*), sarcomeric (*MYH7*, *ACTC1*, *TNNC1*, *TNNI3*) and cytoskeletal (*NRAP*) structure, and development (*TBX20*, *NEK8*).

Mitochondrial function

PPA2 defects often manifest as sudden death or as rapidly progressing DCM (Guimier et al, 2016, 2016). PPA2 is a mitochondrial protein hydrolyzing pyrophosphate to inorganic phosphate, which can be further used for ATP synthesis and other processes. We identified two brothers with rapidly progressing DCM and compound heterozygous mutations in *PPA2*.

Mutations in *TAZ* typically cause Barth syndrome, which includes CMP. TAZ is an acyltransferase responsible for cardiolipin's maturation, the major phospholipid in the inner mitochondrial membrane. In our cohort, two brothers with DCM presented a previously reported pathogenic variant in *TAZ* (D'Adamo et al, 1997).

Protein quality control

Patients with BAG3 mutations, including our patient, present a consistent clinical picture of myofibrillar myopathy, cardiomyopathy, with rigid spine or neuropathy (Selcen et al, 2009). BAG3 is a co-chaperone of Z-disk in muscle cells. It is also involved in apoptosis by interacting with BCL2 and a component of the chaperone-assisted selective autophagy pathway.

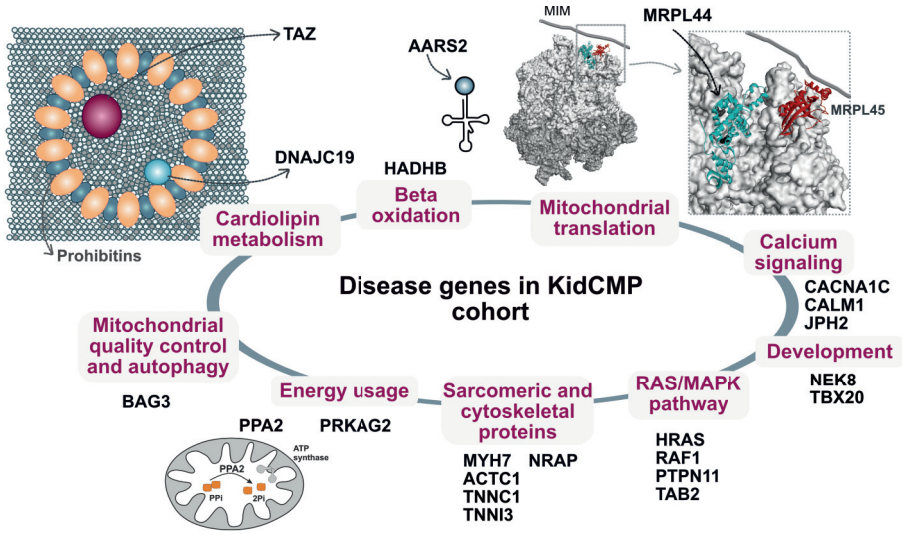


Figure 8. Pathways affected by disease genes in the KidCMP cohort. We uncovered pathogenic variants in sarcomeric and cytoskeletal proteins, mitochondrial and other metabolic proteins, developmental, MAPK pathways, and calcium signaling proteins.

Development

NEK8 kinase regulates the primary cilium function on the cell surface involved in organ development (Frank et al, 2013). It is also connected to the machinery controlling the DNA repair and cell cycle progression (Jackson, 2013). Our patient manifested with HCM and liver cirrhosis but without

kidney involvement, which has been the primary symptom in other patients (Rajagopalan et al, 2016).

TBX20 is a transcription factor essential for heart development (Boogerd Cornelis J. et al, 2018), and its mutations manifest with various cardiac pathologies (Kirk et al, 2007). Our patient presented LVNC without structural heart defects, the typical presentation in other patients.

Ras/MAPK pathway

PTPN11 is a component of the Ras/MAPK pathway involved in cell growth and differentiation. Mutations in this gene cause Noonan syndrome (Takahashi et al, 2005), are typically severe and occur *de novo*, with a typical presentation in our patient.

RAF1 proto-oncogene contributes to the Ras/MAPK pathway, often causes Noonan syndrome with significant HCM (Pandit et al, 2007; Razzaque et al, 2007), but sometimes also can lead to isolated DCM (Dhandapany et al, 2014). Our patients presented *de novo* pathogenic variants with Noonan phenotype.

TAB2 is a developmentally regulated gene, part of MAPK pathways. Our patient with fatal infantile DCM showed a pathogenic variant causing *de novo* TAB2 haploinsufficiency, previously involved in congenital heart disease (Thienpont et al, 2010).

Calcium signaling

JPH2 is a component of junctional complexes between the sarcolemma and sarcoplasmic reticulum in cardiomyocytes, playing important roles in calcium homeostasis and dyad architecture. Mutations in JPH2 associate with HCM and DCM (Landstrom et al, 2007; Sabater-Molina et al, 2016). We found a novel homozygous nonsense mutation in JPH2, causing recessive DCM.

Calmodulin is a highly conserved protein of 149 amino acids encoded by three genes in the human genome (*CALM1*, *CALM2*, and *CALM3*). This redundancy underscores its crucial role in calcium signaling. Our patient was a child with LVNC, infantile spasms, and developmental delay, with a *de novo* mutation in *CALM1*. Another mutation at the same amino acid position has been reported to cause non-sustained ventricular tachycardia, markedly prolonged QTc interval, and cardiac arrest (Crotti et al, 2013).

CACNA1C mediates the influx of calcium into cardiomyocytes. Our patient was diagnosed with prenatal 2:1 heart block and LVNC. Exon 8 and exon 8A of CACNA1C are alternatively spliced, mutually exclusive, and both disease-associated. The patient's mutation resides in exon 8, occurred *de novo*, and was previously associated with Timothy syndrome type 2 (Splawski et al, 2005).

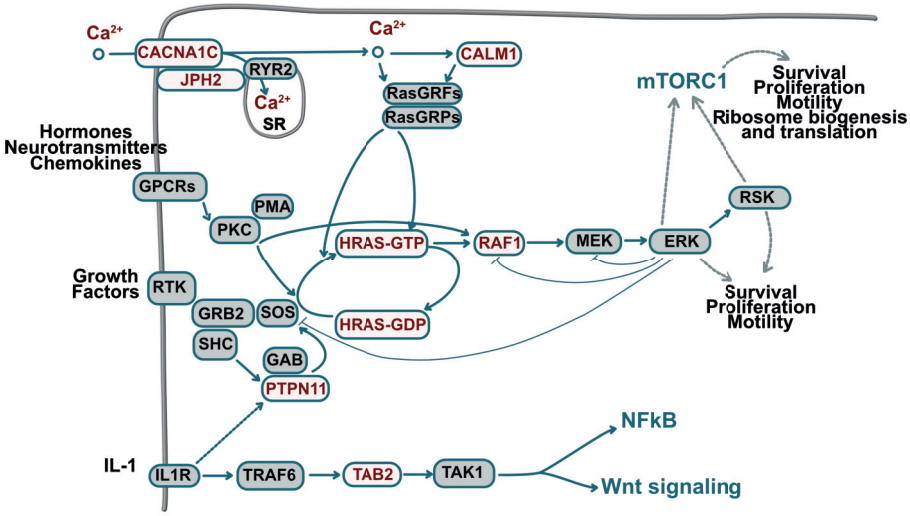


Figure 9. Calcium and MAPK signaling defects in the KidCMP cohort. Proteins marked with red presented pathogenic variants in our patients.

Sarcomeric and cytoskeletal proteins

Sarcomeres are contractile units of the muscle, with a highly organized protein machinery specialized in the generation of force. The sarcomere's primary structural proteins are myosin and actin, which are significant components of the thick and thin filaments, respectively (Sequeira et al, 2014). *MYH7* (Myosin-7) is a frequent cause of cardiomyopathies, and we identified three novel mutations in this gene, one familial and two *de novo*. On the other hand, thin filament mutations in *ACTC1* (Alpha-actin) are also well-known causes of cardiomyopathy, with two novel likely mutations in our cohort.

Troponin complex, consisting of troponin C, troponin I, and troponin T, regulates the interaction between myosin and actin in response to changes in calcium concentration. Troponin C (TNNC1) is the calcium-sensing subunit and drives the complex's conformational changes, which allow myosin-actin interaction. We identified a previously reported *de novo* variant in *TNNC1* (Parvatiyar et al, 2012). Troponin I (TNNI3) is the inhibitory component of the complex, contributing to tropomyosin maintenance in a position that prevents myosin-actin interaction. This gene was found harboring a dominantly inherited pathogenic variant in one of the patients.

NRAP is a cytoskeletal protein involved in anchoring terminal actin filaments to the membrane, tension transmission from myofibrils to the extracellular matrix, and myofibril assembly. Our study involved for the first time *NRAP* in childhood DCM, previously associated with adult-onset DCM (Truszkowska et al, 2017).

5.2.5. SUCCESS OF GENETIC SCREENING WITH IMPLICATIONS FOR GENETIC DIAGNOSIS OF SEVERE CHILDHOOD CARDIOMYOPATHIES

The overall positive genetic diagnosis rate amounted to 39% in our cohort, with 26 patients solved of the total 66 patients. Among the KidCMP's 66 patients, 22 (33%) presented a systemic disorder, of which 16 (72%) had an age of onset <1 year. A genetic diagnosis was reached in 54% of the systemic disease patients and 32% of the isolated cardiomyopathies. Our cohort also showed an increase in genetic diagnosis success with later onset: age <1 year 34% positive DNA diagnosis, 1 to 5 years 38%, 6 to 10 years 60%, 11 to 15 years 60%. Two-thirds of the infantile (<1 year) DNA findings were explained by recessive metabolic genes or *de novo* variants affecting calcium signaling and Ras/MAPK pathways.

The 39% diagnostic rate is significant, considering that no recurrent mutations have been found, but rather family-specific, rare pathogenic variants. Moreover, almost half of the identified mutations are *de novo* (emerged during embryogenesis, or mosaic/germline in a parent), which are challenging to identify using only the index patient data as it was in our case, and would ideally utilize trio screening. Our study brought genetic information to over 25 families and a catalog of mutation types in severe childhood CMPs in Finland.

This study used two CMP gene panels and WES, some patients being investigated with multiple methods. The concordance of positive findings between panels was seven from eight findings (88%) due to a gene not covered in both, and between panels and WES was six from 12 findings (50%), accounting for genes not covered in panels. Given the high number of patients without clarified genetic background, it is evident that many new disease genes are still to be uncovered. Together with increased *de novo* findings, our results suggest that WES or WGS of trios is the most suitable approach to genetic diagnosis of severe childhood CMPs, especially in children <5 years of age. As the diagnosis rate increases in patients >5 years at onset, and the panels are continuously incorporating new disease genes, they might represent a cost-effective approach for children with later CMP onset.

5.3. *TMOD1* IS A NOVEL DISEASE GENE FOR CARDIOMYOPATHY (STUDY III)

5.3.1. CLINICAL PRESENTATION

In this study, we investigated three patients from two families manifesting with childhood-onset CMPs, and who present the same underlying genetic cause of disease.

Patient 1, daughter of non-consanguineous healthy parents, was diagnosed with dilated cardiomyopathy (DCM) and short ventricular tachycardia episodes at the age of 12 years. Cardiac magnetic resonance imaging (cMRI) showed that both right and left atria were mildly dilated, indicating diastolic dysfunction. The right ventricular systolic function was normal. The left ventricle was dilated and decreased in systolic function (ejection fraction EF 38%). Mild late enhancement in the mid-myocardium of the intraventricular septum was observed as a marked dilated cardiomyopathy. Ventricular tachycardia responded to oral treatment with beta-blockers. Heart failure was further treated with enalapril and diuretics. Cardiac function stabilized to the level of mild-to-moderate systolic dysfunction during the 10 years follow-up (NYHA 1-2).

Patient 2 is the daughter of non-consanguineous healthy parents. She was referred to the hospital when she was nine years old due to atrial enlargement observed on an electrocardiogram (ECG), which was incidentally performed as part of a clinical investigation for migraines. cMRI demonstrated the dilatation of both atria, as marked diastolic dysfunction. Right and left ventricular volumes were normal. Right ventricular systolic function was normal, but the left ventricular systolic function was mildly decreased (EF 49%). After a two-year follow-up period, the patient began to feel weak with dyspnea on effort (NYHA 2). Spiroergometry demonstrated decreased oxygen consumption (51%). Warfarin therapy was started to prevent atrial thrombus, and she was listed for heart transplantation, currently waiting for a transplant.

Patient 3, the twin brother of patient 2, was initially asymptomatic and underwent cardiac ultrasound screening at nine years because of his sister's findings. cMRI demonstrated biatrial enlargement but normal ventricular volumes. Right ventricular systolic function was normal, but the left ventricular systolic function was mildly decreased on cMRI (EF 45%). After a two-year follow-up period, he developed atrial tachycardia and non-sustained ventricular tachycardia. He became symptomatic with discomfort and upper-abdominal pain during exercise. Spiroergometry demonstrated decreased oxygen consumption (64%). Beta-blocker and warfarin therapy were started, and he underwent ICD implantation for documented non-sustained ventricular tachycardia. He is currently listed for heart transplantation.

5.3.2. IDENTIFICATION OF *TMOD1* VARIANT R189W

Patient 1 was part of the KidCMP cohort and studied in our lab. Patients 2 and 3, from the second family, were referred by clinicians to Blueprint Genetics for molecular diagnosis. Patients 1 and 2 were first screened with gene panels, with negative findings. WES was further used for all three patients, and a homozygous missense variant in *TMOD1* (Tropomodulin-1, NM_003275, NP_003266, c.565C>T, p. R189W) was independently

identified in the two families as the most likely cause of cardiomyopathy. The variant had a high CADD C-score of 33 and was rated deleterious by SIFT. No homozygotes were found in any public databases, and the minor allele frequency in Finland is 0.00025 (www.sisuproject.fi). The variant amino acid, R189, is highly conserved in species, even in the invertebrates *D. melanogaster* and *C. elegans*.

Tmod1 is an actin-binding protein that stabilizes sarcomeric actin filaments in cardiomyocytes, playing a crucial role in regulating thin filament length and contraction (Gokhin and Fowler, 2011a).

Two actin isoforms have been described in cardiomyocytes: alpha-cardiac actin (ACTC1, >80% in the adult heart), part of thin filaments, and gamma-cytoplasmic actin (ACTG1), which is a component of the Z-disc, surrounds mitochondria, sarcoplasmic reticulum (Gokhin and Fowler, 2011b), and is also found near sarcolemma (Sequeira et al, 2014).

In general, F-actin (filamentous-actin) presents two ends: barbed end with fast polymerization and pointed end with slow polymerization. In cardiomyocytes, CapZ protein caps and anchors to the Z-disc the barbed ends (Papa et al, 1999), while Tropomodulin 1 caps the pointed ends in the A-band (Weber et al, 1994).

5.3.3. R189W RESIDES IN ACTIN-BINDING SITE 2 OF TMOD1 AND IS PREDICTED TO AFFECT THE LOCAL FOLDING

Tmod1 functions as a 'leaky' cap at the pointed end of tropomyosin-coated actin filaments, allowing for the controlled addition and dissociation of actin subunits. Tmod1 harbors two actin- and two tropomyosin-binding sites (Figure 10A). The actin-binding site 2, located in the C-terminal region, contains a leucine-rich repeat (LRR) domain with four and a half LRR motifs. The ascending loops of each LRR motif contact actin. The N- and C- terminal caps of the LRR domain, consisting of short alpha-helical regions, shield the domain's hydrophobic core from solvent exposure (Figure 10B).

The homozygous variant p.R189W is localized to the N-terminal cap of the leucine-rich domain. The actin-binding sites 1 and 2 of human Tmod1 were separately crystalized, bound to actin monomers. Figure 10B depicts the modeling of patient variant into the crystal structure of the human actin-binding site 2 of Tmod1 in complex with an actin monomer (PDB id 4PKI) (Rao et al, 2014b). The model shows that the amino acid change determines the exposure of a hydrophobic residue at the surface of the protein (Figure 10B). This may influence the alpha-helical folding of the N-terminal cap of the LRR domain, its ability to shield the hydrophobic core from the solvent, and ultimately may affect the domain’s affinity for actin.

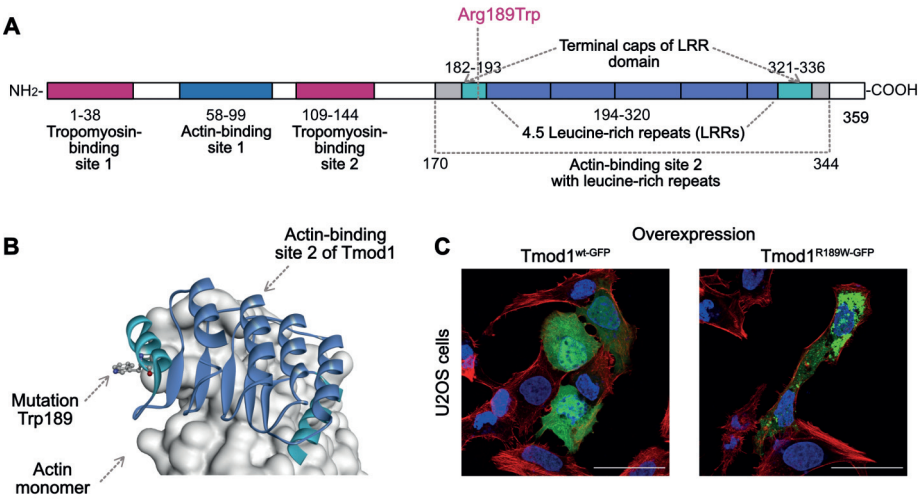


Figure 10. Protein domains of TMOD1, molecular modeling of patient variant, and overexpression in U2OS cells. (**A**) Diagram of protein domains (**B**) Modeling of patient mutation p.Arg189Trp in actin-binding site 2 of Tmod1 in complex with an actin monomer (**C**) Overexpression of Tmod1^{wt}-GFP and Tmod1^{R189W}-GFP in U2OS cells.

5.3.4. TMOD1^{R189W} IS PRONE TO FORM AGGREGATES WHEN IT IS LINKED TO GFP, BUT NOT WITHOUT

To test the functional consequences of Tmod1^{R189W}, including possible local misfolding suggested by the molecular model, we overexpressed GFP-linked wild type and altered protein in U2OS cells. Within one day, Tmod1^{R189W}-GFP showed a propensity to form aggregates. The aggregates were present in different but substantial amounts in the majority of GFP-positive cells. The overexpression of Tmod1^{wt}-GFP or GFP-only (empty-vector) did not induce this phenotype, indicating that the R189W change underlies this presentation.

The availability of a muscle biopsy from Patient 1 enabled us to analyze Tmod1 and Tmod4 levels in skeletal muscle comparing with three control samples. Overall, their levels in the patient muscle were within the control range, but the controls displayed large variations. This large variation in Tmod1 and Tmod4 amounts has been reported for skeletal muscles, probably related to their slow and fast fiber composition variability.

We performed untargeted proteomics from patient skeletal muscle biopsy and 5 controls. The most significant upregulated protein in the patient was Myosin light chain 4 (MYL4) in untargeted data. Other upregulated proteins were: the actin-binding Tropomyosin 1 (TPM1) and Gelsolin (GSN), the metabolic proteins Fatty acid-binding protein 4 (FABP4) and Lactate dehydrogenase 4 (LDHA), and the stress-response proteins Heat shock factor

binding protein 1 (HSPB1) and Alpha-crystallin B chain (CRYAB). The most significantly downregulated protein in the patient was Creatine kinase B (CKB), a protein involved in energy homeostasis. All the protein changes are mirrored at the molecular pathway level by alterations in glycolysis and gluconeogenesis, actin cytoskeleton signaling, and other changes indicated by IPA (Integrated Pathway Analysis). The proteomics data demonstrates that Tmod1 is important in maintaining the structural integrity of the muscle, and Tmod1^{R189W} mutation results in the dysregulation of other actin-binding proteins.

5.3.5. EARLY CARDIOMYOCYTE DEVELOPMENT AND *TMOD1* R189W MUTATION

We considered that studying the patient's cardiomyocytes will enable us to have a more physiologically relevant picture about the mechanism of mutation, how it affects the cellular cytoskeleton even before the disorder may overtly manifest. Therefore, we engineered induced pluripotent stem cells (iPSCs) from patient fibroblasts and further differentiated them to cardiomyocytes (CMs) (Burridge et al, 2015). The cells were purified at day 20 after induction to enrich them in cardiomyocytes and cultured further for maturation, which included 15-days culturing in galactose and fatty acids medium. The cells were plated at day 45 for immunocytochemistry. The patient and control iPSC-CMs were stained to reveal the sarcomeric structure and Tmod1 localization.

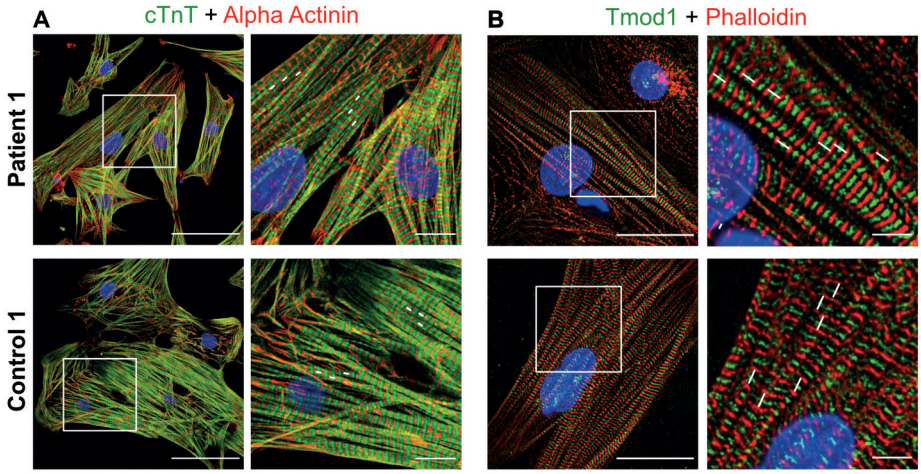


Figure 11. iPSC-CMs from Patient 1 and Control 1.
(**A**) Patient iPSC-CMs have a good sarcomeric development, broadly similar to controls. Scale bars: 50 μm and 10 μm (**B**) Fixation with 'relaxing buffer'. Tmod1 localizes to the developing M-line in the patient and control cells. Scale bars: 25 μm and 5 μm. The white sarcomeric lines are drawn in control cells at the same length as in the patient.

Cardiac troponin T (cTnT) is a common marker for cardiac differentiation and is part of the troponin complex localized along the sarcomeric actin filament. Alpha-actinin is a component of the Z-disk, structure that delimitates the sarcomeres. Figure 11A shows that patient iPSCs differentiated into cardiomyocytes and developed a good sarcomeric structure, broadly similar to controls.

Further, the cells were fixed in relaxing conditions, aiming to open the sarcomeres and reveal the structures in their full length, uncontracted. However, unlike mature cardiomyocytes, phalloidin staining is much weaker along thin filament comparing with the Z-disk area, making the thin filament visualization difficult. Although Tmod1 localized mainly to the developing M-line/A-band at the sarcomeres' middle, it did not separate in relaxing conditions in two rows at the tip of actin filaments, suggesting immature structures in both patient and control cells (Figure 11B).

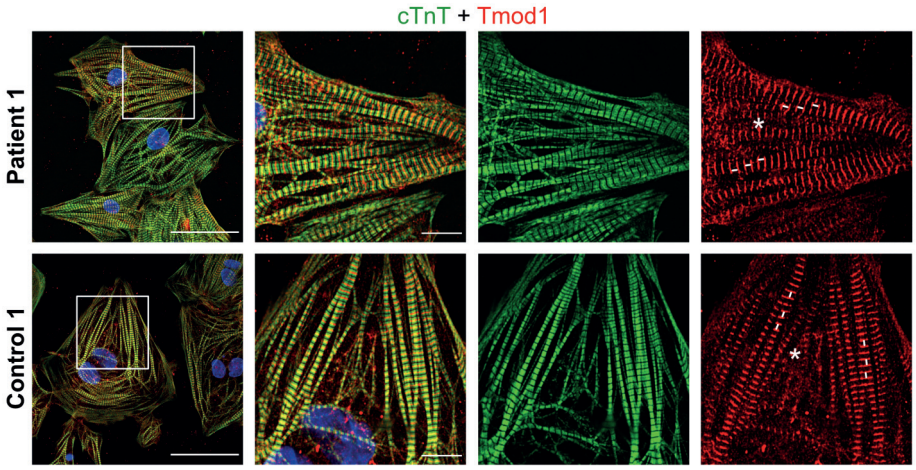


Figure 12. Tmod1 localization in iPSC-CMs from Patient 1 and Control 1. Asterisk: extra-sarcomeric Tmod1. The white sarcomeric lines are drawn in control cells at the same length as in the patient. Scale bars: 50 μm and 10 μm.

A fraction of Tmod1 is found in the cytoplasm (Figure 12), and the cytoplasmic amount seems in general to be inversely related to the degree of sarcomeric development. The heterogeneity of cardiomyocyte differentiation makes subtle comparisons challenging. Occasionally, slightly smaller sarcomeres in the patient were observed, which could constitute a suggestive phenotype (Figure 11B).

5.3.6. *TMOD1* R189W MUTATION CAUSES ISOLATED CARDIOMYOPATHY WITHOUT SKELETAL MUSCLE OR OTHER ORGAN INVOLVEMENT

Tropomodulins (Tmods) are part of a conserved family of proteins, which bind tropomyosin-coated actin filaments at their pointed (slow-growing) end. Tmods promote actin filament stability by regulating actin monomers' addition and dissociation and by enhancing tropomyosin binding. Tmod1 has been detected in mammalian striated muscle, red blood cells, ocular lens fiber cells, and neurons. From the family members, Tmod2 is neuronal-specific; Tmod3 is widely expressed, while Tmod4 is restricted to skeletal muscle (Yamashiro et al, 2012). Leiomodins (Lmods), part of the same family, are structurally and functionally similar to Tmods, but also display important sequence and structure differences, ultimately reflected in their distinct biochemical activities (Boczkowska et al, 2015). Lmod1 is enriched in smooth muscle, while Lmod2 and Lmod3 are expressed in both cardiac and skeletal muscles. There is a preferential co-expression of Tmod4-Lmod3 and Tmod1-Lmod2, especially in skeletal muscle, explained by a gene regulatory network. The current view on these family members is that Lmods nucleate actin filaments, which are subsequently capped and stabilized by Tmods during sarcomere assembly, turnover, and repair (Fowler and Dominguez, 2017).

In skeletal muscle, Tmod1 regulates the capping of thin filament pointed end in slow fibers. In fast skeletal muscle, Tmod4 is associated with sarcomeric thin filament, while Tmod1 with membrane skeleton and costameres. A wide range of compensations between tropomodulin family members has been observed in skeletal muscle. Nevertheless, Tmod1 seems to be the key isoform controlling thin filament length in skeletal muscle, as suggested by Tmod4 knockout mice, as well as by Tmod1 RNAi depletion experiments (Gokhin et al, 2015).

In cardiac muscle, Tmod1 is crucial for thin filament capping. Tmod1 knockout mice present several abnormalities related to fetal heart development, vitelline circulation, and hematopoiesis, with lethality around embryonic day ten (Chu et al, 2003; Fritz-Six et al, 2003). Tmod1 expression, specifically in the heart, under an alpha-myosin heavy chain promoter, rescued lethality and indicated that the knockouts' primary defect is in the myocardium (McKeown et al, 2008). More specifically, in the absence of Tmod1, the cardiomyocyte myofibrils did not become striated, and gaps in F-actin staining were not observed (Fritz-Six et al, 2003). The large compensations and isoform dynamics described for skeletal muscle do not apply to the same extent in the heart. The severe mouse heart phenotype in the absence of Tmod1, together with fewer possibilities for compensation compared with skeletal muscle, support our finding of partially dysfunctional Tmod1 underlying the patients' cardiomyopathies, without the presence of skeletal muscle myopathy.

Tmod1 also plays an important role in the formation of enucleosome during red blood cell maturation, as well as in the architecture of the spectrin-based membrane skeleton found in mature erythrocytes and ocular lens fiber cells (Nowak et al, 2017; Yamashiro et al, 2012). Tmod1 -/- mouse presents spherocytic elliptocytosis (red blood cells surface loss and fragility) (Moyer et al, 2010) and morphological abnormalities of lens fiber cells, although the latter without strong functional consequences as the mice did not develop cataract (Nowak et al, 2009). Our patients did not present hematological or ocular symptoms, supporting further the idea that Tmod1^{R189W} is a partially functional protein.

The compensations and dynamics between Tmod and Lmod isoforms, and their interaction with tropomyosin ultimately determine thin filament length, actomyosin cross-bridge formation, and force generation. The K15N mutation in the TPM1 (Tropomyosin 1) has been described in familial DCM (Hershberger et al, 2010) and was found to negatively affect the binding properties of Tmod1 and Lmod2 with tropomyosin-coated actin filaments and pointed end dynamics (Colpan et al, 2016, 2017), highlighting once more the close relationship between thin filament proteins. Furthermore, an *LMOD2* homozygous nonsense mutation was recently described to cause neonatal DCM (Ahrens-Nicklas et al, 2019). In this framework, we can understand that mutations in any of these thin filament-associated proteins can alter F-actin dynamics and may present in the form of CMP, skeletal muscle myopathy, or in the form of compensatory changes through related proteins.

Although the exact mechanism by which our patients' variant affects the capping at the pointed end remains still open, mutants of the C-terminal region containing the LRR domain exhibited reduced capping activity (Rao et al, 2014b).

Based on the high conservation of the variant amino acid in species, absence of homozygotes in population databases, structural modification predicted to affect the local folding, and the propensity to form aggregates when linked with GFP, we judged *TMOD1* to be a novel gene causing isolated childhood CMP, without skeletal muscle or other organ involvement.

Given the crucial role of actin in cellular architecture and contractility plus its interaction with many proteins involved in assembly/disassembly, anchoring, capping, and turnover of the filaments, it is not surprising that mutations in actin-binding partners cause human disorders. In this study, we add Tmod1, a regulator of actin filament length in cardiomyocytes, to the genetic causes of CMPs that originate from thin filament dysfunction.

6. CONCLUSIONS AND FUTURE PROSPECTS

The studies presented in this thesis are based on genetic discoveries made with NGS. Selected findings were further investigated by functional assays, which extended to patient-derived iPSC-CMs.

Our first study used WES to identify a novel mutation in *HIKESHI*, causing hypomyelinating leukoencephalopathy accompanied by severe deterioration of symptoms during a febrile episode and a malignant ovarian tumor. Previously, six patients from three families of AJ ancestry were described with a founder disease-causing variant in *HIKESHI*. The involvement of Hikeshi in disease has linked its function of heat shock-induced transport of HSP70s into the nucleus with leukoencephalopathy (Edvardson et al, 2016). We further showed that Hikeshi disease mechanisms involve both nuclear chaperone and ER functions. HSP70s play crucial roles in preventing protein aggregation in the CNS, but they also contribute to the disaggregation of formed amyloid fibrils (Wentink et al, 2020). Consequently, the role of Hikeshi in age-associated neurodegenerative disorders remains to be unraveled.

Valuable insights into the critical pathways required for myelin formation and maintenance can be drawn from the genetics of leukoencephalopathies and leukodystrophies (Stadelmann et al, 2019). Besides Hikeshi, other white matter disorders with significant deterioration following febrile illnesses are caused by mutations in eIF2B complex required for protein synthesis [vanishing white matter disorder, reviewed in (Moon and Parker, 2018; Pavitt, 2018)], and nucleoporin NUP214 [acute febrile encephalopathy, (Fichtman et al, 2019)]. These genetic causes strengthen the idea that heat shock response is a *bona fide* pathogenic mechanism and highlight its connections with protein homeostasis and with cytoplasmic-nuclear transport.

In this context, further studies should aim to characterize in detail the Hikeshi-mediated HSP70s nuclear import pathway in myelin disorders, the relationship between ER redox state and the heat shock response, and their possible connection with unfolded protein response (Clayton and Popko, 2016; Lin and Stone, 2020).

Our second project used NGS targeted sequencing panels and WES for the genetic characterization of severe childhood cardiomyopathies in a countrywide cohort. The genetic basis of early manifesting CMPs has not been systematically characterized in Finland. We expected to identify major recessive ancestral mutations enriched in the country, as it is typical for other monogenic disorders (Norio, 2003c). In adult cardiomyopathy patients from Finland, founder heterozygous mutations have been identified in the sarcomeric genes *MYBPC3*, *MYH7*, and *TPM1* (Jääskeläinen et al, 2019), as

well as in *JPH2* involved in calcium handling (Koskenvuo et al, 2018). This was, however, not the case in our study. Instead, we found a remarkably heterogeneous genetic background, each family presenting a different mutation, almost half of them occurring *de novo*. A positive genetic diagnosis was reached in 39% of patients. The significant number of patients that remained unresolved indicates that new disease genes are still to be discovered and point to the importance of whole-genome analysis. Whether those gene defects allow enrichment in the population remains to be seen.

The high genetic heterogeneity observed in early-onset cardiomyopathies underscores the vital importance of many cellular processes and the coordinated action of a myriad of proteins to drive efficient and sustained heart muscle contraction. NGS tools enable a better understanding of cardiomyopathies' genetic causes, which can open possibilities for interventions and offer individualized treatment choices, especially important in major invasive procedures such as cardiac transplants.

Our findings have important implications for a global understanding of childhood cardiomyopathy genetics. They indicate the benefits of NGS as a first-line diagnosis in children considered for a heart transplant, and they trace the molecular pathways affected by disease genes. KidCMP cohort demonstrated the primary progressive nature of specific gene defects, whereas others spontaneously stabilized, despite all manifesting as severe disorders, and most were considered for heart transplantation. Consequently, genotype information provides tools for robust assessment of the natural history of the disease, enabling individualized care and prioritization of patients for cardiac transplantation.

To illustrate this point, the cardiac findings in mitochondrial cardiomyopathies (*TAZ*, *MRPL44*, and *DNAJC19*) stabilized spontaneously in the minority of patients who survived through the early years (Carroll et al, 2013; D'Adamo et al, 1997; Ojala et al, 2012b). A similar finding in our lab was for *TSFM* defects, causing mitochondrial CMP with neurological involvement (Ahola et al, 2014). The potential of spontaneous partial recovery of CMP emphasizes an intensive, conservative treatment strategy to support these patients' heart function through their infantile years.

In our cohort, 60% of patients are still without a molecular diagnosis. There are several factors to consider in our attempts to uncover the 'missing heritability' in rare disorders. First, rare disorders may display both phenotypic heterogeneity (strikingly different phenotypes associated with different variants in the same gene) and genetic heterogeneity (variation in distinct genes leading to similar phenotypes) (Wright et al, 2018). Considering this aspect may be crucial for a successful approach to diagnosis. The success may also depend on the type of test used, WGS trio sequencing proving superior in the diagnostic rate and the time to reach the diagnosis (Fitzgerald et al, 2015).

Although NGS allowed 'reverse phenotyping' in many patients, deep phenotyping becomes essential for diagnosing difficult cases. The concept of

deep phenotyping includes the use of terms defined in the Human Phenotype Ontology database (Köhler et al, 2017; Robinson and Mundlos, 2010), terms representing a standardized description of abnormalities associated with disorders (Köhler et al, 2019). Matchmaking tools, such as GeneMatcher (Sobreira et al, 2015), PhenomeCentral (Buske et al, 2015), and MatchmakerExchange (Philippakis et al, 2015), also help in matching patients with similar rare phenotypes and with the same candidate gene, contributing to positive findings.

Another consideration relates to the mutation spectrum. Presently, SNVs are the type of variation most accurately represented in public databases. There is a need for better methods to detect, compile, and compare indels and SVs across populations. Third-generation sequencing, coupled with bioinformatic advancements, can overcome these challenges. To accurately discover SVs, including repeat expansions, there is a need to develop novel bioinformatic tools (Antaki et al, 2018; Jakubosky et al, 2020). In this direction, machine-learning algorithms are created to merge calls from different SV-calling platforms (Becker et al, 2018). Another novel tool is Expansion Hunter, which uses PCR-free WGS short reads to identify long repeat expansions (Dolzhenko et al, 2017).

Somatic variants are more often discovered in patients resistant to conventional NGS-based diagnostic approaches. Increasing the depth of sequencing and choosing the affected tissue as starting material whenever possible will increase the detection of this subtype of variation.

Moreover, multilocus genetic inheritance can be considered in patients difficult to diagnose. This situation may include atypical clinical presentation due to two or more overlapping genetic conditions, disorders with digenic inheritance (DIDA database of digenic disorders, http://dida.ibsquare.be/) (Gazzo et al, 2016), and oligogenic inheritance that may occur even in disorders classically considered monogenic, such as LongQT syndrome. Genetic modifiers may alleviate or amplify the primary causative variant effect, but their influence is difficult to evaluate in small cohorts. Large-scale projects that combine genotype and phenotype information should shed light on modifiers, penetrance, and expressivity in rare disorders.

Furthermore, unknown causative genes and variants in noncoding regions with unknown effects are important causes of missing diagnoses. Assigning function to poorly characterized protein- and RNA-genes, as well as to functional noncoding elements of the genome, will improve the interpretation of variants in the near future (Ward and Kellis, 2012).

The reference genome and variome resources are key for the accurate discovery of disease variants in rare disorders. The reference genome GRCh38 is of high quality in most regards (Schneider et al, 2017). However, it still has gaps (Miga, 2015) and errors at repetitive and structurally diverse regions (Ameur et al, 2018), as well as many ultra-rare variants (Magi et al, 2015). Even more important for genetic studies, the reference genome does not capture the diversity of our collective human genome of eight billion

people because it is a composition of sequences that originate from only a small number of individuals. The principle of capturing the diversity is of high importance because if a sequence is not represented in the reference, or if it is too divergent from the reference, most algorithms will not detect it. Ongoing efforts are undertaken to append alternative loci to the reference (Church et al, 2015), but presently most tools and pipelines do not make use of them. A future pan-genome resource has the ambition to capture all relatively common sequences and alleles across human populations, which will be represented as a graph-based data structure. Many initiatives in recent years aim to build reference variant databases for different populations. A few examples include the Genome Russia project (Oleksyk et al, 2015a, 2015b), the Ashkenazi-Jewish reference panel (Lencz et al, 2018), and the Silent Genomes Project (www.bcchr.ca/silent-genomes-project) of the Canadian indigenous population.

Incorporating multi-omics layers (from affected tissue or blood) can also bring valuable information. For example, RNA sequencing information enables detecting biased expression levels or splicing effects that are difficult to predict from WES or WGS data (Deelen et al, 2019). Further, proteomics and metabolomics data will shape a better understanding of the pathogenic processes, which can help to identify the responsible gene. Altogether, these combined approaches will promote the discovery of the 'missing heritability' in rare disorders.

Our third project identified *TMOD1* as a novel genetic cause for childhood cardiomyopathies in three patients from two unrelated families, by using WES. Tmod1 binds and stabilizes tropomyosin-coated actin filaments at their pointed end (Weber et al, 1994), regulating the length and stability of the thin filament in cardiomyocytes.

Actin is the most abundant protein in mammalian cells, and the importance of actin-driven processes highlights that the control of actin filament dynamics is crucial (Svitkina, 2018). This is also suggested by compensatory mechanisms shown in different models where Tmod family members have been altered. The manifestation age of *TMOD1* defect was 9-12 years, suggesting that a less severe mutation can have a later onset and indicating *TMOD1* screening also in adult patients. Future studies should establish the biochemical properties of the Tmod1^{R189W} pathogenic variant, how it affects the length of thin filaments in mature cardiomyocytes and the generation of force.

We used iPSC-CM disease modeling, with only suggestive findings in our patient's cells. This brings into the spotlight a discussion regarding the challenges of iPSC-based disease modeling and of strategies to improve cardiomyocyte differentiation outcomes (Parrotta et al, 2019). One challenge relates to adequate controls, as iPSCs exhibit a large heterogeneity upon differentiation (Mirauta et al, 2020) that can confound interpretation. Isogenic cell lines created by CRISPR editing approaches circumvent this issue by providing a homogeneous genetic background (DeBoever et al, 2017;

Kilpinen et al, 2017). However, editing strategies introduce the possibility of off-target effects and the need for NGS screening of the cell lines. Furthermore, stringent criteria remain to be defined for completely reprogrammed iPSC lines, as some retain epigenetic memory, which influences their capacity to differentiate.

Modeling late-onset disorders has been less successful due to recurrent immaturity of derived cells and the lack of native environmental cues (Doss and Sachinidis, 2019). Another confounding effect is the heterogeneous mixture of cells obtained upon differentiation. In cardiomyocyte differentiation, atrial, ventricular, and nodal subtypes co-occur in the same culture, together with other types of cells (Duelen and Sampaolesi, 2017). In our patients with *TMOD1* mutation, the heart atria were especially affected, suggesting that the selection of atrial cardiomyocytes could constitute a future strategy for dissecting the Tmod1 disease mechanism in iPSC-CMs.

Moreover, iPSCs acquire genomic instability (Merkle et al, 2017) with increased passage numbers, manifesting as chromosomal abnormalities, CNVs, and loss of heterozygosity (Henry et al, 2019; Steichen et al, 2019). The genome integrity maintenance is especially relevant for iPSC-CMs because a good differentiation requires a higher passage (Burridge et al, 2015). Engineering cells with one of the 'suicide gene therapy' approaches could represent a future solution to eliminate abnormal or unwanted cells from a mixed culture. An example is the caspase 9 suicide system, inducible by the AP1903 drug, and leading to rapid apoptosis of cells expressing the construct (Zhou et al, 2015).

As efforts to decipher the genetic basis of rare disorders progress towards completion, functional studies remain important to understand the mechanisms by which disruptive variation leads to disease phenotypes. This underlines the importance of tools to model disorders, like iPSC-CMs, with isogenic controls obtained by editing technologies. From the perspective of personalized medicine, we envisage that genotype-phenotype correlations will provide a framework to evaluate disease progression and will enable individualized treatment options.

ACKNOWLEDGMENTS

This thesis work was carried out in Biomedicum-Helsinki, Research Programs Unit, and the Centre of Excellence in Research on Mitochondrial Biogenesis and Disease. I express my gratitude for financial support to the Doctoral Programme in Biomedicine, Finnish Cultural Foundation, Otto Malm, Alfred Kordelin, and Maud Kuistila foundations. I thank Professor Katriina Aalto-Setälä and Professor Johanna Kuusisto for reviewing my thesis, and Professor Holger Prokisch for its public discussion.

My entire admiration for the optimism and exuberance for science goes to Professor Anu Wartiovaara. Many thanks for the opportunity to learn and do so many new and exciting things! Dr Christopher Carroll, my second supervisor, taught me most lab techniques and brought valuable contributions to my manuscripts. Thank you, Chris, for everything!

During the years, we collaborated with many clinicians. I warmly thank all my co-authors and collaborators for their involvement in our projects. Docent Tiina Ojala has been the key person in collecting the KidCMP cohort. Thank you so much! I also thank Juha Koskenvuo from Blueprint Genetics for their contribution to our studies, and I express my appreciation for their dedication to medical genetics.

All members, past and present, of the Wartiovaara research group are thanked for the company and help in the lab. You are experts in your field, and I know that the many hours in the lab will pay off! Big thanks to our lab technicians Anu H, Tuula, Markus, Babette, and program secretaries Maija and Kaisa for all their help. We had many discussions on our exome projects throughout the years, with Pirjo, Virginia, Simo, Risto, Juho, and Anna. Thank you, all! Liliya introduced me to the fascinating world of computational molecular modeling, for which I am very grateful! Anastasiia is warmly thanked for the lively discussions in our office. Moreover, what life would be without the cardiomyopathy team, Nahid and Swagat, thank you so much for discussions and company during long hours in the cell culture room. Similarly, many thanks to my students, Helena and Hanna, for learning and growing together.

The 'Antti Korpin tie' student house was a place where I met many extraordinary people, each of them with a special talent. Among them, my close friends: Riike, Iryna, Mattia, and Fernando. Finally, I warmly thank my family from Romania for their neverending support. I am very fortunate to have you!

Helsinki, August 2020
Catalina Vasilescu

REFERENCES

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7, 248–249.

Ahola, S., Isohanni, P., Euro, L., Brilhante, V., Palotie, A., Pihko, H., Lönnqvist, T., Lehtonen, T., Laine, J., Tyynismaa, H., et al (2014) Mitochondrial EFTs defects in juvenile-onset Leigh disease, ataxia, neuropathy, and optic atrophy. Neurology 83, 743–751.

Ahrens-Nicklas, R.C., Pappas, C.T., Farman, G.P., Mayfield, R.M., Larrinaga, T.M., Medne, L., Ritter, A., Krantz, I.D., Murali, C., Lin, K.Y., et al (2019) Disruption of cardiac thin filament assembly arising from a mutation in LMOD2: A novel mechanism of neonatal dilated cardiomyopathy. Sci Adv 5.

Albakri, A. (2019) Metabolic cardiomyopathy: A review and pooled analysis of pathophysiology, diagnosis and clinical management. Med Clin Arch 3.

Alfares, A., Aloraini, T., Subaie, L.A., Alissa, A., Qudsi, A.A., Alahmad, A., Mutairi, F.A., Alswaid, A., Alothaim, A., Eyaid, W., et al (2018) Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. Genetics in Medicine 20, 1328–1333.

Alvarez Jorge A., Orav E. John, Wilkinson James D., Fleming Lora E., Lee David J., Sleeper Lynn A., Rusconi Paolo G., Colan Steven D., Hsu Daphne T., Canter Charles E., et al (2011) Competing Risks for Death and Cardiac Transplantation in Children With Dilated Cardiomyopathy. Circulation 124, 814–823.

Ameur, A., Che, H., Martin, M., Bunikis, I., Dahlberg, J., Höijer, I., Häggqvist, S., Vezzi, F., Nordlund, J., Olason, P., et al (2018) De Novo Assembly of Two Swedish Genomes Reveals Missing Segments from the Human GRCh38 Reference and Improves Variant Calling of Population-Scale Sequencing Data. Genes 9, 486.

Andreasen, C., Nielsen, J.B., Refsgaard, L., Holst, A.G., Christensen, A.H., Andreasen, L., Sajadieh, A., Haunsø, S., Svendsen, J.H., and Olesen, M.S. (2013) New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. Eur. J. Hum. Genet. 21, 918–928.

Antaki, D., Brandler, W.M., and Sebat, J. (2018) SV2: accurate structural variation genotyping and de novo mutation detection from whole genomes. Bioinformatics 34, 1774–1777.

Anttonen, A.-K., Mahjneh, I., Hämäläinen, R.H., Lagier-Tourenne, C., Kopra, O., Waris, L., Anttonen, M., Joensuu, T., Kalimo, H., Paetau, A., et al (2005) The gene disrupted in Marinesco-Sjögren syndrome encodes SIL1, an HSPA5 cochaperone. Nature Genetics 37, 1309–1311.

Arola, A., Jokinen, E., Ruuskanen, O., Saraste, M., Pesonen, E., Kuusela,

A.L., Tikanoja, T., Paavilainen, T., and Simell, O. (1997) Epidemiology of idiopathic cardiomyopathies in children and adolescents. A nationwide study in Finland. Am. J. Epidemiol. 146, 385–393.

Audano, P.A., Sulovari, A., Graves-Lindsay, T.A., Cantsilieris, S., Sorensen, M., Welch, A.E., Dougherty, M.L., Nelson, B.J., Shah, A., Dutcher, S.K., et al (2019) Characterizing the Major Structural Variant Alleles of the Human Genome. Cell 176, 663-675.e19.

Bacchelli, C., and Williams, H.J. (2016) Opportunities and technical challenges in next-generation sequencing for diagnosis of rare pediatric diseases. Expert Rev. Mol. Diagn. 16, 1073–1082.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. (2002) Recent segmental duplications in the human genome. Science 297, 1003–1007.

Balakier, H., and Pedersen, R.A. (1982) Allocation of cells to inner cell mass and trophectoderm lineages in preimplantation mouse embryos. Developmental Biology 90, 352–362.

Bamshad, M.J., Nickerson, D.A., and Chong, J.X. (2019) Mendelian Gene Discovery: Fast and Furious with No End in Sight. The American Journal of Human Genetics 105, 448–455.

Barnes, D.E., and Lindahl, T. (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. Annu. Rev. Genet. 38, 445–476.

Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233.

Bates, M.G.D., Bourke, J.P., Giordano, C., d'Amati, G., Turnbull, D.M., and Taylor, R.W. (2012) Cardiac involvement in mitochondrial DNA disease: clinical spectrum, diagnosis, and management. Eur Heart J 33, 3023–3033.

Becker, T., Lee, W.-P., Leone, J., Zhu, Q., Zhang, C., Liu, S., Sargent, J., Shanker, K., Mil-homens, A., Cerveira, E., et al (2018) FusorSV: an algorithm for optimally combining data from multiple structural variation detection methods. Genome Biology 19, 38.

Bharucha Tara, Lee Katherine J., Daubeney Piers E.F., Nugent Alan W., Turner Christian, Sholler Gary F., Robertson Terry, Justo Robert, Ramsay Jim, Carlin John B., et al (2015) Sudden Death in Childhood Cardiomyopathy. Journal of the American College of Cardiology 65, 2302–2310.

Biscotti, M.A., Canapa, A., Forconi, M., Olmo, E., and Barucca, M. (2015) Transcription of tandemly repetitive DNA: functional roles. Chromosome Res. 23, 463–477.

Blanpain, C., Horsley, V., and Fuchs, E. (2007) Epithelial Stem Cells: Turning over New Leaves. Cell 128, 445–458.

Blumen, S.C., Astord, S., Robin, V., Vignaud, L., Toumi, N., Cieslik, A., Achiron, A., Carasso, R.L., Gurevich, M., Braverman, I., et al (2012) A rare recessive distal hereditary motor neuropathy with HSJ1 chaperone mutation.

Annals of Neurology 71, 509–519.

Boczkowska, M., Rebowski, G., Kremneva, E., Lappalainen, P., and Dominguez, R. (2015) How Leiomodin and Tropomodulin use a common fold for different actin assembly functions. Nat Commun 6, 8314.

Boeckmann, B., Blatter, M.-C., Famiglietti, L., Hinz, U., Lane, L., Roechert, B., and Bairoch, A. (2005) Protein variety and functional diversity: Swiss-Prot annotation in its biological context. 18.

Bonkowsky, J.L., Nelson, C., Kingston, J.L., Filloux, F.M., Mundorff, M.B., and Srivastava, R. (2010) The burden of inherited leukodystrophies in children. Neurology 75, 718–725.

Boogerd Cornelis J., Zhu Xiaoming, Aneas Ivy, Sakabe Noboru, Zhang Lunfeng, Sobreira Debora R., Montefiori Lindsey, Bogomolovas Julius, Joslin Amelia C., Zhou Bin, et al (2018) Tbx20 Is Required in Mid-Gestation Cardiomyocytes and Plays a Central Role in Atrial Development. Circulation Research 123, 428–442.

Boycott, K.M., Hartley, T., Biesecker, L.G., Gibbs, R.A., Innes, A.M., Riess, O., Belmont, J., Dunwoodie, S.L., Jojic, N., Lassmann, T., et al (2019) A Diagnosis for All Rare Genetic Diseases: The Horizon and the Next Frontiers. Cell 177, 32–37.

Brodehl, A., Ebbinghaus, H., Deutsch, M.-A., Gummert, J., Gärtner, A., Ratnavadivel, S., and Milting, H. (2019) Human Induced Pluripotent Stem-Cell-Derived Cardiomyocytes as Models for Genetic Cardiomyopathies. Int J Mol Sci 20.

Buikema, J.W., and Wu, S.M. (2017) Untangling the Biology of Genetic Cardiomyopathies with Pluripotent Stem Cell Disease Models. Curr Cardiol Rep 19, 30.

Burgess, R. (2016) Stem Cells: A Short Course (John Wiley & Sons)

Burridge, P.W., Holmström, A., and Wu, J.C. (2015) Chemically Defined Culture and Cardiomyocyte Differentiation of Human Pluripotent Stem Cells. Curr Protoc Hum Genet 87, 21.3.1-21.3.15.

Buske, O.J., Girdea, M., Dumitriu, S., Gallinger, B., Hartley, T., Trang, H., Misyura, A., Friedman, T., Beaulieu, C., Bone, W.P., et al (2015) PhenomeCentral: A Portal for Phenotypic and Genotypic Matchmaking of Patients with Rare Genetic Diseases. Human Mutation 36, 931–940.

Carrel, L., and Willard, H.F. (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434, 400–404.

Carroll, C.J., Isohanni, P., Pöyhönen, R., Euro, L., Richter, U., Brilhante, V., Götz, A., Lahtinen, T., Paetau, A., Pihko, H., et al (2013) Whole-exome sequencing identifies a mutation in the mitochondrial ribosome protein MRPL44 to underlie mitochondrial infantile cardiomyopathy. Journal of Medical Genetics 50, 151–159.

Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., and Kondrashov, F.A. (2002) Selection for short introns in highly expressed genes. Nature Genetics 31, 415–418.

Caudle, K.E., Klein, T.E., Hoffman, J.M., Muller, D.J., Whirl-Carrillo, M., Gong, L., McDonagh, E.M., Sangkuhl, K., Thorn, C.F., Schwab, M., et al (2014) Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. Curr Drug Metab 15, 209–217.

Chagastelles, P.C., and Nardi, N.B. (2011) Biology of stem cells: an overview. Kidney International Supplements 1, 63–67.

Chang, S.C., and Brown, C.J. (2010) Identification of regulatory elements flanking human XIST reveals species differences. BMC Mol. Biol. 11, 20.

Chatterjee, N., and Walker, G.C. (2017) Mechanisms of DNA damage, repair and mutagenesis. Environ Mol Mutagen 58, 235–263.

Chen, I.Y., Matsa, E., and Wu, J.C. (2016) Induced pluripotent stem cells: at the heart of cardiovascular precision medicine. Nature Reviews Cardiology 13, 333–349.

Chen, L., Zhou, W., Zhang, L., and Zhang, F. (2014) Genome Architecture and Its Roles in Human Copy Number Variation. Genomics Inform 12, 136.

Chheda, H., Palta, P., Pirinen, M., McCarthy, S., Walter, K., Koskinen, S., Salomaa, V., Daly, M., Durbin, R., Palotie, A., et al (2017) Whole-genome view of the consequences of a population bottleneck using 2926 genome sequences from Finland and United Kingdom. European Journal of Human Genetics 25, 477–484.

Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., et al (2015) The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am J Hum Genet 97, 199–215.

Christenhusz, G.M., Devriendt, K., and Dierickx, K. (2013) To tell or not to tell? A systematic review of ethical reflections on incidental findings arising in genetics contexts. European Journal of Human Genetics 21, 248–255.

Chu, X., Chen, J., Reedy, M.C., Vera, C., Sung, K.-L.P., and Sung, L.A. (2003) E-Tmod capping of actin filaments at the slow-growing end is required to establish mouse embryonic circulation. Am. J. Physiol. Heart Circ. Physiol. 284, H1827-1838.

Chung, C.C.Y., Leung, G.K.C., Mak, C.C.Y., Fung, J.L.F., Lee, M., Pei, S.L.C., Yu, M.H.C., Hui, V.C.C., Chan, J.C.K., Chau, J.F.T., et al (2020) Rapid whole-exome sequencing facilitates precision medicine in paediatric rare disease patients and reduces healthcare costs. The Lancet Regional Health – Western Pacific 1.

Church, D.M., Schneider, V.A., Steinberg, K.M., Schatz, M.C., Quinlan, A.R., Chin, C.-S., Kitts, P.A., Aken, B., Marth, G.T., Hoffman, M.M., et al (2015) Extending reference assembly models. Genome Biol. 16, 13.

Claussnitzer, M., Cho, J.H., Collins, R., Cox, N.J., Dermitzakis, E.T., Hurles, M.E., Kathiresan, S., Kenny, E.E., Lindgren, C.M., MacArthur, D.G., et al (2020) A brief history of human disease genetics. Nature 577, 179–189.

Clayton, B.L.L., and Popko, B. (2016) Endoplasmic reticulum stress and the

unfolded protein response in disorders of myelinating glia. Brain Res. 1648, 594–602.

Cohen, C.C.H., Popovic, M.A., Klooster, J., Weil, M.-T., Möbius, W., Nave, K.-A., and Kole, M.H.P. (2020) Saltatory Conduction along Myelinated Axons Involves a Periaxonal Nanocircuit. Cell 180, 311-322.e15.

Colpan, M., Tolkatchev, D., Grover, S., Helms, G.L., Cort, J.R., Moroz, N., and Kostyukova, A. (2016) Localization of the Binding Interface between Leiomodin-2 and α-Tropomyosin. Biochim Biophys Acta 1864, 523–530.

Colpan, M., Ly, T., Grover, S., Tolkatchev, D., and Kostyukova, A.S. (2017) The cardiomyopathy-associated K15N mutation in tropomyosin alters actin filament pointed end dynamics. Archives of Biochemistry and Biophysics 630, 18–26.

Conrad, B., and Antonarakis, S.E. (2007) Gene duplication: a drive for phenotypic diversity and cause of human disease. Annu Rev Genomics Hum Genet 8, 17–35.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al (2010) Origins and functional impact of copy number variation in the human genome. Nature 464, 704–712.

Correia, C., Koshkin, A., Duarte, P., Hu, D., Teixeira, A., Domian, I., Serra, M., and Alves, P.M. (2017) Distinct carbon sources affect structural and functional maturation of cardiomyocytes derived from human pluripotent stem cells. Sci Rep 7, 8590.

Couvillion, M.T., Soto, I.C., Shipkovenska, G., and Churchman, L.S. (2016) Synchronized mitochondrial and cytosolic translation programs. Nature 533, 499–503.

Cox, G.F., Sleeper, L.A., Lowe, A.M., Towbin, J.A., Colan, S.D., Orav, E.J., Lurie, P.R., Messere, J.E., Wilkinson, J.D., and Lipshultz, S.E. (2006) Factors associated with establishing a causal diagnosis for children with cardiomyopathy. Pediatrics 118, 1519–1531.

Crotti, L., Johnson, C.N., Graf, E., De Ferrari, G.M., Cuneo, B.F., Ovadia, M., Papagiannis, J., Feldkamp, M.D., Rathi, S.G., Kunic, J.D., et al (2013) Calmodulin mutations associated with recurrent cardiac arrest in infants. Circulation 127, 1009–1017.

D'Adamo, P., Fassone, L., Gedeon, A., Janssen, E.A., Bione, S., Bolhuis, P.A., Barth, P.G., Wilson, M., Haan, E., Orstavik, K.H., et al (1997) The X-linked gene G4.5 is responsible for different infantile dilated cardiomyopathies. Am J Hum Genet 61, 862–867.

Dallabona, C., Diodato, D., Kevelam, S.H., Haack, T.B., Wong, L.-J., Salomons, G.S., Baruffini, E., Melchionda, L., Mariotti, C., Strom, T.M., et al (2014) Novel (ovario) leukodystrophy related to AARS2 mutations. Neurology 82, 2063–2071.

Davey, K.M., Parboosingh, J.S., McLeod, D.R., Chan, A., Casey, R., Ferreira, P., Snyder, F.F., Bridge, P.J., and Bernier, F.P. (2006) Mutation of DNAJC19,

a human homologue of yeast inner mitochondrial membrane co-chaperones, causes DCMA syndrome, a novel autosomal recessive Barth syndrome-like condition. J. Med. Genet. 43, 385–393.

De Luca, M., Aiuti, A., Cossu, G., Parmar, M., Pellegrini, G., and Robey, P.G. (2019) Advances in stem cell research and therapeutic development. Nature Cell Biology 21, 801–811.

DeBoever, C., Li, H., Jakubosky, D., Benaglio, P., Reyna, J., Olson, K.M., Huang, H., Biggs, W., Sandoval, E., D'Antonio, M., et al (2017) Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene Expression in Human Induced Pluripotent Stem Cells. Cell Stem Cell 20, 533-546.e7.

Deelen, P., van Dam, S., Herkert, J.C., Karjalainen, J.M., Brugge, H., Abbott, K.M., van Diemen, C.C., van der Zwaag, P.A., Gerkes, E.H., Zonneveld-Huijssoon, E., et al (2019) Improving the diagnostic yield of exome-sequencing by predicting gene–phenotype associations using large-scale gene expression analysis. Nat Commun 10, 2837.

Deng, X., Berletch, J.B., Nguyen, D.K., and Disteche, C.M. (2014) X chromosome regulation: diverse patterns in development, tissues and disease. Nat. Rev. Genet. 15, 367–378.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

Dhandapany, P.S., Razzaque, M.A., Muthusami, U., Kunnoth, S., Edwards, J.J., Mulero-Navarro, S., Riess, I., Pardo, S., Sheng, J., Rani, D.S., et al (2014) RAF1 mutations in childhood-onset dilated cardiomyopathy. Nature Genetics 46, 635–639.

Díez Villanueva, P., Sanz-Ruiz, R., Núñez García, A., Fernández Santos, M.E., Sánchez, P.L., and Fernández-Avilés, F. (2012) Functional Multipotency of Stem Cells: What Do We Need from Them in the Heart? (Hindawi)

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al (2012) Landscape of transcription in human cells. Nature 489, 101–108.

Dolzhenko, E., Vugt, J.J.F.A. van, Shaw, R.J., Bekritsky, M.A., Blitterswijk, M. van, Narzisi, G., Ajay, S.S., Rajan, V., Lajoie, B.R., Johnson, N.H., et al (2017) Detection of long repeat expansions from PCR-free whole-genome sequence data. Genome Res. 27, 1895–1903.

Doss, M.X., and Sachinidis, A. (2019) Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications. Cells 8.

Duelen, R., and Sampaolesi, M. (2017) Stem Cell Technology in Cardiac Regeneration: A Pluripotent Stem Cell Promise. EBioMedicine 16, 30–40.

von der Ecken, J., Heissler, S.M., Pathan-Chhatbar, S., Manstein, D.J., and

Raunser, S. (2016) Cryo-EM structure of a human cytoplasmic actomyosin complex at near-atomic resolution. Nature 534, 724–728.

Edvardson, S., Cinnamon, Y., Ta-Shma, A., Shaag, A., Yim, Y.-I., Zenvirt, S., Jalas, C., Lesage, S., Brice, A., Taraboulos, A., et al (2012) A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. PLoS ONE 7, e36458.

Edvardson, S., Kose, S., Jalas, C., Fattal-Valevski, A., Watanabe, A., Ogawa, Y., Mamada, H., Fedick, A.M., Ben-Shachar, S., Treff, N.R., et al (2016) Leukoencephalopathy and early death associated with an Ashkenazi-Jewish founder mutation in the Hikeshi gene. J. Med. Genet. 53, 132–137.

El-Hattab, A.W., and Scaglia, F. (2016) Mitochondrial Cardiomyopathies. Front. Cardiovasc. Med. 3.

Elliott, P., Andersson, B., Arbustini, E., Bilinska, Z., Cecchi, F., Charron, P., Dubourg, O., Kühl, U., Maisch, B., McKenna, W.J., et al (2008) Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases. Eur. Heart J. 29, 270–276.

ENCODE Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE) PLoS Biol. 9, e1001046.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

Falkenberg, M., Larsson, N.-G., and Gustafsson, C.M. (2007) DNA replication and transcription in mammalian mitochondria. Annu. Rev. Biochem. 76, 679–699.

Faust, O., Abayev-Avraham, M., Wentink, A.S., Maurer, M., Nillegoda, N.B., London, N., Bukau, B., and Rosenzweig, R. (2020) HSP40 proteins use class-specific regulation to drive HSP70 functional diversity. Nature 1–6.

Fernandez-Marmiesse, A., Gouveia, S., and Couce, M.L. (2018) NGS Technologies as a Turning Point in Rare Disease Research , Diagnosis and Treatment. Curr. Med. Chem. 25, 404–432.

Fichtman, B., Harel, T., Biran, N., Zagairy, F., Applegate, C.D., Salzberg, Y., Gilboa, T., Salah, S., Shaag, A., Simanovsky, N., et al (2019) Pathogenic Variants in NUP214 Cause "Plugged" Nuclear Pore Channels and Acute Febrile Encephalopathy. The American Journal of Human Genetics 105, 48–64.

Finsterer, J. (2008) Histiocytoid Cardiomyopathy: A Mitochondrial Disorder. Clinical Cardiology 31, 225–227.

Fitzgerald, T.W., Gerety, S.S., Jones, W.D., van Kogelenberg, M., King, D.A., McRae, J., Morley, K.I., Parthiban, V., Al-Turki, S., Ambridge, K., et al (2015) Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223–228.

Flaus, A., and Owen-Hughes, T. (2011) Mechanisms for ATP-dependent chromatin remodelling: The means to the end. FEBS J 278, 3579–3595.

Fowler, V.M., and Dominguez, R. (2017) Tropomodulins and Leiomodins: Actin Pointed End Caps and Nucleators in Muscles. Biophys. J. 112, 1742–1760.

Frank, V., Habbig, S., Bartram, M.P., Eisenberger, T., Veenstra-Knol, H.E., Decker, C., Boorsma, R.A.C., Göbel, H., Nürnberg, G., Griessmann, A., et al (2013) Mutations in NEK8 link multiple organ dysplasia with altered Hippo signalling and increased c-MYC expression. Hum. Mol. Genet. 22, 2177–2185.

Fraser, J., Williamson, I., Bickmore, W.A., and Dostie, J. (2015) An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiol. Mol. Biol. Rev. 79, 347–372.

Frésard, L., and Montgomery, S.B. (2018) Diagnosing rare diseases after the exome. Cold Spring Harb Mol Case Stud 4, a003392.

Fritz-Six, K.L., Cox, P.R., Fischer, R.S., Xu, B., Gregorio, C.C., Zoghbi, H.Y., and Fowler, V.M. (2003) Aberrant myofibril assembly in tropomodulin1 null mice leads to aborted heart development and embryonic lethality. J. Cell Biol. 163, 1033–1044.

Fuchs, M., Poirier, D.J., Seguin, S.J., Lambert, H., Carra, S., Charette, S.J., and Landry, J. (2009) Identification of the key structural motifs involved in HspB8/HspB6-Bag3 interaction. Biochem. J. 425, 245–255.

Gammage, P.A., and Frezza, C. (2019) Mitochondrial DNA: the overlooked oncogenome? BMC Biology 17, 53.

Gardner, R.L. (1968) Mouse Chimaeras obtained by the Injection of Cells into the Blastocyst. Nature 220, 596–597.

Gazzo, A.M., Daneels, D., Cilia, E., Bonduelle, M., Abramowicz, M., Van Dooren, S., Smits, G., and Lenaerts, T. (2016) DIDA: A curated and annotated digenic diseases database. Nucleic Acids Res 44, D900–D907.

Germain, P.-L., Ratti, E., and Boem, F. (2014) Junk or functional DNA? ENCODE and the function controversy. Biol Philos 29, 807–831.

Giacomelli, E., Meraviglia, V., Campostrini, G., Cochrane, A., Cao, X., van Helden, R.W.J., Krotenberg Garcia, A., Mircea, M., Kostidis, S., Davis, R.P., et al (2020) Human-iPSC-Derived Cardiac Stromal Cells Enhance Maturation in 3D Cardiac Microtissues and Reveal Non-cardiomyocyte Contributions to Heart Disease. Cell Stem Cell 26, 862-879.e11.

Gilbert-Barness, E. (2004) Metabolic Cardiomyopathy and Conduction System Defects in Children. Ann Clin Lab Sci 34, 15–34.

Gilmore, A.P. (2005) Anoikis. Cell Death & Differentiation 12, 1473–1477.

Gintant Gary, Burridge Paul, Gepstein Lior, Harding Sian, Herron Todd, Hong Charles, Jalife José, Wu Joseph C., and null null (2019) Use of Human Induced Pluripotent Stem Cell–Derived Cardiomyocytes in Preclinical Cancer Drug Cardiotoxicity Testing: A Scientific Statement From the American Heart Association. Circulation Research 125, e75–e92.

Gokhin, D.S., and Fowler, V.M. (2011a) Tropomodulin capping of actin

filaments in striated muscle development and physiology. J. Biomed. Biotechnol. 2011, 103069.

Gokhin, D.S., and Fowler, V.M. (2011b) The sarcoplasmic reticulum: Actin and tropomodulin hit the links. Bioarchitecture 1, 175–179.

Gokhin, D.S., Ochala, J., Domenighetti, A.A., and Fowler, V.M. (2015) Tropomodulin 1 directly controls thin filament length in both wild-type and tropomodulin 4-deficient skeletal muscle. Development 142, 4351–4362.

Goldmann, J.M., Wong, W.S.W., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A.B., Glusman, G., Vissers, L.E.L.M., Hoischen, A., Roach, J.C., et al (2016) Parent-of-origin-specific signatures of de novo mutations. Nat Genet 48, 935–939.

Goloborodko, A., Marko, J.F., and Mirny, L.A. (2016) Chromosome Compaction by Active Loop Extrusion. Biophysical Journal 110, 2162–2168.

Goodier, J.L. (2016) Restricting retrotransposons: a review. Mob DNA 7, 16.

Götz, A., Tyynismaa, H., Euro, L., Ellonen, P., Hyötyläinen, T., Ojala, T., Hämäläinen, R.H., Tommiska, J., Raivio, T., Oresic, M., et al (2011) Exome Sequencing Identifies Mitochondrial Alanyl-tRNA Synthetase Mutations in Infantile Mitochondrial Cardiomyopathy. Am J Hum Genet 88, 635–642.

Grampa, V., Delous, M., Zaidan, M., Odye, G., Thomas, S., Elkhartoufi, N., Filhol, E., Niel, O., Silbermann, F., Lebreton, C., et al (2016) Novel NEK8 Mutations Cause Severe Syndromic Renal Cystic Dysplasia through YAP Dysregulation. PLoS Genet 12, e1005894.

Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A., Nussbaum, R.L., O'Daniel, J.M., Ormond, K.E., et al (2013) ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. Genet Med 15, 565–574.

Grigoryev, S.A. (2018) Chromatin Higher-Order Folding: A Perspective with Linker DNA Angles. Biophys J 114, 2290–2297.

Gruen, M., and Gautel, M. (1999) Mutations in beta-myosin S2 that cause familial hypertrophic cardiomyopathy (FHC) abolish the interaction with the regulatory domain of myosin-binding protein-C. J. Mol. Biol. 286, 933–949.

Guimier, A., Gordon, C.T., Godard, F., Ravenscroft, G., Oufadem, M., Vasnier, C., Rambaud, C., Nitschke, P., Bole-Feysot, C., Masson, C., et al (2016) Biallelic PPA2 Mutations Cause Sudden Unexpected Cardiac Arrest in Infancy. The American Journal of Human Genetics 99, 666–673.

Guo Yuxuan, and Pu William T. (2020) Cardiomyocyte Maturation. Circulation Research 126, 1086–1106.

Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E., Watkins, P.C., Ottina, K., Wallace, M.R., Sakaguchi, A.Y., et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306, 234–238.

Haack, T.B., Danhauser, K., Haberberger, B., Hoser, J., Strecker, V., Boehm, D., Uziel, G., Lamantea, E., Invernizzi, F., Poulton, J., et al (2010) Exome

sequencing identifies ACAD9 mutations as a cause of complex I deficiency. Nature Genetics 42, 1131–1134.

Hancks, D.C., and Kazazian, H.H. (2012) Active human retrotransposons: variation and disease. Curr. Opin. Genet. Dev. 22, 191–203.

Hansen, J.J., Dürr, A., Cournu-Rebeix, I., Georgopoulos, C., Ang, D., Nielsen, M.N., Davoine, C.-S., Brice, A., Fontaine, B., Gregersen, N., et al (2002) Hereditary spastic paraplegia SPG13 is associated with a mutation in the gene encoding the mitochondrial chaperonin Hsp60. Am. J. Hum. Genet. 70, 1328–1332.

Heather, J.M., and Chain, B. (2016) The sequence of sequencers: The history of sequencing DNA. Genomics 107, 1–8.

Heim, P., Claussen, M., Hoffmann, B., Conzelmann, E., Gärtner, J., Harzer, K., Hunneman, D.H., Köhler, W., Kurlemann, G., and Kohlschütter, A. (1997) Leukodystrophy incidence in Germany. Am J Med Genet 71, 475–478.

Henry, M.P., Hawkins, J.R., Boyle, J., and Bridger, J.M. (2019) The Genomic Health of Human Pluripotent Stem Cells: Genomic Instability and the Consequences on Nuclear Organization. Front. Genet. 9, 623.

Hershberger, R.E., Norton, N., Morales, A., Li, D., Siegfried, J.D., and Gonzalez-Quintana, J. (2010) Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1 and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. Circ Cardiovasc Genet 3, 155–161.

Hicks, J.K., Shealy, A., Schreiber, A., Coleridge, M., Noss, R., Natowicz, M., Moran, R., Moss, T., Erwin, A., and Eng, C. (2018) Patient Decisions to Receive Secondary Pharmacogenomic Findings and Development of a Multidisciplinary Practice Model to Integrate Results Into Patient Care. Clin Transl Sci 11, 71–76.

Hiippala, A., Vasilescu, C., Tallila, J., Alastalo, T.-P., Paetau, A., Tyni, T., Suomalainen, A., Euro, L., and Ojala, T. (2016) The rare Costello variant HRAS c.173C>T (p.T58I) with severe neonatal hypertrophic cardiomyopathy. Am. J. Med. Genet. A 170, 1433–1438.

Hijikata, A., Yura, K., Ohara, O., and Go, M. (2015) Structural and functional analyses of Barth syndrome-causing mutations and alternative splicing in the tafazzin acyltransferase domain. Meta Gene 4, 92–106.

Hillier, L.W., Coulson, A., Murray, J.I., Bao, Z., Sulston, J.E., and Waterston, R.H. (2005) Genomics in C. elegans: So many genes, such a little worm. Genome Res. 15, 1651–1660.

Hollander, S.A., Addonizio, L.J., Chin, C., Lamour, J.M., Hsu, D.T., Bernstein, D., and Rosenthal, D.N. (2013) Abdominal complaints as a common first presentation of heart failure in adolescents with dilated cardiomyopathy. The American Journal of Emergency Medicine 31, 684–686.

Hood, L., and Rowen, L. (2013) The Human Genome Project: big science transforms biology and medicine. Genome Med 5, 79.

Hurst, L.D. (2009) Fundamental concepts in genetics: genetics and the

understanding of selection. Nat. Rev. Genet. 10, 83–93.

Imamoto, N. (2018) Heat stress-induced nuclear transport mediated by Hikeshi confers nuclear function of Hsp70s. Curr. Opin. Cell Biol. 52, 82–87.

Imamoto, N., and Kose, S. (2012) Heat-shock stress activates a novel nuclear import pathway mediated by Hikeshi. Nucleus 3, 422–428.

International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437, 1299–1320.

International HapMap Consortium, Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., et al (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–861.

Jääskeläinen, P., Vangipurapu, J., Raivo, J., Kuulasmaa, T., Heliö, T., Aalto-Setälä, K., Kaartinen, M., Ilveskoski, E., Vanninen, S., Hämäläinen, L., et al (2019) Genetic basis and outcome in a nationwide study of Finnish patients with hypertrophic cardiomyopathy. ESC Heart Failure 6, 436–445.

Jackson, P.K. (2013) Nek8 Couples Renal Ciliopathies to DNA Damage and Checkpoint Control. Molecular Cell 51, 407–408.

Jakubosky, D., Smith, E.N., D'Antonio, M., Jan Bonder, M., Young Greenwald, W.W., D'Antonio-Chronowska, A., Matsui, H., i2QTL Consortium, Stegle, O., Montgomery, S.B., et al (2020) Discovery and quality analysis of a comprehensive set of structural variants and short tandem repeats. Nat Commun 11, 2928.

James, C.A., Bhonsale, A., Tichnell, C., Murray, B., Russell, S.D., Tandri, H., Tedford, R.J., Judge, D.P., and Calkins, H. (2013) Exercise Increases Age-Related Penetrance and Arrhythmic Risk in Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy–Associated Desmosomal Mutation Carriers. Journal of the American College of Cardiology 62, 1290–1297.

Jefferies, J.L., Wilkinson, J.D., Sleeper, L.A., Colan, S.D., Lu, M., Pahl, E., Kantor, P.F., Everitt, M.D., Webber, S.A., Kaufman, B.D., et al (2015) Cardiomyopathy Phenotypes and Outcomes for Children With Left Ventricular Myocardial Noncompaction: Results From the Pediatric Cardiomyopathy Registry. Journal of Cardiac Failure 21, 877–884.

Kaessmann, H., Vinckenbosch, N., and Long, M. (2009) RNA-based gene duplication: mechanistic and evolutionary insights. Nat. Rev. Genet. 10, 19–31.

Kalia, S.S., Adelman, K., Bale, S.J., Chung, W.K., Eng, C., Evans, J.P., Herman, G.E., Hufnagel, S.B., Klein, T.E., Korf, B.R., et al (2017) Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med 19, 249–255.

Kantor, P.F., Lougheed, J., Dancea, A., McGillion, M., Barbosa, N., Chan, C., Dillenburg, R., Atallah, J., Buchholz, H., Chant-Gambacort, C., et al (2013) Presentation, Diagnosis, and Medical Management of Heart Failure in

Children: Canadian Cardiovascular Society Guidelines. Canadian Journal of Cardiology 29, 1535–1552.

Kashi, K., Henderson, L., Bonetti, A., and Carninci, P. (2016) Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1859, 3–15.

Kellis, M., Wold, B., Snyder, M.P., Bernstein, B.E., Kundaje, A., Marinov, G.K., Ward, L.D., Birney, E., Crawford, G.E., Dekker, J., et al (2014) Defining functional DNA elements in the human genome. Proc. Natl. Acad. Sci. U.S.A. 111, 6131–6138.

Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F.P., Culley, O.J., et al (2017) Common genetic variation drives molecular heterogeneity in human iPSCs. Nature 546, 370–375.

Kimura, M. (1968) Evolutionary Rate at the Molecular Level. Nature 217, 624–626.

Kindel, S.J., Miller, E.M., Gupta, R., Cripe, L.H., Hinton, R.B., Spicer, R.L., Towbin, J.A., and Ware, S.M. (2012) Pediatric cardiomyopathy: importance of genetic and metabolic evaluation. J. Card. Fail. 18, 396–403.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46, 310–315.

Kirk, E.P., Sunde, M., Costa, M.W., Rankin, S.A., Wolstein, O., Castro, M.L., Butler, T.L., Hyun, C., Guo, G., Otway, R., et al (2007) Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. Am. J. Hum. Genet. 81, 280–291.

van der Knaap, M.S., and Bugiani, M. (2017) Leukodystrophies: a proposed classification system based on pathological changes and pathogenetic mechanisms. Acta Neuropathol 134, 351–382.

Köhler, S., Vasilevsky, N.A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., Baynam, G., Bello, S.M., Boerkoel, C.F., Boycott, K.M., et al (2017) The Human Phenotype Ontology in 2017. Nucleic Acids Res. 45, D865–D876.

Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J.O.B., Danis, D., Gourdine, J.-P., Gargano, M., Harris, N.L., Matentzoglu, N., McMurry, J.A., et al (2019) Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Research 47, D1018–D1027.

Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al (2012) Rate of de novo mutations and the importance of father's age to disease risk. Nature 488, 471–475.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011) Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 7, e1002384.

Kose, S., Furuta, M., and Imamoto, N. (2012) Hikeshi, a Nuclear Import Carrier for Hsp70s, Protects Cells from Heat Shock-Induced Nuclear Damage. Cell 149, 578–589.

Koskenvuo, J.W., Leivo, K., Vanninen, S., Seppala, E.H., Aalto-Setala, K., Pitkanen, O., Suursalmi, P., Annala, A.-P., Anttila, I., Alastalo, T.-P., et al (2018) P6325JPH2 p.(Thr161Lys) is a Finnish founder mutation associating to hypertrophic cardiomyopathy with or without systolic heart failure and conduction abnormalities. Eur Heart J 39.

Kossinova, O., Malygin, A., Krol, A., and Karpova, G. (2013) A novel insight into the mechanism of mammalian selenoprotein synthesis. RNA 19, 1147–1158.

Krumm, A., and Duan, Z. (2019) Understanding the 3D genome: Emerging impacts on human disease. Seminars in Cell & Developmental Biology 90, 62–77.

Kussauer, S., David, R., and Lemcke, H. (2019) hiPSCs Derived Cardiac Cells for Drug and Toxicity Screening and Disease Modeling: What Micro-Electrode-Array Analyses Can Tell Us. Cells 8.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al (2001) Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Landstrom, A.P., Weisleder, N., Batalden, K.B., Bos, J.M., Tester, D.J., Ommen, S.R., Wehrens, X.H.T., Claycomb, W.C., Ko, J.-K., Hwang, M., et al (2007) Mutations in JPH2-encoded junctophilin-2 associated with hypertrophic cardiomyopathy in humans. J. Mol. Cell. Cardiol. 42, 1026–1035.

Lappalainen, T., Scott, A.J., Brandt, M., and Hall, I.M. (2019) Genomic Analysis in the Age of Human Genome Sequencing. Cell 177, 70–84.

Lee Teresa M., Hsu Daphne T., Kantor Paul, Towbin Jeffrey A., Ware Stephanie M., Colan Steven D., Chung Wendy K., Jefferies John L., Rossano Joseph W., Castleberry Chesney D., et al (2017) Pediatric Cardiomyopathies. Circulation Research 121, 855–873.

Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., Bergeron, J., Borchers, C.H., Corthals, G.L., Costello, C.E., et al (2011) The human proteome project: current state and future direction. Mol. Cell Proteomics 10, M111.009993.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.

Lelieveld, S.H., Spielmann, M., Mundlos, S., Veltman, J.A., and Gilissen, C. (2015) Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. Hum. Mutat. 36, 815–822.

Lencz, T., Yu, J., Palmer, C., Carmi, S., Ben-Avraham, D., Barzilai, N., Bressman, S., Darvasi, A., Cho, J.H., Clark, L.N., et al (2018) High-depth

whole genome sequencing of an Ashkenazi Jewish reference panel: enhancing sensitivity, accuracy, and imputation. Hum Genet 137, 343–355.

Li, C., and Williams, S.M. (2013) Human Somatic Variation: It's Not Just for Cancer Anymore. Curr Genet Med Rep 1, 212–218.

Li, J., and Liu, C. (2019) Coding or Noncoding, the Converging Concepts of RNAs. Front. Genet. 10.

Lieberman-Aiden, E., Berkum, N.L. van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science 326, 289–293.

Lin, W., and Popko, B. (2009) Endoplasmic reticulum stress in disorders of myelinating cells. Nat. Neurosci. 12, 379–385.

Lin, W., and Stone, S. (2020) Unfolded protein response in myelin disorders. Neural Regen Res 15, 636–645.

Lindberg, I., Shorter, J., Wiseman, R.L., Chiti, F., Dickey, C.A., and McLean, P.J. (2015) Chaperones in Neurodegeneration. J. Neurosci. 35, 13853–13859.

Lipshultz, S.E., Sleeper, L.A., Towbin, J.A., Lowe, A.M., Orav, E.J., Cox, G.F., Lurie, P.R., McCoy, K.L., McDonald, M.A., Messere, J.E., et al (2003) The incidence of pediatric cardiomyopathy in two regions of the United States. N. Engl. J. Med. 348, 1647–1655.

Lipshultz, S.E., Cochran, T.R., Briston, D.A., Brown, S.R., Sambatakos, P.J., Miller, T.L., Carrillo, A.A., Corcia, L., Sanchez, J.E., Diamond, M.B., et al (2013) Pediatric cardiomyopathies: causes, epidemiology, clinical course, preventive strategies and therapies. Future Cardiol 9, 817–848.

Lipshultz Steven E., Law Yuk M., Asante-Korang Alfred, Austin Eric D., Dipchand Anne I., Everitt Melanie D., Hsu Daphne T., Lin Kimberly Y., Price Jack F., Wilkinson James D., et al (2019) Cardiomyopathy in Children: Classification and Diagnosis: A Scientific Statement From the American Heart Association. Circulation 140, e9–e68.

Luo, S., Valencia, C.A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J., et al (2018) Biparental Inheritance of Mitochondrial DNA in Humans. Proc Natl Acad Sci USA 115, 13039–13044.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161, 1012–1025.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al (2012) A systematic survey of loss-of-function variants in human protein-coding genes. Science 335, 823–828.

Magen, D., Georgopoulos, C., Bross, P., Ang, D., Segev, Y., Goldsher, D., Nemirovski, A., Shahar, E., Ravid, S., Luder, A., et al (2008) Mitochondrial hsp60 chaperonopathy causes an autosomal-recessive neurodegenerative disorder linked to brain hypomyelination and leukodystrophy. Am. J. Hum.

Genet. 83, 30–42.

Magi, A., D'Aurizio, R., Palombo, F., Cifola, I., Tattini, L., Semeraro, R., Pippucci, T., Giusti, B., Romeo, G., Abbate, R., et al (2015) Characterization and identification of hidden rare variants in the human genome. BMC Genomics 16, 340.

Majo, F., Rochat, A., Nicolas, M., Jaoudé, G.A., and Barrandon, Y. (2008) Oligopotent stem cells are distributed throughout the mammalian ocular surface. Nature 456, 250–254.

Maroilley, T., and Tarailo-Graovac, M. (2019) Uncovering Missing Heritability in Rare Diseases. Genes (Basel) 10.

Martello, G., and Smith, A. (2014) The Nature of Embryonic Stem Cells. Annual Review of Cell and Developmental Biology 30, 647–675.

McKenna William J., Maron Barry J., and Thiene Gaetano (2017) Classification, Epidemiology, and Global Burden of Cardiomyopathies. Circulation Research 121, 722–730.

McKeown, C.R., Nowak, R.B., Moyer, J., Sussman, M.A., and Fowler, V.M. (2008) Tropomodulin1 is required in the heart but not the yolk sac for mouse embryonic development. Circ. Res. 103, 1241–1248.

McWilliams, T.G., and Suomalainen, A. (2019) Mitochondrial DNA can be inherited from fathers, not just mothers. Nature 565, 296–297.

Merkle, F.T., Ghosh, S., Kamitaki, N., Mitchell, J., Avior, Y., Mello, C., Kashin, S., Mekhoubad, S., Ilic, D., Charlton, M., et al (2017) Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. Nature 545, 229–233.

Miga, K.H. (2015) Completing the human genome: the progress and challenge of satellite DNA assembly. Chromosome Res. 23, 421–426.

Mirauta, B.A., Seaton, D.D., Bensaddek, D., Brenes, A., Bonder, M.J., Kilpinen, H., HipSci Consortium, Agu, C.A., Alderton, A., Danecek, P., et al (2020) Population-scale proteome variation in human induced pluripotent stem cells. Elife 9.

Miska, E.A., and Ferguson-Smith, A.C. (2016) Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. Science 354, 59–63.

Moindrot, B., and Brockdorff, N. (2016) RNA binding proteins implicated in Xist-mediated chromosome silencing. Semin. Cell Dev. Biol. 56, 58–70.

Moon, S.L., and Parker, R. (2018) EIF2B2 mutations in vanishing white matter disease hypersuppress translation and delay recovery during the integrated stress response. RNA 24, 841–852.

Morell, P., and Quarles, R.H. (1999a) Characteristic Composition of Myelin. Basic Neurochemistry: Molecular, Cellular and Medical Aspects. 6th Edition.

Morell, P., and Quarles, R.H. (1999b) Developmental Biology of Myelin. Basic Neurochemistry: Molecular, Cellular and Medical Aspects. 6th Edition.

Morita, H., Rehm, H.L., Menesses, A., McDonough, B., Roberts, A.E.,

Kucherlapati, R., Towbin, J.A., Seidman, J.G., and Seidman, C.E. (2008) Shared genetic causes of cardiac hypertrophy in children and adults. N. Engl. J. Med. 358, 1899–1908.

Moyer, J.D., Nowak, R.B., Kim, N.E., Larkin, S.K., Peters, L.L., Hartwig, J., Kuypers, F.A., and Fowler, V.M. (2010) Tropomodulin 1-null mice have a mild spherocytic elliptocytosis with appearance of tropomodulin 3 in red blood cells and disruption of the membrane skeleton. Blood 116, 2590–2599.

Musunuru Kiran, Sheikh Farah, Gupta Rajat M., Houser Steven R., Maher Kevin O., Milan David J., Terzic Andre, and Wu Joseph C. (2018) Induced Pluripotent Stem Cells for Cardiovascular Disease Modeling and Precision Medicine: A Scientific Statement From the American Heart Association. Circulation: Genomic and Precision Medicine 11, e000043.

Nelis, M., Esko, T., Mägi, R., Zimprich, F., Zimprich, A., Toncheva, D., Karachanak, S., Piskácková, T., Balascák, I., Peltonen, L., et al (2009) Genetic structure of Europeans: a view from the North-East. PLoS ONE 4, e5472.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461, 272–276.

Ngo, H.B., Lovely, G.A., Phillips, R., and Chan, D.C. (2014) Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation. Nat Commun 5, 3077.

Norio, R. (2003a) Finnish Disease Heritage II: population prehistory and genetic roots of Finns. Hum Genet 112, 457–469.

Norio, R. (2003b) The Finnish Disease Heritage III: the individual diseases. Hum. Genet. 112, 470–526.

Norio, R. (2003c) Finnish Disease Heritage I: Hum Genet 112, 441–456.

Norio, R., Nevanlinna, H.R., and Perheentupa, J. (1973) Hereditary diseases in Finland; rare flora in rare soul. Ann Clin Res 5, 109–141.

Nowak, R.B., Fischer, R.S., Zoltoski, R.K., Kuszak, J.R., and Fowler, V.M. (2009) Tropomodulin1 is required for membrane skeleton organization and hexagonal geometry of fiber cells in the mouse lens. J. Cell Biol. 186, 915–928.

Nowak, R.B., Papoin, J., Gokhin, D.S., Casu, C., Rivella, S., Lipton, J.M., Blanc, L., and Fowler, V.M. (2017) Tropomodulin 1 controls erythroblast enucleation via regulation of F-actin in the enucleosome. Blood 130, 1144–1155.

Nugent, A.W., Daubeney, P.E.F., Chondros, P., Carlin, J.B., Cheung, M., Wilkinson, L.C., Davis, A.M., Kahler, S.G., Chow, C.W., Wilkinson, J.L., et al (2003) The epidemiology of childhood cardiomyopathy in Australia. N. Engl. J. Med. 348, 1639–1646.

Nugraha, B., Buono, M.F., Boehmer, L. von, Hoerstrup, S.P., and Emmert, M.Y. (2019) Human Cardiac Organoids for Disease Modeling. Clinical

Pharmacology & Therapeutics 105, 79–85.

Ohta, T. (1973) Slightly Deleterious Mutant Substitutions in Evolution. Nature 246, 96–98.

Ojala, M., Prajapati, C., Pölönen, R.-P., Rajala, K., Pekkanen-Mattila, M., Rasku, J., Larsson, K., and Aalto-Setälä, K. (2015a) Mutation-Specific Phenotypes in hiPSC-Derived Cardiomyocytes Carrying Either Myosin-Binding Protein C Or α-Tropomyosin Mutation for Hypertrophic Cardiomyopathy (Hindawi)

Ojala, T., Polinati, P., Manninen, T., Hiippala, A., Rajantie, J., Karikoski, R., Suomalainen, A., and Tyni, T. (2012a) New mutation of mitochondrial DNAJC19 causing dilated and noncompaction cardiomyopathy, anemia, ataxia, and male genital anomalies. Pediatric Research 72, 432–437.

Ojala, T., Polinati, P., Manninen, T., Hiippala, A., Rajantie, J., Karikoski, R., Suomalainen, A., and Tyni, T. (2012b) New mutation of mitochondrial DNAJC19 causing dilated and noncompaction cardiomyopathy, anemia, ataxia, and male genital anomalies. Pediatric Research 72, 432–437.

Ojala, T., Nupponen, I., Saloranta, C., Sarkola, T., Sekar, P., Breilin, A., and Tyni, T. (2015b) Fetal left ventricular noncompaction cardiomyopathy and fatal outcome due to complete deficiency of mitochondrial trifunctional protein. Eur. J. Pediatr. 174, 1689–1692.

Oleksyk, T.K., Brukhin, V., and O'Brien, S.J. (2015a) Putting Russia on the genome map. Science 350, 747–747.

Oleksyk, T.K., Brukhin, V., and O'Brien, S.J. (2015b) The Genome Russia project: closing the largest remaining omission on the world Genome map. Gigascience 4.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017) ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science 357.

Özdemir, V., Dove, E.S., Gürsoy, U.K., Şardaş, S., Yıldırım, A., Yılmaz, Ş.G., Ömer Barlas, I., Güngör, K., Mete, A., and Srivastava, S. (2017) Personalized medicine beyond genomics: alternative futures in big data-proteomics, environtome and the social proteome. J Neural Transm (Vienna) 124, 25–32.

Paik, D.T., Chandy, M., and Wu, J.C. (2020) Patient and Disease–Specific Induced Pluripotent Stem Cells for Discovery of Personalized Cardiovascular Drugs and Therapeutics. Pharmacol Rev 72, 320–342.

Palazzo, A.F., and Gregory, T.R. (2014) The Case for Junk DNA. PLOS Genetics 10, e1004351.

Palazzo, A.F., and Lee, E.S. (2015) Non-coding RNA: what is functional and what is junk? Front. Genet. 6.

Pandit, B., Sarkozy, A., Pennacchio, L.A., Carta, C., Oishi, K., Martinelli, S., Pogna, E.A., Schackwitz, W., Ustaszewska, A., Landstrom, A., et al (2007) Gain-of-function RAF1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. Nat. Genet. 39, 1007–1012.

Pang, L. (2020) Toxicity testing in the era of induced pluripotent stem cells: A perspective regarding the use of patient-specific induced pluripotent stem cell–derived cardiomyocytes for cardiac safety evaluation. Current Opinion in Toxicology 23–24, 50–55.

Papa, I., Astier, C., Kwiatek, O., Raynaud, F., Bonnal, C., Lebart, M.C., Roustan, C., and Benyamin, Y. (1999) Alpha actinin-CapZ, an anchoring complex for thin filaments in Z-line. J. Muscle Res. Cell. Motil. 20, 187–197.

Parikh, S., Bernard, G., Leventer, R.J., van der Knaap, M.S., van Hove, J., Pizzino, A., McNeill, N.H., Helman, G., Simons, C., Schmidt, J.L., et al (2015) A clinical approach to the diagnosis of patients with leukodystrophies and genetic leukoencephelopathies. Mol. Genet. Metab. 114, 501–515.

Parrotta, E.I., Scalise, S., Scaramuzzino, L., and Cuda, G. (2019) Stem Cells: The Game Changers of Human Cardiac Disease Modelling and Regenerative Medicine. Int J Mol Sci 20.

Parvari, R., Hershkovitz, E., Grossman, N., Gorodischer, R., Loeys, B., Zecic, A., Mortier, G., Gregory, S., Sharony, R., Kambouris, M., et al (2002) Mutation of TBCE causes hypoparathyroidism– retardation–dysmorphism and autosomal recessive Kenny–Caffey syndrome. Nature Genetics 32, 448–452.

Parvatiyar, M.S., Landstrom, A.P., Figueiredo-Freitas, C., Potter, J.D., Ackerman, M.J., and Pinto, J.R. (2012) A Mutation in TNNC1-encoded Cardiac Troponin C, TNNC1-A31S, Predisposes to Hypertrophic Cardiomyopathy and Ventricular Fibrillation. J Biol Chem 287, 31845–31855.

Pavitt, G.D. (2018) Regulation of translation initiation factor eIF2B at the hub of the integrated stress response. Wiley Interdiscip Rev RNA 9, e1491.

Pearce, S.F., Rebelo-Guiomar, P., D'Souza, A.R., Powell, C.A., Van Haute, L., and Minczuk, M. (2017) Regulation of Mammalian Mitochondrial Gene Expression: Recent Advances. Trends Biochem Sci 42, 625–639.

Pelak, K., Shianna, K.V., Ge, D., Maia, J.M., Zhu, M., Smith, J.P., Cirulli, E.T., Fellay, J., Dickson, S.P., Gumbs, C.E., et al (2010) The Characterization of Twenty Sequenced Human Genomes. PLoS Genet 6.

Peltonen, L., Jalanko, A., and Varilo, T. (1999) Molecular Genetics the Finnish Disease Heritage. Hum Mol Genet 8, 1913–1923.

Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., et al (2015) The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. Human Mutation 36, 915–921.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al (2012) A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. Cell 151, 1617–1632.

Portin, P., and Wilkins, A. (2017) The Evolving Definition of the Term "Gene." Genetics 205, 1353–1364.

Pöyhönen, P., Hiippala, A., Ollila, L., Kaasalainen, T., Hänninen, H., Heliö, T., Tallila, J., Vasilescu, C., Kivistö, S., Ojala, T., et al (2015) Cardiovascular magnetic resonance findings in patients with PRKAG2 gene mutations. J Cardiovasc Magn Reson 17, 89.

Rahman, K.M.Z., Mamada, H., Takagi, M., Kose, S., and Imamoto, N. (2017) Hikeshi modulates the proteotoxic stress response in human cells: Implication for the importance of the nuclear function of HSP70s. Genes Cells 22, 968–976.

Rajagopalan, R., Grochowski, C.M., Gilbert, M.A., Falsey, A.M., Coleman, K., Romero, R., Loomes, K.M., Piccoli, D.A., Devoto, M., and Spinner, N.B. (2016) Compound heterozygous mutations in NEK8 in siblings with end-stage renal disease with hepatic and cardiac anomalies. American Journal of Medical Genetics Part A 170, 750–753.

Ralph, P., and Coop, G. (2013) The geography of recent genetic ancestry across Europe. PLoS Biol. 11, e1001555.

Rao, J.N., Madasu, Y., and Dominguez, R. (2014a) Mechanism of actin filament pointed-end capping by tropomodulin. Science 345, 463–467.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al (2014b) A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159, 1665–1680.

Rapezzi, C., Arbustini, E., Caforio, A.L.P., Charron, P., Gimeno-Blanes, J., Heliö, T., Linhart, A., Mogensen, J., Pinto, Y., Ristic, A., et al (2013) Diagnostic work-up in cardiomyopathies: bridging the gap between clinical phenotypes and final diagnosis. A position statement from the ESC Working Group on Myocardial and Pericardial Diseases. Eur Heart J 34, 1448–1458.

Razzaque, M.A., Nishizawa, T., Komoike, Y., Yagi, H., Furutani, M., Amo, R., Kamisago, M., Momma, K., Katayama, H., Nakagawa, M., et al (2007) Germline gain-of-function mutations in RAF1 cause Noonan syndrome. Nature Genetics 39, 1013–1017.

Rehm, H.L. (2013) Disease-targeted sequencing: a cornerstone in the clinic. Nat Rev Genet 14, 295–300.

Rice, G., Patrick, T., Parmar, R., Taylor, C.F., Aeby, A., Aicardi, J., Artuch, R., Montalto, S.A., Bacino, C.A., Barroso, B., et al (2007) Clinical and molecular phenotype of Aicardi-Goutieres syndrome. Am. J. Hum. Genet. 81, 713–725.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17, 405–424.

Richter, J.D., and Sonenberg, N. (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature 433, 477–480.

te Riele, A.S.J.M., James, C.A., Sawant, A.C., Bhonsale, A., Groeneweg, J.A., Mast, T.P., Murray, B., Tichnell, C., Dooijes, D., van Tintelen, J.P., et al

(2015) Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy in the Pediatric Population: Clinical Characterization and Comparison With Adult-Onset Disease. JACC: Clinical Electrophysiology 1, 551–560.

Robinson, P.N., and Mundlos, S. (2010) The human phenotype ontology. Clin. Genet. 77, 525–534.

Rocchetti, M., Sala, L., Dreizehnter, L., Crotti, L., Sinnecker, D., Mura, M., Pane, L.S., Altomare, C., Torre, E., Mostacciuolo, G., et al (2017) Elucidating arrhythmogenic mechanisms of long-QT syndrome CALM1-F142L mutation in patient-specific induced pluripotent stem cell-derived cardiomyocytes. Cardiovasc. Res. 113, 531–541.

Roche, M.I., and Berg, J.S. (2015) Incidental Findings with Genomic Testing: Implications for Genetic Counseling Practice. Curr Genet Med Rep 3, 166–176.

Ronaldson-Bouchard, K., Ma, S.P., Yeager, K., Chen, T., Song, L., Sirabella, D., Morikawa, K., Teles, D., Yazawa, M., and Vunjak-Novakovic, G. (2018) Advanced maturation of human cardiac tissue grown from pluripotent stem cells. Nature 556, 239–243.

Ronaldson-Bouchard, K., Yeager, K., Teles, D., Chen, T., Ma, S., Song, L., Morikawa, K., Wobma, H.M., Vasciaveo, A., Ruiz, E.C., et al (2019) Engineering of human cardiac muscle electromechanically matured to an adult-like phenotype. Nature Protocols 14, 2781–2817.

Rosenbaum, A.N., Agre, K.E., and Pereira, N.L. (2020) Genetics of dilated cardiomyopathy: practical implications for heart failure management. Nat Rev Cardiol 17, 286–297.

Sabater-Molina, M., Navarro, M., Sáez, E.G.-M., Garrido, I., Pascual-Figal, D., Carrillo, J.G., and Blanes, J.R.G. (2016) Mutation in JPH2 cause dilated cardiomyopathy. Clinical Genetics 90, 468–469.

Saelaert, M., Mertes, H., De Baere, E., and Devisch, I. (2018) Incidental or secondary findings: an integrative and patient-inclusive approach to the current debate. European Journal of Human Genetics 26, 1424–1431.

Saelaert, M., Mertes, H., Moerenhout, T., De Baere, E., and Devisch, I. (2019) Criteria for reporting incidental findings in clinical exome sequencing – a focus group study on professional practices and perspectives in Belgian genetic centres. BMC Medical Genomics 12, 123.

Salzer, J.L. (1997) Clustering Sodium Channels at the Node of Ranvier: Close Encounters of the Axon–Glia Kind. Neuron 18, 843–846.

Sanger, F., and Coulson, A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94, 441–448.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. PNAS 74, 5463–5467.

Sawant, A.C., Te Riele, A.S.J.M., Tichnell, C., Murray, B., Bhonsale, A., Tandri, H., Judge, D.P., Calkins, H., and James, C.A. (2016) Safety of American Heart Association-recommended minimum exercise for

desmosomal mutation carriers. Heart Rhythm 13, 199–207.

Sawyer, S.L., Hartley, T., Dyment, D.A., Beaulieu, C.L., Schwartzentruber, J., Smith, A., Bedford, H.M., Bernard, G., Bernier, F.P., Brais, B., et al (2016) Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. Clin Genet 89, 275–284.

Schägger, H., and von Jagow, G. (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Analytical Biochemistry 199, 223–231.

Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P.A., Murphy, T.D., Pruitt, K.D., Thibaud-Nissen, F., Albracht, D., et al (2017) Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. 27, 849–864.

Schoenfelder, S., and Fraser, P. (2019) Long-range enhancer–promoter contacts in gene expression control. Nature Reviews Genetics 20, 437–455.

Schübeler, D. (2015) Function and information content of DNA methylation. Nature 517, 321–326.

Schwartz, M., and Vissing, J. (2002) Paternal Inheritance of Mitochondrial DNA. New England Journal of Medicine 347, 576–580.

Selcen, D., Muntoni, F., Burton, B.K., Pegoraro, E., Sewry, C., Bite, A.V., and Engel, A.G. (2009) Mutation in BAG3 causes severe dominant childhood muscular dystrophy. Annals of Neurology 65, 83–89.

Senderek, J., Krieger, M., Stendel, C., Bergmann, C., Moser, M., Breitbach-Faller, N., Rudnik-Schöneborn, S., Blaschek, A., Wolf, N.I., Harting, I., et al (2005) Mutations in SIL1 cause Marinesco-Sjögren syndrome, a cerebellar ataxia with cataract and myopathy. Nature Genetics 37, 1312–1314.

Sequeira, V., Nijenkamp, L.L.A.M., Regan, J.A., and van der Velden, J. (2014) The physiological role of cardiac cytoskeleton and its alterations in heart failure. Biochimica et Biophysica Acta (BBA) - Biomembranes 1838, 700–722.

Sharpless, N.E., and DePinho, R.A. (2007) How stem cells age and why this makes us grow old. Nature Reviews Molecular Cell Biology 8, 703–713.

Shendure, J., and Akey, J.M. (2015) The origins, determinants, and consequences of human mutations. Science 349, 1478–1483.

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., and Waterston, R.H. (2017) DNA sequencing at 40: past, present and future. Nature 550, 345–353.

Sherman, M.Y., and Gabai, V.L. (2015) Hsp70 in cancer: back to the future. Oncogene 34, 4153–4161.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40, W452–W457.

Simons, B.D., and Clevers, H. (2011) Strategies for Homeostatic Stem Cell

Self-Renewal in Adult Tissues. Cell 145, 851–862.

Sivakumar, A., de las Heras, J.I., and Schirmer, E.C. (2019) Spatial Genome Organization: From Development to Disease. Front. Cell Dev. Biol. 7.

Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015) GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. Human Mutation 36, 928–930.

Soellner, L., Begemann, M., Mackay, D.J.G., Grønskov, K., Tümer, Z., Maher, E.R., Temple, I.K., Monk, D., Riccio, A., Linglart, A., et al (2017) Recent Advances in Imprinting Disorders. Clin. Genet. 91, 3–13.

Song, J., Kose, S., Watanabe, A., Son, S.-Y., Choi, S., Hong, H., Yamashita, E., Park, I.Y., Imamoto, N., and Lee, S.J. (2015) Structural and functional analysis of Hikeshi, a new nuclear transport receptor of Hsp70s. Acta Crystallogr D Biol Crystallogr 71, 473–483.

Splawski, I., Timothy, K.W., Decher, N., Kumar, P., Sachse, F.B., Beggs, A.H., Sanguinetti, M.C., and Keating, M.T. (2005) Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. Proc. Natl. Acad. Sci. U.S.A. 102, 8089–8096; discussion 8086-8088.

Stadelmann, C., Timmler, S., Barrantes-Freer, A., and Simons, M. (2019) Myelin in the Central Nervous System: Structure, Function, and Pathology. Physiological Reviews 99, 1381–1431.

Stark, Z., Tan, T.Y., Chong, B., Brett, G.R., Yap, P., Walsh, M., Yeung, A., Peters, H., Mordaunt, D., Cowie, S., et al (2016) A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. Genetics in Medicine 18, 1090–1096.

Stassart, R.M., Möbius, W., Nave, K.-A., and Edgar, J.M. (2018) The Axon-Myelin Unit in Development and Degenerative Disease. Front. Neurosci. 12.

Steichen, C., Hannoun, Z., Luce, E., Hauet, T., and Dubart-Kupperschmitt, A. (2019) Genomic integrity of human induced pluripotent stem cells: Reprogramming, differentiation and applications. WJSC 11, 729–747.

Stewart, J.B., and Larsson, N.-G. (2014) Keeping mtDNA in shape between generations. PLoS Genet. 10, e1004670.

Strachan, T., and Read, A. (2018) Human Molecular Genetics (Garland Science)

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H.-Y., et al (2015) An integrated map of structural variation in 2,504 human genomes. Nature 526, 75–81.

Sulem, P., Helgason, H., Oddson, A., Stefansson, H., Gudjonsson, S.A., Zink, F., Hjartarson, E., Sigurdsson, G.T., Jonasdottir, A., Jonasdottir, A., et al (2015) Identification of a large set of rare complete human knockouts. Nat Genet 47, 448–452.

Sulonen, A.-M., Ellonen, P., Almusa, H., Lepistö, M., Eldfors, S., Hannula, S., Miettinen, T., Tyynismaa, H., Salo, P., Heckman, C., et al (2011) Comparison

of solution-based exome capture methods for next generation sequencing. Genome Biology 12, R94.

Svitkina, T. (2018) The Actin Cytoskeleton and Actin-Based Motility. Cold Spring Harb Perspect Biol 10, a018267.

Takahashi, K., and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Takahashi, K., Kogaki, S., Kurotobi, S., Nasuno, S., Ohta, M., Okabe, H., Wada, K., Sakai, N., Taniike, M., and Ozono, K. (2005) A novel mutation in the PTPN11 gene in a patient with Noonan syndrome and rapidly progressive hypertrophic cardiomyopathy. Eur J Pediatr 164, 497–500.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131, 861–872.

Takeda, S., Yamashita, A., Maeda, K., and Maéda, Y. (2003) Structure of the core domain of human cardiac troponin in the Ca(2+)-saturated form. Nature 424, 35–41.

Tarkowski, A.K. (1959) Experiments on the Development of Isolated Blastomeres of Mouse Eggs. Nature 184, 1286–1287.

Tennyson, C.N., Klamut, H.J., and Worton, R.G. (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nature Genetics 9, 184–190.

The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073.

The Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45, 1113–1120.

Theunissen, T.W., and Jaenisch, R. (2014) Molecular Control of Induced Pluripotency. Cell Stem Cell 14, 720–734.

Thienpont, B., Zhang, L., Postma, A.V., Breckpot, J., Tranchevent, L.-C., Van Loo, P., Møllgård, K., Tommerup, N., Bache, I., Tümer, Z., et al (2010) Haploinsufficiency of TAB2 causes congenital heart defects in humans. Am. J. Hum. Genet. 86, 839–849.

Towbin, J.A., Lowe, A.M., Colan, S.D., Sleeper, L.A., Orav, E.J., Clunie, S., Messere, J., Cox, G.F., Lurie, P.R., Hsu, D., et al (2006) Incidence, Causes, and Outcomes of Dilated Cardiomyopathy in Children. JAMA 296, 1867–1876.

Trounson, A., and DeWitt, N.D. (2016) Pluripotent stem cells progressing to the clinic. Nature Reviews Molecular Cell Biology 17, 194–200.

Truszkowska, G.T., Bilińska, Z.T., Muchowicz, A., Pollak, A., Biernacka, A., Kozar-Kamińska, K., Stawiński, P., Gasperowicz, P., Kosińska, J., Zieliński, T., et al (2017) Homozygous truncating mutation in NRAP gene identified by

whole exome sequencing in a patient with dilated cardiomyopathy. Scientific Reports 7, 3362.

Tu Chengyi, Chao Benjamin S., and Wu Joseph C. (2018) Strategies for Improving the Maturity of Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes. Circulation Research 123, 512–514.

Vanderver, A., Prust, M., Tonduti, D., Mochel, F., Hussey, H.M., Helman, G., Garbern, J., Eichler, F., Labauge, P., Aubourg, P., et al (2015) Case definition and classification of leukodystrophies and leukoencephalopathies. Mol. Genet. Metab. 114, 494–500.

Vanderver, A., Simons, C., Helman, G., Crawford, J., Wolf, N.I., Bernard, G., Pizzino, A., Schmidt, J.L., Takanohashi, A., Miller, D., et al (2016) Whole exome sequencing in patients with white matter abnormalities. Ann Neurol 79, 1031–1037.

Vanderver, A., Bernard, G., Helman, G., Sherbini, O., Boeck, R., Cohn, J., Collins, A., Demarest, S., Dobbins, K., Emrick, L., et al Randomized Clinical Trial of First-Line Genome Sequencing in Pediatric White Matter Disorders. Annals of Neurology n/a.

Veltman, J.A., and Brunner, H.G. (2012) De novo mutations in human genetic disease. Nature Reviews Genetics 13, 565–575.

Venkatesh, S., and Workman, J.L. (2015) Histone exchange, chromatin structure and the regulation of transcription. Nature Reviews Molecular Cell Biology 16, 178–189.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al (2001) The Sequence of the Human Genome. Science 291, 1304–1351.

Vilariño-Güell, C., Rajput, A., Milnerwood, A.J., Shah, B., Szu-Tu, C., Trinh, J., Yu, I., Encarnacion, M., Munsie, L.N., Tapia, L., et al (2014) DNAJC13 mutations in Parkinson disease. Hum. Mol. Genet. 23, 1794–1801.

Wakao, S., Kitada, M., Kuroda, Y., Ogura, F., Murakami, T., Niwa, A., and Dezawa, M. (2012) Morphologic and Gene Expression Criteria for Identifying Human Induced Pluripotent Stem Cells. PLOS ONE 7, e48677.

Walters, B., and Thompson, S.R. (2016) Cap-Independent Translational Control of Carcinogenesis. Front. Oncol. 6.

Wang, G., McCain, M.L., Yang, L., He, A., Pasqualini, F.S., Agarwal, A., Yuan, H., Jiang, D., Zhang, D., Zangi, L., et al (2014) Modeling the mitochondrial cardiomyopathy of Barth syndrome with iPSC and heart-on-chip technologies. Nat Med 20, 616–623.

Wang, Q., Pierce-Hoffman, E., Cummings, B.B., Alföldi, J., Francioli, L.C., Gauthier, L.D., Hill, A.J., O'Donnell-Luria, A.H., Karczewski, K.J., and MacArthur, D.G. (2020) Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. Nature Communications 11, 2539.

Ward, L.D., and Kellis, M. (2012) Evidence of abundant purifying selection in humans for recently acquired regulatory functions. Science 337, 1675–1678.

Webber Steven A., Lipshultz Steven E., Sleeper Lynn A., Lu Minmin, Wilkinson James D., Addonizio Linda J., Canter Charles E., Colan Steven D., Everitt Melanie D., Jefferies John Lynn, et al (2012) Outcomes of Restrictive Cardiomyopathy in Childhood and the Influence of Phenotype. Circulation 126, 1237–1244.

Weber, A., Pennise, C.R., Babcock, G.G., and Fowler, V.M. (1994) Tropomodulin caps the pointed ends of actin filaments. J. Cell Biol. 127, 1627–1635.

Wentink, A.S., Nillegoda, N.B., Feufel, J., Ubartaitė, G., Schneider, C.P., De Los Rios, P., Hennig, J., Barducci, A., and Bukau, B. (2020) Molecular dissection of amyloid disaggregation by human HSP70. Nature 1–6.

Whiffin, N., Karczewski, K.J., Zhang, X., Chothani, S., Smith, M.J., Evans, D.G., Roberts, A.M., Quaife, N.M., Schafer, S., Rackham, O., et al (2020) Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. Nature Communications 11, 2523.

Wilkinson, J.D., Lowe, A.M., Salbert, B.A., Sleeper, L.A., Colan, S.D., Cox, G.F., Towbin, J.A., Connuck, D.M., Messere, J.E., and Lipshultz, S.E. (2012) Outcomes in children with Noonan syndrome and hypertrophic cardiomyopathy: a study from the Pediatric Cardiomyopathy Registry. Am Heart J 164, 442–448.

Williamson, J.M., and Lyons, D.A. (2018) Myelin Dynamics Throughout Life: An Ever-Changing Landscape? Front. Cell. Neurosci. 12.

Wright, C.F., FitzPatrick, D.R., and Firth, H.V. (2018) Paediatric genomics: diagnosing rare disease in children. Nature Reviews Genetics 19, 253–268.

Xu, X.Q., Soo, S.Y., Sun, W., and Zweigerdt, R. (2009) Global Expression Profile of Highly Enriched Cardiomyocytes Derived from Human Embryonic Stem Cells: Global Expression Profile of hESC-Cardiomyocytes. STEM CELLS 27, 2163–2174.

Yamashiro, S., Gokhin, D.S., Kimura, S., Nowak, R.B., and Fowler, V.M. (2012) Tropomodulins: Pointed-end capping proteins that regulate actin filament architecture in diverse cell types. Cytoskeleton 69, 337–370.

Yang, X., Pabon, L., and Murry, C.E. (2014) Engineering Adolescence: Maturation of Human Pluripotent Stem Cell–Derived Cardiomyocytes. Circ Res 114, 511–523.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al (2007) Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science 318, 1917–1920.

Yu, Z.-H., Zhang, R.-Y., Walls, C.D., Chen, L., Zhang, S., Wu, L., Liu, S., and Zhang, Z.-Y. (2014) Molecular Basis of Gain-of-Function LEOPARD Syndrome-Associated SHP2 Mutations. Biochemistry 53, 4136–4151.

Zheng, D., and Gerstein, M.B. (2007) The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? Trends Genet. 23, 219–224.

Zheng, H., and Xie, W. (2019) The role of 3D genome organization in development and cell differentiation. Nature Reviews Molecular Cell Biology 20, 535–550.

Zhou, Q., Li, H., Li, H., Nakagawa, A., Lin, J.L.J., Lee, E.-S., Harry, B.L., Skeen-Gaar, R.R., Suehiro, Y., William, D., et al (2016) Mitochondrial endonuclease G mediates breakdown of paternal mitochondria upon fertilization. Science 353, 394–399.

Zhou, X., Di Stasi, A., and Brenner, M.K. (2015) iCaspase 9 Suicide Gene System. Methods Mol Biol 1317, 87–105.

ISBN 978-951-51-6873-3 (PRINT)
ISBN 978-951-51-6874-0 (ONLINE)
ISSN 2342-3161 (PRINT)
ISSN 2342-317X (ONLINE)
http://ethesis.helsinki.fi

HELSINKI 2020