From the Department of Molecular Medicine and Surgery
Karolinska Institutet, Stockholm, Sweden

GENETIC STUDIES OF RARE SKELETAL DISORDERS – TO SOLVE THE UNSOLVED

Dominyka Batkovskyte

Karolinska Institutet

Stockholm 2024

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2024

ISBN 978-91-8017-284-4

Genetic Studies of Rare Skeletal Disorders – To Solve the Unsolved

Thesis for Doctoral Degree (Ph.D.)

By

Dominyka Batkovskyte

The thesis will be defended in public at Eva & Georg Klein Lecture Hall, Karolinska Institutet, Stockholm, on May 8, 2024, 9:30 AM

Principal Supervisor:
Associate Professor Giedre Grigelioniene
Karolinska Institutet
Department of Molecular Medicine and Surgery

Co-supervisor(s):
Associate Professor Fulya Taylan
Karolinska Institutet
Department of Molecular Medicine and Surgery

Anna Hammarsjö, PhD
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Ann Nordgren
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Anna Lindstrand
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Andrei Chagin
Gothenburg University
Department of Internal Medicine and Clinical Nutrition

Opponent:
Professor Ravi Savarirayan
Murdoch Children's Research Institute
University of Melbourne

Examination Board:
Professor Maria Eriksson
Karolinska Institutet
Department of Biosciences and Nutrition

Professor Paul Gerdhem
Uppsala University
Department of Surgical Sciences

Professor Pétur B. Júlíusson
University of Bergen
Department of Clinical Science

Abstract

Congenital skeletal disorders, also called skeletal dysplasias, constitute a diverse group of rare genetic conditions that occur in approximately 1 per 3000-5000 births. Skeletal dysplasias exhibit both clinical and molecular heterogeneity, with the phenotype of the affected individuals varying based on the severity of the disease. While skeletal dysplasias primarily affect cartilage and bone, they often include extraskeletal manifestations, including internal organ abnormalities. Diagnosis of skeletal disorders includes clinical characterization, radiographic pattern recognition, and genetic testing. Despite the advancements in genome sequencing technologies, there remains a gap in understanding the genetic and phenotypic characteristics of some of the skeletal disorders, and new types of skeletal dysplasias continue to be identified.

The work in this thesis focused on identifying genetic causes and molecular mechanisms of rare skeletal disorders. Using genome sequencing the objective was to identify new genes or novel variants and improve the clinical and genetic characterization in the individuals with skeletal disease of unknown genetic background. The research goals were achieved through four constituent studies included in this thesis.

Study I identified *TOMM7* as a novel disease-causing gene in a patient with syndromic short stature and developmental delay. The molecular pathogenesis of *TOMM7* (c.73T>C, p.Trp25Arg) variant was studied in a mouse model. Tomm7 mutant mice showed growth restriction, reduced chondrocyte proliferation, and lipoatrophy caused by mitochondrial dysfunction.

Study II identified *ADAMTSL2* as a genetic cause of previously genetically uncharacterized skeletal dysplasia cortical dysostosis Al-Gazali type and connected it to the spectrum of ADAMTSL2-related disorders. This study included a cohort of nine individuals characterized by intrauterine growth restriction, distinct radiological patterns, and early mortality. Using primary dermal fibroblasts from the affected individual, we showed abnormal organization of the fibrillin-1 microfibrils.

Study III identified disease-causing variants in the *RAB34* gene in a fetus with complex malformations, including skeletal abnormalities consistent with short-rib thoracic dysplasia, and a combination of rarely occurring pre- and postaxial polydactyly.

Study IV focused on investigating gene dosage abnormalities in type I collagen genes *COL1A1* and *COL1A2* in eight patients from five unrelated families with Osteogenesis Imperfecta (OI). A complex structural variant, including deletion and duplication in *COL1A2*, was resolved by long-read genome sequencing in a patient with progressively deforming OI, characterized by multiple fractures and short stature.

This thesis has contributed to the field by broadening the genetic and phenotypic understanding of several congenital skeletal disorders. Two genes previously not associated with human phenotype, *TOMM7* and *RAB34*, were described. Furthermore, a genetic cause was identified for the previously unresolved Al-Gazali skeletal dysplasia. Genome sequencing serves as a first-tier research tool for identifying novel candidate genes and variants; however, linking these variants to specific diseases depends on a range of factors, including knowledge of gene and protein function and detailed phenotypic characterization of patient groups.

List of scientific papers included in the thesis

I. **A hypomorphic variant in the translocase of the outer mitochondrial membrane complex subunit *TOMM7* causes short stature and developmental delay**
Young C, Batkovskyte D, Kitamura M, Shvedova M, Mihara Y, Akiba J, Zhou W, Hammarsjö A, Nishimura G, Yatsuga S, Grigelioniene G, Kobayashi T. *HGG Advances, 2022 Oct 4;4(1):100148. doi: 10.1016/j.xhgg.2022.100148. PMID: 36299998*

II. **Al-Gazali skeletal dysplasia constitutes the lethal end of ADAMTSL2-related disorders**
Batkovskyte D, McKenzie F, Taylan F, Simsek-Kiper PO, Nikkel SM, Ohashi H, Stevenson RE, Ha T, Cavalcanti DP, Miyahara H, Skinner SA, Aguirre MA, Akçören Z, Utine GE, Chiu T, Shimizu K, Hammarsjö A, Boduroglu K, Moore HW, Louie RJ, Arts P, Merrihew AN, Babic M, Jackson MR, Papadogiannakis N, Lindstrand A, Nordgren A, Barnett CP, Scott HS, Chagin AS, Nishimura G, Grigelioniene G.
Journal of Bone and Mineral Research, 2023 May;38(5):692-706. doi: 10.1002/jbmr.4799. Epub 2023 Mar 27. PMID: 36896612

III. **Compound heterozygous variants in *RAB34* in a rare skeletal ciliopathy syndrome**
Batkovskyte D, Komatsu M, Hammarsjö A, Pooh R, Shimokawa O, Ikegawa S, Grigelioniene G, Nishimura G, Yamada T.
Clinical Genetics, 2024 Jan;105(1):87-91. doi: 10.1111/cge.14419. Epub 2023 Aug 24. PMID: 37619988

IV. **Structural variants in *COL1A1* and *COL1A2* in Osteogenesis Imperfecta**
Batkovskyte D, Swolin-Eide D, Hammarsjö A, Bilgrav Sæther K, Thunström S, Lundin J, Eisfeldt J, Lindstrand A, Nordgren A, Åström E, Grigelioniene G.
Manuscript, 2024

Contents

1 INTRODUCTION 1
1.1 HUMAN GENETICS 2
1.1.1 Patterns of inheritance 2
1.1.2 Genetic diagnostics 3
1.1.3 Consequences of genetic variants 4
1.2 CONGENITAL SKELETAL DISORDERS 6
1.2.1 What does it mean to solve the unsolved? 7
1.2.2 Clinical diagnostics of skeletal disorders 7
1.3 DEVELOPMENT OF HUMAN SKELETON 10
1.3.1 Transcription factors and signaling pathways regulating bone development 11
1.3.2 Extracellular matrix proteins in skeletal development 13
1.3.3 Cilia proteins in skeletal development 14
1.3.4 Mitochondria proteins in skeletal development 15
2 RESEARCH AIMS 17
3 MATERIALS AND METHODS 19
4 ETHICAL CONSIDERATIONS 23
5 RESULTS AND DISCUSSION 25
Study I - A hypomorphic variant in *TOMM7* leads to short stature and developmental delay 26
Study II - Biallelic variants in *ADAMTSL2* cause Al-Gazali skeletal dysplasia 29
Study III – Compound heterozygous variants in *RAB34* cause atypical skeletal ciliopathy 31
Study IV – Structural variants in *COL1A1/COL1A2* in Osteogenesis Imperfecta 33
6 CONCLUSIONS 37
7 POINTS OF PERSPECTIVE 39
8 ACKNOWLEDGMENTS 41
9 REFERENCES 43

List of abbreviations

aCGH	Array comparative genomic hybridization
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motif
ADAMTSL	ADAMTS-like
ATP	Adenosine triphosphate
CADD	Combined annotation dependent depletion score
COL1A1	Collagen type I alpha 1
COL1A2	Collagen type I alpha 2
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
FBN	Fibrillin
FGF	Fibroblast growth factor
GA	Gestational age
GoF	Gain-of-function
GPHYSD	Geleophysic dysplasia
GRCh37	Genome Reference Consortium Human Build 37
GS	Genome sequencing
HDF	Human dermal fibroblasts
HH	Hedgehog signaling pathway
LoF	Loss-of-function
lrGS	Long-read genome sequencing
MAF	Minor allele frequency
MLPA	Multiplex ligation-dependent probe amplification
mRNA	Messenger RNA
OCR	Oxygen consumption rate
OFDS	Oral-facial-digital syndrome
OI	Osteogenesis Imperfecta
OMIM	Online Mendelian Inheritance in Man

PCR	Polymerase chain reaction
RAB	Ras-associated binding protein
RNA	Ribonucleic acid
SDS	Standard deviation score
SNV	Single nucleotide variant
SRTD	Short-rib thoracic dysplasia
SV	Structural variant
TDO	Tricho-dento-osseous syndrome
TGFβ	Transforming growth factor beta
TOMM	Translocase of the outer mitochondrial membrane
TOP	Termination of pregnancy
UTR	Untranslated region
WHO	World Health Organization
µCT	Micro-computed tomography

"Look beyond what you see."
Rafiki from The Lion King

1 INTRODUCTION

The concept of heredity may have already been observed in ancient civilizations that speculated about the passing of traits from one generation to another. Hippocrates suggested that "invisible seeds" were given off from all organs in parental bodies to later contribute to the formation of the new body. It wasn't until the 19th century that systematic and scientific approaches were applied to the study of genetics. What we think of inheritance today is the path largely paved by Gregor Mendel and his experiments with pea plants in the mid-1800s. Mendel's observations that certain traits were predictable in the next generation and dominant over other traits laid the foundation for modern genetics. The discovery of the structure of the DNA molecule has further increased our understanding of the genetic code of life. Comprising four nucleotide bases — adenine (A), thymine (T), cytosine (C), and guanine (G) arranged in a 3 billion base pair sequence – the DNA molecule serves a universal language that spans across species, revealing the unity in the diversity of life.

A monumental collaborative effort of the Human Genome Project completed in 2003 marked a significant breakthrough, providing a comprehensive and near-complete map of human genes. The advances in technology, such as massively parallel or next-generation sequencing, have enabled large-scale genomic studies. Research has shifted towards understanding the functional implications of genes, the role of epigenetics, and the genetic basis of complex diseases. Human genetics research continues to evolve rapidly, with breakthroughs in gene editing and personalized medicine. The journey from early observations to modern genomic medicine highlights the relentless curiosity about life and the pursuit of knowledge that defines the field of human genetics.

1.1 HUMAN GENETICS

The human genome is a complete set of genetic material encoded in the DNA and organized within 23 pairs of chromosomes. Genes, the functional units within the genome, dictate the synthesis of proteins that govern the structure and function of cells. Only ~1.5% of the human genome is coding for proteins, while the remaining 98% of the genome consists of non-coding regions, such as regulatory elements, non-coding RNAs, repetitive DNA sequences, and other elements. The human genome contains approximately 20 000 protein-coding genes, but the number of functionally important elements is most likely much larger.

On the DNA level, humans are ~99.9% identical to each other, but the remaining 0.1% is what distinguishes us from one another. Single nucleotide variants (SNVs) are alternative forms of the DNA sequence produced by spontaneous events in the cellular machinery, also called mutations. The majority of DNA variants are neutral and do not affect the phenotype. Variants that are present in more than 1-5% of the population are considered to be common. Minor allele frequency (MAF) measures the frequency of the less common allele in the population. A rare variant is usually defined as being present in less than 1% of the general population (MAF < 1 %), but for many rare disorders, a disease-causing variant may be completely absent in publicly available genomic databases (1). It has to be taken into account though that these cutoff values are not completely stringent and while common variants are not typically associated with disease risk, they can have an impact on the population as a whole (2, 3).

1.1.1 Patterns of inheritance

According to Mendel's observations, some of the genetic traits or characters are present or absent depending on their genotype at a unique chromosomal location, called locus. If a specific feature or character is determined by variation in a unique chromosomal location, it is called Mendelian or monogenic. The dominant trait will be transmitted and present in the next generation (inherited in an autosomal dominant manner), while the recessive trait will only be present if it is inherited from both paternal alleles (autosomal recessive manner). The genes located on the X chromosome will be inherited in either an X-linked dominant or X-linked recessive manner. It is important to mention that the majority of human traits are much more complex and do not follow Mendelian

laws. Even traits like height and weight depend on multiple loci, whereas behavioral characteristics are strongly influenced by environmental factors (2, 4).

1.1.2 Genetic diagnostics

Genetic diagnostics has undergone a remarkable evolution from its early days to the present, driven by advances in technology, methodology, and our understanding of the genetic basis of diseases. From the early days of identifying chromosomal abnormalities through karyotyping or fluorescent in situ hybridization (FISH) to high-throughput sequencing technologies that constitute the foundation for genetic diagnosis today. Massively parallel sequencing (MPS) (sometimes also called next-generation sequencing, NGS) significantly accelerated the pace of genetic diagnostics by enabling the rapid sequencing of entire genomes, exomes, or targeted gene panels. Genome sequencing is in particular advantageous because of the ability to detect most types of variants in a single test approach. In addition to single nucleotide variants, small insertions and deletions, or copy number variation, whole genome analysis gives a full overview of non-coding regions, structural variants, and repeat expansions, enabling the researchers to identify causative variants in intronic, intergenic, and regulatory regions (5).

The integration of bioinformatics tools and data analysis pipelines has become essential for interpreting the vast amounts of sequencing data generated in genetic diagnostics. Typically, a whole exome analysis identifies 20,000 different variants whilst in whole genome analysis the number of different variants can be up to 5 million per individual (6). Bioinformatic tools provide the annotation, filtering, and prioritization of genetic variants, which enables clinicians and researchers to generate candidate gene lists for further investigation.

Recent advancements in genetic diagnostics include the adoption of RNA sequencing (RNA-seq), long-read sequencing technologies, proteomics, methylation analysis, and the application of artificial intelligence and machine learning algorithms for predicting variant pathogenicity, disease risk, or treatment outcomes (7, 8). By integrating information about gene expression, transcript and protein structure, epigenetic changes, and genomic variations, clinicians and researchers can gain deeper insights into the molecular mechanisms underlying diseases, leading to improved diagnosis, prognosis, and personalized treatment strategies. However, challenges such as cost, data

analysis complexity, and the need for specialized expertise remain barriers to the widespread adoption of these technologies.

1.1.3 Consequences of genetic variants

One of the fundamental aspects of studying genetic variants is understanding their effect on the protein level. A variant can have a quantitative effect, i.e. causing a complete loss, reduction, or increase of the gene product without changing its functional characteristics. The function of the protein remains the same, however, its amounts are below or above the normal. On the other hand, a variant can have a qualitative effect, and affect the structural or regulatory properties of the protein or change its function. Altering or diminishing the function of the protein does not always lead to disease. The two questions that are important to answer are: how a variant affects a gene product or protein, and how a variant affects the person carrying it.

The first question can be answered by reviewing the principal examples of the consequences of genetic variants on the protein:

Loss-of-function (LoF) variants:

- Complete LoF variants (also called amorphic or null) – leading to complete loss of the functional gene product.
- Hypomorphic variants – causing only a partial loss of gene product and often leading to a milder phenotype than complete LoF.

In some cases, LoF variants can cause haploinsufficiency, a situation, where an amount of gene product from the non-affected allele is not sufficient for a healthy phenotype.

Gain-of-function (GoF) variants:

- Hypermorphic variants – leading to an increase in the activity of a wild-type protein. The protein is produced but it functions inappropriately (for example, a receptor signaling even in the absence of its ligand) or constitutively active protein.
- Neomorphic variants – leading to the acquisition of a new function.

A dominant-negative effect refers to mutations that unfavorably change the function of the wild-type protein. The abnormal protein resulting from a

dominant-negative variant not only fails to function correctly but also actively interferes with the wild-type form. Proteins that build multimeric structures, such as type I collagen, are particularly sensitive to dominant-negative effects (2).

The answer to a second question on what a variant does to a person is rather complex. The concept of genotype-phenotype correlation, which aims to describe the connection between specific genetic variants and the traits resulting from those variants, has been a long-pursued goal in molecular genetics. While it holds true for some conditions, the relationship between genotype and phenotype is not always a direct one-to-one correlation. Individuals with the same genotype can exhibit variable phenotypes due to differences in genetic makeup, gene expression, and environmental factors. Many diseases show a phenotype continuum, ranging from the most severe to milder phenotypes. It is not uncommon that variants in the same gene can manifest in different phenotypes, and that proteins have both canonical (standard roles within established pathways) and non-canonical (alternative roles or within less established pathway) roles in an organism (4). Novel discoveries often challenge the status quo in biology and constantly contribute to increased knowledge in the field. The overall impact of protein dysfunction and its contribution to disease depends on multiple factors, such as compensatory mechanisms in the cell, individual differences, as well as environment, and stochastic events.

1.2 CONGENITAL SKELETAL DISORDERS

Historical evidence depicting short-statured or very tall individuals shows that skeletal disorders have been part of humankind's history for many years. Antique art remains from ancient Greece and Egypt show short-statured and disproportionate individuals. The Roman emperor Maximinus I was portrayed as a man of incredible size, and perhaps with features of acromegaly (9). A selection of portraits by Spanish painter Diego Velázquez in the 17th century depicts a set of individuals likely with chondrodysplasia or pseudoachondroplasia phenotypes (10).

The systematic classification of congenital skeletal disease families began in the beginning of the twentieth century when more and more clinical entities were recognized based on their specific radiological appearance (11). The identification of a unique collection of radiological signs, also called radiographic patterns, in an affected individual constitutes the basis of diagnostics of skeletal disorders even today. When genetics took a rapid development from the middle of the twentieth century and onwards it enabled pediatricians and physicians to confirm the diagnosis, establish genotype-phenotype correlations, and decipher molecular mechanisms of congenital skeletal disorders (12). As of 2024, there are 771 diseases associated with pathogenic variants in 552 genes listed in the 'Nosology of genetic skeletal disorders' (13). The 41 disease groups in the Nosology enable clinicians and molecular geneticists to recognize genotype and phenotype similarities across different conditions, while also illustrating the complexity of human disease genetics. Genetic variants in a vast array of proteins, including enzymes, scaffolding proteins, signal transductors, cilia proteins, extracellular matrix components, membrane transporters, and others can lead to skeletal disease. Furthermore, disease-causing variants can be located in non-protein coding genes, as evident by miR140-related spondyloepiphyseal dysplasia (OMIM #618618) or the long non-coding RNA RMRP-related cartilage-hair hypoplasia (OMIM #250250) (13, 14). The multitude of clinical entities also highlights the diversity of skeletal dysplasia phenotypes, characterized by either short or abnormally tall stature, multiple bone deformities, absence, or presence of additional digits, variations in bone mineral density, and involvement of other organ systems. The phenotypic heterogeneity of skeletal dysplasia is further evident in terms of survival, with some genetic skeletal disorders leading to perinatal or early lethality while others are fully compatible with life and resulting in only minor difficulties.

1.2.1 What does it mean to solve the unsolved?

The estimated frequency of congenital skeletal disorders is around 1 in 4000-5000 births, which makes it a subgroup of rare diseases. In the European Union, rare diseases are defined as conditions affecting fewer than 1 in 2000 individuals, with over 400 million people globally living with such conditions (15, 16). While individually rare, collectively rare diseases are common and form a group of 7000 different conditions. Despite significant progress in genomic technology, only 30-40% of individuals living with rare genetic diseases receive accurate diagnosis. Several challenges contribute to this diagnostic gap. For rare conditions, the small number of affected individuals makes clinical and genetic recognition of the disease challenging (17). Moreover, it is now recognized that many of the genetic diseases do not follow traditional Mendelian patterns of inheritance and it is likely that mechanisms such as epigenetic changes, tissue-specific gene expression, and oligogenic inheritance contribute to disease. Additionally, despite the massive amounts of generated data, interpreting it is still a limiting factor. The limited functional evidence for the affected protein makes the variant interpretation and association with the phenotype very challenging. Lastly, inequitable access to genetic testing remains a challenge in improving diagnostic outcomes. Limited access to specialized healthcare services in underserved areas and inconsistent health insurance coverage create a significant barrier to those living with a rare disease (15, 16). Therefore, the future challenge lies in developing out-of-the-box as well as more accessible approaches to solve the yet unsolved cases of rare diseases, including congenital skeletal disorders.

1.2.2 Clinical diagnostics of skeletal disorders

Skeletal dysplasia can already be suspected in utero during a routine ultrasound examination when short limbs are observed. A detailed two-dimensional ultrasound enables accurate measurements of long bones and helps to assess the presence of thoracic hypoplasia or any spinal abnormalities (18). Furthermore, a three-dimensional ultrasound can enable visualization of dysmorphic facial features and positioning of the fingers and toes. In some cases, low-dose fetal computed tomography (CT) can be used when precise details of bone structure are crucial for diagnosis (18). Fetal diagnosis is challenging due to the small size of the fetus, fetal motion, and limitations of

certain imaging techniques. However, an early diagnosis of severe conditions helps provide a prognosis and assists the family to make informed decisions.

During childhood, early detection of an abnormal growth pattern can help in the rapid establishment of a diagnosis. This is done by following the growth curve of the child and comparing it to the reference standard, such as WHO Child Growth Standards or population-specific reference. The concept of disproportion is important when diagnosing skeletal dysplasia and distinguishing it from an idiopathic short stature. Newborn children often have a distinct appearance, with larger heads and longer trunks in relation to their shorter extremities, giving them a seemingly uneven proportion. However, at around 10 years of age, the length between the extended hands should become approximately the same as the height of a child. By the age of 11, the upper-to-lower segment ratio should be about 1.0, and when standing, the hands should lie at 1/3-1/2 the distance down the thighs. Notably, the proximal segment (humerus/femur) should be longer than the middle segment (forearm/lower leg), which in turn should be longer than the distal segments (hands/feet). Disproportionate appearance arises when there are deviations in these proportions, such as limbs or trunk being shorter than normal (19). The three key examples of limb shortening are rhizomelia – shortening of the proximal bones (humerus and femur), mesomelia – shortening of the middle segment (radius and ulna, or tibia and fibula), and acromelia – shortening of the distal segments (hand and feet bones) (**Figure 1**).

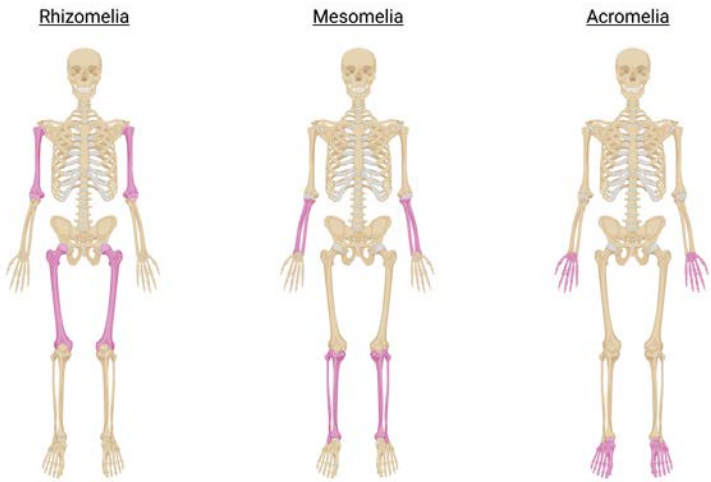

Figure 1. Examples of typical bone shortening. The shortening of all three segments is called micromelia. *Created with BioRender.com*

Most congenital skeletal diseases have specific dysmorphology patterns and unique presentation making the radiographic examination a fundamental step in the diagnostics of skeletal disorders. A skeletal survey should include images of the skull (anterior and lateral), spine (anterior and lateral), thorax, pelvis, one upper and one lower limb, and one hand if the abnormalities are symmetrical. It was recognized in the 1980s that different skeletal dysplasias share similarities in their radiographic presentation (pattern) and can be grouped into "skeletal dysplasia families" (20). The development of genetic and molecular analysis methods has further validated the concept that congenital skeletal disorders sharing similarities in their radiological pattern will most likely have similar molecular pathogenesis. Thus, the sharp eyes of an experienced physician can pinpoint unique features in an X-ray image and generate a good clinical hypothesis for further genetic investigation (21).

Moreover, biochemical parameters, such as levels of calcium and phosphate, magnesium, bone-specific alkaline phosphatase (ALP), vitamin D, parathyroid hormone (PTH), bone mineral density measurements, and bone fracture history can also indicate the health status of the individual's skeletal system.

It is important to note that the diagnosis should be based on the integration of all available findings, rather than a single observation. The overall combination of the external morphology, radiographic features, biochemical parameters, and family history is necessary for accurate diagnosis of congenital skeletal disorders, which then has to be confirmed by genetic analysis.

1.3 DEVELOPMENT OF HUMAN SKELETON

The human skeleton consists of 206 bones and has an important structural, mechanical, and homeostatic function. Formation of the skeleton initiates at 6-7 weeks of gestational period and continues up until 25 years of age (22). Human bones are formed through two distinct developmental processes – intramembranous and endochondral ossification. Intramembranous ossification occurs when bone-forming cells osteoblasts differentiate directly from mesenchymal stem cells, and accounts for the formation of craniofacial bones and clavicles (23). Endochondral ossification drives the formation of all remaining bones in the body and occurs via the differentiation of mesenchymal stem cells into cartilage-forming cells chondrocytes, which later mineralize to form the long bone (23, 24). The cartilaginous structure responsible for elongation of long bones enabling children to grow is called the growth plate (**Figure 2**)

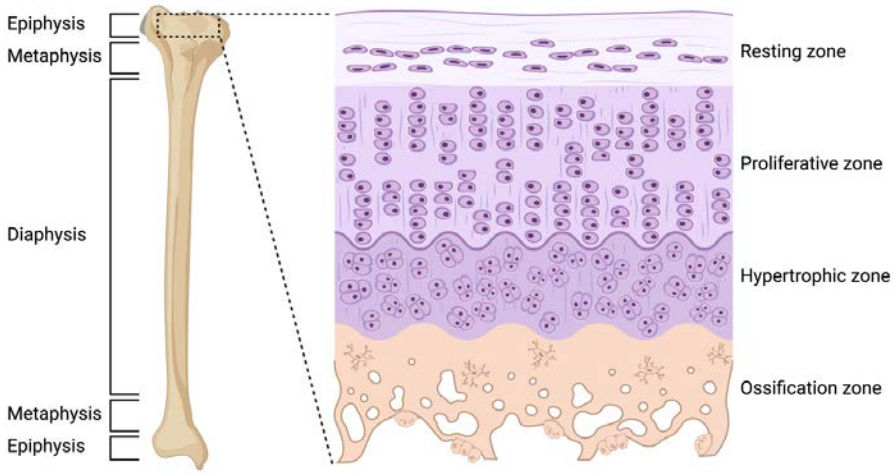

Figure 2. Structure of a long bone and epiphyseal growth plate. *Created with BioRender.com*

The growth plate is located at the proximal and distal ends of the long bones called epiphyses and remains active until puberty. The cells of the growth plate, chondrocytes, are arranged in three different zones, each of them having different cellular properties important for cell division, differentiation, and growth (24). The least differentiated chondrocytes reside in the resting zone and mainly provide cells to the proliferative zone. In the proliferative zone, cells undergo a rapid division and align in columns. Upon exiting the cell cycle, chondrocytes differentiate into hypertrophic chondrocytes occupying the hypertrophic zone.

The development of the bone tissue from cartilage occurs when hypertrophic chondrocytes commit to apoptosis or trans-differentiation into bone-forming cells osteoblasts, followed by invasion of blood vessels, and production of bone matrix (24). Although bone may seem an inert structural element in the musculoskeletal system, it constitutes a dynamic tissue that is continuously remodeling in response to mechanical and physiological stimuli. This inherent capacity is vital for the maintenance of bone strength and homeostasis.

1.3.1 Transcription factors and signaling pathways regulating bone development

Development of the skeleton is a tightly regulated process, dependent on multiple transcription factors and signaling pathways involved in limb bud patterning, chondrocyte differentiation, secretion of matrix proteins, and osteoblast differentiation (25). Some of the key molecular players in skeletal development are reviewed below.

Transcription factor SOX9 is a master regulator of chondrogenesis expressed by early skeletal precursors. It activates the expression of chondrocyte differentiation markers and acts together with SOX5 and SOX6 to ensure proper chondrocyte differentiation (23, 26). Loss-of-function mutations in *SOX9* lead to campomelic dysplasia (OMIM #114290), and the evidence from mouse studies suggests that loss of Sox9 leads to defects in initial mesenchymal condensations, premature chondrocyte hypertrophy, and mineralization (25).

Transcription factor RUNX2 is a major regulator of osteoblast differentiation and is highly expressed by prehypertrophic chondrocytes during development. Loss-of-function mutations in *RUNX2* lead to cleidocranial dysplasia (CCD, #119600), the main features of which are absent clavicle bones and delayed ossification of the skull bones. Gain-of-function mutations in *RUNX2* lead to metaphyseal dysplasia with maxillary hypoplasia (OMIM #156510) (23).

Transcription factor Osterix (OSX or SP7) regulates osteoblast differentiation and maturation. In the presence of OSX, RUNX2-positive cells are osteoblast precursors with low expression of SOX9 and thus can differentiate into osteoblasts. OSX absence in RUNX2-positive cells keeps them in the chondrogenic state (23). Disease-causing variant in *SP7* lead to a moderate form of Osteogenesis Imperfecta (OMIM #613849).

The Hedgehog (HH) signaling pathway is important for morphogenesis and patterning of many tissues, including the skeleton. Sonic hedgehog (SHH) acts in the early epithelial–mesenchymal transition of skeletal mesenchyme, promoting the formation of vertebrae and ribs, as well as ensuring digit patterning (23). Pathogenic variants in *SHH* or its associated transcription factor GLI3 cause polydactyly or syndactyly phenotypes. Another member of the HH family, the Indian hedgehog (IHH) is expressed in the growth plate and together with parathyroid hormone-related peptide (PTHrP) forms a negative feedback loop, promoting chondrocyte proliferation. Autosomal dominant mutations in *IHH* lead to brachydactyly type 1A (OMIM #112500), and recessive mutations lead to acrocapitofemoral dysplasia (OMIM #607778), showing dual IHH role in both limb patterning and chondrocyte proliferation/differentiation (25).

Fibroblast growth factor (FGF) family proteins act via cell surface FGF receptors FGF1-4 and act in activating multiple signaling pathways, such as JAK-STAT, PKC, MAPK, and PI3K (23). Two hotspot *FGFR3* gain-of-function mutations lead to the most common non-lethal skeletal dysplasia, achondroplasia. Interestingly, mutations located in extracellular and intracellular domains of the FGFR3 protein cause a lethal thanatophoric dysplasia. This phenomenon, where mutations in different parts of the protein cause a spectrum of phenotypes, depends on the degree of overactivation and can be associated with the location of the mutations within the protein. It has been shown that FGFR3 controls bone growth by suppressing chondrocyte differentiation and proliferation via STAT1 and MAPK pathways (25).

The importance of the transforming growth factor beta (TGFβ) pathway has been studied via multiple animal knockout models of the family proteins, including the receptors and SMAD proteins (27). The common feature in these models is affected endochondral and intramembranous ossification. TGFβ is essential for the development of cartilage cells by promoting the mesenchymal condensations and their differentiation into chondrocytes. In the osteoblast lineage, TGFβ is essential for proliferation and early differentiation of osteoprogenitor cells. Disease-causing variants in TGFβ receptors, as well as SMAD2/3 proteins, lead to Loeys-Dietz syndrome, associated with tall stature (23). The pathogenic mechanisms of several skeletal dysplasias are thought to be through abnormalities in TGFβ signaling and, for example, TGFβ inactivating antibodies are currently being tested in the clinical trials for Osteogenesis Imperfecta (22).

In addition to the above-mentioned signaling pathways, several other major cellular signaling pathways are crucial for bone development. The BMP pathway is closely intertwined with TGFβ and is especially important for patterning of the distal limbs. Mutations in WNT pathway members, WNT ligands, FZD receptors, and LRP co-receptors have also been associated with multiple congenital skeletal disorders (23, 25, 26).

1.3.2 Extracellular matrix proteins in skeletal development

The extracellular matrix (ECM) has a well-established structural function in multiple cells and its key importance for skeletal tissue development is reflected by the magnitude of mutations in ECM-relevant genes leading to skeletal dysplasias.

Collagens are one of the most common ECM proteins in the human body and play a key role in the homeostasis of various connective tissues, including the skeleton. The signature amino acid sequence Gly-X-Y is important for the proper assembly and structure of the triple-helical collagen protein (28). Disease-causing variants in type I collagen (*COL1A1* or *COL1A2*) account for ~85% of Osteogenesis Imperfecta cases, and the majority of the variants are missense substitutions of the glycine residues (29). Copy number variants or structural variants (SVs) leading to quantitative changes of type I collagen have also been reported in the literature (30, 31, 32, 33). The most common feature of OI is bone fragility leading to frequent fractures. The severity of the disease varies depending on the variant, with qualitative defects of type I collagen generally being more severe than the quantitative (also known as the dominant-negative effect). In other cases, mutations that interfere with type I collagen production and affect osteoblast differentiation can lead to the disease phenotype (34). In addition, disease-causing variants in other genes essential for type I collagen synthesis or function can also cause OI phenotype. Currently, there are 23 genes associated with Osteogenesis Imperfecta (13).

Abnormalities in type II collagen (*COL2A1*), the main collagen of the hyaline cartilage, lead to a group of skeletal dysplasias called type II collagenopathies, such as spondyloepiphyseal dysplasia congenita (SEDC) or Kniest dysplasia. The main clinical presentation of individuals with type II collagenopathies is short stature, short neck, and trunk, with additional features, such as spinal or joint abnormalities, and midface hypoplasia (35).

A group of skeletal dysplasias called acromelic dysplasias are caused by pathogenic variants in genes encoding secreted ECM proteins, such as FBN1, ADAMTSL2, ADAMTS10, ADAMTS17, LTBP2, and LTBP3. Collectively they share characteristic features such as short stature, brachydactyly, joint stiffness, pseudomuscular build, and thick skin (36). One of the conditions in the acromelic dysplasia group is geleophysic dysplasia (GPHYSD), caused by pathogenic variants in the *ADAMTSL2* gene. The affected individuals are characterized by severe short stature, small hands and feet, limited joint movement, joint contractures, unique facial features, tip-toe walking, and thickened skin. Skeletal features include wide proximal phalanges, cone-shaped epiphyses, delay in bone age, short tubular bones, and vertebral abnormalities (36, 37). In addition to the skeletal abnormalities, recurrent respiratory and cardiac problems are common and typically result in premature mortality (36).

1.3.3 Cilia proteins in skeletal development

Cilia are cellular membrane-bound organelles located on most types of the eukaryotic cells. There are two types of cilia – motile and non-motile cilia. Motile cilia are located on sperm cells, epithelial cells in the airway and middle ear, oviduct cells, and ependymal cells lining the brain ventricle and are associated with generating fluid flow or movement within the fluids (38, 39). Non-motile cilia, also called the primary cilia, function as key sensory organelles on the cell surface and have a crucial function in regulating cell signaling and transduction, contributing to overall tissue development and organogenesis (39). The primary cilia line almost all cells in the human body. Dysfunction of motile or non-motile cilia leads to ciliopathies, a disease group with a wide range of morphological features, involving nearly every organ. Skeletal ciliopathies are a subgroup of ciliopathies that exhibit distinct skeletal features and are classified based on their radiographic presentation. Typical skeletal abnormalities in skeletal ciliopathies include narrow thorax, short ribs, and short tubular bones, metaphyseal abnormalities, hypoplastic pelvis, trident-shaped acetabulum, brachydactyly with cone-shaped epiphyses, and polydactyly. Additionally, variable non-skeletal manifestations are often present, including renal cystic dysplasia, retinopathy, malformations of the central nervous system, heart, and genitalia. Currently, pathogenic variants in at least 30 genes are associated with skeletal ciliopathies, with the majority inherited in an autosomal recessive manner. Examples of common clinical entities are short-rib thoracic dysplasias,

chondroectodermal dysplasia (Ellis-van Creveld syndrome), Meckel syndrome, and some of the subtypes of the orofaciodigital syndromes (13, 40). Skeletal ciliopathies are complex and can involve a wide range of clinical manifestations, making accurate diagnosis and management challenging. Accompanying abnormalities of internal organs are common, and phenotypes range from severe lethal conditions to milder abnormalities compatible with survival. Most of the genes associated with skeletal ciliopathies encode proteins of the intraflagellar transport machinery (IFT) or dynein motor components. IFT has an important role in regulating the Hedgehog (Hh) signaling pathway. Thus, developmental abnormalities in syndromic skeletal ciliopathies often are a result of compromised Hh signaling (38). Multiple small GTPases, including the Rab family, have been shown to have a critical role in ciliary formation, function, transport, and signaling (41).

1.3.4 Mitochondria proteins in skeletal development

A mitochondrion is an organelle present in almost all eukaryotic cells. It is crucial for energy production and conversion, regulation of cellular proliferation, storage of calcium ions, and several other metabolic functions (42). Mitochondria has its own mitochondrial genome (mtDNA) that is inherited maternally. However, mtDNA comprises only a very small part of DNA, and thus many genes located in the nuclear genome encode proteins involved in mitochondrial function. Mitochondrial diseases are heterogeneous and can lead to a wide variety of phenotypes, including neurological, respiratory, renal, muscular, and skeletal abnormalities (42). Mitochondrial bioenergetics and quality control mechanisms are important for stem cell regulation in bone homeostasis, and abnormal energy metabolism as well as oxidative stress can contribute to malfunctioning stem cells in bones (43). Furthermore, mitophagy, a removal process of damaged mitochondria, plays a key role in osteoblast and osteoclast proliferation, differentiation, and function. Mitophagy dysregulation can lead to an increase in damaged mitochondria, which in turn leads to oxidative stress and apoptosis of the bone cells (44, 45). Pathogenic variants in several genes coding for mitochondrial proteins have been associated with growth failure and skeletal dysplasia phenotype in humans. Some of the examples include disease-causing variants in *PISD* leading to spondyloepimetaphyseal dysplasia (SEMD) with microcephaly, *IARS2* leading to short stature with mild dysmorphic features, also

called CAGSS syndrome, and *AIFM1* in spondyloepimetaphyseal dysplasia with leukodystrophy (13).

2 RESEARCH AIMS

The overall aim of this thesis was to study genetic causes and molecular pathogenesis of genetic skeletal disorders and to increase the understanding of the clinical course and molecular mechanisms associated with the disease.

The specific goals of the thesis were:

- To identify novel genetic causes of congenital skeletal disorders
- To further expand the genetic and phenotypic spectrum of known skeletal dysplasias
- To study the effects of the identified disease-causing variants in primary human cells or animal models

3 MATERIALS AND METHODS

Study participants

Study participants were patients referred to the Department of Clinical Genetics and Genomics at Karolinska University Hospital, Stockholm, Sweden because of the suspected congenital skeletal disorder and were invited to join the research study "Clinical and Molecular Studies of Congenital Skeletal Diseases" at Karolinska University Hospital, Stockholm, Sweden. International collaborations were facilitated by the International Skeletal Dysplasia Society (ISDS), Nordic Skeletal Dysplasia Council, and other previously established collaborations between the referring physicians. Clinical data, radiograms, growth parameters, and other relevant data were collected from the patients' journal files by the referring physicians. Informed written consent for participation in the study and the publication of images was obtained from patients, their parents, and/or legal guardians.

Genome sequencing

DNA extraction from peripheral blood, dermal fibroblasts, saliva, or other tissue was done at the Department of Clinical Genetics and Genomics, Karolinska University Hospital, or referring physicians' sites according to standard procedures.

Genome sequencing has been a first-line test in routine clinical diagnostics at Karolinska University Hospital since 2015. DNA samples were submitted for genome sequencing on Illumina HiSeq X Ten or Illumina NovaSeq 6000 with a median coverage of 30× at Clinical Genomics Science for Life Laboratory (SciLifeLab, Stockholm, Sweden). Genome sequencing data was processed as described in Stranneheim et al. (46). Briefly, the bioinformatic analysis includes detection of single nucleotide variants (SNVs), insertions and deletions (INDELs), structural variants (SVs), uniparental disomy (UPD), and repeat expansions. Called and annotated variants are prioritized by a rank model including multiple parameters, such as inheritance pattern, variant conservation, population and local allele frequencies, Combined Annotation Dependent Depletion (CADD) score, and predicted impact on the protein.

Variant segregation was done using Sanger sequencing according to standard protocols and samples were sequenced at the KIGene core facility (Karolinska Institutet).

Long read genome sequencing (lrGS) on DNA sample from patient 1 in Study IV was performed using Oxford Nanopore Technologies (ONT) PromethION (NGI Stockholm, Sweden). lrGS analysis was done using an in-house developed pipeline LOMPE (https://github.com/kristinebilgrav/LOMPE).

Variant analysis and interpretation

Initial genetic analysis was done by analyzing the in-house developed Skeletal Dysplasia panel (671 genes as of 2024) and OMIM morbid gene panel (4537 genes). The research-setting analysis included analysis of all variants in the genome sequencing data. The variants were prioritized according to inheritance pattern, allele frequency, and predicted effect on the protein. For each individual case, a candidate gene list was compiled, and an in-depth literature review was done to evaluate the feasibility of the identified variants for further investigation.

Mouse model in Study I

The mouse model and all experimental data in Study I was generated at Massachusetts General Hospital by the collaborating authors. Detailed methodological description is available in Young C et al. (47). The animal study was approved by the Institutional Animal Care and Use Committee (IACUC) at Massachusetts General Hospital (#2009N000217).

Human dermal fibroblast culture

Primary human dermal fibroblasts (HDF) derived from a patient skin biopsy were used in Studies II and IV. For controls, commercially available human neonatal dermal fibroblasts (HDFn, purchased from Thermo Fisher Scientific, cat. no. C-004-5C and Sigma, cat. no. 106-05N) and human adult dermal fibroblasts (HDFa, purchased from ATCC, cat. no. PCS-201-012) were used. Cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with high-glucose, GlutaMAX (Thermo Fisher Scientific), 10% fetal bovine serum (FBS) (Thermo

Fisher Scientific), 1× non-essential amino acid solution (NEAA) (Thermo Fisher Scientific), and 0.2% primocin (InvivoGen) at 37°C and 5% CO_2.

RNA extraction and cDNA synthesis for splice variant analysis and qPCR

RNA from HDF was extracted using TRIzol™ Reagent (Invitrogen) and Direct-zol™ RNA extraction kit (Zymo Research). RNA concentration and quality were determined using Qubit (Thermo Fisher Scientific). cDNA was synthesized using Maxima First Strand cDNA synthesis kit (Thermo Fisher Scientific). RT-PCR was performed according to standard procedures. qPCR expression analysis was done using the SsoAdvanced Universal SYBR® Green kit (Bio-Rad) on Bio-Rad 96QX Thermocycler according to the manufacturer's recommendations. Relative expression was calculated using $\Delta\Delta$ C_t method.

Droplet digital PCR (ddPCR)

Probes against *ADAMTSL2* (target gene, FAM-labeled) and *HBB* (reference gene, HEX-labeled) were purchased from Bio-Rad and used with 2X ddPCR Supermix for Probes (No dUTP) according to the manufacturer's recommendations. A maximum of 1300 ng of RNA was used for generating cDNA. Droplets were generated by a QX100 Automated Droplet Generator (Bio-Rad) using DG8 Cartridge (Bio-Rad), and thermal cycling was performed according to the manufacturer's recommendations in C1000 Thermal Cycler (Bio-Rad). The droplets were scanned by QX200 Droplet Reader (Bio-Rad), and the analysis was done using QuantaSoft PRO Analysis Software (Bio-Rad).

Western blot

Western blot experiments were performed according to standard procedures. Briefly, protein from HDF was extracted using RIPA lysis buffer (Thermo Fisher Scientific) supplemented with Halt™ protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific). Protein concentration was measured using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). Protein lysates were run on 4%-15% gradient Mini-PROTEAN TGX™ precast gels (Bio-Rad) and transferred onto PVDF or nitrocellulose membrane. Membrane was blocked in 5% BSA for 1 hour at

room temperature followed by primary antibody incubation at 4°C overnight. Proteins were visualized using Clarity Western ECL substrate (Bio-Rad), and membrane was developed using LI-COR C Di-Git blot scanner or ChemiDoc MP Imaging System (Bio-Rad). Quantification was done using ImageJ.

Immunofluorescence

HDF were cultured in 24-well black plates with flat coverslip bottom or µ-slide 8-well chamber slides (ibidi). Fixation was done for 10 minutes at 37°C using Image-iT Fixative Solution (Thermo Fisher Scientific). Cells were permeabilized for 10 minutes at room temperature using 1× eBioscience Permeabilization Buffer (Thermo Fisher Scientific). Cells were blocked in Serum-Free Protein Block (DAKO) for 1 hour at room temperature followed by primary antibody incubation at 4°C overnight. Secondary antibody incubation was done for one hour at room temperature. Nuclei were stained using DAPI solution (Thermo Fisher Scientific). Cells were mounted in ProLong Diamond Antifade Mountant (Thermo Fisher Scientific), and images were taken on LSM710 confocal microscope (Zeiss) at Karolinska Institutet Biomedicum Imaging Core Facility (BIC). Quantification was done using ImageJ software.

4 ETHICAL CONSIDERATIONS

Genetic testing raises complex ethical and legal issues, including concerns about privacy, consent, and potential misuse of genetic information. Ensuring that genetic studies are conducted ethically and with respect to individual autonomy is crucial. All studies in this thesis complied with the ethical regulations stated by the Swedish Ethical Review Authority (Etikprövningsmyndigheten), local ethical committees in the collaborating institutions, and the Declaration of Helsinki. Written informed consent was obtained from patients, their parents, and/or legal guardians.

When any type of genetic testing is performed an individual must be informed of all possible outcomes and implications associated on a personal and family level. In some cases, genetic testing can lead to ethical challenges, such as disclosure of parental relationship or revealing the genetic status of other family members without their willingness to know. Results obtained from genetic testing that are not related to the primary condition of the individual are called incidental or secondary findings. Secondary findings occur in ~1–4% cases of genetic testing (48). Using massively parallel sequencing in clinical diagnostics significantly increases the risk of detecting such variants. Some of the genes, for example, *TP53* or *BRCA1* are known to be associated with increased risk of cancer and early detection of variants in these genes can lead to actionable disease prevention. If, during the clinical routine analysis of the skeletal dysplasia panel a significant variant is detected (such as in *BRCA1* or *CFTR*), these findings are reported to the family. Extended genetic analysis often includes healthy parents and siblings. DNA or RNA sampling from healthy family members allows the investigators to exclude rare variants shared in the family and examine the Mendelian inheritance patterns, altogether leading to a more accurate diagnosis. Moreover, clinical experience shows that providing accurate genetic diagnosis is very important for the psychological well-being of the family as well as disease management. To make sure that all the data is safe, the samples are coded and pseudonymized.

To claim the variant's pathogenicity, one needs to provide functional evidence. This is usually achieved by using primary cells from the affected individuals and/or animal models. Both approaches raise ethical discussions but are an inseparable part of translational research. We address these issues by carefully approaching sampling from humans, for example, the skin biopsy is usually taken

as a part of the planned surgery, which is common for skeletal dysplasia patients. Animal models allow more in-depth mechanistic investigations of detected variants. Moreover, it also enables researchers to access bone and cartilage tissue, which is rarely available to receive from human patients. The mouse model used in Study I was developed by using a novel genome editing technology i-Gonad, which significantly reduces the number of experimental animals needed (49, 50).

5 RESULTS AND DISCUSSION

The main results of this thesis are:

- Identification of two novel genes previously not associated with human phenotype: *TOMM7* (Study I) and *RAB34* (Study III)
- Expanding the phenotypic and genotypic spectrum of ADAMTSL2-related skeletal dysplasia by adding Al-Gazali skeletal dysplasia to the lethal end of Geleophysic Dysplasia type I spectrum (Study II)
- Further genetic characterization of type I collagen-related Osteogenesis Imperfecta (Study IV)

A summary of the most important findings in each study is presented throughout the upcoming pages. More detailed results can be found in the corresponding publications of this thesis.

Study I - A hypomorphic variant in *TOMM7* leads to short stature and developmental delay

In Study I, we identified *TOMM7* as a disease-causing gene in a male infant with syndromic short stature and developmental delay (47). The patient exhibited progressive growth failure during the postnatal period (65.5 cm (-5.4 SDS) at 15 months of age, SDS calculated according to WHO Child Growth Standards). The main clinical findings are summarized in **Figure 3**. The patient had a narrow thorax, nystagmus, hypotonia, and progressive breathing abnormalities. Post-mortem tissue analysis revealed prominent brown adipose tissue and lipid accumulation in the liver. The skeletal abnormalities included small facial bones, thin ribs, platyspondyly, slender tubular bones with large epiphyses, small metacarpals, and short middle phalanges. The patient deceased at the age of 2 years and 7 months after developing pneumonia and respiratory failure.

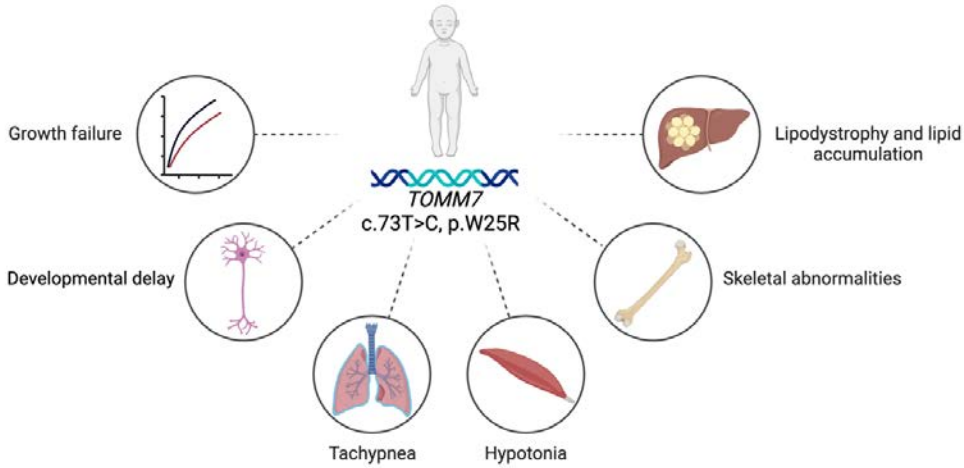

Figure 3. Illustrative summary of main findings in the patient with disease-causing variant in *TOMM7*. *Created with BioRender.com*

Genome sequencing analysis identified a homozygous missense variant in the *TOMM7* gene (NM_019059.5:c.73T>C, p.(Trp25Arg)). To study the functional consequences of the *TOMM7* c.73T>C, p.W25R variant, a mouse model (mutation-specific knock-in *Tomm7*$^{R/R}$ and functionally null *Tomm7*$^{-/-}$) was developed by collaborators at the Harvard University and Massachusetts General Hospital. The main findings from the mouse model are summarized in the figure below:

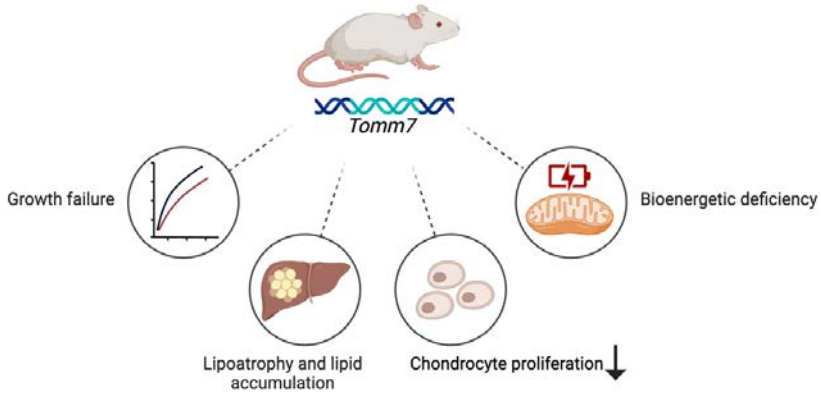

Figure 4. Illustrative summary of main findings in *Tomm7*-mutant mice.
Created with BioRender.com

Tomm7$^{R/R}$-mutant mice grew relatively normally until 8 weeks of age but developed rapidly progressing fat loss, followed by death at 10 weeks of age. Even more severe growth failure was observed in *Tomm7*$^{-/-}$ mice leading to death at 3-4 weeks of age. Lipoatrophy and lipid accumulation were seen in the liver of Tomm7-mutant mice, consistent with the fat loss observed in the patient. µCT analysis of the spine showed mild shortening and hypomineralization of the vertebral bodies in Tomm7-mutant mice compared to wild-type controls. Analysis of the tibial growth plates showed reduced chondrocyte density and proliferation in Tomm7 mutants compared to controls. The general phenotype observed in knock-in mice (*Tomm7*$^{R/R}$) was milder compared to the *Tomm7*$^{-/-}$ mice, suggesting that the identified TOMM7 p.W25R variant is hypomorphic and leads to a partial loss-of-function of the Tomm7 protein. Investigation of mitochondrial function revealed upregulation of mitochondrial dysfunction markers Gdf15 and Fgf21, and an increase in oxygen consumption rate (OCR), with normal responses to electron transport chain (ETC) and ATP synthase inhibitors, suggestive of uncoupling between oxidation and ATP synthesis (47).

Overall, Study I shows that TOMM7 contributes to bone growth, adipose tissue homeostasis, and mitochondrial bioenergetics in mice, making it a new candidate gene for human syndromic conditions with skeletal abnormalities.

Just after the publication of our study, a paper by Garg et al. was published describing *TOMM7* as a candidate disease gene in a 21-year-old male with autosomal recessive progeroid syndrome (51). The affected individual was homozygous for *TOMM7* c.86C>T, p.Pro29Leu variant, which is located in close proximity to the variant in our case. Clinical features of the affected individual included postnatal growth failure, short stature, learning disabilities, dysmorphic features, mandibular hypoplasia, and partial loss of fat tissue, all consistent with the features identified in our study. The patient in our study died in early childhood, while the individual in Garg A et al. has reached adulthood. Interestingly, the affected individual in Garg A et al. had an older sister with similar dysmorphic features who died at 10 years of age due to a likely respiratory infection; however, the genetic data of the sister were not available. Even though the survival age of the affected patients with *TOMM7* variants is variable, the remarkable similarities of the clinical features in both cases suggest that they belong to the same disease group. Mitochondrial diseases lead to a complex combination of multiple clinical features, and the individual susceptibility to infections may impact survival.

The molecular pathogenesis studied in Garg A et al. is also consistent with the findings in our study, including an increase in the oxygen consumption rate (OCR), suggestive of mitochondrial bioenergetic deficiency. No changes were seen in PINK1/parkin-mediated mitophagy in both studies, strengthening the hypothesis that the mitochondrial bioenergetic dysfunction contributes to the patients' phenotype observed in both cases. While we did not observe significant changes in electron transport chain components (ETC), Garg et al. showed upregulation for some of the ETC members. The increase in OCR observed in both cases suggests the possible dissociation between oxidation and ATP synthesis in the mitochondria, but the exact molecular mechanism by which variants in *TOMM7* lead to mitochondrial dysfunction remains to be understood.

Study II - Biallelic variants in *ADAMTSL2* cause Al-Gazali skeletal dysplasia

Study II was an international collaboration between multiple clinical centers worldwide where we described a cohort of nine patients with Al-Gazali skeletal dysplasia and pathogenic variants in the *ADAMTSL2* gene (52). The first case of Al-Gazali skeletal dysplasia was described in 1996 and two more individuals with remarkably similar clinical and skeletal features were described in 2011, yet without a known genetic cause (53, 54). Throughout the years, two more affected individuals were identified by our research team, however, the genetic cause was not determined. In early 2020, we were approached by a collaborator who had a patient with features of Al-Gazali skeletal dysplasia and possibly pathogenic variants in the *ADAMTSL2* gene. It prompted us to re-analyze genome data from other individuals and it quickly became evident why it took so long to find the genetic cause behind Al-Gazali skeletal dysplasia. The second half of the *ADAMTSL2* gene shares a 98% sequence similarity with a pseudogene region, creating read alignment challenges when using the GRCh37/hg19 reference genome (55).

All nine patients with Al-Gazali skeletal dysplasia described in Study II shared unique clinical and radiographic features that are summarized in **Figure 5**. The condition is associated with early mortality, and only two out of nine patients survived past one year of age. Biallelic variants in the *ADAMTSL2* were spread across the gene and none of the patients shared the same variants.

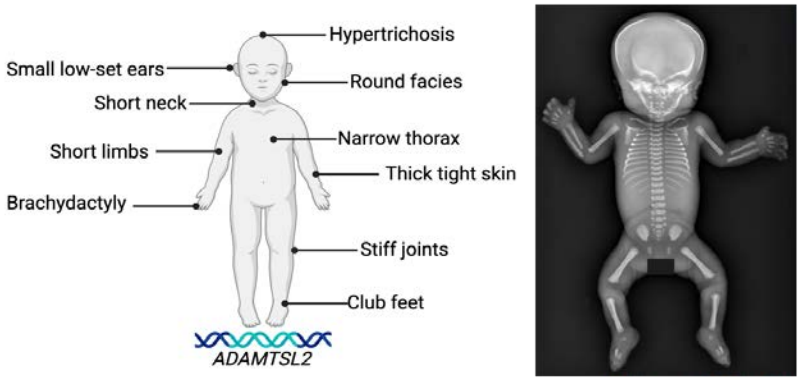

Figure 5. Left: Illustrative summary of clinical features of Al-Gazali skeletal dysplasia patients. Some patients had internal abnormalities of the heart and lungs (pulmonary stenosis). **Right: X-ray image of Al-Gazali skeletal dysplasia patient.** Note osteosclerosis, large fontanelle, broad ribs, broad tubular bones, delayed ossification of the pubic bones, and brachydactyly. *Created with BioRender.com*

*ADAMTSL2 i*s known to cause another skeletal dysplasia, geleophysic dysplasia type 1 (GPHYSD1, OMIM #231050), which is an acromelic dysplasia characterized by short stature, joint and skin stiffness, dysmorphic facial features, and skeletal abnormalities (37, 56). One of the most intriguing questions was deciphering the similarities and differences between GPHYSD1 and Al-Gazali skeletal dysplasia. The genetic heterogeneity was evident in both conditions, as disease-causing variants were spread across the whole gene, and there were no correlations between the diseases (56). Individuals with GPHYSD1 and Al-Gazali skeletal dysplasia share similar clinical features, however patients with Al-Gazali skeletal dysplasia have distinct radiographic hallmarks (shown in the figure above), which are not commonly seen in GPHYSD1. The phenotypic continuum observed in GPHYSD1 and Al-Gazali skeletal dysplasia cases suggests that both conditions belong to the same disease family.

Previous studies from GPHYSD1 animal and cell models suggest the involvement of TGFβ signaling and ECM in GPHYSD1 pathogenesis, but the exact mechanism by how variants in *ADAMTSL2* lead to disease has not yet been understood (36, 57, 58, 59). Primary dermal fibroblast from patient 1 were used in this study, and no changes in TGFβ signaling was observed in the patient's cells compared to controls. However, notable differences were seen in the morphology of the FBN1 microfibrillar network, suggestive of ECM involvement in the disease pathogenesis.

While this thesis was being written, a study presenting some answers to why there is phenotypic heterogeneity in ADAMTSL2-related disorders was published by Camarena et al. (60). Camarena et al. performed a comprehensive analysis of different allelic combinations of pathogenic *ADAMTSL2* variants in a murine model and showed that certain *ADAMTSL2* variants lead to a stronger damaging effect on ADAMTSL2 protein secretion, which, in turn, leads to a more severe phenotype in mutant mice. Moreover, the authors of the study did not observe any changes in TGFβ signaling, which further supports the likely hypothesis that ECM involvement plays a significant role in the pathogenesis of ADAMTSL2-related disorders (60).

In summary, Study II contributed to the field by identifying the genetic cause of Al-Gazali skeletal dysplasia and adding it to the family of ADAMTSL2-related skeletal disorders.

Study III – Compound heterozygous variants in *RAB34* cause atypical skeletal ciliopathy

In Study III, we described the clinical and genetic findings of a fetus with a skeletal ciliopathy and compound heterozygous variants in the *RAB34* gene (61). The initial suspicion of a complex malformation syndrome was raised during the ultrasound scan at GA 13+2, revealing multiple abnormalities, including posterior neck edema, micrognathia, microtia, nasal hypoplasia, bilateral cleft, short limbs, and polydactyly. The pregnancy was terminated at GA 17+5. The summary of the main clinical findings is shown in **Figure 6**. Postmortem radiographic analysis of the affected fetus revealed a narrow thorax, handlebar clavicles, poorly developed iliac bones, and polydactyly – features consistent with a short-rib-thoracic dysplasia (SRTD).

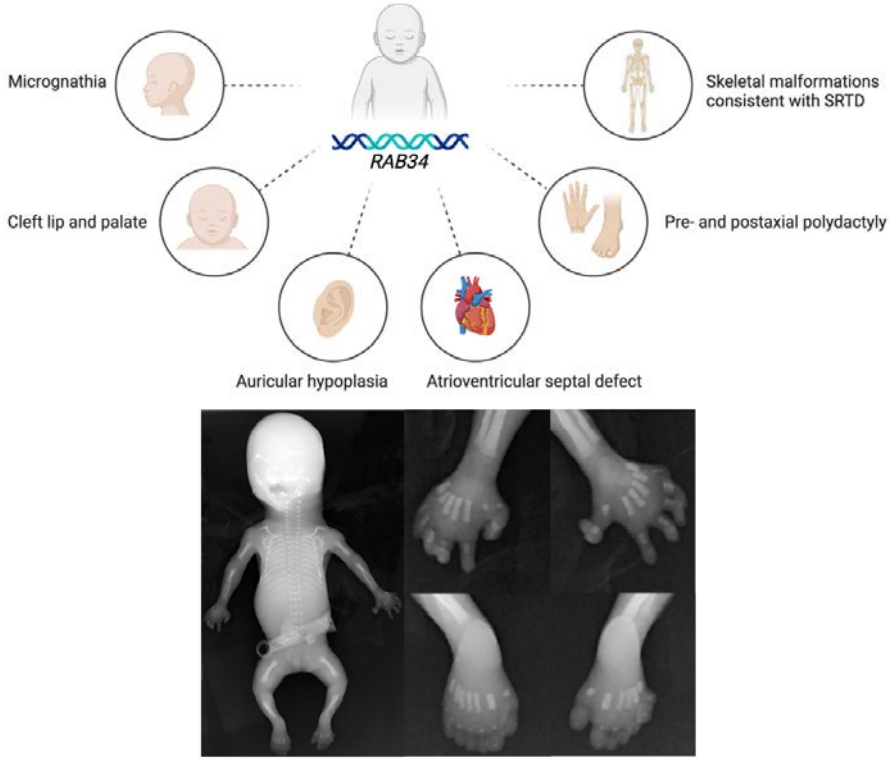

Figure 6. Top: Illustrative summary of the malformations caused by compound heterozygous variants in the *RAB34* gene in the affected fetus (SRTD – short-rib thoracic dysplasia). **Bottom: X-ray images of the affected fetus.** Note handlebar-shaped clavicles, narrow thorax, underdeveloped ilia, short tubular bones, and the presence of additional digits and toes (pre- and postaxial polydactyly). *Created with BioRender.com*

The initial clinical family trio exome sequencing did not reveal any candidate genes and the family was referred to Karolinska University Hospital for genome sequencing. Extended analysis of the genome sequencing data identified compound heterozygous variants in the *RAB34* (NM_031934.6, hg19) gene: a maternal c.254T>C, p.(Ile85Thr) and a paternal c.691C>T, p.(Arg231*). Both identified variants were rare in the general population, reported 17 times and 27 times respectively, in heterozygous state in gnomAD v4.0.0.

RAB34 was not previously associated with a human phenotype. However, literature analysis revealed that a *Rab34* knockout mouse model (*Rab34*$^{-/-}$) generated by Xu et al. exhibited remarkably similar phenotype to the affected fetus in our study, including polydactyly and cleft lift and palate, hallmark features of a skeletal ciliopathy (62). The evidence of Rab34 impact on the skeletal phenotype is limited due to the early lethality of the Rab34-null mice. However, molecular studies suggest that Rab34 is important for cilia function, and loss of *Rab34* leads to a disrupted Hedgehog (Hh) signaling pathway, which is one of the key pathways in early skeletal development and limb patterning (62, 63, 64).

While our study was undergoing review, four more individuals with oral-facial-digital syndrome (OFDS) and disease-causing variants in *RAB34* were described by Bruel et al. (65). The limb and facial malformations were generally similar to the patient in our study, but the oral abnormalities were more complex in the individuals described by Bruel et al. OFDS constitutes a heterogenous group of phenotypes, of which OFDS type 2 (OMIM #252100) and type 4 (OMIM #258860) are listed as genetic skeletal disorders (13). The genetic and phenotypic heterogeneity is a common phenomenon in skeletal ciliopathies. For example, pathogenic variants in *KIAA0753* lead to skeletal dysplasia with overlapping features of several ciliopathies, including Joubert, Jeune, and OFD syndromes (40). However, the clinical features can vary from case to case, with different degrees of skeletal involvement, absence, or presence of oral clefting, and different types of polydactylies (66).

Thus, it is most likely that disease-causing variants in the *RAB34* gene lead to a spectrum of ciliopathy phenotypes, including SRTD-like as seen in the affected fetus in our study, and OFDS-like as described by Bruel et al.

Study IV – Structural variants in *COL1A1/COL1A2* in Osteogenesis Imperfecta

In Study IV, we described eight Swedish patients from five unrelated families with Osteogenesis Imperfecta (OI) caused by structural variants (SVs) in the *COL1A1* and *COL1A2* genes. OI, also called Brittle Bone Disease, is a heterogenous disorder characterized by fragility of the bones, skeletal malformations, short stature, and extraskeletal features including blue/grey sclerae, dental abnormalities, instability of the joints, and in some cases coagulation deficiency (67). This study focused on the analysis of gene dosage abnormalities in DNA samples from OI patients where no sequence variants were detected in known OI genes.

Six out of eight patients from three unrelated families had OI type I phenotype, which is characterized by recurring fractures and low bone mass, and is generally considered the mildest form of OI. Structural variants were detected by the combination of multiplex ligation-dependent probe amplification (MLPA), array comparative genomic hybridization (aCGH), and genome sequencing, and breakpoint junctions were validated by Sanger sequencing. Three patients had deletions in the 17q21 region, where the *COL1A1* gene is located, and one family had an in-frame deletion in the *COL1A2* gene.

The remaining two patients in Study IV had a more complex clinical and genetic presentation. One patient (patient 5 in the study) had a syndromic phenotype of tricho-dento-osseous (TDO) syndrome and OI caused by a 2.1 Mb deletion in chromosome 17. The deletion included 48 genes, and of those *DLX3* and *COL1A1* were associated with TDO and OI phenotypes, respectively. The clinical features consistent with OI were bone fragility and blue sclerae, while curly hair, dental abnormalities, high bone density, and hypoplastic nails were consistent with TDO syndrome (68, 69).

The most intriguing finding in this study was a complex genomic rearrangement in the *COL1A2* gene identified in a patient with OI type III (patient 1 in the study), which was resolved by a combination of short-read and long-read genome sequencing. Initially, MLPA and aCGH analysis suggested a partial deletion in *COL1A2*. However, genome sequencing revealed a combination of 15 kb deletion and 8 kb duplication affecting exons 1-31 of the *COL1A2* gene **(Figure 7)**, and long-read genome sequencing confirmed that the rearrangement was monoallelic.

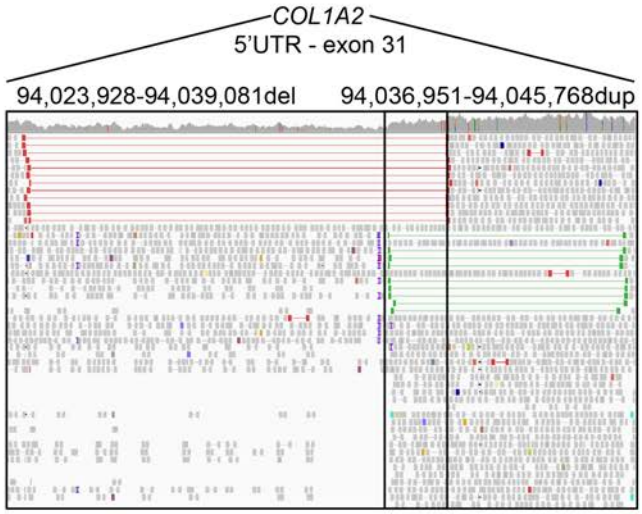

Figure 7. Integrative Genomics Viewer (IGV) snapshot of short-read sequencing data in patient 1. Red-colored reads indicate the deletion (5'UTR-exon 19), and green-colored reads indicate the duplication (exon 13-31). The genomic coordinates correspond to the GRCh37 reference genome.

The first deletion breakpoint occurred in the 5'UTR, before the initiation codon ATG, suggesting that transcription of the affected allele was likely absent. This was confirmed by the *COL1A2* gene and protein expression analysis in the primary dermal fibroblasts from the affected patient, where both reduced COL1A2 mRNA and protein expression were observed. It has long been postulated that *COL1A2* haploinsufficiency does not lead to OI phenotype in humans. However, previous evidence in the literature suggests that monoallelic loss of *COL1A2* can lead to OI phenotype in some cases (70, 71). Thus, this study further expands the genetic and phenotypic spectrum of patients with reported *COL1A2* haploinsufficiency.

In summary, Study IV provides detailed clinical phenotype and multiple follow-up data of the affected patients and further increases the knowledge of clinical and genetic heterogeneity of OI caused by structural variants in type I collagen genes.

6 CONCLUSIONS

The studies in this thesis focused on the genetic and clinical characterization of rare skeletal disorders. Using genome sequencing, two genes previously not associated with human phenotype, *TOMM7* and *RAB34*, were described. Furthermore, a genetic cause was identified for the previously unresolved entity known as skeletal dysplasia Al-Gazali type. This thesis has contributed to the field by expanding the genetic and phenotypic spectrum of several congenital skeletal disorders and highlighting several important points as summarized below:

- Genome sequencing is a first-tier research tool for identifying novel candidate genes and variants. In particular, a family trio-based sequencing facilitates the analysis of genome sequencing data.
- Variant identification and association with disease is dependent on multiple factors, such as knowledge about gene and protein function, as well as meticulous phenotypic characterization of patient groups.
- Functional studies for the characterization of new genes and proteins are of great importance in solving unsolved cases. In Study I, a previously published *Tomm7*$^{-/-}$ mouse model was an initial hint for the potential pathogenicity of the identified *TOMM7* variant. A mutation-specific mouse model developed in Study I enabled further molecular explanation of TOMM7 involvement in the disease phenotype. In Study III, the malformations in a previously published *Rab34*$^{-/-}$ mouse model had a striking resemblance to the phenotypic abnormalities of the patient. Primary dermal fibroblasts derived from patients, used in Studies II and IV, enabled direct investigation of molecular consequences associated with pathogenic variants.
- International collaboration and identification of other patients with similar genetic and/or clinical presentation is crucial. Study II included nine patients from seven clinical centers worldwide. The limitation of Study I and III was the sample size of n=1. However, almost simultaneously, independent studies with similar findings were published.
- Disease-causing variants can be located in poorly covered, repetitive, or non-coding sequence regions, as shown in Study II by highlighting the pseudogene region of the *ADAMTSL2* gene. Coverage discrepancies in some cases can be resolved by using the GRCh38/hg38 reference genome instead of GRCh37/hg19.

- Studies of known clinical entities, such as Osteogenesis Imperfecta in Study IV, can lead to unexpected findings, as shown by the complex genomic rearrangement in the *COL1A2* gene; and increased knowledge of overlapping phenotypes, as shown by the TDO/OI patient.
- Congenital skeletal diseases manifest across a spectrum of phenotypes, ranging from mild to severe. In particular, Study II added Al-Gazali skeletal dysplasia to the most severe spectrum of ADAMTSL2-related Gelepohysic dysplasia type 1.

7 POINTS OF PERSPECTIVE

Continued research into the genetic and molecular mechanisms underlying congenital skeletal disorders holds promise for the development of targeted therapies and interventions. Collaborative efforts among clinicians, researchers, and advocacy groups are essential for advancing understanding, improving diagnostics, and expanding treatment options for affected individuals. Additionally, raising awareness and promoting inclusivity is essential for supporting individuals with congenital skeletal disorders and their families, fostering a more inclusive and accessible society for all.

Nowadays generating massive amounts of data from various sequencing platforms is almost effortless. However, a persistent challenge in improving diagnostic yield is variant interpretation. Clinical experience shows that 15-20% of patients have variants of unknown significance (VUS) in disease-associated genes. Moreover, many of the variants identified in genome sequencing data occur in genes with poorly characterized function and no prior association with human phenotypes. Therefore, functional studies, using cell or animal models are imperative for understanding the pathogenicity of identified variants. Additionally, it is very likely that various mechanisms, such as epigenetic changes, variation in the non-coding sequences, enhancers, genetic mosaicism, or oligogenic inheritance can contribute to disease phenotypes. Understanding these more complex mechanisms of genetic inheritance and developing tools to interpret and characterize non-Mendelian inheritance patterns would significantly increase diagnostic rates for patients with genetic diseases. Furthermore, the integration of multi-omics data, including genomics, transcriptomics, proteomics, and metabolomics, holds immense potential for unraveling the complex networks involved in skeletal development and disease.

Studying rare diseases means that patient numbers are very low, and quite often n=1. Identifying another patient with a suspected deleterious variant in the same gene and similar phenotype can often provide sufficient evidence for establishing the diagnosis. Tools such as GeneMatcher or Matchmaker Exchange facilitate the connections between researchers and clinicians, increasing the likelihood of finding matches between patients with similar genetic profiles and clinical presentations (72, 73). Integration and increased usage of large-scale collaborative efforts and data-sharing initiatives are essential for overcoming the challenges caused by the rarity and genetic heterogeneity of these conditions.

Another key perspective involves the translation of genomic research findings into tangible clinical applications. Understanding the molecular mechanisms of genetic diseases is essential for driving pharmaceutical progress. Several skeletal dysplasias have successfully entered clinical trials leading to disease-specific drugs available on the market for Osteogenesis Imperfecta, achondroplasia, X-linked hypophosphataemic rickets, hypophosphatasia, and fibrodysplasia ossificans progressiva. Furthermore, a better understanding of fundamental mechanisms not only facilitates the development of novel treatments but also fosters drug repurposing, as evidenced by the clinical trials of carbamazepine use for the treatment of metaphyseal chondrodysplasia Schmid type (22).

The human reference genome was built by sequencing individuals of European genetic ancestry leading to a lack of diversity and limited understanding of genetic variation within different populations and creating challenges in understanding how disease-causing variants impact various ethnic groups. Worldwide efforts are now being made to extend the data for other populations and create ethnically more diverse human reference genome. The National Institutes of Health (NIH) *All of Us* research program is currently building one of the most diverse datasets and preliminary results indicate that genetic variation may reflect real differences in the prevalence of certain diseases among different ancestral groups (74). Similarly, the Human Pangenome Reference Consortium aims to capture the full spectrum of genetic diversity within the human population (75). Thus, the future holds great promise in addressing the challenges and contributing to a better understanding of human genetic makeup.

In addition to the promising prospects, research can also be influenced by a "publish or perish" mentality, fostering unnecessary competition, and regrettably instances of unreproducible or fabricated data. This aspect of scientific research is an unfortunate reality, yet is imperative for each of us, as researchers, to initiate change from within. By fostering a culture of integrity, collaboration, and transparency, we can collectively address these challenges and preserve authentic scientific curiosity. It is upon us to prioritize the pursuit of knowledge with honesty and integrity, ensuring that our contributions guarantee reliability and trustworthiness in both, the scientific and general community.

8 ACKNOWLEDGMENTS

No matter what one's journey entails, the people one encounters along the way are often the most pivotal aspect. I am very thankful to everyone I have had the privilege to meet along the way of my PhD.

Especially I would like to acknowledge:

Giedre Grigelioniene, my main supervisor. We agreed and disagreed, we celebrated successes and learned from the mistakes. Thank you for supporting me in all the additional ventures outside of my PhD; I would not have been where I am today without your encouragement.

My great team of co-supervisors:

Fulya Taylan, for setting an example of knowledge, hard work, and dedication, and for always willing to answer my questions. **Anna Hammarsjö**, for all your help throughout these years, and for being a source of calmness and assurance. **Andrei Chagin**, for sharing your passion for bones, evolution, and the joy of new discoveries. **Ann Nordgren**, for sharing your enthusiasm for patient care and undiagnosed diseases. **Anna Lindstrand**, for sharing your excitement about the newest technologies and structural variants.

I would also like to acknowledge all the patients and their family members who participated in the studies; this would not have been possible without your engagement. Furthermore, all the studies in this thesis result from the collective effort and I would like to extend my sincere thank you to all the collaborators involved in the projects.

I was fortunate to have the most incredible colleagues, and I would like to sincerely thank everyone at J10:20 in BioClinicum, both past and current members, who contributed to a wonderful working environment. **Jakob**, thank you for always being there to talk both science and life. Having you as a PhD-buddy, and, most importantly, as a friend has been an absolute privilege. Your contribution to the success of this thesis is immeasurable. **Anton**, thank you for teaching me the cell culture with such dedication and kindness; I will always be grateful for that. Thank you, **Clara, Isabel, Jesper, Hillevi, Raquel, Sailan, Kristine, Marlene, Carolina, Blaž, Alice, Nina, Katja, Mastoureh, Benedicte, Sofia, Esmee, Angelica, Karl, Ida, Alexandra, Fan, Fatemah, Richard, Zahra, Daniel, Emma, Leily, Dania, Larry, Jessica, Jose, Karthick**.

Throughout my PhD I had the opportunity to be a doctoral student representative, a role that has been an extremely rewarding learning experience. I would like to thank **The Course and Programme Committee**, especially **Anna Gustafsson** and **Ingeborg van der Ploeg**; **MMK Research Education Board**, with special thanks to **Catharina Lavebratt** and **Cecilia Österholm Corbascio**, as well as **The Doctoral Students' Association**, for their trust, collaboration, and the shared goal of improving the doctoral education.

Thank you to the administration at the Department of Molecular Medicine and Surgery, especially **Ann-Britt Wikström** and **Susanne Forsberg**. **Jan-Erik Kaarre**, for the help with all IT-related matters.

Lastly, I would like to acknowledge the people who were beyond the office and lab walls, yet played a crucial role in supporting, encouraging, and motivating me. Thank you for taking an interest in my research, for sharing in the joyful moments, and for standing by me during the most challenging days.

Laura, Emelie, **Ilke**, for your unconditional and sincere friendship, I cherish it greatly. The remaining OG Queens **Linda, Giorgia, Helena, Emma**, for being a source of inspiration and motivation, regardless of where in the world you are.

Manuela och **Peter**, tack för det varma välkomnandet till familjen.

Dan, thank you for being the rock I can lean on, for believing in me, and for bringing so much joy into my life.

Thank you, **Mom** and **Dad**, for instilling in me the value of education and hard work. Thank you to my brother **Marius** and my sister **Izabele**, for always being a source of encouragement and happiness. It all becomes a little bit easier knowing that, no matter what happens, I will always have your full support.

9 REFERENCES

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

2. Strachan T, Read AP. Human molecular genetics. Fifth edition. ed. Boca Raton, Florida: CRC Press; 2019.

3. Bagger FO, Borgwardt L, Jespersen AS, Hansen AR, Bertelsen B, Kodama M, et al. Whole genome sequencing in clinical practice. BMC Med Genomics. 2024;17(1):39.

4. Zschocke J, Byers PH, Wilkie AOM. Mendelian inheritance revisited: dominance and recessiveness in medical genetics. Nat Rev Genet. 2023;24(7):442-63.

5. Murdock DR, Rosenfeld JA, Lee B. What Has the Undiagnosed Diseases Network Taught Us About the Clinical Applications of Genomic Testing? Annual Review of Medicine. 2022;73(1):575-85.

6. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet. 2018;19(5):253-68.

7. Ketkar S, Burrage LC, Lee B. RNA Sequencing as a Diagnostic Tool. Jama. 2023;329(1):85-6.

8. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381(6664):eadg7492.

9. Boccuto L, Neri G. Overgrowth in myth and art. Am J Med Genet C Semin Med Genet. 2021;187(2):176-81.

10. Botto LD. Skeletal dysplasias in art and antiquities: A cultural journey through genes, environment, and chance. Am J Med Genet C Semin Med Genet. 2021;187(2):199-212.

11. Offiah AC, Hall CM. The radiologic diagnosis of skeletal dysplasias: past, present and future. Pediatr Radiol. 2020;50(12):1650-7.

12. Spranger JW, Brill PW, Hall C, Nishimura G, Superti-Furga A, Unger S. Bone Dysplasias: An Atlas of Genetic Disorders of Skeletal Development: Oxford University Press; 2018 01 Nov 2018.

13. Unger S, Ferreira CR, Mortier GR, Ali H, Bertola DR, Calder A, et al. Nosology of genetic skeletal disorders: 2023 revision. Am J Med Genet A. 2023.

14. Grigelioniene G, Suzuki HI, Taylan F, Mirzamohammadi F, Borochowitz ZU, Ayturk UM, et al. Gain-of-function mutation of microRNA-140 in human skeletal dysplasia. Nat Med. 2019;25(4):583-90.

15. Stark Z, Scott RH. Genomic newborn screening for rare diseases. Nat Rev Genet. 2023;24(11):755-66.

16. Kernohan KD, Boycott KM. The expanding diagnostic toolbox for rare genetic diseases. Nat Rev Genet. 2024.

17. Smith CIE, Bergman P, Hagey DW. Estimating the number of diseases – the concept of rare, ultra-rare, and hyper-rare. iScience. 2022;25(8):104698.

18. Nishimura G, Handa A, Miyazaki O, Tsujioka Y, Murotsuki J, Sawai H, et al. Prenatal diagnosis of bone dysplasias. Br J Radiol. 2023;96(1147):20221025.

19. Legare JM, Basel D. What the pediatric endocrinologist needs to know about skeletal dysplasia, a primer. Front Pediatr. 2023;11:1229666.

20. Spranger J. Pattern recognition in bone dysplasias. Prog Clin Biol Res. 1985;200:315-42.

21. Handa A, Grigelioniene G, Nishimura G. Skeletal Dysplasia Families: A Stepwise Approach to Diagnosis. Radiographics. 2023;43(5):e220067.

22. Sabir A, Irving M. Clinical trials in skeletal dysplasia: a paradigm for treating rare diseases. British Medical Bulletin. 2021;139(1):16-35.

23. Guasto A, Cormier-Daire V. Signaling Pathways in Bone Development and Their Related Skeletal Dysplasia. Int J Mol Sci. 2021;22(9).

24. Chagin AS, Newton PT. Postnatal skeletal growth is driven by the epiphyseal stem cell niche: potential implications to pediatrics. Pediatr Res. 2020;87(6):986-90.

25. Yip RKH, Chan D, Cheah KSE. Chapter Twelve - Mechanistic insights into skeletal development gained from genetic disorders. In: Olsen BR, editor. Current Topics in Developmental Biology. 133: Academic Press; 2019. p. 343-85.

26. Long F, Ornitz DM. Development of the endochondral skeleton. Cold Spring Harb Perspect Biol. 2013;5(1):a008334.

27. Costantini A, Guasto A, Cormier-Daire V. TGF-β and BMP Signaling Pathways in Skeletal Dysplasia with Short and Tall Stature. Annu Rev Genomics Hum Genet. 2023;24:225-53.

28. Bielajew BJ, Hu JC, Athanasiou KA. Collagen: quantification, biomechanics, and role of minor subtypes in cartilage. Nat Rev Mater. 2020;5(10):730-47.

29. Robinson ME, Rauch F. Mendelian bone fragility disorders. Bone. 2019;126:11-7.

30. Bardai G, Lemyre E, Moffatt P, Palomo T, Glorieux FH, Tung J, et al. Osteogenesis Imperfecta Type I Caused by COL1A1 Deletions. Calcif Tissue Int. 2016;98(1):76-84.

31. Mannstadt M, Lin AE, Le LP. Case records of the Massachusetts General Hospital. Case 24-2014. A 27-year-old man with severe osteoporosis and multiple bone fractures. N Engl J Med. 2014;371(5):465-72.

32. van Dijk FS, Huizer M, Kariminejad A, Marcelis CL, Plomp AS, Terhal PA, et al. Complete COL1A1 allele deletions in osteogenesis imperfecta. Genet Med. 2010;12(11):736-41.

33. Numbere N, Weber DR, Porter G, Jr., Iqbal MA. A 235 Kb deletion at 17q21.33 encompassing the COL1A1, and two additional secondary copy number variants in an infant with type I osteogenesis imperfecta: A rare case report. Mol Genet Genomic Med. 2020;8(6):e1241.

34. Jovanovic M, Guterman-Ram G, Marini JC. Osteogenesis Imperfecta: Mechanisms and Signaling Pathways Connecting Classical and Rare OI Types. Endocr Rev. 2022;43(1):61-90.

35. Handa A, Grigelioniene G, Nishimura G. Radiologic Features of Type II and Type XI Collagenopathies. Radiographics. 2021;41(1):192-209.

36. Stanley S, Balic Z, Hubmacher D. Acromelic dysplasias: how rare musculoskeletal disorders reveal biological functions of extracellular matrix proteins. Ann N Y Acad Sci. 2020.

37. Le Goff C, Morice-Picard F, Dagoneau N, Wang LW, Perrot C, Crow YJ, et al. ADAMTSL2 mutations in geleophysic dysplasia demonstrate a role for ADAMTS-like proteins in TGF-beta bioavailability regulation. Nat Genet. 2008;40(9):1119-23.

38. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev Mol Cell Biol. 2017;18(9):533-47.

39. Mitchison HM, Valente EM. Motile and non-motile cilia in human pathology: from function to phenotypes. J Pathol. 2017;241(2):294-309.

40. Hammarsjö A, Pettersson M, Chitayat D, Handa A, Anderlid BM, Bartocci M, et al. High diagnostic yield in skeletal ciliopathies using massively parallel genome sequencing, structural variant screening and RNA analyses. J Hum Genet. 2021;66(10):995-1008.

41. Blacque OE, Scheidel N, Kuhns S. Rab GTPases in cilium formation and function. Small GTPases. 2018;9(1-2):76-94.

42. Gorman GS, Chinnery PF, DiMauro S, Hirano M, Koga Y, McFarland R, et al. Mitochondrial diseases. Nat Rev Dis Primers. 2016;2:16080.

43. Zheng CX, Sui BD, Qiu XY, Hu CH, Jin Y. Mitochondrial Regulation of Stem Cells in Bone Homeostasis. Trends Mol Med. 2020;26(1):89-104.

44. Dobson PF, Dennis EP, Hipps D, Reeve A, Laude A, Bradshaw C, et al. Mitochondrial dysfunction impairs osteogenesis, increases osteoclast activity, and accelerates age related bone loss. Sci Rep. 2020;10(1):11643.

45. Behera J, Ison J, Tyagi A, Mbalaviele G, Tyagi N. Mechanisms of autophagy and mitophagy in skeletal development, diseases and therapeutics. Life Sci. 2022;301:120595.

46. Stranneheim H, Lagerstedt-Robinson K, Magnusson M, Kvarnung M, Nilsson D, Lesko N, et al. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. Genome Med. 2021;13(1):40.

47. Young C, Batkovskyte D, Kitamura M, Shvedova M, Mihara Y, Akiba J, et al. A hypomorphic variant in the translocase of the outer mitochondrial membrane complex subunit TOMM7 causes short stature and developmental delay. HGG Adv. 2023;4(1):100148.

48. Sapp JC, Facio FM, Cooper D, Lewis KL, Modlin E, van der Wees P, et al. A systematic literature review of disclosure practices and reported outcomes for medically actionable genomic secondary findings. Genet Med. 2021;23(12):2260-9.

49. Ohtsuka M, Sato M, Miura H, Takabayashi S, Matsuyama M, Koyano T, et al. i-GONAD: a robust method for in situ germline genome engineering using CRISPR nucleases. Genome Biol. 2018;19(1):25.

50. Gurumurthy CB, Sato M, Nakamura A, Inui M, Kawano N, Islam MA, et al. Creation of CRISPR-based germline-genome-engineered mice without ex vivo handling of zygotes by i-GONAD. Nat Protoc. 2019;14(8):2452-82.

51. Garg A, Keng WT, Chen Z, Sathe AA, Xing C, Kailasam PD, et al. Autosomal recessive progeroid syndrome due to homozygosity for a TOMM7 variant. J Clin Invest. 2022;132(23).

52. Batkovskyte D, McKenzie F, Taylan F, Simsek-Kiper PO, Nikkel SM, Ohashi H, et al. Al-Gazali Skeletal Dysplasia Constitutes the Lethal End of ADAMTSL2-Related Disorders. J Bone Miner Res. 2023;38(5):692-706.

53. Al-Gazali L, Devadas K. and Hall C. M. A new lethal neonatal short limb dwardism. Clin Dysmorphol. 1996(5):154-64.

54. Grigelioniene G, Papadogiannakis N, Conner P, Geiberger S, Nishikawa M, Nakayama M. Extending the phenotype of lethal skeletal dysplasia type al Gazali. Am J Med Genet A. 2011;155A(6):1404-8.

55. Li H, Dawood M, Khayat MM, Farek JR, Jhangiani SN, Khan ZM, et al. Exome variant discrepancies due to reference-genome differences. Am J Hum Genet. 2021;108(7):1239-50.

56. Allali S, Le Goff C, Pressac-Diebold I, Pfennig G, Mahaut C, Dagoneau N, et al. Molecular screening of ADAMTSL2 gene in 33 patients reveals the genetic heterogeneity of geleophysic dysplasia. J Med Genet. 2011;48(6):417-21.

57. Hubmacher D, Wang LW, Mecham RP, Reinhardt DP, Apte SS. Adamtsl2 deletion results in bronchial fibrillin microfibril accumulation and bronchial epithelial dysplasia--a novel mouse model providing insights into geleophysic dysplasia. Dis Model Mech. 2015;8(5):487-99.

58. Hubmacher D, Taye N, Balic Z, Thacker S, Adams SM, Birk DE, et al. Limb- and tendon-specific Adamtsl2 deletion identifies a role for ADAMTSL2 in tendon growth in a mouse model for geleophysic dysplasia. Matrix Biol. 2019;82:38-53.

59. Delhon L, Mahaut C, Goudin N, Gaudas E, Piquand K, Le Goff W, et al. Impairment of chondrogenesis and microfibrillar network in Adamtsl2 deficiency. FASEB J. 2019;33(2):2707-18.

60. Camarena V, Williams MM, Morales AA, Zafeer MF, Kilic OV, Kamiar A, et al. ADAMTSL2 mutations determine the phenotypic severity in geleophysic dysplasia. JCI Insight. 2024;9(5).

61. Batkovskyte D, Komatsu M, Hammarsjö A, Pooh R, Shimokawa O, Ikegawa S, et al. Compound heterozygous variants in RAB34 in a rare skeletal ciliopathy syndrome. Clin Genet. 2024;105(1):87-91.

62. Xu S, Liu Y, Meng Q, Wang B. Rab34 small GTPase is required for Hedgehog signaling and an early step of ciliary vesicle formation in mouse. J Cell Sci. 2018;131(21).

63. Ganga AK, Kennedy MC, Oguchi ME, Gray S, Oliver KE, Knight TA, et al. Rab34 GTPase mediates ciliary membrane formation in the intracellular ciliogenesis pathway. Curr Biol. 2021;31(13):2895-905.e7.

64. Stuck MW, Chong WM, Liao JC, Pazour GJ. Rab34 is necessary for early stages of intracellular ciliogenesis. Curr Biol. 2021;31(13):2887-94.e4.

65. Bruel AL, Ganga AK, Nosková L, Valenzuela I, Martinovic J, Duffourd Y, et al. Pathogenic RAB34 variants impair primary cilium assembly and cause a novel oral-facial-digital syndrome. Hum Mol Genet. 2023.

66. Sabir AH, Sheikh J, Gowda V, Wallis C, Singham S, Durve D, et al. KIAA0753-related skeletal ciliopathy: a ninth case, extending the phenotype and reporting a novel variant. Clin Dysmorphol. 2021;30(3):142-6.

67. Marini JC, Forlino A, Bächinger HP, Bishop NJ, Byers PH, Paepe A, et al. Osteogenesis imperfecta. Nat Rev Dis Primers. 2017;3:17052.

68. Harbuz R, Bilan F, Couet D, Charraud V, Kitzis A, Gilbert-Dussardier B. Osteogenesis imperfecta, tricho-dento-osseous syndrome and intellectual disability: a familial case with 17q21.33-q22 (COL1A1 and DLX3) deletion and 7q32.3-q33 duplication resulting from a reciprocal interchromosomal insertion. Am J Med Genet A. 2013;161a(10):2504-11.

69. Yokoi T, Saito T, Nagai JI, Kurosawa K. 17q21.32-q22 Deletion in a girl with osteogenesis imperfecta, tricho-dento-osseous syndrome, and intellectual disability. Congenit Anom (Kyoto). 2019;59(2):51-2.

70. Costantini A, Skarp S, Kämpe A, Mäkitie RE, Pettersson M, Männikkö M, et al. Rare Copy Number Variants in Array-Based Comparative Genomic Hybridization in Early-Onset Skeletal Fragility. Front Endocrinol (Lausanne). 2018;9:380.

71. Tsai MC, Chou YY, Li CY, Wang YC, Yu HW, Chen CH, et al. New Structural and Single Nucleotide Mutations in Type I and Type II Collagens in Taiwanese Children With Type I and Type II Collagenopathies. Front Genet. 2021;12:594285.

72. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36(10):928-30.

73. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Hum Mutat. 2015;36(10):915-21.

74. Venner E, Patterson K, Kalra D, Wheeler MM, Chen YJ, Kalla SE, et al. The frequency of pathogenic variation in the All of Us cohort reveals ancestry-driven disparities. Commun Biol. 2024;7(1):174.

75. Liao WW, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, et al. A draft human pangenome reference. Nature. 2023;617(7960):312-24.