Translational Research of Mendelian Disorders

Applications of Cutting-Edge Sequencing Techniques and Molecular Tools

SANNA GUDMUNDSSON
Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Rudbecklaboratoriet, Dag Hammerskjölds väg 20, Uppsala, Friday, 3 May 2019 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Joris Veltman (Institute of Genetic Medicine, Newcastle University, United Kingdom).

Abstract

Up to 8% of all live-born children are affected with a congenital disorder. Some are Mendelian disorders of known etiology, but many are of undetermined genetic cause and mechanism, limiting diagnosis and treatment. This project aims to investigate the underlying causes of unresolved Mendelian disorders, and especially syndromes associated with intellectual disability, by using cutting-edge sequencing techniques and molecular tools in a translational setting that intends to directly benefit affected families.

In Paper I, we report the first keratitis-ichthyosis-deafness syndrome patient presenting with reversion of disease phenotype, a phenomenon known as revertant mosaicism. Third-generation sequencing and a cell assay were used to pin-point the mechanism of the somatic variants giving rise to healthy looking skin in the patient. In Paper II, we describe a novel approach to investigate parental origin, gonadal mosaicism, and estimate recurrence risk of disease in two families. Third-generation sequencing was used for haplotype phasing and detection of low-frequency variants in paternal sperm. The recurrence risk in future offspring in the families affected with Noonan syndrome and Treacher Collins syndrome was determined to be 40% and <0.1% respectively. In Paper III, we describe a novel variant in a patient affected with Cornelia de Lange Syndrome, primarily associated with intellectual disability. The affected gene is linked to an extremely rare form of the syndrome, with limited cases described in the literature, usually associated with mild symptoms. Investigation of rare intellectual disability syndromes was continued in Paper IV, by clinical and genetic characterization of six affected males with a likely pathogenic variant in the TAF1 gene. By creating the first TAF1 orthologue knockout we revealed that taf1 is essential for life and that lack of functional taf1 during embryonic development in zebrafish primarily impacts expression of genes in pathways associated with neurodevelopment.

By progressive translational research, using state-of-the-art methodology, this project has illuminated the implication of revertant and gonadal mosaicism in disease (Papers I-II), as well as two extremely rare intellectual disability syndromes (Papers III-IV). In total, five families affected with five different disorders have gained clinical and genetic diagnosis and/or further understanding of prognosis and recurrence risk. The study has led to improved understanding of disease etiology and basic developmental processes, enabling development of new therapies and improved care of future patients.

Keywords: translational research, Mendelian disorders, intellectual disability, sequencing technologies

Sanna Gudmundsson, Department of Immunology, Genetics and Pathology, Medicinsk genetik och genomik, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

© Sanna Gudmundsson 2019

ISSN 1651-6206
ISBN 978-91-513-0595-0
urn:isbn:se:uu:diva-379363 (http://urn.kb.se/resolve?urn=urn:isbn:se:uu:diva-379363)
Till mina föräldrar,
Anna-Berit & Anders

To my parents,
Anna-Berit & Anders
Main supervisor
Prof. Marie-Louise Bondeson, PhD
Dept. of Immunology, Genetics and Pathology,
Uppsala University, Sweden

Co-supervisors
Prof. Niklas Dahl, MD, PhD
Dept. of Immunology, Genetics and Pathology,
Uppsala University, Sweden

Dr. Maria Wilbe, PhD
Dept. of Immunology, Genetics and Pathology,
Uppsala University, Sweden

Faculty opponent
Prof. Joris Veltman, PhD
Institute of Genetic Medicine,
Newcastle University, United Kingdom

Review board members
Prof. Dan Larhammar, PhD
Dept. of Neuroscience,
Uppsala University, Sweden

Prof. Zeynep Tümer, MD, PhD
Kennedy Center, Dept. of Clinical Genetics,
Copenhagen University Hospital, Denmark

Associate Prof. Tomas Bergström, PhD
Dept. of Animal Breeding and Genetics,
Swedish University of Agricultural Sciences, Sweden

Associate Prof. Cecilia Gunnarsson, MD, PhD
Dept. of Clinical and Experimental Medicine,
Linköping University, Sweden

Dr. Jessica Nordlund, PhD
Dept. of Medical Sciences,
Uppsala University, Sweden
List of Papers

This thesis is based on the following papers, referred to in the text by their Roman numerals. Reprints were made with permission from the respective publishers.


* Shared authorship
Additional Publications


## Contents

Introduction ................................................................................................................. 11  
The human genome, a topic of the 21\textsuperscript{st} century ........................................... 11  
Genetic variants drive evolution and Mendelian disease ................................. 12  
Genetic variation comes in all shapes and sizes ...................................................... 13  
\textit{De novo} variants and the effect of parental age ................................ 13  
Recessive variants and consanguinity ................................................................. 15  
Epigenetic factors and sex can affect disease outcome .................................... 15  
Mosaicism affects recurrence risk and can revert disease .............................. 17  
Somatic mosaicism ............................................................................................... 17  
Gonadal mosaicism .............................................................................................. 17  
Revertant mosaicism ........................................................................................... 18  

Investigated Mendelian Disorders ...................................................................... 21  
Keratitis-ichthyosis-deafness syndrome ............................................................... 21  
Treacher Collins syndrome .................................................................................. 21  
Noonan syndrome .................................................................................................. 22  
Cornelia de Lange syndrome .................................................................................. 22  
X-linked intellectual disability and \textit{TAF1} ...................................................... 23  

Methodology ............................................................................................................. 25  
Ethical approval ....................................................................................................... 25  
Genome sequencing, the key to the code .............................................................. 25  
First-generation sequencing .................................................................................. 26  
Next-generation sequencing .................................................................................. 26  
Third-generation sequencing .................................................................................. 27  
Interpreting sequencing variants ......................................................................... 27  
Publicly available sequencing data ....................................................................... 28  
\textit{In silico} predictions .......................................................................................... 28  
Molecular tools to illuminate underlying mechanisms ......................................... 30  
Protein detection \textit{in vivo} and \textit{in vitro} ..................................................................... 30  
Frequency determination with Droplet Digital\textsuperscript{TM} PCR ........................ 30  
Investigation of X-chromosome inactivation ......................................................... 30  
Gene editing with CRISPR/Cas9 in zebrafish ......................................................... 31  

Relevance and Aim .................................................................................................. 33  

Results and Discussion .......................................................................................... 35  

Paper I: Revertant mosaicism repairs skin lesions in a patient with keratitis-ichthyosis-deafness syndrome by second-site mutations in connexin 26 .................................................................35
  Result ........................................................................................................35
  Discussion ................................................................................................37

Paper II: A novel approach using long-read sequencing and ddPCR to investigate gonadal mosaicism and estimate recurrence risk in two families with developmental disorders ........................................41
  Result ........................................................................................................41
  Discussion ................................................................................................43

Paper III: A novel RAD21 p.(Gln592del) variant expands the clinical description of Cornelia de Lange syndrome type 4 – review of the literature ..............................................................................45
  Result ........................................................................................................45
  Discussion ................................................................................................46

Paper IV: TAF1, associated with intellectual disability in humans, is essential for life and regulates neurodevelopmental processes in zebrafish ..........................................................................49
  Results .......................................................................................................49
  Discussion ................................................................................................50

Concluding Remarks and Future Perspectives ............................................55

Svensk populärvetenskaplig sammanfattning .................................................57
  Introduktion till det humana genomet och orsaken till mendelsks sjukdom .................................................................57
  Avhandlingsarbetets relevans och syfte .................................................58
  Forskningsresultat .................................................................................58
  Slutsats ...................................................................................................60

Acknowledgements ....................................................................................61

References ..................................................................................................63
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACMG</td>
<td>American College of Medical Genetics and Genomics</td>
</tr>
<tr>
<td>ASIC</td>
<td>acid-sensing (proton-gated) ion channel</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>brdt</td>
<td>bromodomain testis-specific protein</td>
</tr>
<tr>
<td>CACNA1G</td>
<td>calcium channel, voltage-dependent, T type, alpha 1G subunit</td>
</tr>
<tr>
<td>Cas9</td>
<td>CRISPR-associated 9</td>
</tr>
<tr>
<td>CCN</td>
<td>cyclin</td>
</tr>
<tr>
<td>CdLS</td>
<td>Cornelia de Lange syndrome</td>
</tr>
<tr>
<td>CNV</td>
<td>copy number variations</td>
</tr>
<tr>
<td>CRISPR</td>
<td>clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>Cx</td>
<td>connexin</td>
</tr>
<tr>
<td>ddPCR</td>
<td>droplet digital PCR</td>
</tr>
<tr>
<td>GJB2</td>
<td>gap junction beta 2</td>
</tr>
<tr>
<td>gnomAD</td>
<td>Genome Aggregation Database</td>
</tr>
<tr>
<td>HGMD</td>
<td>Human Gene Mutation Database</td>
</tr>
<tr>
<td>ID</td>
<td>intellectual disability</td>
</tr>
<tr>
<td>indels</td>
<td>insertions or deletions</td>
</tr>
<tr>
<td>kb</td>
<td>kilo base pairs</td>
</tr>
<tr>
<td>kcnj</td>
<td>voltage-gated Channel subfamily J</td>
</tr>
<tr>
<td>KID</td>
<td>keratitis-ichthyosis-deafness</td>
</tr>
<tr>
<td>mbpa</td>
<td>myelin basic protein a</td>
</tr>
<tr>
<td>NGS</td>
<td>next-generation sequencing</td>
</tr>
<tr>
<td>NS</td>
<td>Noonan syndrome</td>
</tr>
<tr>
<td>o/e</td>
<td>observed/expected</td>
</tr>
<tr>
<td>padj</td>
<td>adjusted p-value</td>
</tr>
<tr>
<td>PAE</td>
<td>paternal age effect</td>
</tr>
<tr>
<td>PLA-WB</td>
<td>proximity ligation-based western blot</td>
</tr>
<tr>
<td>polr2</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>PTPN11</td>
<td>tyrosine-protein phosphatase non-receptor type 11</td>
</tr>
<tr>
<td>RM</td>
<td>revertant mosaicism</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SMRT</td>
<td>single molecule real-time</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphisms</td>
</tr>
<tr>
<td>SNV</td>
<td>single nucleotide variants</td>
</tr>
<tr>
<td>SweGen</td>
<td>SweGen Variant Frequency browser</td>
</tr>
<tr>
<td>TADs</td>
<td>topological associated domains</td>
</tr>
<tr>
<td>TAF1</td>
<td>TBP-associated factor 1</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-box binding protein</td>
</tr>
<tr>
<td>TCOF1</td>
<td>treacle ribosome biogenesis factor 1</td>
</tr>
<tr>
<td>TCS</td>
<td>Treacher Collins syndrome</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>VUS</td>
<td>variant of uncertain significance</td>
</tr>
<tr>
<td>WES</td>
<td>whole-exome sequencing</td>
</tr>
<tr>
<td>WGS</td>
<td>whole-genome sequencing</td>
</tr>
<tr>
<td>wt</td>
<td>wild-type</td>
</tr>
<tr>
<td>XCI</td>
<td>X-chromosome inactivation</td>
</tr>
<tr>
<td>XLID</td>
<td>X-linked intellectual disability</td>
</tr>
</tbody>
</table>
Introduction

The human genome, a topic of the 21st century

The DNA structure was discovered in 1953 (Watson and Crick, 1953), but it was not until 2001 that the first draft of the human genome was presented, requiring more than five years of hands-on work and $100 million to compile (Lander et al., 2001; Venter et al., 2001). Eighteen years later, the human genome of an individual is accessible in a few days, at less than $1000. It is an ongoing technical revolution that has completely changed the scope of human genetic research (https://www.genome.gov/27541954/dna-sequencing-costs-data/, accessed March 25, 2019). The ability to sequence DNA in a cost- and time-efficient manner has not only resulted in increased disease gene discovery, but also improved diagnostic yield by implementation of advanced genetic analysis into standard clinical care (Taylor et al., 2015; Veltman and Brunner, 2012; Vissers et al., 2016). Today, the haploid human genome is estimated to be 3.1 billion base pairs (bp) in length, of which approximately 1% encode the protein-coding genes (https://www.ensembl.org/Homo_sapiens/, accessed March 25, 2019). The 99%, non-coding part of the genome, is far less understood, but it is known to be important for gene regulation.

The basic theory of gene regulation and RNA has been known since the 1960s. However, precise knowledge of a specific transcript’s expression patterns has become available within the last decades due to a burst in sequencing techniques, including methods such as single-cell and long-read sequencing. Today we know that there are over 200,000 human transcripts and that their expression differs depending on both time-point and cell type, allowing humans to grow from a single zygote to the complex structure of an adult (Mortazavi et al., 2008; Shapiro et al., 2013). By investigating the underlying mechanisms of gene regulation, we have begun to gain insight into the non-coding part of the genome. For example, it was recently established that gene expression is regulated by the spatial organization of the genome within the cell nucleus, within so-called topological associated domains (TADs) (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012).
The final piece of the central dogma, the protein-coding function of the genome, was cracked in the early 1960s (Nirenberg et al., 1963). Today, we have most likely identified close to all human protein-coding genes, counted up to 20,000. The precise preservation of the protein sequence is often required for stable protein function, and therefore the exome has been highly conserved throughout evolution. Genetic variants that alter the localization, expression or function of proteins are the main cause of Mendelian disorders and the vast majority lies in the protein-coding genome (Chong et al., 2015). Clinically recognized Mendelian phenotypes are estimated to affect 0.4% of all live-born children. However, all congenital disorders are reported to affect up to 8% of all live births worldwide (Baird et al., 1988; Chong et al., 2015; Deciphering Developmental Disorders, 2017; Sheridan et al., 2013). Thus, genetic variants of the human genome are a major source of severe disease.

**Genetic variants drive evolution and Mendelian disease**

Genetic variation occurs due to endogenous processes like replication errors, as well as exogenous mutagenic processes like tobacco smoke (Alexandrov et al., 2013; Rahbari et al., 2016), at a rate of 1–2*10^-8 variants per nucleotide per generation (Campbell et al., 2012; Kondrashov, 2003; Rahbari et al., 2016; Roach et al., 2010). Most genetic variants are neutral, but some decrease fitness, and others increase fitness. Consequently, genetic variants drive both environmental adaptation processes, i.e. evolution and genetic disease. Complex genetic disorders are mostly caused by a set of common genetic variants in combination with adverse environmental risk factors. In contrast, the vast majority of Mendelian disorders are monogenetic, and the outcome is generally not affected by environmental factors (Chong et al., 2015). Variants implicated in Mendelian disorders are mostly rare and have a severe impact on protein function, in contrast to SNVs in complex disorders that mostly are non-coding, and, when coding, have a less severe impact on protein function (Thomas and Kejariwal, 2004). Mendelian variants can be divided into gain-of-function or loss-of-function variants, depending on mechanism of action. A gain-of-function variant alters protein function or expression and are often dominant (Paper I). Loss-of-function variants inhibits expression of the protein, and classically loss of both copies by homozygosity or compound heterozygosity is associated with disease (recessive). However, a recent study confirmed that heterozygous loss-of-function variants (resulting in haploinsufficiency) are as common as heterozygous gain-of-function variants in developmental disorders (Deciphering Developmental Disorders, 2017).
Genetic variation comes in all shapes and sizes

Protein-coding nonsynonymous single-nucleotide variants (SNVs) that result in an altered protein sequence (missense or nonsense variants) are the most prevalent cause of Mendelian disorders (Papers I, II and IV) (Thomas and Kejariwal, 2004). Coding synonymous SNVs tend to have a neutral effect on the protein-coding sequence and are thus seldom associated with Mendelian disorders. However, there is the important exception of surmised harmless synonymous variants that affect splicing and can cause disease by introducing alternative transcripts (Paper II) (Cummings et al., 2017).

Small insertions or deletions (indels) can, like SNVs, cause disease by altering the function of a protein (Paper III). However, the majority of indels will shift the open reading frame and result in a premature stop codon, making them highly deleterious (Paper IV). One-third of Mendelian disorders are estimated to be caused by frame shift, nonsense or splice variants that result in a premature termination codon. In case of protein-truncating variants, disease occurs due to expression of a truncating protein that has escaped degradation via nonsense-mediated decay, by haploinsufficiency or by loss of protein expression (Kurosaki and Maquat, 2016; Rivas et al., 2015).

Structural genetic variants include translocations, inversions, duplications, and deletions of more than 1 kilo bp (kb). Substantial structural variants, like copy number variations (CNVs) spanning more than 1 Mega bp, are rare in the healthy* population, indicative of their deleteriousness and implication in disease (Itsara et al., 2009). Structural variants can, like SNVs and indels, cause disease by affecting the protein-coding sequence, e.g. by deletion of a one or a set of genes. Non-coding structural variants have recently been highlighted to cause Mendelian disorders by disrupting TADs. Specifically, TADs disrupted by structural variation have been suggested to cause limb malformations (Lupianez et al., 2015) and disruption by expansion of short tandem repeats has been linked to fragile X syndrome (Sun et al., 2018).

De novo variants and the effect of parental age

De novo variants have by definition occurred in the germline of the parents or in the zygote state of the offspring. Clinically, a variant is considered de novo if it is not detected in parental DNA but is found in approximately 50% of the offspring’s DNA. Commonly, Sanger sequencing on DNA from blood is used for diagnosis and thus de novo variants can potentially include parentally inherited low-level mosaic variants, variants occurring in the zygote, and post-zygotic variants that results in mosaicism in the proband (further discussed below) (Acuna-Hidalgo et al., 2015; Forsberg et al., 2017). Each generation is estimated to gain 0.02 de novo CNVs, 2.9–9 de novo indels, and 44–82 de

*In this thesis “healthy” is used to describe a tissue, an individual or a population that is not reported to be affected by symptoms of severe congenital disorder(s).
*novo* SNVs, of which on average 1–2/100 will deposit in the protein-coding region (Acuna-Hidalgo et al., 2016; Crow, 2000). Since *de novo* variants have not undergone a natural selection process, they are on average more deleterious than variants that have been inherited throughout generations. Thus, *de novo* variants are a major cause of severe Mendelian disorders (Acuna-Hidalgo et al., 2016). This was recently highlighted in a cohort of 4293 undiagnosed and an additional 3287 previously described patients with severe developmental disorders, where 42% of patients were reported to be affected due to a *de novo* variant (Deciphering Developmental Disorders, 2017).

The quantity of *de novo* variants in offspring has been observed to increase with parental and especially paternal age (Francioli et al., 2015; Kong et al., 2012; Penrose, 1955; Wong et al., 2016). Recent estimations concluded an increase of 0.91–2.87 *de novo* variants of paternal origin/paternal year (Goldmann et al., 2016; Kong et al., 2012; Rahbari et al., 2016; Wong et al., 2016). The increase has mainly been linked to the continuous reproduction of spermatogonia throughout male life, leading to an accumulation of replication errors (Fig. 1) (Crow, 2000). There are also reports of a minor but existing maternal age effect of 0.24–0.51 *de novo* variants per maternal year (Goldmann et al., 2016; Wong et al., 2016). Interestingly, unlike paternal *de novo* variants, maternal *de novo* variants are enriched for C>G transversions and cluster at chromosomes 8, 9 and 16, suggesting that the occurrence mechanisms of maternal and paternal variants differ (Goldmann et al., 2018; Goldmann et al., 2016; Rahbari et al., 2016). Oocytes do not go through mitosis in adult females (Fig. 1), but instead the meiotic gene conversion rate is higher in females compared to males (2.2:1), and is reported to increase with aging oocytes and consequently maternal age (Goldmann et al., 2018; Halldorsson et al., 2016). Aging oocytes also have an exponential age-related risk of nondisjunction, associated with aneuploidies like Down syndrome (MIM 190685) that has a prevalence of 1/1300 at maternal age 20 but 1/30 at maternal age 45 (Newberger, 2000).

The effect of paternal age has also been studied in light of a small group of disorders, referred to as paternal age effect (PAE) disorders (Goriely and Wilkie, 2012), including, for example, Apart (MIM 101200), Costello (MIM 218040), Noonan syndrome (NS; MIM 169350) and Achondroplasia (MIM 100800). The disorders occur more frequently than expected by chance, almost exclusively on the paternal allele, and are mainly caused by gain-of-function variants in the RAS/MAPK pathway. The elevated occurrence is explained by a suggested positive selective advantage during spermatogenesis that results in increased levels of mutant sperm cells over time, a process referred to as selfish spermatogonial selection (Paper II) (Goriely and Wilkie, 2012; Shinde et al., 2013). The number of variants associated with spermatogonial selection was recently increased from 6 to 61, of which 80% were variants in genes of the RAS/MAPK pathway (Maher et al., 2018).
In summary, the parental contribution of *de novo* variants in the offspring is skewed to a ratio of 1:3.6, maternal: paternal (Goldmann et al., 2016), underlining paternal age as a major risk factor for severe Mendelian disorders.

**Recessive variants and consanguinity**

Autosomal recessive disorders are caused by two variants that affect the same autosomal locus. Both alleles need to be affected for disease to occur, and the mechanism is often loss-of-function. Commonly, both variants are parentally inherited, resulting in homozygosity or compound heterozygosity in the offspring (Martin et al., 2018b). X-linked recessive disorders have a slightly different inheritance pattern, with a higher prevalence in males, since only one affected allele is required for disease-penetrance due to X-chromosome hemizygosity. Heterozygous female carriers are protected against recessive X-linked disorders by inactivation of the disease-causing allele, i.e. skewed X-chromosome inactivation (XCI) (Fieremans et al., 2016) or by diploid expression of the locus. A classic example is X-linked loss-of-function red-green color blindness (MIM 303800), affecting 8% of males but only 0.5% of females (Deeb, 2005).

The risk for recessive disorders increases in isolated populations in which endogamous marriages have led to enrichment of rare founder variants. Consanguineous marriages, customary for about 1.1 billion people around the world, are also a risk factor for recessive disorders. Studies show that compared to the general population, first-degree cousins have a 2% increased risk of having offspring with a congenital malformation (mainly recessive disorders). This means that statistically, 8% of consanguineous couples have a 25% risk of having affected offspring. Of note, this highlights that 92% of first-degree cousin couples do not have an increased risk of having offspring with Mendelian disorders, compared to non-consanguineous couples of the same population (Sheridan et al., 2013). However, in consanguineous couples from families with a history of parental relatedness, originating from a population with endogamous marriages, the risk is elevated (Hamamy et al., 2011). The combined risk of endogamous and consanguineous marriages was demonstrated by Martin et al. when studying the prevalence of recessive forms of developmental disorders in a cohort of 6040 probands. In the patient group of European ancestry, 3.6% were affected due to recessive variants, whereas in the patient group with Pakistani ancestry, in whom consanguinity is elevated, notably 31% were affected due to recessive variants (Martin et al., 2018b).

**Epigenetic factors and sex can affect disease outcome**

Even with the monogenic inheritance pattern seen in most Mendelian disorders, the outcome of a specific variant can vary. Variable expressivity of
disease phenotypes and reduced penetrance of disorders obstruct interpretation of genetic variants and stress the need to provide disease prognosis with caution (Cooper et al., 2013; Tuke et al., 2018). Many different mechanisms have been suggested to impact the outcome of a disease-associated variant, such as differential allelic expression, environmental factors, modifier genes, additional genetic variants (Paper I), sex, and epigenetic changes (Paper IV) (Cooper et al., 2013; Posey et al., 2019).

Epigenetic modifications can alter disease expressivity or reduce disease penetrance by altering the expression of a variant. One group of such disorders is imprinting disorders, where the penetrance depends on parental origin of the affected allele, e.g. Angelman syndrome (MIM 105830) and Prader-Willi syndrome (MIM 176270) (Kalsner and Chamberlain, 2015; Kishino et al., 1997). XCI is also an epigenetic trait that can affect disease outcome in females. Skewed XCI can give rise to X-linked recessive disease in heterozygous females by inactivation of the wild-type (wt) allele (Viggiano et al., 2017), but also protect from dominant X-linked disorders by inactivation of the disease-causing allele, a phenomenon often seen in intellectual disability (ID; Paper IV) (Fieremans et al., 2016).

ID is also a specific example of a disorder with a sex-related variance in occurrence, where males have a 40% higher incidence compared to females. The mechanism of skewed sex ratio in ID is not well understood, but a female protective model has been suggested (Vissers et al., 2016). The model is partly based on observations of a higher mutational burden in females with ID compared to males, indicating that females require more severe alterations to be affected. Also, CNVs causing ID in males have been reported to be inherited from asymptomatic mothers (Jacquemont et al., 2014). Variants on the X-chromosome have been a natural target for ID research due to the skewed gender ratio, and about 15% of ID genes identified today are X-linked (Neri et al., 2018). A recent study investigated patients affected with developmental disorders (5659 males and 4200 females) and demonstrated that de novo X-linked ID (XLID) is equally prevalent in females and males, 6% and 7% respectively (Martin et al., 2018a). Males, however, have the additional burden of their non-affected mother’s and grandmother’s de novo variants (Paper IV), and as a result, 10-12% of all male ID cases are estimated to be X-linked. However, the majority of ID cases in males are not X-linked and XLID cannot explain the 40% excess (Vissers et al., 2016). This indicates that there are protective mechanisms in females yet to be discovered. Of interest, Tukiainen et al. recently reported escape of XCI in 23% of 186 investigated X-linked genes in females and highlighted that this might contribute to phenotypic diversity between the sexes (Tukiainen et al., 2017). Hypothetically, the biallelic expression of some escape genes could compensate for variants that are disease-causing in males, and thus, result in a protective effect and reduced penetrance of ID variants.
Mosaicism affects recurrence risk and can revert disease

Mosaicism refers to the existence of two or more genetically distinct cells within one soma originating from the same zygote. Mosaicism is mostly harmless and is, like inherited variants, part of a natural variation. However, there are examples of when mosaicism increases the risk for disease in offspring, causes disease, and even reverts disease phenotypes.

Somatic mosaicism

Variants occurring after the zygote state will be present in that first clone and in all descending cells in that cell line, leading to somatic mosaicism (Fig. 1). In that sense, all individuals are mosaic, as spontaneous post-zygotic variants occur at each cell division from early embryonic development throughout adult life (Forsberg et al., 2017). If the variant arose early in embryonic development, the proportion of mutant cells can be so high that the variant is interpreted as a de novo variant. Acuna-Hidalgo et al. reported that 6.5% of 107 probands with de novo variants were somatic mosaic, probably due to early post-zygotic occurrence of the variant (Acuna-Hidalgo et al., 2015). Distinguishing between inherited, zygotic and post-zygotic de novo variants is important as it can have an effect on disease penetrance, phenotype and recurrence risk. For example, the aneuploidy disorders Down syndrome and Turner syndrome (MIM 300082) are reported to result in milder phenotypes in mosaic patients. There are also examples of mosaic women screened positive for Turner syndrome (45,X) that go on to have a normal reproductive lifespan and no cardiovascular complications (Papavassiliou et al., 2015; Tuke et al., 2018). Another syndrome in which the incidence of mosaicism is central is Proteus syndrome (MIM 176920), presenting with overgrowth and hyperplasia of various organs and tissues by activating variants in the v-akt murine thymoma viral oncogene homolog gene (Lindhurst et al., 2011). Inherited Proteus variants, or variants occurring during early development, are lethal. Thus, all living patients with Proteus syndrome are mosaics and the phenotypic presentation varies depending on when and where the variant occurred during development.

Gonadal mosaicism

Gonadal mosaicism or gonosomal (gonad and soma) mosaicism is mosaicism that includes the germ cells. Gonadal mosaicism arises due to post-zygotic variants in an embryonic cell that later differentiates into germ cells. All germ cells derived from the mutant clone will carry the variant. Thus the proportion of gonadal mosaicism will depend on when the variant arose (Fig. 1) (Forsberg et al., 2017).
As mentioned earlier, a considerable number of \textit{de novo} variants arise due to parental gonadal mosaicism. Studies investigating assumed \textit{de novo} variants identify the variant in 4\% (Campbell et al., 2014) and 8.3\% (Myers et al., 2018) of parental DNA, suggestive of gonadal mosaicism. The reports are likely to be underestimates as some gonadal mosaicism cannot be detected in blood (Paper II), which was the primary source of DNA in both studies. However, in patients with high level gonadal mosaicism (>25\%) the variant has generally occurred in early embryogenesis in a progenitor cell that later gave rise to both blood and germ cells and can therefore be detected in blood. A suggested exception to this is PAE disease variants that can reach high gonadal frequencies without being detectable in blood (Paper II) due to the positive selective advantage, as discussed above.

Since some \textit{de novo} variants are a result of gonadal mosaicism the recurrence risk in future offspring in families with disease-associated \textit{de novo} variants is estimated to 1\% population-wide (Rahbari et al., 2016). However, this estimate does not transfer well to individual couples, since parents with gonadal mosaicism are likely to have a recurrence risk higher than 1\%. Additionally, parents of a child affected due to a zygotic or post-zygotic variant have the same recurrence risk as the general population, which is far less than 1\%. Hence, it is important to define the true source of genetic variants to improve genetic counseling (Paper II).

Revertant mosaicism

Revertant mosaicism (RM) occurs when the pathogenic effect of a germline variant is reverted by a second genetic event. RM can occur by a back mutation of the pathogenic variant or by introduction of a variant that inhibits the disease-causing mechanism, e.g. a truncating SNV or mitotic recombination (Lim et al., 2017). The phenomenon was first described in a patient suffering from Lesch-Nyhan syndrome (Yang et al., 1988). Further investigations have been performed by Jonkman et al. in the skin disorder epidermolysis bullosa (Jonkman et al., 1997), in which RM occurs in about 30\% of patients (Jonkman and Pasmooij, 2009). In these cases, RM gives rise to healthy-looking spots of skin that grow in size due to a positive selective advantage that results in clonal expansion of reverted cells. RM has been suggested as a possible therapeutic target (Lim et al., 2017), and successful transplantation of reverted cells was demonstrated in a epidermolysis bullosa patient in 2006 (Mavilio et al., 2006). There are single cases of successful transplantation of endogenous revertant skin patches in epidermolysis bullosa (Gostynski et al., 2014), but as of yet there are no applications used in a routine clinical setting (Uitto et al., 2016).
Figure 1. A post-zygotic variant will be inherited by all descending clones of that cell line and give rise to mosaicism. If the variant occurred in early embryogenesis the variant might be present in both somatic and gonadal tissue (pink). A variant that occurs in a primordial germ cell will give rise to gonadal mosaicism and not be detectible in somatic tissues like blood (blue). The constant renewal of spermatogonia throughout male life results in accumulation of gonadal replication-error variants with time (red, yellow). Oocytes do not replicate after birth but can acquire e.g. meiotic variants. The parental contribution of de novo variants in the offspring is skewed to a ratio of 1:3.6, maternal: paternal.
Mendelian disorders are often rare, severe and of early onset making diagnosis challenging but crucial. To date, the genetic locus and molecular basis has been described for 5498 Mendelian disorders. However, for 1757 likely Mendelian disorders neither a genetic locus nor a disease mechanism has been identified, preventing genetic diagnosis. For another 1568 disorders the genetic locus is known but the molecular basis is not understood, limiting treatment and development of new therapies (https://www.omim.org/statistics/entry, accessed March 25, 2019). Therefore, translational research of Mendelian disorders remains a prioritized area of research. In papers I-IV, we investigated the etiology of five disorders introduced below.

Keratitis-ichthyosis-deafness syndrome
In Paper I we investigated keratitis-ichthyosis-deafness (KID) syndrome (MIM 148210), an autosomal dominant disorder giving rise to eye inflammation (keratitis), red and scaly skin (ichthyosis), and impaired hearing (Grob et al., 1987). The disorder has only been described in about 100 patients, mostly affected due to missense variant in gap junction beta 2 (GJB2) that encodes the gap junction channel protein connexin (Cx) 26. KID syndrome is caused by gain-of-function variants in Cx26 that give rise to dysfunctional gap junction channels (Garcia et al., 2016). Currently, treatment of KID syndrome is limited to symptomatic relief (Bondeson et al., 2006). Two gain-of-function variants have been associated with a lethal form of KID syndrome, of which one (p.Gly45Glu) is prevalent in the Japanese population but hindered from expression by co-expression of a downstream in cis nonsense variant (Ogawa et al., 2014). Recessive loss-of-function variants in Cx26 is associated with hearing loss (MIM 22029) (Chang, 2015).

Treacher Collins syndrome
In Paper II we investigated Treacher Collins syndrome (TCS; MIM 154500), a developmental disorder characterized by craniofacial anomalies, affecting 1:50,000 live births (Vincent et al., 2016). The major concern for affected children is respiratory failure due to the abnormalities affecting the respiratory
system (Tse, 2016). TCS is an autosomal dominant disorder that in most cases (>60%) is caused by variants in treacle ribosome biogenesis factor 1 (TCOF1), encoding the treacle protein responsible for formation of bone and other facial tissues (Vincent et al., 2016). Variants in POLR1C and POLR1D have recently also been associated with the disorder and demonstrated to be expressed in facial tissues during zebrafish development (Lau et al., 2016; Noack Watt et al., 2016).

Noonan syndrome

In Paper II we also investigated a family affected with NS (MIM 163950), a clinically and genetically heterogeneous disorder with an estimated prevalence of 1:1,000–2,500 live births (Roberts et al., 2013). NS symptoms vary between patients, but often include distinct facial features, craniofacial abnormalities, cardiovascular abnormalities, musculoskeletal abnormalities, cutaneous lesions and in some cases mild ID (Aoki et al., 2016; Roberts et al., 2013). NS has been associated with several genes of the RAS/MAPK pathway, and about 50% of NS patients have autosomal dominant variants in Tyrosine-protein phosphatase non-receptor type 11 (PTPN11) (Roberts et al., 2013). The majority of PTPN11 variants give rise to NS by a gain-of-function mechanism whereby the SHP-2 protein (encoded by PTPN11) has an activating role on the RAS/MAPK pathway (Pannone et al., 2017). This affects the downstream intracellular mechanism that controls cell survival, proliferation, differentiation, migration, and adhesion. Further, somatic PTPN11 variants have been associated with myeloid and lymphoid malignancies, and an increased risk for cancer is reported in NS patients (Aoki et al., 2016; Roberts et al., 2013). Occurrence of NS has been shown to increase with paternal age (the PAE) (Goriely and Wilkie, 2012; Maher et al., 2018).

Cornelia de Lange syndrome

In Paper III we investigated Cornelia de Lange syndrome (CdLS), a heterogeneous developmental disorder divided into five different types depending on the affected gene, estimated to affect 1:10,000–30,000 live births. CdLS manifests in cognitive impairment, growth delay, limb malformations, organ deviations, and characteristic facial features, such as long eyelashes and thick arched eyebrows. The severity of the syndrome is linked to the affected gene. NIPBL is associated with the most severe and common form, accounting for 60% of genetically diagnosed patients (Kline et al., 2018). Around 30% of CdLS patient lack a genetic diagnosis, which hampers prognosis and prediction of recurrence (Boyle et al., 2015)
RAD21 has been associated with a rare form of CdLS, type 4 (MIM 614701) including heterozygous deletions (Deardorff et al., 2012; Pereza et al., 2015), frameshift variants (Boyle et al., 2017; Minor et al., 2014), a splice donor variant, an in-frame deletion (Ansari et al., 2014), and missense variants (Deardorff et al., 2012; Martinez et al., 2017). Like all proteins associated with CdLS, RAD21 is part of the cohesin complex. RAD21 forms the cohesin ring together with SMC1A and SMC3 that joins sister chromatids during cell division, regulates DNA repair, and controls transcriptional processes by folding DNA into TADs (Ji et al., 2016). Disturbed gene regulation is suggested to cause the developmental phenotype seen in CdLS (Dorsett, 2007).

X-linked intellectual disability and TAF1

ID is characterized by significant cognitive impairments defined as IQ<70 and is one of the most prevalent congenital disorders with a worldwide prevalence of 1–2:100. It is a heterogeneous group of disorders that spans in severity (mild-profound) and can include other symptoms such as epilepsy, autism, and/or congenital malformations (Vissers et al., 2016). The X-chromosome is enriched for genes associated with ID, harboring 15% of all identified ID genes (Deng et al., 2014; Neri et al., 2018). XLID accounts for about 10–12% of ID in males (Vissers et al., 2016), and approximately 150 XLID genes have been described (Neri et al., 2018).

In Paper IV we investigated the TATA-box binding protein (TBP)-associated factor 1 (TAF1) gene, that has just recently been associated with syndromic XLID (MIM 300966) (Hu et al., 2016). The syndrome is extremely rare and reported to manifest in mild–severe ID, postnatal growth retardation, delayed gross motor development, delayed speech and language development, and facial features such as prominent supraorbital ridges, long face, low-set and protruding ears and a high palate. In at least three families the variant has been inherited from asymptomatic heterozygous mothers that present with skewed XCI (Hurst, 2018; O’Rawe et al., 2015). TAF1 is the largest unit in the transcription factor II D (TFDII) complex (Fig. 2), of which several other components, e.g. TBP (Rooms et al., 2006), TAF2 (Hellman-Aharony et al., 2013), TAF6 (Alazami et al., 2015) and TAF13 (Tawamie et al., 2017) have been associated with ID. Taf1

Figure 2. TAF1 is a key unit of the transcription initiation complex that is involved in transcription of the vast majority of mRNA genes (Warfield et al., 2017).
expression levels have been shown to be elevated in mice during early embryonic development (Jambaldorj et al., 2012), however, the role of TAF1 during early embryogenesis and its implication in neurodevelopment is still elusive.
Methodology

Novel more sophisticated methods tend to open new doors and improve research by allowing one to investigate more refined questions. The introduction of microarrays and next-generation sequencing (NGS) enhanced resolution compared to previously used karyotyping and targeted fluorescence in situ hybridization methods, which increased disease discovery and diagnostic yield (Acuna-Hidalgo et al., 2016; Bamshad et al., 2011; Veltman and Brunner, 2012). With increased use of sequencing techniques, the availability of large reference sets has grown, enabling better interpretation of sequencing variants. Together with molecular methods that investigate variant function, we have acquired a tool-box that allow us to dig deeper in to the details of Mendelian disease-biology. In Papers I-IV we took advantage of cutting-edge sequencing technique and molecular tools to elucidate Mendelian disease etiology. The main methods are presented below.

Ethical approval

The local ethics committee for human research in Uppsala, Sweden has approved all studies: Dnr 2012/523 (Paper I), Dnr 2012/321 (Papers II-IV), prior to initiation. All clinical investigations and genetic analyses have been conducted in accordance with the guidelines of the Declaration of Helsinki. Patients have been enrolled via Clinical Genetics, Academic Hospital, Uppsala, Sweden, and informed consent was obtained prior initiation. Patient DNA and RNA have been extracted and handled according to standard protocols at Clinical Genetics, Rudbeck Laboratory, Uppsala, Sweden. Animal experimental procedures have been approved by the local ethics committee for animal research in Uppsala, Sweden: permit number C161/4 (Paper IV).

Genome sequencing, the key to the code

The ability to read the genetic sequence of our genome is crucial in order to identify genetic variants. The invention of NGS led to a burst in genome sequencing techniques that improved our abilities to sequence the genome beyond imagination. The last 20 years have been a journey that has taken us from sequencing of single DNA fragments (first-generation), to whole-genome
short-sequencing (next-generation), to today’s long and deep third-generation techniques that generates sequencing reads at single molecule level without breaking the DNA strand. These rapid developments have improved genetic research and patient care. For example the diagnostic yield of patients with ID has increased from around 15% in the early 1990s to a notable 55–70% currently, partly because of the introduction of micro-arrays but also largely because of implementation of advanced exome sequencing pipelines in the clinic (Paper III) (Vissers et al., 2016).

First-generation sequencing
Sanger sequencing was developed 40 years ago and was the first method that allowed precise sequencing of DNA and RNA (cDNA) (Sanger et al., 1977). The method is still widely used for fast and cost-efficient amplification of specific targets <1 kb, both in clinic and research (Papers I–IV). However, because of limited throughput and informative data (e.g. allele frequencies and haplotype information) the method is often replaced by next- and third-generation sequencing techniques (Papers I–IV).

Next-generation sequencing
Genetic research was completely transformed with NGS that enabled cost-effective sequencing of whole-exome, whole-genome and whole-RNA sequencing in less than a week. The techniques generate short sequencing reads (<250 nucleotides), sufficient for SNV and indel variant detection. The introduction of whole-exome sequencing (WES) enabled identification of disease-causing variants, and especially de novo variants, for which previously used linkage analysis was insufficient, and WES is now used in routine clinical assessment (Paper III) (Bamshad et al., 2011; Gilissen et al., 2012; Hu et al., 2016; Martinez et al., 2017; Veltman and Brunner, 2012). Enrichment of the protein-coding region in WES is retrieved by a PCR amplification step, which has the drawback of introducing errors that, along with sequencing artifacts generated by the sequencing method itself, cannot be distinguished from a true variant. This is overcome by increased sequencing depth, but normally not to a level so that low-frequency variants can be discriminated from artifacts. PCR amplification also limits the sequencing of GC-rich regions, resulting in a reduced total coverage. In these cases, whole-genome sequencing (WGS) can be more sufficient since it does not have to be amplification-based, along with the advantage of recovering intronic regions. However, WES has been a gold standard for investigating Mendelian disorders because of the low price, low requirement for input DNA, and coverage of almost all protein sequences (Sims et al., 2014). NGS-variant detection is limited by the short read length that obstructs alignment of repetitive regions and retrieval of haplotype
information, as well as the low sequencing depth that limits low-frequency variant detection.

Analysis of RNA expression by NGS of whole cDNA (transcriptomics) has made it possible to recover a snapshot of gene expression at certain time-points in an organism, tissue, or a specific cell of interest. Targeted sequencing of mRNA of biologically distinct populations (e.g. healthy and affected) and then analysis of differentially expressed genes is often used to illuminate dysregulation of pathways that might be involved in disease. Differential expression analysis depends on read counts of a gene or transcript, rather than variant detection, resulting in other challenges compared to WES and WGS. For one, comparison of fold-time change between lowly and highly expressed genes of different lengths might be difficult, and also gene expression in individual samples is affected by individual variation and environment (Sims et al., 2014). This can partly be addressed by including sufficient numbers of replicates and using established tools and methods for alignment, generation of counts and differential expression analysis (Costa-Silva et al., 2017; Merico et al., 2010; Schurch et al., 2016).

Third-generation sequencing
Novel sequencing chemistry, i.e. third-generation sequencing, has enabled sequencing of native DNA and RNA (cDNA) molecules without prior fragmentation or amplification, allowing increased read length and read depth, as well as sequencing of epigenetic markers (van Dijk et al., 2018). One such method is single molecule real-time (SMRT) sequencing on RSII by Pacific Biosciences. SMRT sequencing generates long (>20kb) and deep sequencing with an accuracy of 99.999%. In contrast to short sequencing methods, SMRT sequencing can thus distinguish between alleles and detect low-frequency variants (Nakano et al., 2017). The technology provides tremendous possibilities for investigating full allele sequences in whole-genome data, with the high cost as a disadvantage. At a lower cost, sequencing of a PCR amplified region can be performed, with the disadvantage of the introduction of PCR artifacts but the advantages of allele-specific full-length reads with a 0.5% variant detection sensitivity (Papers I and II).

Interpreting sequencing variants
Evaluating the pathogenicity of a variant can be challenging. To enable worldwide uniformed classification of sequence variants the American College of Medical Genetics and Genomics (ACMG), the Association for Molecular Pathology and the College of American Pathologists have created official guidelines for variant interpretation. Accordingly, genetic variants are now classified into five categories: benign, likely benign, uncertain significance, likely
pathogenic and pathogenic. The guidelines address how classification should be conducted using population databases, bioinformatic tools, segregation data as well as important complementary functional data (Richards et al., 2015).

Publicly available sequencing data

Population databases of the healthy population are powerful reference datasets that paint a picture of the variant architecture in populations not affected by Mendelian disorders. Even if population databases cannot be assumed to only contain individuals not affected by genetic disease, and likely does contain cases of reduced penetrance, they have facilitated variant interpretation. The 1000 Genomes Project, launched in 2008, was the pioneer project, creating a publicly available reference sequencing database of sequencing variants (Genomes Project et al., 2010). Today, the dataset is complemented by the genome Aggregation Database (gnomAD), a publicly available database providing variant data from >140,000 healthy individuals (Lek et al., 2016). The databases enable interpretation of variants by looking at population allele frequencies and variant density in the sequence of interest. Moreover, gnomAD provides pLI and Z-scores reflecting the observed number of variants compared to the expected (o/e) of protein-coding genes. The SweGen Variant Frequency browser (SweGen) is a similar Swedish initiative generated by the Science for Life Laboratory where WGS data from 1000 healthy Swedish individuals have been collected (Ameur et al., 2017). Publicly available databases have been a key source of information in all studies of this research project.

In silico predictions

In silico computational predictive programs are commonly used to evaluate a variants effect (Papers I-IV). The impact of a missense variant is based on for example, conservation, location within the protein sequence and the biochemical effect (Richards et al., 2015). PhyloP (Rhead et al., 2010) and SIFT (Kumar et al., 2009) are bioinformatic tools that estimates the deleterious effects depending on the sequence conservation, which reflects the sensitivity to genetic change. Similarly, the MutationTaster tools estimates the pathogenic potential of a variant and also include variant data from disease databases such as ClinVar and the Human Gene Mutation Database (Schwarz et al., 2014). ACMG report that most algorithms have a 65–80% accurate prediction rate, however with a tendency to overestimate missense variants as deleterious, and stresses the need to use several independent tools (Richards et al., 2015). Collectively, in silico models are a good complement to assess the function of sequencing variants. However, these are only predictive and molecular studies are essential to unravel the true function of variants of uncertain significance (VUS).
<table>
<thead>
<tr>
<th>Databases and in silico tools</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHOPCHOP</td>
<td><a href="http://chopchop.cbu.uib.no/">http://chopchop.cbu.uib.no/</a></td>
</tr>
<tr>
<td>Cosmic</td>
<td><a href="http://cancer.sanger.ac.uk/cosmic">http://cancer.sanger.ac.uk/cosmic</a></td>
</tr>
<tr>
<td>DECIPHER</td>
<td><a href="https://decipher.sanger.ac.uk/">https://decipher.sanger.ac.uk/</a></td>
</tr>
<tr>
<td>ENSEMBL</td>
<td><a href="https://www.ensembl.org/">https://www.ensembl.org/</a></td>
</tr>
<tr>
<td>ExAC</td>
<td><a href="http://exac.broadinstitute.org/">http://exac.broadinstitute.org/</a></td>
</tr>
<tr>
<td>GnomAD</td>
<td><a href="http://gnomad.broadinstitute.org/">http://gnomad.broadinstitute.org/</a></td>
</tr>
<tr>
<td>GO2MSIG</td>
<td><a href="http://www.go2msig.org/cgi-bin/prebuilt.cgi?taxid=7955">http://www.go2msig.org/cgi-bin/prebuilt.cgi?taxid=7955</a></td>
</tr>
<tr>
<td>HGMD</td>
<td><a href="http://www.hgmd.cf.ac.uk/">http://www.hgmd.cf.ac.uk/</a></td>
</tr>
<tr>
<td>MutationTaster</td>
<td><a href="http://www.mutationtaster.org/">http://www.mutationtaster.org/</a></td>
</tr>
<tr>
<td>OMIM</td>
<td><a href="https://www.omim.org/">https://www.omim.org/</a></td>
</tr>
<tr>
<td>PANTHER</td>
<td><a href="http://www.pantherdb.org/">http://www.pantherdb.org/</a></td>
</tr>
<tr>
<td>SIFT</td>
<td><a href="http://sift.jcvi.org/">http://sift.jcvi.org/</a></td>
</tr>
<tr>
<td>SweGen</td>
<td><a href="https://swegen-exac.nbis.se/">https://swegen-exac.nbis.se/</a></td>
</tr>
<tr>
<td>UCSC</td>
<td><a href="https://genome.ucsc.edu/">https://genome.ucsc.edu/</a></td>
</tr>
</tbody>
</table>
Molecular tools to illuminate underlying mechanisms

Only by modeling VUS’ molecular link to disease can one provide confident evidence of its implication in disease. The mechanism of the disorder can be investigated by using in vitro or in vivo systems, or by studying the effect of a VUS in patient-derived tissues.

Protein detection in vivo and in vitro

A VUS’ effect on protein expression can be investigated in vivo by e.g. a protein expression assay in a cell-based system. Cloning of the hypothesized disease-causing allele into an expression vector, and/or using site-directed mutagenesis to introduce a variant of interest, is a successful approach to study properties such as cellular localization (Paper I) and protein morphology using microscopy.

Proximity ligation-assay-based (PLA) Western blot (WB) is an in vitro assay for specific detection of low-level proteins in solution. In contrast to WB, the PLA-WB technique entails a secondary antibody that allows amplification of the protein signal by rolling-circle amplification, which increases the detection sensibility by up to 16 times (Paper I). By using two or multiple different primary antibodies, a higher specificity can be achieved compared to regular WB (Liu et al., 2011).

Frequency determination with Droplet Digital™ PCR

The frequency of a variant or the expression level of a gene (RNA level) can, as previously mentioned, be sufficiently measured by genome sequencing. However, NGS and third-generation sequencing can be costly, especially if running several samples. In this case, Droplet Digital PCR (ddPCR) is a useful complementary method. ddPCR allows absolute quantification of targeted DNA or cDNA by preforming 20,000 parallel PCRs in oil droplets. The main advantage of ddPCR is the possibility to relatively cheaply and quickly detect low-frequency variants (Paper II). Compared to e.g. real-time PCR, with a detection rate of 1%, ddPCR can detect alleles down to a frequency of 0.001% (Hindson et al., 2011). A drawback compared to NGS is the requirement for target specific probes, which demands knowledge of the target of interest prior to the experiment.

Investigation of X-chromosome inactivation

Detection of skewed XCI is possible by investigating the methylation pattern of X-linked microsatellite markers. The androgen receptor and the retinitis pigmentosa 2 genes contain markers with 80% and 90% heterozygosity in the female population and are thus suitable (Paper IV) (Allen et al., 1992;
The method is limited in providing exact ratios of skewed XCI when using fragment length analysis as the allele detection method. However, it is still a convenient tool because of the simple pipeline using standard methylation sensitive digestion enzymes and PCR amplification.

Gene editing with CRISPR/Cas9 in zebrafish

Procreating a disorder in a model system is a sufficient way to demonstrate pathogenicity of a variant but also to obtain detailed information about gene and disease mechanisms. The possibility to model specific genetic variants drastically improved when the clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR-associated (Cas) 9 system was discovered in 2012 (Jinek et al., 2012). CRISPR/Cas9 has revolutionized molecular genetics by enabling site-directed gene-editing in vitro, in cells and in model organisms (Adli, 2018). The method utilizes the viral defense mechanism of bacteria whereby a single guide RNA aligns to the complementary DNA locus and Cas9 introduces a double-strand break (Jinek et al., 2012). The break is repaired by error-prone non-homologous end joining, often introducing a random indel (Paper IV), or by homology-directed repair where a DNA template is used to repair the strand, enabling introduction of a specific DNA sequence. In in vivo gene editing, the CRISPR/Cas9 system is delivered to a cell that gives rise to a mosaic population. Subsequently, a stable line is established by clonal selection or crossing (Fig. 3). Even though CRISPR/Cas9 has paved the way for disease-modeling it is not flawless. Limitations are, for example, the Cas9’s need for a protospacer-adjacent motif (PAM), e.g. NGG, in the native DNA strand for cleavage, and the specificity of the 20 nucleotide single guide RNA (Adli, 2018). Several recent studies have highlighted the risk for off-target effects (Kosicki et al., 2018), one out of many aspects that will need thorough investigation before germline-directed CRISPR/Cas9 gene therapy can be clinically implemented (Karimian et al., 2019).

The zebrafish (Danio rerio) is a model organism that can be used in combination with CRISPR/Cas9. It has the advantage of spawning large clutches of external transparent embryos and a fast generation time of 3 months. With orthologues for 70% of the human genes, and 80% of human disease genes, the zebrafish is an excellent model organism for human development (Howe et al., 2013). Limitations lies in deviations from human genome structure, such as lack of sex chromosomes. As with any model organism, the zebrafish cannot be assumed to recapitulate the process of a human and interpretations need to be presented with care. Nevertheless, zebrafish have proven successful for studying neurodevelopment with a remarkable degree of conservation compared to humans, both in brain structure and in expression of developmental genes (Sakai et al., 2018).
Figure 3. Schematic overview of the Crispr/Cas9 system. The 20 nucleotide long single guide RNA (sgRNA) aligns to the DNA strand. Cas9 introduces a double-strand break 3–4 nucleotides upstream of the protospacer-adjacent motif (PAM) site, in this case NGG. In paper IV the endogenous non-homologous end joining repair mechanism was utilized to introduce random insertions and deletions in zebrafish embryos. Mosaic F0 were raised and incrossed to generate a heterozygous offspring (F1) that could be crossed with each other to generate a (F2) population of homozygous, heterozygous and wild-type zebrafish.
Relevance and Aim

The project aims to illuminate the etiology of unresolved cases of Mendelian disorders by applying cutting-edge sequencing techniques and molecular tools. The investigations are translational and are intended to directly benefit the affected families by clarifying the genetic cause of their disorders. A genetic diagnosis is crucial to confirm the clinical diagnosis, allow physicians to provide information regarding the course of the disease, provide the best possible treatment, and enable family planning. A recent study by Krabbenborg et al. showed that a diagnosis aids the parents to become more accepting of the situation, cope with feelings of guilt, deal with the outside world and adapt care and activities for the child’s needs (Krabbenborg et al., 2016). Further, molecular genetic investigation of disease is a successful approach to gain knowledge of underlying disease mechanisms, which is key for future development of new therapies and understanding of general biological developmental processes.

To date, the genetic cause and molecular basis is known for 5498 Mendelian disorders. However, for thousands of disorders knowledge of disease- etiology is incomplete, leaving families with limited diagnosis, prognosis, and treatments, and hampering development of new therapies.

During this project we aimed to:

- Characterize novel syndromes, genes and genetic variants in families affected with unresolved Mendelian disorders, with a focus on neurodevelopmental disorders manifesting in ID.
- Enable delivery of diagnosis, prognosis, improved care and possibility for family planning in affected families.
- Enhance knowledge of molecular mechanisms in normal developmental processes by investigating underlying causes of disease.
Results and Discussion

**Paper I:** Revertant mosaicism repairs skin lesions in a patient with keratitis-ichthyosis-deafness syndrome by second-site mutations in connexin 26

We aimed to investigate the mechanism giving rise to healthy-looking spots of skin in a patient affected with KID syndrome.

**Result**

The patient presented with skin lesions, hearing deficiency and keratitis since early childhood, suggestive of KID syndrome. A recurrent *GJB2* c.148G>A, p.Asp50Asn variant (NM_004004.5) confirmed the clinical diagnosis (Bondeson et al., 2006). At the age of 20, the patient developed healthy-looking spots of skin within her erythrokeratodermic skin lesions on the inside of her thighs (Fig. 4A). Within a few years the spots had grown in size and number (Fig. 4B) and spread to her hands.

![Figure 4](image-url)

**Figure 4.** Patient feature and schematic illustration of a gap junction channel. (A) The inside of the patient’s thigh displayed healthy-looking spots within the affected area of skin. (B) After a few years the spots had grown in size and number. (C) The patient developed squamous cell carcinoma. (D) Schematic picture of two hemi-channels connecting two cells and forming a gap junction channel. One hemichannel obtained six Cx26 units. The disease disease-causing variant (green) and second-site, somatic variants (blue) are marked.
Two biopsies from the affected tissue and two biopsies from healthy-looking spots were investigated by SMRT sequencing. A 4.1 kb (DNA) and 1033 bp (cDNA) region of \textit{GJB2} was sequenced, covering the protein-coding sequence of 678 bp, to a depth of >10,000 reads. We detected a total of five somatic variants present on both DNA and RNA level in frequencies of 2.4–12.5% in skin biopsied from the healthy-looking spots (Table 2; Fig. 4D blue). All variants were found \textit{in cis} with the disease-causing p.Asp50Asn variant. No somatic variants were identified in biopsies from affected tissue.

Three of the somatic variants, p.Gly21Arg, p.Asp46Asn, and p.Ser138Asn, have been associated with autosomal hearing loss in previous studies (Bazazzadegan et al., 2011; Rabionet et al., 2006; Snoeckx et al., 2005). Two VUS, p.Asp46Ala and p.Ala148Asp, were identified. They are not reported in the gnomAD (Lek et al., 2016) or SweGen (Ameur et al., 2017) public databases (March 25, 2019), and are predicted as disease-causing (MutationTaster), deleterious (SIFT) and conserved (PhyloP) by \textit{in silico} prediction tools (Table 2).

\textbf{Table 2:} \textit{In silico} predictions of the five somatic variants. Three variants had previously been implicated in autosomal recessive hearing loss.

<table>
<thead>
<tr>
<th>Protein pos.</th>
<th>MutationTaster</th>
<th>SIFT</th>
<th>PhyloP, conservation</th>
<th>Reported in hearing loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.Gly21Arg</td>
<td>disease-causing</td>
<td>deleterious 0</td>
<td>highly (5.94)</td>
<td>Rabionet et al. 2006</td>
</tr>
<tr>
<td>p.Asp46Asn</td>
<td>disease-causing</td>
<td>deleterious 0</td>
<td>highly (5.94)</td>
<td>Bazazzadegan et al. 2011</td>
</tr>
<tr>
<td>p.Asp46Ala</td>
<td>disease-causing</td>
<td>deleterious 0</td>
<td>highly (4.89)</td>
<td>\textit{In this report}</td>
</tr>
<tr>
<td>p.Ser138Asn</td>
<td>polymorphism</td>
<td>tolerated (0.2)</td>
<td>weakly (1.09)</td>
<td>Snoeckx et al. 2005</td>
</tr>
<tr>
<td>p.Ala148Asp</td>
<td>disease-causing</td>
<td>deleterious 0.01</td>
<td>moderately (2.38)</td>
<td>\textit{In this report}</td>
</tr>
</tbody>
</table>

The effect of the somatic variants on Cx26 protein (encoded by \textit{GJB2}) was investigated by transfection of fluorescently-tagged Cx26 protein with the p.Asp50Asn variant, as well as Cx26 protein with the p.Asp50Asn variant and all the five somatic variants expressed individually. Cx26 p.Asp50Asn formed gap junction channel plaque in the same manner as wt Cx26 (Fig. 5A). By contrast, Cx26 p.Asp50Asn with somatic variants did not form gap junction channel plaques (Fig. 5B–F). The results indicate that Cx26 Asp50Asn with the somatic variants identified in the patient does not contribute to formation of gap junction channels, which reverts the dominant negative effect of Cx26 p.Asp50Asn. The vague green fluorescent signal implied that Cx26 p.Asp50Asn with somatic variants was intracellularly expressed (Fig. 5B–F), which was confirmed with the PLA-WB assay, detecting low expression of Cx26 p.Asp50Asn with somatic variants in transfected HeLa cells.
Figure 5. Transfection results displaying wt Cx26, Cx26 p.Asp50Asn, and Cx26 p.Asp50Asn expressing all somatic variants individually. Arrows mark gap junction channel formation. (A) Cx26 p.Asp50Asn (green) and wt Cx26 (red) formed gap junction channels in a similar way. The scale bar: 15 µm. (B-F) When expressing wt Cx26 and Cx26 p.Asp50Asn with additional somatic variants only expression of wt Cx26 could be noted.

Discussion

Five somatic variants were identified within reverted tissue, independently present in cis with the disease-causing variant. Patient skin cells with nullifying variants in cis with the disease-causing variants likely proliferate under positive selective advantage because of the restored gap junction channel function, resulting in reversion of the skin phenotype and RM. Nullifying somatic variants likely also occur in trans with the disease-causing variant. However, clones that do not express wt Cx26 protein at the cell surface are suggested to be under negative selection due to enhanced disturbance of gap junction channel function. Investigation of protein expression suggests that Cx26 with secondary somatic variants is intracellularly detained after translation. Hypothetically, for example, posttranslational processes are hampered, and transportation is hindered during oligomerization to hexameric units within the endoplasmic reticulum (Ahmad and Evans, 2002; Johnstone et al., 2012).

KID syndrome is thought to arise due to a gain-of-function mechanism whereby lost ability to regulate hemichannel activity results in hyperactive “leaky” gap junction channels (Garcia et al., 2015; Sanchez and Verselis, 2014). Missense variants in GJB2 are also the most common cause of recessive non-syndromic hearing impairment, caused by loss of Cx26 function (Zazo Seco et al., 2017). Understandably, three out of five somatic variants identified in this study have previously been associated with recessive hearing impairment, shedding light on the mechanism of RM in our patient. RM is often seen in congenital skin disorders, such as epidermolysis bullosa.
(Jonkman and Pasmooij, 2009; Jonkman et al., 1997), but this is the first time, to our knowledge, that RM is reported in a patient with KID syndrome.

We sought to shed a light to the mechanism driving the mutagenic process in the patient by investigating mutational signatures (Alexandrov et al., 2013). Based on the five somatic variants ultraviolet (UV) light is a likely source of occurrence (signature 7, retrieved from patients with melanoma). Like other KID syndrome patients, the patient also developed squamous cell carcinoma, the most common form of skin cancer caused by uncontrolled growth of squamous cells in the epidermis of the skin. Interestingly, the skin cancer occurred on the inside of the patient’s thigh, in the same area as the reverted spots (Fig. 4C). Squamous cell carcinoma normally occurs in elderly people because of extensive sun exposure and is thus UV-induced (Parekh and Seykora, 2017). Hypothetically, patients with KID syndrome could be at a higher risk of accumulating UV-induced variants due to reduced UV protection in the chronically inflamed skin. Another suggestion is that the somatic variants and the cancer occur due to the inflammation itself (ichthyosis), and that they are not linked to exposure to UV light. Inflammatory processes have been shown to give rise to cancer by promoting tumor cell proliferation and inhibiting apoptosis through inflammation related signaling molecules like cytokines and chemokines. The link to UV light through signature 7 would then suggestively be because the inflammation giving rise to melanoma is mostly UV-induced (Maru et al., 2014). A third hypothesis is that the occurrence of somatic variants and cancer in patients with KID syndrome is directly linked to altered expression of GJB2. Elevated levels of Cx26 have been associated with cancer and Cx has been suggested as a biomarker and a possible target for cancer therapy. Specifically, expression of GJB2 has been implicated in pancreatic cancer (Zhu et al., 2017), and in the tumorigenesis of breast cancer, in which knockdown of GJB2 led to reduction in tumor proliferation (Shettar et al., 2018). The investigation of signature mutations in this study was limited to the five somatic variants identified in the patient. The occurrence of somatic variants would possibly be better understood if a larger set of somatic variants was investigated. Investigation of an increased number of biopsies would possibly reveal additional secondary variants resulting in loss-of-function. Those variants are likely to be implicated in autosomal recessive hearing loss, and further investigation could thus provide a map over genotype-phenotype correlations in GJB2 that could be of clinical use. However, further investigation must be ethically weighted against the uncongenial situation of the patient.

RM is often referred to as natural gene therapy and the mechanism has the possibility to supply insight into potent therapeutic tools. Attempts to transplanting reverted tissue into epidermolysis patients have been made but are still far from being implemented in routine treatment (Gostynski et al., 2014; Lim et al., 2017; Mavilio et al., 2006). Dominant skin disorders like KID syndrome, could potentially be treated by different applications of CRISPR/Cas9: introducing truncating variants (Jinek et al., 2012), suppressing the mutant
allele (Long et al., 2015) or introducing back-mutations (Komor et al., 2016). Ideally, gene therapy would be introduced already at the embryonic stage to revert disease in all tissues. At this point, the introduction of variants in embryos for reproduction is prohibited, due to ethical issues like the limited knowledge of its effect on the genome and future generations. Most countries also prohibit research-based gene editing of human embryos. Instead, since somatic variants in adult tissue have been proven to restore cell function and revert disease, somatic gene therapy could be sufficient in rapidly dividing cells (van den Akker et al., 2018). Nevertheless, off-target effects and disease-specific consequences of such approaches, in this case for example the upcoming of skin cancer, would need to be carefully investigated before such interventions can be proven safe for recipient patients.

In conclusion, we investigated a patient with KID syndrome who developed healthy-looking skin due to RM. We identified five nullifying somatic variants in two patient biopsies and demonstrated that they inhibit expression of mutant Cx26 at the cell surface. This is the first time, to our knowledge, that RM has been demonstrated to revert the skin phenotype in a patient with KID syndrome.
Paper II: A novel approach using long-read sequencing and ddPCR to investigate gonadal mosaicism and estimate recurrence risk in two families with developmental disorders

We aimed to determine parental origin, risk of recurrence and presence of gonadal mosaicism in two families: (1) to determine if the oocyte donor could be used for future donations, and (2) aid prenatal diagnosis and establish suspected gonadal mosaicism.

Result

The first family gave birth to a boy affected with suspected Treacher Collins syndrome (TCS) after assisted reproduction involving oocyte donation (Fig. 6A). The second family had terminated two pregnancies due to fetal abnormalities discovered during ultrasound and suffered two miscarriages (Fig. 6B). NS was suspected based on clinical assessment of the two affected fetuses.

In family one, directed Sanger sequencing of a previously TCS-associated TCOF1 gene on patient and father blood DNA revealed a VUS in TCOF1 c.3156C>T, p.Gly1052Gly (NM_001135244), only present in the boy. The VUS had been associated with TCS previously, but the effect had not been further characterized (Bowman et al., 2012). TCOF1 c.3156C>T is not reported in the gnomAD (Lek et al., 2016) or SweGen (Ameur et al., 2017) databases (March 25, 2019), and is predicted to create a new splice donor site by MutationTaster. The alleles were separated by TA-cloning of cDNA generated from patient blood RNA, confirming that a novel splice donor site was created. The splice donor site affected two TCOF1 transcripts, resulting in expression of two additional transcripts with premature stop codons (Fig. 7).

In family two, Sanger sequencing of fetus II:2 was performed on a NS panel covering five NS associated genes, PTPN11, SOS1, RAF1, KRAS and SHOC2. Sequencing revealed a recurrent PTPN11 c.923A>C, p.Asn308Thr variant (NM_002834.3) previously reported to cause NS (rs121918455) (Tartaglia et al., 2006).
Figure 7. The TOCF1 c.3165C>T variant introduced a splice donor site in exon 19, affecting two transcripts (A) In transcript 1, the c.3156C>T variant resulted in a deletion of 29 base pair and a frameshift of 24 amino acids led to a premature stop codon in exon 20. (B) In transcript 2, the c.3156C>T variant resulted in an immediate stop in the first amino acid of exon 21.

A haplotype analysis was performed on available parent and index patient DNA using SMRT sequencing. In family one, sequencing a 3.8 kb region in the father and the affected offspring revealed four informative SNPs (rs79012265, rs77558738, rs2295223 and rs8004246) as well as the disease-causing variant. The patient was heterozygous for all SNPs whereas the father was homozygous. The disease-causing variant had occurred on patient allele 1 (Table 3, grey), indicating that the variant had occurred on the father’s allele and thus could not have originated from the oocyte donor. A 3.9 kb region including the PTPN11 variant was sequenced in family two (trio). However, the region lacked informative markers and haplotype phasing could not be performed using this method.

Gonadal mosaicism was investigated by SMRT deep sequencing of paternal sperm DNA in duplicates, utilizing the advantage of discovering low-frequency reads. For the father of the patient affected with TCS, one mutant TCOF1 read could be detected, resulting in a frequency below the detection rate for SMRT sequencing. The PTPN11 variant was detected in 36.6% and 37.1% of the reads in paternal sperm DNA. The sequencing frequencies were confirmed using ddPCR, which established that the mutant allele could not be detected in the father’s sperm or blood, but was, as expected, detected in 49.7% of the heterozygous index patient blood. In the family affected with NS, ddPCR detected the variant in 38.0% of the father’s sperm. Sperm from family one was used as negative control for the ddPCR assay for family two (Fig. 8).

Table 3: Haplotype analysis of family one with SMRT sequencing revealed four informative markers within the 3.8kb region. The TCOF1 c.3156C>T variant (red) was present on the paternal haplotype (grey) in the index patient.

<table>
<thead>
<tr>
<th>Sample</th>
<th>rs79012265</th>
<th>rs77558738</th>
<th>c.3156C&gt;T</th>
<th>rs2295223</th>
<th>rs8004246</th>
<th>Reads</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient allele 1</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>269</td>
<td>51.9%</td>
</tr>
<tr>
<td>Patient allele 2</td>
<td>A</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>249</td>
<td>48.1%</td>
</tr>
<tr>
<td>Father allele</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>1839</td>
<td>100%</td>
</tr>
</tbody>
</table>
Figure 8. Occurrence of gonadal mosaicism was assessed with SMRT sequencing and Droplet Digital PCR (ddPCR). No gonadal mosaicism could be detected for the variant causing Treacher Collins syndrome that had occurred on the paternal allele (red). Gonadal mosaicism was detected in 38% of paternal sperm in the family affected with Noonan syndrome (blue).

Discussion

The variant causing TCS in family one had occurred on the paternal allele. Gonadal mosaicism was not detected in paternal sperm, determining the recurrence risk to be <0.1%. Presence of low-level gonadal mosaicism (<0.1%) in the father is possible, however, ddPCR performed on the father’s DNA from blood and sperm in triplicates detected no mutant-positive droplets, suggestive of no gonadal mosaicism. The one mutant read detected with SMRT sequencing is below detection frequency of the method and can be a result of PCR artifacts or a single positive sperm cell, for example. The ddPCR results suggest that the variant detected in the patient occurred in a single sperm cell, in the zygote or post-zygotically. Hence, most likely the family has no elevated risk compared to the rest of the population for recurrence of TCS when using the same oocyte donor and the father’s sperm. Additionally, future families can use the same oocyte donor without elevated risk for TCS.

For the family affected with NS, we identified paternal gonadal mosaicism and estimate the recurrence risk to be around 40%. However, this can potentially vary with time depending on when during spermatogenesis the variant has occurred. NS and variants in PTPN11 are reported to be associated with the PAE and mutant sperm cells are suggested to gain a selective advantage (Goriely and Wilkie, 2012; Yoon et al., 2013). Thus, there might be a risk of
an increased ratio of mutant sperm with time, if the selective advantage causes the mutant cell population to grow. Variants associated with the PAE, including *PTPN11*, have also been suggested to predispose to cancer (Goriely et al., 2009).

Establishing presence of gonadal mosaicism is essential for parents of children affected with *de novo* variants in order to deliver the accurate recurrence risk of future offspring. To date a general risk of 1% is given to the parents, but for parents with gonadal mosaicism the risk of recurrence can be as high as 50%. Previous studies have shown that 4–8% of determined *de novo* variants are in fact parental mosaic variants present in parental blood, indicative of parental germline mosaicism (Campbell et al., 2014; Myers et al., 2018). Further, the risk for paternal gonadal mosaicism will increase with age for PAE disorders like RASopathies, including NS. The results of this study indicates that it might be effective to screen paternal germline DNA in families with *de novo* PAE-associated variants to establish the risk of recurrence in future offspring. This study provides a successful approach to do so. For example, the recurrence risk for non-affected parents with a previous child with NS is estimated to be 5% (Sharland et al., 1993). Hypothetically, screening of paternal sperm DNA on a ddPCR based panel including previously described PAE-associated variants (Maher et al., 2018) could be clinically established. This would aid clinical counseling and enable disease prevention for families with gonadal mosaicism, as well as reduce unnecessary distress in families that do not have gonadal mosaicism.

Conclusively, we used a novel approach to investigate the origin and recurrence risk of two *de novo* variants utilizing third-generation sequencing and ddPCR. We established that both variants had occurred on the paternal allele and that paternal gonadal mosaicism was present in the family affected with NS (approximately 40%), but not in the family affected with TCS (<0.1%). The analysis assisted genetic counseling, the course of action and/or prenatal diagnosis, highlighting the possibilities of, and the need to bring available techniques into a clinical practice.
Paper III: A novel RAD21 p.(Gln592del) variant expands the clinical description of Cornelia de Lange syndrome type 4 – review of the literature

We aimed to characterize the etiology of a suspected Mendelian disorder in a 15-month year old boy with an unknown developmental disorder.

Result

A boy of two healthy non-related parents presented with an unknown congenital disorder resulting in growth and developmental delay, distinct facial morphology and organ malformations since early childhood (Fig. 9A). A clinical WES test was performed on blood-derived DNA from the family trio. Analysis revealed a VUS, a three-nucleotide deletion in RAD21 c.1774_1776del, p.(Gln592del) (NM_006265). The VUS was confirmed de novo by Sanger sequencing of trio DNA from blood. Clinical assessment suggested CdLS type 4 (CdLS4).

The variant is not reported in reference databases (gnomAD or SweGen, March 25, 2019) and gnomAD reports a Z-score of 2.64 for RAD21 missense variants (o/e = 0.6). The deletion is located at an evolutionary conserved locus (PhyloP score 4.2) and the alteration is predicted to be disease-causing (MutationTaster) and deleterious (PROVEAN: –11.124). The deletion is located in the C-terminal domain of RAD21 protein that facilitates formation of the cohesin ring by binding to structural maintenance of chromosomes 1A (SMC1A) protein. In silico modeling demonstrated that Gln592, together with downstream amino acids Arg590 and Lys591 are directly linked to SMC1A residues Glu1191, Glu1192, Glu1198. According to the in silico model, deletion of RAD21-Glu592 resulted in a positional shift of RAD21-Lys591 and

Figure 9. The predicted effect of RAD21 p.Gln592del variant. (A) The patient presented with low-set, protruding ears, thick and highly arched eyebrows, long and prominent eyelashes, short nose and micrognathia. (B) Wild-type RAD21-SMC1A binding site, structure and electrostatic charge (blue: positive, red: negative). (C) Deletion of p.Gln592 is predicted to result in a shift of Lys591 losing its interaction with SMC1A.
thereby a lost connection to SMC1A residues Glu1191 and Glu1192 (Fig. 9B-C). Further, Glu592 is located in the same alpha helix as Lys605 that is important for stabilization of the ATPs-dependent locus that catalyzes opening of the cohesin ring (Marcos-Alcalde et al., 2017). Deletion of Glu592 is predicted to result in a conformational change that affects Lys604 (previously Lys605) and possibly the catalytic locus.

**Discussion**

CdLS is divided into five subtypes depending on the affected gene, with great phenotypic variability between the different subtypes. About 30% of patients clinically diagnosed with CdLS are without a genetic diagnosis (Boyle et al., 2015). Likely there are patients with CdLS worldwide without both clinical and genetic diagnoses, so called uncharacterized cases, especially for mild forms like type 4. Lack of diagnosis obstructs proper genetic counseling of affected families as disease prognosis and recurrence risk cannot be established. *RAD21* has previously been associated with a mild form of CdLS. However, since only eight different alterations and 13 patients have been described in the literature complete knowledge of symptomatology is lacking, emphasizing the importance of gathering clinical and genetic findings (Kline et al., 2018).

We assessed a novel *RAD21* variant according to ACMG guidelines (Richards et al., 2015), and a diagnosis could be provided to the background of the clinical findings overlapping with previous patients, bioinformatic predictions, and *in silico* modeling of the variant’s impact on protein structure. As the variant could not be detected in DNA from parental blood it was established to be de novo. However, since only parental DNA from blood was screened and Sanger sequencing has a detection sensitivity of about 10%, it cannot be excluded that there might be a risk of parental mosaicism. Hence, a recurrence risk of 1% was handed to the family. However, parental gonadal mosaicism resulting in CdLS in the offspring is reported as low (0.89%) (Kline et al., 2018). Since the variant was not present in >10% of paternal blood and *RAD21* has not, to the best of our knowledge, been associated with the PAE (Maher et al., 2018), the risk for recurrence is likely low.

By structural modeling we could predict the variant’s effect on protein structure. However, the study is limited in molecular observations regarding the variant’s molecular effect as well as novel insight into CdLS disease mechanism. The RAD21 protein is known to form the cohesin ring by linking to the SMC1A and SMC3 head domains. The cohesin ring is responsible for maintaining the connection between the sister chromatids during cell division (Nasmyth and Haering, 2009), and cleavage of RAD21 will later allow their separation (Lin et al., 2016). Investigation of altered sister chromatid separation in CdLS has been contradictory, but generally premature chromatid separation has not been proven to be a sufficient marker for CdLS, nor is it
thought to be the cause of the syndrome (Castronovo et al., 2009; Kaur et al., 2005). Instead, the cohesin complex’s involvement in regulation of chromatin architecture and thus transcriptional control and DNA repair is thought to give rise to CdLS (Liu and Krantz, 2009; Nasmyth and Haering, 2009). It has been suggested that haploinsufficiency of CdLS genes disturbs gene regulation during embryonic development, giving rise to the symptoms seen in CdLS. This was recently investigated in a mice model that was haploinsufficient for the orthologue to human \textit{NIPBL}, the most prevalent gene in CdLS. In line with suggested disease-mechanism of CdLS, the \textit{Nipbl}^{+/−} mice showed global decrease in cohesin binding and reduced promotor-enhancer activity for genes important for development (Newkirk et al., 2017).

In conclusion, we report a novel one amino acid deletion in \textit{RAD21}, likely causing extremely rare CdLS4, and we summarize previously described cases. The clinical and genetic investigations will aid future patients. However, with few reported cases affected by \textit{RAD21} variants, and the high heterogeneity observed for CdLS, we suggest that clinical prognosis of CdLS4 should be carefully delivered.
**Paper IV:** *TAF1*, associated with intellectual disability in humans, is essential for life and regulates neurodevelopmental processes in zebrafish

We aimed to investigate the role of *TAF1* during embryonic development by assessment of the first complete *taf1* knockout model, using zebrafish. We also sought to illuminate clinical and genetic features of a family affected by a *TAF1* c.3568C>T, p.(Arg1190Cys) variant.

**Results**

Six related males in a five-generation family presented with syndromic XLID, including ID, delayed speech and language development, as well as characteristic facial features (Fig. 10A). A *TAF1* c.3568C>T, p.(Arg1190Cys) variant was identified by X-exome sequencing (Hu et al., 2016). Sanger sequencing confirmed that the variant segregated according to the inheritance pattern in 17 individuals: five asymptomatic hemizygous females, of which four were mothers of affected males, two affected hemizygous males as well as ten non-affected, not carrying the variant (Fig. 10B, red). Analysis of X-chromosome methylation and RNA expression revealed fully skewed XCI in all five carrier females (Fig. 10C).

The *TAF1* c.3568C>T, p.(Arg1190Cys) variant affects a conserved site (PhyloP, GERP++), is predicted to be disease-causing (MutationTaster) and deleterious (SIFT), and is not present in the gnomAD (Lek et al., 2016) or SweGen population databases (Ameur et al., 2017) (March 25, 2019). Further, *TAF1* has a Z-score of 5.49 (o/e = 0.44) and pLI of 1 (o/e = 0.0) (Lek et al., 2016).

**Figure 10.** Clinical and genetic characteristics of the investigated family. (A) Clinical features of III:2 and IV:5 display characteristic facial features such as long face, prominent supraorbital ridges, deep-set eyes, low-set protruding ears, and pointed chin. (B) Segregation analysis confirmed *TAF1* c.3568C>T heterozygosity in five asymptomatic female carriers, hemizygosity in two affected males, and only wild-type allele in ten non-affected (red). (C) Female carriers (*n* = 5) presented with fully skewed X-chromosome inactivation, investigated by fragment length analysis of polymorphic methylation markers (left) and Sanger sequencing of the *TAF1* c.3568C>T variant on DNA (top, right) and RNA (bottom, right).
Knockout of TAF1 orthologue (taf1) in a zebrafish model was generated using the CRISPR/Cas9 system to investigate taf1 implication in gene regulation during early embryonic development. The single guide RNA targeted exon 8, resulting in a 10 bp deletion, a frameshift of 8 amino acids, and a premature stop codon. At 3 days post fertilization (dpf) homozygous taf1<sup>-/-</sup> fish showed general developmental delay with features such as reduced length, underdeveloped ears, eyes, tectum and cartilage, short pectoral fins and dorsally bent body axis. Heart and cerebral edema as well as blood-filled cavities could also be noted (Fig. 11A-B) indicating early embryonic lethality (Kimmel et al., 1995). There was no measurable difference between heterozygous and wt embryos (Fig. 11B).

Transcriptome analysis at 3 dpf revealed 6628 genes as differentially expressed in taf1<sup>-/-</sup> compared to siblings (taf1<sup>++/+</sup> and taf1<sup>-/+</sup>), at significance level adjusted p-value (padj) <0.01 using DESeq2 paired differential gene expression analysis. Gene set enrichment analysis and PANTHER overexpression analysis was used to investigate enrichment of genes of certain pathways. Gene set enrichment analysis, visualized by enrichment map, on all differentially expressed genes (n = 6628), suggested upregulation of genes associated with chromatin and DNA assembly pathways, and muscle development, as well as downregulation of genes associated with ion channel pathways, and sensory and visual perception. The result was replicated by PANTHER overexpression analysis on genes more than 4-fold overexpressed (n = 258), showing 54.8 enrichment for chromatin assembly genes (Fig. 11C, right), and genes more than 4-fold underexpressed (n = 612), showing 7.5 enrichment for neuromuscular synaptic transmission (Fig. 11C, left). Overexpression analysis of all genes revealed 2.2 enrichment for neuromuscular synaptic transmission. V-plot of all differentially expressed genes highlighted RNA polymerase II (polr2) subunit a, bromodomain testis-specific protein (brdt) gene, several histone genes (e.g. hist2h3c) and 5.8 ribosomal RNA (5.8 rRNA) as upregulated, and myelin basic protein a (mbpa) gene as well as potassium voltage-gated Channel subfamily J member (kcnj) 12b as downregulated.

Discussion

Affected individuals IV:5 and V:4 were included in a previous XLID cohort study in which the novel XLID candidate gene TAF1 was identified (Hu et al., 2016). No clinical information nor molecular investigation of the family was presented. We described six males with symptoms similar to recent publications (Hurst, 2018; O’Rawe et al., 2015) and additional features such as large hands and feet, short necks and deep-set eyes. Five asymptomatic female carriers had completely skewed XCI, as described in three families previously (Hurst, 2018; O’Rawe et al., 2015). Usually, XCI follows a normal distribution with an average of 50% maternal and 50% paternal X-chromosome expression. Extreme skewing >95% is rare in
Figure 11. Zebrafish taf1<sup>-/-</sup> model. (A) 3 days post fertilization (dpf) taf1<sup>+/+</sup> (right) zebrafish embryos show heart and ventricle edema, blood filled cavities, bend body axis, and general underdevelopment (UD) including short pectoral fins, reduced length and underdeveloped cartilage, eyes, and ears. (B) Phenotype quantification demonstrated difference between taf1<sup>+/+</sup> and taf1<sup>-/-</sup> but not between taf1<sup>+/+</sup> and taf1<sup>-/+</sup>, using Student t-test, p-value adjusted with Bonferroni correction. (C) V-plot of 6628 differentially expressed genes (padj < 0.01) highlighted kcnj12b (blue), mbpa (pink) brdt (purple), polr2a (green), histone genes (red) and rRNA (yellow). Top association from PANTHER overexpression analysis of top downregulated (blue), all dysregulated (black) and top upregulated (red) is displayed.

the healthy population (0.8%) (Amos-Landgraf et al., 2006). However, female asymptomatic carriers of recessive XLID variants often present with skewed XCI, thought to appear because of selective advantage of wt cells or disadvantage of mutant cells, likely due to reduced cell viability or cell proliferation (Plenge et al., 2002; Tzschach et al., 2015). Muers et al. investigated the
mechanism behind skewed XCI in Alpha-thalassemia XLID syndrome using a mouse model and demonstrated that the skewed XCI was not present at embryonic day 8. Interestingly, the skewed XCI was shown to occur during gastrulation, embryonic day 11–15 at different time-points in different tissues. The results indicated that skewed XCI in Alpha-thalassemia XLID does not occur due to a general defect in proliferation rate or cell viability, but because of different selective mechanisms in different cell lines (Muers et al., 2007). TAF1 has previously been associated with proliferation processes and regulation of cyclin (CCN) D1 (Hilton et al., 2005; Kloet et al., 2012). A recent article reported reduced expression of CCND1 and CCNA2 in a neuroblastoma cell line overexpressing TAF1 p.Ser1600Gly, a variant also detected in a patient with XLID. Further, they associate the reduced expression of cyclin genes to reduced proliferation (Hurst, 2018). This observation was replicated in our dataset presenting downregulation of both ccnd1 (log2 -0.46, padj = 1.2E-07), as well as ccna2 (log2 -0.43, padj = 8.9E-04), underlining a possible mechanism causing skewed XCI. However, as emphasized by Muers et al., the precise mechanism of skewed inactivation can be complex and does not necessarily occur due to a general reduction in proliferation abilities of mutant cells (Muers et al., 2007). Apart from ccnd1 and ccna2 our dataset reveals striking upregulation of ccna1 (log2 3.6, padj = 2.2E-16) and to a lesser extent, dysregulation of other cyclins, including B1, B3, C, E2, G1, L1 and T2. Additional studies are needed for further conclusions about the time-point and selective mechanism giving rise to skewed XCI in XLID associated with TAF1.

TAF1 has repeatedly been associated with neurodevelopmental disorders (Hu et al., 2016; Hurst, 2018; O’Rawe et al., 2015), but to our knowledge, no investigation of TAF1’s relation to early neurogenesis has been performed. A previous study in mice showed elevated levels of Taf1 during embryonic development, suggesting a key role during early development (Jambaldorj et al., 2012). The lethal outcome in taf1^-/- zebrafish embryos suggests that TAF1 is essential for life, which is further strengthened by observations in gnomAD population dataset in which TAF1 is depleted from loss-of-function variants and has a pLI score of 1 (o/e = 0) (Lek et al., 2016). We sought to highlight pathways that might be correlated with the features noted in the patients with TAF1 variants by transcriptomic analysis of taf1^-/- zebrafish knockouts during embryonic development.

Among the top downregulated genes (< -2 log2, n = 612) PANTHER overexpression analysis revealed enrichment (7.5 fold) of genes involved in neuromuscular synaptic transmission (Fig. 11C, left), mostly due to downregulated GABA receptor genes. The association was strengthened by the gene set enrichment analysis visualized with enrichment map that highlighted ion transportation pathways as downregulated. A recent study specifically investigated altered expression of 86 neuronal ion-channel genes in SH-SY5Y neuronal cells treated with TAF1 siRNAs. Out of 86 investigated genes eight
showed differential expression, four of which overlap with the 6628 differentially expressed genes in our study: calcium channel, voltage-dependent, T type, alpha 1G subunit (CACNA1G), acid-sensing (proton-gated) ion channel (ASIC) 1, ASIC2, and KCNJ14 (Hurst, 2018). Variants in a paralogue, KCNJ11, have been associated with neurological features before (Gloyn et al., 2006), and interestingly another paralogue, KCNJ12 (kcnj12b), is our top underexpressed gene, log2 -6.1, padj = 6.8e-04 (Fig. 11C, blue). In contrast to the study by Hurst et al., our analysis revealed dysregulation of 106 ion channel genes. The low number of differentially expressed genes in Hurst study (8/86) and small overlap with our datasets (4/6628), could be because of the different time-points, tissues, and method of downregulation. Using adult neurons that have expressed TAF1 during development limits the investigation of compensatory mechanisms and the effect of TAF1 depletion during embryogenesis when XLID likely is established. Another highlighted downregulated gene in our dataset was mbpa (Fig. 11C, pink), responsible for CNS myelination (Popko et al., 1987) and associated with the neurological disorder multiple sclerosis in humans (Sheremata et al., 1974).

The top upregulated genes (>2 log2, n = 258) were strongly enriched (54.5 fold) for genes associated with chromatin assembly, mainly due to upregulated histone genes (Fig. 11C, some marked in red). The association was replicated by gene enrichment analysis, visualized with enrichment map, highlighting chromatin and DNA assembly pathways. Histones have a key function in transcriptional regulation and neuronal development. Histone genes have repeatedly been associated with neurological symptoms (Kleefstra et al., 2012; Zech et al., 2016) and ID syndromes, such as Kabuki syndrome caused by haploinsufficiency of genes responsible for chromatin opening (Lederer et al., 2012; Ng et al., 2010).

The polr2a gene was one of the top upregulated genes (Fig. 11C, green). The gene encodes the largest unit of RNA polymerase II that directly interacts with TFIIID and thus TAF1, to form the preinitiation complex. Several units of the transcription preinitiation complex have been associated with neurological phenotypes and ID (Alazami et al., 2015; Hellman-Aharony et al., 2013; Rooms et al., 2006; Tawamie et al., 2017). Apart from ID, homologous genes of the preinitiation complex have been proven important for gametogenesis (Goodrich and Tjian, 2010). For example, truncating variants in TAF4b have been associated with azoospermia (Ayhan et al., 2014) and TAF7L with spermatogenic failure (Akinloye et al., 2007). TAF1 has been shown to be colocalized with testis-specific TAFs (Metcalf and Wassarman, 2007), and BRDT orthologue (tBRD-1) in drosophila spermatocytes (Leser et al., 2012). In line with this, one of the top overexpressed genes is zebrafish brdt (Fig. 11C, purple). BRDT has been shown to be specifically expressed in human testis and associated with spermatogenic failure (MIM 617644) (Li et al., 2017).

The top overexpressed gene in our dataset was 5.8 rRNA, log2 7.9 (Fig. 11C, yellow). TAF1 is generally not described to regulate polymerase I
induced rRNA transcription. However, one previous study showed that $TAF1$ also regulates rRNA transcription by binding to upstream binding factor (Lin et al., 2002). However, to the best of our knowledge, this has not been replicated, and generally a complex of TFII B, TAF1A, TAF1B, TAF1C, TAF1D and TAF12, and not TAF1, is described to induce rRNA transcription. However, the mechanism of class-specific transcription still remains poorly understood (Engel et al., 2017). Ribosomal RNA is pointed out as a possible regulator of mRNA transcription, similarly to histone genes (Locati et al., 2017). Hypothetically, overexpression of $rRNA$ in mutant fish could be associated with altered transcription activity.

Collectively, the transcriptome analysis shows that $taf1$ knockout during embryonic development results in differential expression of 6628 genes, primarily associated with pathways important for neurodevelopment. We hypothesize that XLID in patients with $TAF1$ variants could be caused by dysregulation of several genes of chromatin and ion channel pathways. A specific gene or pathway is difficult to highlight at this point. The model is limited, as the missense variants found in affected patients most likely do not result in complete loss-of-function, and thus further studies are needed to pinpoint the specific disease-mechanism.

The zebrafish is a widely used model organism for investigation of developmental processes with similarities to humans in both brain structure and gene expression during early development (Howe et al., 2013). Apart from this, the fast generation time, the production of large clutches of transparent embryos and the 76% protein sequence similarity to human $TAF1$ emphasized zebrafish to be a suitable model. Two lines of mutant zebrafish were created, one targeting exon 8 (described above) and one targeting exon 7. By comparing the two lines we could conclude that there are no evident off-target effects as phenotypes of both strains were similar. The zebrafish model is inadequate for investigation of X-chromosome related questions, like skewed XCI, as the zebrafish does not harbor specific sex chromosomes.

In summary, six males presented with syndromic XLID, suggested to be caused by $TAF1$ c.3568C>T, p.(Arg1190Cys) hemizygosity, supported by observed skewed XCI in female carriers and in silico predictions. Further, loss of $taf1$ during embryonic development resulted in upregulation of chromatin assembly pathways and downregulation of ion channel pathways in zebrafish embryos 3 dpf. To the best of our knowledge, this is the first illumination of pathways regulated by $taf1$ during embryonic development, providing a molecular link to neurological anomalies seen in patients with $TAF1$ variants.
Concluding Remarks and Future Perspectives

If you know the question, 
you know half.

– Herb Boyer

This project has investigated revertant and gonadal mosaicism as well as two extremely rare ID syndromes with state-of-the-art sequencing technologies and molecular tools. The research has assisted delivery of diagnosis, prognosis, improved treatment and recurrence risk in five investigated families. Further, the insight from our work has improved understanding of disease etiology, basic biological mechanisms, and thus contributed to improved care of future patients, the possibility to develop new therapies and increased understanding of fundamental human developmental processes. However, continued research is needed to fully address the topics investigated in Papers I-IV.

Specifically, it would be fruitful to investigate the occurrence mechanism of somatic variants in the patient affected with KID syndrome described in Paper I. Also, the link between KID syndrome, connexin 26, and cancer is not well understood and should be thoroughly investigated in order to enable prevention of skin cancer in future patients with KID syndrome.

An observation that still puzzles us is the high frequency of mutant sperm in one of the fathers in Paper II. A 40% occurrence in sperm would require 80% heterozygosity for the disease-causing variant in spermatocytes. Such a high mutation rate indicates occurrence during early embryogenesis. However, since the variant is not detected in blood this is not a likely the explanation. As of now we are in a collaboration with Professor Anne Goriely and Professor Andrew Wilkie to investigate this mechanism, possibly linked to PAE and selfish spermatogonial selection.

CdLS4, investigated in Paper III, has only been described in 13 patients, with 8 different alterations, hampering diagnosis, prognosis, and treatment. We are currently taking part in a combined international effort in which previously reported and newly identified individuals with RAD21 variants are investigated. The unified report covers 10–15 novel RAD21 variants and patients (Kline et al., 2018), as well as previously described patients with CdLS4, which will aid diagnosis and genetic counseling of future patients.

In Paper IV we describe a family affected with a rare, recently discovered ID syndrome in which sufficient diagnosis and prognosis is lacking due to limited reported cases. Reports of additional patients with TAF1 variants causing XLID are needed as they will provide improved perspectives of the disorder, aid genetic counseling and likely increase the number of disease-causing
variants. Our work shows molecular evidence for TAF1’s association with neurodevelopmental function during early embryogenesis. However, the zebrafish knockout model does not explore the specific mechanism of the TAF1 variants associated with XLID, which should be further investigated. We believe that the time-point is crucial as the malformations likely manifest during early embryogenesis. Therefore, investigation of adult patient cells has not been highly prioritized. However, transcriptomics of an alternative model system e.g. in mice or induced pluripotent stem cells from patient-derived fibroblasts might point to a pathogenic mechanism of TAF1 variants and possibly identify future therapeutic targets.

During the course of this project we have initiated investigations of more than 12 affected families, many of which are still under examination. In these cases, initial WES has not resulted in any candidate genetic variants and thus additional WGS and further analysis is now being performed. The aspiration is that WGS might detect coding variants in regions not covered by WES or structural variants as well as non-coding intra- and intergenic variants (Lelieveld et al., 2015), with the challenge of non-coding variant interpretation yet to be overcome.

Research during the last two decades has provided us with a map of the human genome and sufficient NGS techniques that allow genetic diagnosis of 30–70% of patients affected with Mendelian disorders (Chong et al., 2015; Clark et al., 2018; Jalkh et al., 2019; Vissers et al., 2016). Yet, for many affected families a diagnosis is still missing or limited. Many discoveries remain in order to diagnose all families affected with uncharacterized Mendelian disease. Another important focus of future genetic research is to provide sufficient therapeutics for affected patients. As of today, treatment for Mendelian disorders is mostly limited to symptomatic relief and only 0.1% of disorders have approved treatments (Austin and Dawkins, 2017). However, new molecular tools such as CRISPR/Cas9 show promising results for curing Mendelian disorders by direct editing of the human genome (Zeng et al., 2018). Today, most researchers worldwide agree that human gene editing lies many discoveries ahead and that there are several technical and ethical challenges that needs to be thoroughly addressed before clinical trials can be initiated. Yet, I believe that future genetic research will master clinical applications of gene editing tools like Crispr/Cas9 and by that, enable curative therapies for future children affected with Mendelian disorders. The 21st century is the era for medical genetic research.
Introduktion till det humana genomet och orsaken till mendelsk sjukdom

I kroppens alla cellkärnor finns 23 par kromosomer som består av den DNA-kod som reglerar alla våra biologiska processer. En uppsättning av kromosomerna härstammar från mamma och en från pappa och sammanlagt kallas de 46 kromosomerna för det humana genomet. DNA översätts till RNA som sedan utgör mallen för skapandet av protein. Tillsammans styr DNA, RNA och protein alla kroppens funktioner och möjliggör en människas komplexa utveckling från en enda cell till en vuxen individ (Fig. 12).

Vårt DNA förändras ständigt genom att en eller flera molekyler byts ut, tas bort, byter plats eller vänder håll. Förändringarna kallas genetiska varianter och har oftast inte någon märkbar effekt, men ibland kan de leda till en förbättrade funktion och ibland till en försämrad funktion. En sådan förbättrad funktion är det som driver evolutionen framåt och genetiska varianter är således livsviktigt för ett en art skall utvecklas. Genetiska varianterna som resulterar i en försämrad funktion kan ibland vara så allvarliga att de ger upphov

![Diagram](image)

**Figur 12.** Schematisk sammanfattning av hur en genetisk variant kan ge upphov till sjukdom. (A) I varje cellkärna transkriberas DNA till RNA som sedan translateras till protein. I detta fall har basen A på position 13 (c.13) bytts mot ett C. Detta gör att den nya RNA-koden CCU ger upphov till den nya byggstenen Prolin (Pro), istället för Treonin (Thr) som tidigare kodades av ACU. (B) Den genetiska varianten har uppstått i spermien DNA (rosa) och ger upphov till ett protein som får en annan struktur vilket leder till sjukdom. Den maternella kopian från äggcellen (vit) är fortfarande intakt och ger upphov till ett funktionellt protein.

Avhandlingsarbetets relevans och syfte


Avhandlingsarbetet har kartlagt de bakomliggande orsakerna till ouppklärade fall av misstänkt mendelsk sjukdom med ett speciellt fokus på syndrom med utvecklingsstörning. Genom kliniska, genetiska och molekylära studier har vi undersökt sjudomsetiologin hos fem olika familjer drabbade av framförallt utvecklingsstörningssyndrom. Vi har använt de mest aktuella sekvenseringsteknikerna, som läser av DNA och RNA-koden, samt nyligen uppfunna molekylära verktyg som kan modulera genetiska varianters effekt. Arbetena har sammanställts i fyra artiklar som granskats och publicerats i internationella vetenskapliga tidskrifter.

Forskningsresultat

I arbete I har vi undersökt en kvinna som har en medfödd hudsjukdom, keratitis-ichthyosis-deafness (KID) syndrom, men som i vuxen ålder utvecklade fläckar med frisk hud. Vi fastställde att de friska fläckarna uppstod genom en
mekanism som kallas reverterande mosaicism, där genetiska varianter som uppstår i den vuxna kroppen, somatiska varianter, förhindrar sjukdomsmekanismen. KID syndrom orsakas av en genetisk variant i en av gen-kopiorna som kodar för jonkanaler. Varianten leder till en störd saltbalans i cellerna som uppstår på grund av dysfunktionella jonkanaler. De somatiska varianterna vi återfann i den friska huden visade sig hindra de dysfunktionella kanalernas transport till celllytan. Det resultera i att endast jonkanaler som bildats från den friska genkopian transporteras till celllytan, och huden fick en normaliserad funktion. Detta är den första rapporten som beskriver att reverterande mosaicism kan revetera symtomen i en patient med KID syndrom. Fynden är viktiga för bland annat framtidiga möjligheter att nyttja kroppens egna mekanismer i genterapi.

**I arbete II** presenterar vi ett nytt tillvägagångssätt för att undersöka varianter ursprung, återupprepningsrisken av sjukdom, samt förekomsten av gonadal mosaicism, alltså att ett flertal könsstall bär på den genetiska varianten som ger sjukdom. Vi undersökte två familjer med barn som drabbats av utvecklingssyndrom. Den första familjen hade använt sig av en äggdonator och fått ett barn med Treacher Collins syndrom. Vi fastställde att varianten inte var nedärvd från äggdonatorn och att den inte finns indikationer på förekomst av gonadal mosaicism hos pappan. Sannolikt var variansen ny hos barnet. Uppreppningsrisken i familjen kunde bestämmas till mindre än 0,1% och äggdonatorn kan användas vid framtida graviditeter. Den andra familjen hade drabbats av två missfall och genomgått två aborter på grund av misstänkt Noonan syndrom. Vi kunde fastställa förekomst av gonadal mosaicism hos pappan och att återupprepningsrisken i familjen bedömdes till cirka 40%. Informationen om variantens ursprung var viktigt för att möjliggöra prenataldiagnostik vid framtida graviditeter.


**I arbete IV** har vi undersökt en familj i fem generationer där sex män drabbats av ett tidigare okänd utvecklingsstörningssyndrom. Vi identifierade en variant i en gen som de senaste åren associerats till neuronala utvecklingsavvikelser och utvecklingsstörning. Vi skapade en zebrafiskmodell där vi eliminerade den aktuella genen genom att klippa sönder dem med CRISPR/Cas9-metoden. Genom att sekvensera alla proteinkodande RNA molekyler i 3 dagar
gamla zebrafiskembryon undersökte vi vilka gener som var upp- och nedreglerade. Analysen visade att det fanns ett förändrat uttryck av gener som är involverade i signalvägar som styr hjärnans utveckling. Detta är första gången man studerat hur förändringar i den aktuella genen påverkar embryonalutvecklingen och kan leda till neuronal avvikelse. Vidare kunde vi lämna en genetisk diagnos till de drabbade i familjen, samt upplysa de fem kvinnor som bar på varianten att deras risk för återupprepning av utvecklingsstörningssyndromet är 50% hos pojkfoster och att fosterdiagnostik finns tillgänglig om de önskar.

**Slutsats**

Avhandlingsarbetet har demonstrerat nya applikationer av de mest aktuella sequensmetoderna, bidragit till en ökad förståelse av sjukdomsmekanismer som reverterande och gonadal mosaic samt kartlagt två extremt ovanliga utvecklingsstörningssyndrom. Den ökade kunskapen möjliggör förbättrad diagnostik, information om sjukdomsförloppet, familjeplanering och bidrar till möjligheten att utveckla nya behandlingar. Studierna har utförts i en transnationell miljö där resultaten från forskningen kan tillämpas i klinisk rutindiagnostik. Specifikt, har studierna resulterat i att fem familjer, drabbade av fem olika syndrom har erhållit kliniska och genetiska diagnoser och/eller fördjupad förståelse för sin sjukdomsbild, prognos, upprepningsrisk och bakomliggande sjukdomsmekanismer.
I would like to thank all the families that participate in this project. Thank you for the generosity with your time and your stories. It is only because of your collaboration, interest and will to help, that we can continue the search for answers. You are my true source of inspiration.

I would also like to thank the Sävstaholm Foundation for their incredible work for people affected by intellectual disability, and for funding my work for the majority of my four years as a PhD student.

Thank you, group leader and main supervisor, Prof. Marie-Louise Bonde-son. I am forever grateful that you welcomed me to the group with – literally – open arms. Your positive spirit makes work a joy, you support has been invaluable. Thank you, co-supervisor, colleague and friend Dr. Maria Wilbe, for stimulating discussions and support. I especially enjoyed brainstorming science with you, you taught me so much. I want to thank my co-supervisor, Prof. Niklas Dahl for valuable feedback on my scientific work and for good discussions during journal club hour. A big thank you to Prof. Göran Annerén, my favorite MD and bonus-supervisor, for finding the families and then inviting me to fully be a part of all aspects of the projects. I am so glad I got to work with you throughout my PhD studies. Thank you, Dr. Sara Ekvall and Dr. Anna-Maja Molin for great collaborations, especially on the TAF1 project that was a gold nugget that stayed with all three of us throughout or combined PhD period of 15 years. Thank you Josefine Johansson, I believe we make the best team and I will miss working with you tremendously! You have been a great colleague, friend, and support, you always bring that extra sparkle to work.

Thank you, Assistant Prof. Lars Forsberg, Assistant Prof. Joakim Klar, Dr. Ammar Zaghloul, Assistant Prof. Jens Schuster and Assistant Prof. Lars Feuk for making time for my questions. My everyday PhD studies had not been the same without you. You have truly helped me in my scientific development with specific hands-on help, inspiring scientific discussions and life advice in general. Further, a big thank you to the corridor of Medical Genetics and Genomics at the Dept. of Immunology, Genetics and Pathology, it has been great fun working – and not working – with all of you.

All great science is a product of a team effort and I want to express my gratitude to all the co-authors of papers I-IV for valuable and inspiring collaborations. Especially, I would like to thank a few collaborators who took that extra time and I was privileged enough to learn from. Thank you, Dr.
Adam Ameur, Prof. Hans Törmä, assistant Prof. Ludmil Alexandrov, Dr. Beata Filipek-Górniok, Prof. Johan Ledin, Prof. Paulino Gómez-Puertas, Dr. Íñigo Marcos-Alcalde and Prof. Vera M. Kalscheuer, Prof. Gary Bader as well as his team.

I want to thank the staff at Klinisk Genetik, Rudbeck Laboratory for always being helpful during my experiments. I would also like to acknowledge the Science for Life Laboratory platforms: The National Genomics Infrastructure; Uppsala Genome Centre, the SNP&SEQ Technology Platform, the Zebrafish facility, and the National Bioinformatics Infrastructure Sweden for your great expertise and collaborative skills.

Thank you, Christina Magnusson and Helene Norlin for always being around for urgent administrative tasks and for looking after all PhD students at the department.

My colleagues, buddies, friends – some future PhDs, some already doctors. Thank you for the exchange of knowledge and your support and encouragement, Vero, Svea, Loora, Bene, Ida, Adam, Ambrin, Magdalena, Marcus, Mitra, Anja, Amanda, Malin, Zafar, Jamil and many more of you who I had the fortune to cross roads with. Also, a big thank you to all the hard-working heroes of the PhD council.

My parents, Anna-Berit Gudmundsson and Anders Johansson, thank you for both genetics & environment, for giving me a platform to grow, learn and do what I love. You have taught me about doing things for the right reasons and making a difference. You have pushed me to go further, encouraged me to aim higher, and convinced me I can do more. You are my most faithful fans and the fact that you think I can do it all, makes me consider to at least try. I am forever grateful.

Eric Gudmundsson, my husband, my best friend and greatest support. Thank you for keeping me sane these four years, accompanying me through times of doubt and celebrating my victories. Thank you for making sure I eat, sleep and look up from the computer every now and then. Thank you for taking me on lots of adventures that gave me the energy and inspiration to start again. You bring perspective to my world and make life in- and outside research worth living. This thesis is for my parents, but the rest of my life will be for you.
References


Zech, M., Boesch, S., Maier, E.M., Borggrefe, I., Vill, K., Laccone, F., Pilshofer, V.,

Zeng, Y., Li, J., Li, G., Huang, S., Yu, W., Zhang, Y., Chen, D., Chen, J., Liu, J., and

Zhu, T., Gao, Y.F., Chen, Y.X., Wang, Z.B., Yin, J.Y., Mao, X.Y., Li, X., Zhang, W.,
Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1551*

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-379363