

Aus dem
Department für Diagnostische Labormedizin
Institut für Medizinische Genetik
und Angewandte Genomik

**Molekulargenetische Analysen zur Bestimmung
der Pathogenität von Varianten in Kandidatengenen
für seltene genetische Erkrankungen**

**Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Medizin**

**der Medizinischen Fakultät
der Eberhard Karls Universität
zu Tübingen**

vorgelegt von

Ehrhardt, Anna Sophia

2026

Dekanin: Professorin Dr. S. Y. Brucker

1. Berichterstatter: Professor Dr. O. Rieß

2. Berichterstatter: Professor Dr. K. M. Niyazi

Tag der Disputation: 10.03.2026

Inhaltsverzeichnis

Inhaltsverzeichnis	III
I Abkürzungsverzeichnis	VII
II Abbildungsverzeichnis	XI
III Tabellenverzeichnis	XII
1 Einleitung	1
1.1 Seltene Erkrankungen	1
1.2 Seltene genetische Erkrankungen.....	2
1.3 Die diagnostische Odyssee und der Wunsch nach evidenzbasierter Medizin	4
1.4 <i>Rare genetic diseases</i> und <i>Next generation sequencing</i>	5
1.5 Fragestellung dieser Arbeit.....	7
1.6 Untersuchungskollektiv.....	8
2 Material	15
2.1 Primäre Fibroblastenzelllinien	15
2.2 Chemikalien, Größenmarker, Kits	15
2.2.1 Chemikalien	15
2.2.2 Größenmarker	17
2.2.3 Kits.....	17
2.3 Puffer und Lösungen	17
2.4 Primer	19
2.5 Antikörper	19
2.5.1 Primärantikörper	19
2.5.2 Sekundärantikörper	19
2.6 Geräte.....	19
2.7 Software und Datenbanken	21
2.7.1 Software.....	21
2.7.2 Datenbanken	21

2.7.3	<i>In silico</i> -Tools	22
2.8	Verbrauchsmaterialien	22
3	Methoden	24
3.1	DNA- und RNA-Sequenzierungen mittels <i>Next Generation Sequencing</i>	24
3.1.1	Vorbereitung einer NGS-Bibliothek (<i>library preparation</i>)	24
3.1.2	Vorbereitung vor dem Sequenzierlauf (<i>template preparation</i>)	26
3.1.3	Sequenzierung durch Synthese (<i>sequencing by synthesis</i>)	27
3.2	Bioinformatische Aufarbeitung der Rohdaten der Sequenzierungsreaktionen	28
3.2.1	Zuordnung zum Referenzgenom (<i>mapping</i>)	29
3.2.2	Auflistung und Annotation der Varianten (<i>variant calling und variant annotation</i>)	30
3.2.3	Kopienzahlvarianten (<i>Copy number variations</i>)	30
3.2.4	Eltern-Kind- und Familienanalysen	30
3.2.5	Bioinformatische Pipeline für die Rohdaten der RNA-Seq	31
3.3	Auswertung der Genom-, Exom- und Transkriptom-Sequenzierungen ...	31
3.3.1	Verwendete Software	31
3.3.2	Filterung der vorhandenen Varianten	32
3.3.3	Bewertung der gefilterten Varianten	36
3.3.4	X-Inaktivierungsanalyse	38
3.4	Aufklärung unklarer Befunde aus den Transkriptom-Daten mittels PCR aus cDNA	38
3.4.1	cDNA-Synthese	39
3.4.2	Design der Primersequenzen	40
3.4.3	Polymerasekettenreaktion	42
3.4.4	Agarosegelelektrophorese und Detektion der PCR-Produkte	43
3.5	Zellkultur	44
3.5.1	Kultivierung primärer Fibroblastenzelllinien	44
3.5.2	Passagieren von adhärennten Zelllinien	45
3.5.3	Herstellung und Auftauen kryokonservierter Fibroblasten	45
3.5.4	Ernte von Zellen	46

3.6	Qualitativer und semiquantitativer Nachweis der Kandidatengenexpression auf Proteinebene.....	47
3.6.1	Proteinextraktion aus Zellpellets.....	47
3.6.2	Bradford-Test.....	48
3.6.3	SDS-Polyacrylamidgelelektrophorese	50
3.6.4	Western Blot	52
3.6.5	Immundetektion	54
3.6.6	Stripping und Reprobing	56
3.6.7	Messung der Signalintensitäten der detektierten Banden	57
3.6.8	Statistische Auswertung der Signalintensitäten.....	58
4	Ergebnisse	59
4.1	DNA- und RNA-Sequenzierungen.....	59
4.2	Identifizierung einer Mutation in <i>DBR1</i> bei Individuum MS002	66
4.2.1	MS002: Phänotyp	66
4.2.2	Auswertung der DNA- und RNA-Sequenzierungen.....	68
4.2.3	Qualitative und semiquantitative Analyse der Proteinexpression mittels Western Blot.....	71
4.3	Identifizierung einer Mutation in <i>SEC31B</i> bei Individuum MS001	74
4.3.1	MS001: Phänotyp	74
4.3.2	Auswertung der Exom-Sequenzierung in Kombination mit den RNA-Seq-Daten.....	74
4.3.3	Qualitative Analyse der Proteinexpression mittels Western Blot.....	79
4.4	Identifizierung einer Mutation in <i>OSCP1</i> bei Familie 01	83
4.4.1	Familie 01: Phänotyp	83
4.4.2	Auswertung der Genom-Sequenzierungen in Kombination mit den RNA-Seq-Daten.....	84
4.4.3	Aufklärung des unklaren Befundes aus der RNA-Seq mittels PCR aus cDNA.....	91
4.5	Identifizierung einer Mutation in <i>HUWE1</i> bei Individuum MS005.....	92
4.5.1	MS005: Phänotyp	92
4.5.2	Auswertung der Trio-Exom-Sequenzierung in Kombination mit den RNA-Seq-Daten.....	92
4.5.3	X-Inaktivierungsanalyse.....	96

5	Diskussion	101
5.1	Diskussion der Beispiele aus dem Ergebnisteil	102
5.1.1	Identifizierung einer Mutation in <i>DBR1</i> bei Individuum MS002.....	102
5.1.2	Identifizierung einer Mutation in <i>SEC31B</i> bei Individuum MS001.....	107
5.1.3	Identifizierung einer Mutation in <i>OSCP1</i> bei Familie 01	111
5.1.4	Identifizierung einer Mutation in <i>HUWE1</i> bei Individuum MS005	113
5.2	Gibt es eine ideale Pipeline für ungelöste Fälle?.....	115
5.2.1	Phänotypisierung	115
5.2.2	Wahl des molekulargenetischen Verfahren zur Detektion genomischer Varianten	117
5.2.3	RNA-Seq.....	120
5.2.4	Multiomics Approach.....	121
5.2.5	Reanalyse der Daten	122
5.3	Schlussfolgerung und Ausblick.....	123
6	Zusammenfassung	126
7	Literaturverzeichnis	128
	Erklärungen zum Eigenanteil.....	141
	Danksagung	142
	Anhang.....	143
	Qualitätsparameter der DNA-Sequenzierungen	143
	Qualitätsparameter RNA-Sequenzierungen	144

I Abkürzungsverzeichnis

*	Stopcodon
1000g	<i>1000 Genomes Project</i>
A (Aminosäure)	Alanin
A (DNA oder RNA)	Adenin
aa	Aminosäuren
<i>ABCA4</i>	<i>ATP binding cassette subfamily A member 4</i>
<i>ACMG</i>	<i>American College of Medical Genetics and Genomics</i>
<i>ANEC</i>	<i>acute necrotizing encephalopathy of childhood</i>
<i>Array-CGH</i>	<i>array-based comparative genomic hybridization</i>
ASD	Autismus Spektrum Störungen
bp	<i>base pairs (Basenpaare)</i>
BSA	<i>bovine serum albumin</i>
C	Cytosin
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
<i>CCDC40</i>	<i>coiled-coil domain 40 molecular ruler complex subunit</i>
CD4+	Zellen, welche das Oberflächenprotein CD4 exprimieren
CD8+	Zellen, welche das Oberflächenprotein CD8 exprimieren
cDNA	<i>complementary DNA</i>
CNV(s)	<i>copy number variation(s)</i>
comp.-het.	compound-heterozygot
<i>COPII</i>	<i>coat protein complex II</i>
CPT-Blut	Blut, welche mit BD Vacutainer CPT™ System abgenommen wird
<i>DBR1</i>	<i>RNA lariat debranching enzyme</i>
DD	Differentialdiagnose
DMSO	Dimethylsulfoxid
DNA	<i>deoxyribonucleic acid</i>
<i>DNAH11</i>	<i>dynein axonemal heavy chain 11</i>
<i>DNAH5</i>	<i>dynein axonemal heavy chain 5</i>
DPBS	<i>dulbecco's phosphate buffered saline</i>

E	Glutaminsäure
E. coli	Escherichia coli
E(Nummer)	Exon
EDTA	Ethylendiamintetraacetat
<i>EIF4EBP2</i>	<i>eukaryotic translation initiation factor 4E binding protein 2</i>
EM	Erstmanifestation
ES	<i>exome sequencing</i>
EU	Europäische Union
<i>FASTKD3</i>	<i>FAST kinase domain-containing protein 3, mitochondrial</i>
FBS	<i>fetal bovine serum</i>
<i>FBXO11</i>	<i>F-Box Protein 11</i>
G	Guanin
<i>GANC</i>	<i>neutral alpha-glucosidase C</i>
GAPDH	Glycerinaldehyd-3-phosphat-Dehydrogenase
gDNA	genomische DNA
gnomAD	<i>Genome Aggregation Database</i>
GRCh37/hg19	<i>Homo sapiens genome assembly GRCh37 (hg19)</i> vom Genome Reference Consortium
HEK-Zellen	Zellen der humanen embryonalen Niere (<i>human embryonic kidney</i>)
HSV1	Herpes-simplex-Virus-1
<i>HUWE1</i>	<i>HECT, UBA and WWE domain containing E3 ubiquitin protein ligase 1</i>
I	Isoleucin
IBV	Influenza-B
ID	<i>intellectual disability</i>
IFN- α/β	Interferon α/β
IGV	<i>Integrative Genomics Viewer</i>
Indels	Insertionen und Deletionen
IRDIRC	<i>The International Rare Diseases Research Consortium</i>
L	Leucin
LDS	Lithiumdodecylsulfat

LoF	<i>loss of function</i>
LOH	<i>loss of heterozygosity</i>
MAF	<i>minor allele frequency</i>
mRNA	<i>messenger RNA</i>
MRT	Magnetresonanztomographie
<i>MUC5B</i>	<i>mucin 5B, oligomeric mucus/gel-forming</i>
N	Asparagin
NGS	<i>next generation sequencing</i>
NMD	<i>nonsense mediated decay</i>
NOC	<i>neuroocular syndrome</i>
<i>NOR1</i>	<i>oxidored-nitro domain-containing protein 1</i>
<i>NSUN4</i>	<i>NOP2/Sun RNA methyltransferase 4</i>
<i>NTNG1</i>	<i>netrin-G1</i>
o/e-Score	<i>observed/expected-Score</i>
OMIM	<i>Online mendelian inheritance in Man</i>
<i>OSCP1</i>	<i>organic solute carrier protein 1</i>
P	Perzentile
P (Aminosäure)	Prolin
<i>PAICS</i>	<i>phosphoribosylaminoimidazol carboxylase</i>
PCD	<i>primary ciliary dyskinesia</i>
PCR	<i>polymerase chain reaction</i>
pLI	<i>probability of being loss-of-function intolerant</i>
pLOF	<i>predicted Loss-of-Function</i>
<i>PRR12</i>	<i>proline rich 12</i>
Q	Glutamin
R	Arginin
RCT(s)	<i>randomized controlled trial(s)</i>
RGD(s)	<i>rare genetic disease(s)</i>
RNA	<i>ribonucleic acid</i>
RNA-seq	Sequenzierung der mRNA
RNasen	Ribonukleasen
RPE-Zellen	Zellen des Retinalen Pigmentepithels
RSV	Respiratorisches Synzytial-Virus

RT	Reverse Transkription
RT-qPCR	<i>reverse transcription quantitative polymerase chain reaction</i>
RTT	Rett-Syndrom
S	Serin
SARS-CoV2	<i>severe acute respiratory syndrome coronavirus 2</i>
<i>SDHB</i>	<i>succinate dehydrogenase complex iron sulfur subunit B</i>
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamid-Gel-Elektrophorese
<i>SEC23</i>	<i>SEC23, COPII coat complex component</i>
<i>SEC24A</i>	<i>SEC24 homolog A, COPII coat complex component</i>
<i>SEC31A</i>	<i>SEC31 homolog A, COPII coat complex component</i>
<i>SEC31B</i>	<i>SEC31 homolog B, COPII coat complex component</i>
SNP(s)	<i>single nucleotide polymorphism(s)</i>
SOLVE-RD	<i>Solve-rare disease Consortium</i>
<i>SQRDL</i>	<i>sulfide quinone oxidoreductase</i>
T (Aminosäure)	Threonin
T (DNA oder RNA)	Tyrosin
TLR3	<i>toll like receptor 3</i>
U	Uracil
US/USA	<i>United States (of America)</i>
UTR(s)	<i>untranslated region(s)</i>
UV-Licht	ultraviolettes Licht
V (Aminosäure)	Valin
VUS	<i>variant(s) of unknown significance</i>
W	Tryptophan
WES	<i>Whole Exome Sequencing</i>
WGS	<i>Whole Genome Sequencing</i>
WT	Wildtyp
XLID	<i>X-linked intellectual disability</i>
Y	Tyrosin

Für die Nomenklatur zu den Angaben der Varianten siehe *HGVS Nomenclature* (den Dunnen et al., 2016).

II Abbildungsverzeichnis

Abbildung 3.1: Aufbau einer Blotting-Kassette.....	53
Abbildung 4.1: Exemplarisches MRT im Axialschnitt (T2-gewichtet) von MS002 auf Höhe der Pons/Abgang VIII. Hirnnerv.....	67
Abbildung 4.2: Western Blots zur semiquantitativen Bestimmung von DBR1 in MS002 und Kontrollen.....	72
Abbildung 4.3: Boxplots der normalisierten Signalintensitäten für DBR1	73
Abbildung 4.4: Western Blots zum Nachweis von SEC31B mit Proteinlysaten aus Fibroblasten unter Verwendung von zwei unterschiedlichen Primärantikörpern	81
Abbildung 4.5: Western Blots zum Nachweis von SEC31B mit Proteinlysaten aus Fibroblasten (MS001 und Kontrollen 1-3) sowie aus HEK-Zellen, CPT-Blut und RPE-Zellen	82
Abbildung 4.6: Stammbaum der Familie 01	83
Abbildung 4.7: Spleißdefekt in <i>OSCP1</i> bei Familie 01	88
Abbildung 4.8: Schematische Darstellung des Spleißdefekts bei Familie 01 im Vergleich zum Wildtyp.....	89
Abbildung 4.9: Sashimi-Plot für <i>HUWE1</i> bei Individuum MS005	95

III Tabellenverzeichnis

Tabelle 1.1: Zusammenfassung der Phänotypen für Individuen mit Organerkrankungen u/o Metabolischen Erkrankungen	10
Tabelle 1.2: Zusammenfassung der Phänotypen für Individuen mit Neurologischen Erkrankungen.....	11
Tabelle 3.1: <i>MegSAP pipeline</i> : Verwendete Tools und Versionen zum Zeitpunkt der Datenanalyse	29
Tabelle 3.2: Filterungsschritte für kleine Varianten (Original-Wortlaut aus dem Programm GSVar)	34
Tabelle 3.3: Filterungsschritte für Eltern-Kind- und Familienanalysen (Original-Wortlaut aus dem Programm GSVar)	35
Tabelle 3.4: Ansatz für die Elimination genomischer DNA	39
Tabelle 3.5: Ansatz für die Retrotranskription.....	40
Tabelle 3.6: Überblick über die wichtigsten Auswahlkriterien beim Primerdesign.....	41
Tabelle 3.7: PCR-Ansatz	42
Tabelle 3.8: Programmierung des Thermocyclers G-Storm GS1 für die Touchdown-PCR.....	43
Tabelle 3.9: Pipettierschema der BSA-Standardreihe für den Bradford-Test	48
Tabelle 3.10: Pipettierschema der Probenansätze für den Bradford-Test	49
Tabelle 3.11: Ansätze für eine 8 % oder 10 %ige Trenngellösung	51
Tabelle 3.12: Ansatz für eine 6 %ige Sammelgellösung.....	51
Tabelle 3.13: Beispielansatz eines Primärantikörper-Gemisches.....	55
Tabelle 3.14: Ansatz für die Lösung mit Sekundärantikörpern	56
Tabelle 3.15: Detektionswellenlängen und Belichtungszeiten	56
Tabelle 4.1: Anzahl der verbleibenden Varianten nach den beschriebenen Filterungsschritten in Proben der einzelnen Individuen....	61
Tabelle 4.2: Anzahl der verbleibenden Varianten nach den beschriebenen Filterungsschritten für die durchgeführten Multi-Analysen	62
Tabelle 4.3: Anzahl der verbleibenden Varianten nach den beschriebenen Filterungsschritten für die durchgeführten Trio-Analysen.....	63
Tabelle 4.4: Übersicht über Kandidatengene, Auswahlgründe und Erkenntnisse aus RNA-Seq-Daten	65

Tabelle 4.5: Potenzielle Kandidatengene für Individuum MS002.....	68
Tabelle 4.6: Potenzielle Kandidatengene von Individuum MS001	75
Tabelle 4.7: Interessante Varianten für Familie 01.....	85
Tabelle 4.8: Potenzielle Kandidatengene für MS005	93
Tabelle 4.9: X-Inaktivierungsanalyse bei Individuum MS005 (für Varianten mit coverage >50 in der RNA-Seq)	98

1 Einleitung

1.1 Seltene Erkrankungen

Eine seltene Erkrankung ist eine Erkrankung, welche verglichen mit häufigen Erkrankungen in der Gesellschaft lediglich eine kleine Gruppe von Menschen betrifft. In der Europäischen Union (EU) gilt eine Prävalenz unter 1 : 2.000 als selten (Europäische Union, 1999). Diese Definition ist jedoch keineswegs allgemeingültig, sondern unterscheidet sich weltweit je nach zugrundeliegender rechtlicher Grundlage. So ist bspw. in den USA eine seltene Erkrankung mit einer absoluten Anzahl von unter 200.000 Betroffenen beziffert (*Anmerkung:* bezogen auf den US Amerikanischen Staat; entspricht einer Prävalenz von ~ 1 : 1.630; United States Congress, 2002, Richter et al., 2015).

Einige Krankheiten zeigen zudem eine Häufung in bestimmten Populationen oder geographischen Regionen, sodass die Begriffsdefinition „selten“ in den meisten Fällen kontextabhängig ist (Haendel et al., 2020). Ein Paradebeispiel hierfür stellt die Sichelzellanämie dar – ~ 75 % der Neugeborenen weltweit mit Sichelzellanämie kommen in Subsahara-Afrika zur Welt und sorgen dort für eine extreme Häufung, während die Erkrankung in Deutschland selten ist (< 25 / 100000 Geburten; Kato et al., 2018).

Eine international gleichermaßen anerkannte Definition gibt es derzeit zu der Gruppe der seltenen Erkrankungen nicht, tatsächlich konnten Richter et al., 2015 in ihrem Review 296 Definitionen von 1109 Organisationen aus über 32 Ländern identifizieren. Die Cut-offs der Prävalenzen schwanken international zwischen 5 und 76 / 100.000 Einwohner (bezogen auf die rechtsgültigen Verordnungen der Länder). Hinzu kommen tlw. noch andere zusätzliche Kriterien, welche für die Anerkennung als seltene Erkrankung erfüllt sein müssen (z. B. Krankheitsschwere wie bspw Lebensbedrohlichkeit).

Trotz dieser hohen Diskrepanz herrscht jedoch Einigkeit darüber, dass Seltene Erkrankungen zwar individuell rar sind, die Gesamtheit jedoch eine fundamentale sozioökonomische Belastung für die Weltbevölkerung darstellt (Ferreira, 2019). Neben dem individuellen Leid für die einzelnen Betroffenen, belasten seltene Erkrankungen die Gesundheitssysteme jährlich mit enormen Kosten. Studien gehen davon aus, dass jedes Kind mit einer angeborenen genetischen Erkrankung Kosten in Höhe von 5.000.000 US-Dollar während seines Lebens verursacht (Chong et al., 2015).

Laut Bundesgesundheitsministerium leben schätzungsweise vier Millionen Menschen in Deutschland mit einer Seltenen Erkrankung (Bundesministerium für Gesundheit, 2023). In der EU geht man von ca. 30 Millionen Betroffenen aus. Diese Zahlen decken sich mit einer Studie von Nguengang Wakap et al., welche anhand einer Analyse der Datenbank Orphanet die kumulative Punktprävalenz von Seltenen Erkrankungen untersucht hat (Nguengang Wakap et al., 2020). Die KollegInnen kalkulierten eine Punktprävalenz von 3.482,3 – 5.910,3 pro 100 000 (~3,5 – 5,9 %) in der Globalbevölkerung – somit sind nach dieser Schätzung 17,8 – 30,3 Millionen Europäer und 262,9 – 446,2 Millionen Menschen weltweit von einer Seltenen Erkrankung betroffen.

1.2 Seltene genetische Erkrankungen

Seltene *genetische* Erkrankungen (*rare genetic diseases*, RGDs) stellen eine bedeutsame Untergruppe der seltenen Erkrankungen dar. Genetische Defekte sind für ca. 70 % der seltenen Erkrankungen verantwortlich und machen somit den Großteil der seltenen Erkrankungen aus (Nguengang Wakap et al., 2020). Insgesamt werden jährlich ca. 8 Mio. Kinder weltweit geboren, welche an einer ernstesten genetischen Erkrankung leiden (Chong et al., 2015).

Schätzungen der exakten Häufigkeiten von seltenen Erkrankungen durch die aktuelle Fachliteratur geben absolute Zahlen zwischen 5.000 – 10.000 heraus (Haendel et al., 2020). In diesen Bezifferungen sind RGDs inkludiert,

jedoch ist die absolute Anzahl bislang unklar und befindet sich durch den stetigen Fortschritt der Wissenschaft permanent im Wandel.

Für eine genaue Betrachtung dieser Dynamik lohnt sich ein Blick in zwei der bedeutsamsten Datenbanken im Bereich der seltenen Erkrankungen – Orphanet und OMIM.

Orphanet, erstmalig im Jahr 1997 durch das *Institut National de la Santé et de la Recherche Médicale* und dem französischen Gesundheitsministerium veröffentlicht, katalogisiert seltene Erkrankungen und klassifiziert jeden klinisch einzigartigen Eintrag in einem multi-hierarchischen System mit eindeutigen Kennziffern (sog. Orpha-Nummern; Rath et al., 2012). Orphanet ist die bis dato umfangreichste Datenbank für seltene Erkrankungen und bietet sowohl ein Inventar als auch eine freizugängliche Online-Enzyklopädie.

Die Datenbank *Online Mendelian Inheritance in Man (OMIM)* hingegen stellt seit ihrer Erstveröffentlichung 1966 (damals als Buchversion *Mendelian Inheritance in Man* von Dr. V. McKusick publiziert) das Standard-Kompendium für genetische Erkrankungen aller Art dar und erfasst somit auch alle seltenen genetischen Erkrankungen (Amberger et al., 2015). Die Datenbank behandelt die aktuelle Fachliteratur und katalogisiert Geno- und Phänotypen jeweils in separaten Einträgen mit einzigartiger Identifikationsnummer (MIM-Nummer). Dies ermöglicht die Zuordnung mehrerer Genotypen zu einem Phänotyp (und *vice versa*), um phänotypische Vielfalt und genetische Heterogenität aufzuschlüsseln.

Jährlich kommen ca. 250 – 300 Einträge in den beiden Datenbanken hinzu, welche der Gruppe der RGDs zugeordnet werden können (Chong et al., 2015, Boycott et al., 2017). Die Zahlen unterscheiden sich zwischen Orphanet und OMIM aufgrund der unterschiedlichen Struktur der Datenhierarchie geringfügig (Boycott et al., 2017).

Obwohl diese Daten verdeutlichen, dass das Verständnis der Erbkrankheiten v. a. im letzten Jahrzehnt enorm zugenommen hat und weiter stetig wächst, ist die Ursache eines Großteils der vermuteten genetischen Erkrankungen weiterhin unklar. Der weite Weg, welchen die

Wissenschaft noch zu gehen hat, um dem Anspruch alle seltenen genetischen Erkrankungen aufzuklären, gerecht zu werden, wird umso deutlicher, wenn man sich die absoluten Zahlen der bisher noch nicht kausal einer RGD zugeordneten Gene anschaut. Chong et al. beschreiben in ihrem Review eine Gesamtzahl von 16.290 potenziellen Kandidatengenen für mendelsche Phänotypen, von denen 5.960 bereits in *Loss of function*-Studien im Mausmodell als letal beschrieben worden sind, jedoch bisher noch keine menschliche Phänotypen beobachtet worden sind (Chong et al., 2015). 3.341 Phänotypen hingegen sind in OMIM im Jahr 2019 beschrieben, bei denen noch keine molekulargenetische Grundlage herausgefunden werden konnte (Ferreira, 2019).

1.3 Die diagnostische Odyssee und der Wunsch nach evidenzbasierter Medizin

Das nach wie vor insuffiziente Verständnis der RGDs verdeutlicht sich in der sogenannten *diagnostischen Odyssee*, welche die PatientInnen häufig durchlaufen müssen. So ist es keine Seltenheit, dass PatientInnen zum Teil mehrere Jahre (in Europa im Durchschnitt fast fünf Jahre) auf die abschließende genetische Diagnose warten und bis zu acht MedizinerInnen konsultieren, bis ihr Fall aufgeklärt werden kann (Faye et al., 2024). 75 % der Betroffenen geben zudem an, dass eine primäre Diagnosestellung inkorrekt war (Faye et al., 2024).

Trotz der mittlerweile routinemäßigen Anwendung von *Whole Exome Sequencing* (WES) im klinischen Expertensetting können ~ 50 % der Fälle mit vermuteten genetischen Erkrankungen im Kontext einer *normalen* genetischen Sprechstunde weiterhin nicht genetisch diagnostiziert werden (Zurek et al., 2021).

Problematisch an dieser niedrigen Diagnoserate in Verbindung mit der beschriebenen Odyssee bis eine Diagnose letztendlich feststeht, ist die signifikant schlechtere gesundheitliche Versorgung der Betroffenen (Kole und Faurisson, 2009). Lange Zeit wurden RGDs aufgrund ihrer individuellen Seltenheit zudem in der Forschung vernachlässigt. Vor diesem Kontext forderten mehrere Forschungsgesellschaften und

PatientInnenorganisationen im letzten Jahrzehnt zunehmend, die Aufklärung der zugrundeliegenden Mechanismen von seltenen Erkrankungen voranzutreiben, um Präventionsmechanismen, diagnostische Algorithmen und Therapien zu entwickeln (Kole und Faurisson, 2009, Boycott et al., 2017). Übergeordnetes Ziel stellt die bessere medizinische Versorgung der PatientInnen durch evidenzbasierte medizinische Algorithmen, die Entwicklung von individualisierten Therapien und damit die Reduktion des Leides der einzelnen Betroffenen und deren Familien, sowie der sozioökonomischen Belastung dar. Die RGDs sind durch diese Forderungen zunehmend in das wissenschaftliche Interesse gerückt und Studien, welche die Diagnoserate verbessern, stellen den Grundstein für diese ambitionierten Ziele dar und sind daher dringend gefragt.

1.4 *Rare genetic diseases* und *Next generation sequencing*

Die Technologie des *Next generation sequencing* (NGS) hat seit seiner Einführung die Aufklärungsrate im Vergleich zu *konventioneller* genetischer Testung revolutioniert und den „Werkzeugkasten“ des klinischen Genetikers zur Diagnosestellung weiter verfeinert. Derzeit kann mit Hilfe von Exom- oder Genomsequenzierungen eine Aufklärungsrate von bis zu 52 % im Bereich der Fälle mit vermuteter RGD erreicht werden (Zhong et al., 2021). Bei ungeklärten Fällen kann diese Aufklärungsrate jedoch auch deutlich geringer ausfallen (Boycott et al., 2017).

Ursachen für diese niedrigen Diagnoseraten scheinen zum einen typische Schwierigkeiten in der Etablierung und Auswertung von methodisch komplexen Pipelines zu sein. Hier liegen von der Beschreibung des Phänotyps, über die Wahl der NGS-Methodik bis hin zur Validierung, ob ein Kandidatengen tatsächlich krankheitsursächlich ist, viele Tücken in der molekulargenetischen Diagnosestellung. Zum anderen sorgt die Komplexität des menschlichen Organismus dafür, dass die Auswirkungen von genomischen Varianten häufig noch nicht vollständig verstanden werden und daher die Detektion von Varianten mit NGS-Plattformen sehr

gut gelingt, deren Interpretation jedoch auf alleiniger Basis genomischer Daten limitiert ist (Lohmann und Klein, 2014).

Vor diesem Hintergrund steht die Forschungscommunity vor der Aufgabe neue Ansätze zu entwickeln, um die große Masse an *variants of unknown significance* (VUS), welche zwangsläufig aus groß angelegten genomischen Sequenzierstudien hervorgehen, zu bewerten.

Die Sequenzierung der Gesamtheit der mRNA (das sog. Transkriptom) stellt eine solche Möglichkeit dar, um Varianten direkt auf Expressions- oder Prozessierungsunterschiede der prä-mRNA zu überprüfen. Dies ist besonders interessant, da bei einem Großteil (ggf. bis zu 60 %) der Mutationen Spleißeffekte als ursächlich für deren Pathogenität vermutet werden (Wang und Cooper, 2007, Shiraishi et al., 2022). Außerdem verbessert sich auch die Interpretation von seltenen intronischen Varianten, welche mittlerweile in bis zu 30 % der Fälle als krankheitsauslösend vermutet werden und die größten Schwierigkeiten in der Interpretation bieten (Gonorazky et al., 2019).

In dieser Arbeit werden 16 WGS/WES-negative Fälle mit Verdacht auf eine RGD untersucht und mittels der Ebene des Transkriptoms (RNA-seq) ergänzt, um diese molekulargenetischen Methoden in Bezug auf die Kandidatengensuche zu evaluieren. Die RNA-seq wurde auf Basis der genomischen Varianten qualitativ ausgewertet. Für vielversprechende Kandidatengene wurden weitere klassische Laboruntersuchungen (v.a. Western Blots) angeschlossen, um zusätzliche Evidenz für die Kandidatengene zusammenzutragen.

Die genomischen Sequenzierungen ergaben im Durchschnitt 535 Varianten (Range 321 – 879) nach einer frequenzbasierten Filterung ($MAF < 0,1\%$ in der gnomAD-Datenbank), welche weiteren Filterungsschritten und einer manuellen Bewertung unterzogen wurden (zu ausführlicher Methodik s. Abschnitt 3).

In zehn Fällen ergab sich nach ausführlicher Auswertung des WGS oder WES bereits ein Verdachtsmoment für ein Kandidatengen. Bei diesen PatientInnen konnte die RNA-Seq in drei Fällen eine präferentielle Expression einer Missense-Variante und in einem Fall eine derangierte X-Inaktivierung, sowie eine herabgesetzte Expression des entsprechenden Genes aufdecken und war somit in 40 % der Fälle mit bereits vorhandenem Kandidatengen hilfreich.

In zusätzlich drei Fällen (18 % der eingeschlossenen Individuen) wurde durch die RNA-Seq bei Nachweis von Spleißdefekten ein neues Kandidatengen postuliert.

Insgesamt kann der Analyse des Transkriptoms demnach ein Zusatznutzen von ca. 44 % in der hier vorgestellten Kohorte für die Festlegung des Kandidatengens zugeschrieben werden.

Eine Variante konnte durch Durchführung der RNA-Seq als wahrscheinlich pathogen (Klasse 4 nach ACMG) eingeordnet werden. Anhand von vier Kandidatengenen (entsprechend fünf Individuen) werden die Vor- und Nachteile der verwendeten Methodik und die möglichen Gründe für diese weiterhin hohe Detektion an VUS ausführlich diskutiert. Außerdem sollen Ausblicke gegeben werden, welche Schritte in der Auflösung der hier vorgestellten Fälle und allgemein von ungelösten Fällen im Bereich der RGDs in den nächsten Jahren als entscheidend angesehen werden müssen.

1.5 Fragestellung dieser Arbeit

Genaue epidemiologischen Daten zu etablieren, sowie die Grundlagen und Therapiemöglichkeiten der seltenen Erkrankungen zu erforschen, steht in vielen Ländern, wie in den vorherigen Absätzen beschrieben, im Fokus der wissenschaftlichen Bemühungen.

Die vorliegende Arbeit beschäftigt sich mit der Frage des diagnostischen Nutzens von neuen molekulargenetischen Methoden (insbesondere der Analyse der Gesamtheit der mRNA) in Kombination mit NGS-Daten (*Whole Exome Sequencing* und *Whole Genome Sequencing*) von Individuen mit

vermuteten seltenen genetischen Erkrankungen und bisher negativer Diagnostik (WES- oder WGS-negative Fälle). Durch die erneute qualitative Auswertung der NGS-Daten in Kombination mit der zusätzlichen Ebene der mRNA-Sequenzierung sollten Kandidatengene für diese bisher ungelösten Fälle etabliert werden und anschließend weiterführende Untersuchungen (Western Blots und Validierungen durch „klassische“ molekulargenetische Verfahren) im hauseigenen Forschungslabor angeschlossen werden. Bestenfalls führt diese Sequenz zu einer Validierung (Klassifizierung als *likely pathogenic* oder *pathogenic* nach ACMG-Kriterien) oder einem Ausschluss des vorliegenden Kandidatengens als Krankheitsentität. Im Anschluss sollen die Vor- und Nachteile der vorgestellten diagnostischen Schritte herausgearbeitet und diskutiert werden. Hierdurch soll die Frage beantwortet werden, ob und wenn ja welche Schritte einer idealen Pipeline in diesem Bereich der genetischen Forschung angehören sollten.

1.6 Untersuchungskollektiv

Im Rahmen dieser Arbeit wurde ein Kollektiv von insgesamt 16 Individuen auf das Vorliegen einer genetischen Erkrankung untersucht. In dem Kollektiv befanden sich neben sieben Individuen ohne Hinweise auf weitere betroffene Familienmitglieder auch insgesamt fünf Familien mit bis zu drei betroffenen Geschwistern (nur tlw. in dieser Studie eingeschlossen). Bei einigen Familien bestand eine Konsanguinität (s. Tabelle 1.1 und Tabelle 1.2) der Eltern der Betroffenen.

Im Vorfeld wurden alle betroffenen Individuen von einem/r ÄrztIn am Institut für Medizinische Genetik und Angewandte Genomik Tübingen oder an einem Partnerinstitut klinisch untersucht. Außerdem wurde eine ausführliche Anamnese erhoben und Befunde von behandelnden ÄrztInnen gesichtet und zusammengetragen (u. a. auch Befunde von auswärtigen humangenetischen Laboren).

Bei allen rekrutierten Betroffenen stand zum Zeitpunkt dieser Arbeit eine definitive genetische Diagnose aus – der Verdacht einer genetischen Erkrankung sollte mit molekulargenetischen Methoden erhärtet bzw. im

besten Fall sogar gelöst werden. Aus diesem Grund ist das phänotypische Spektrum der eingeschlossenen Individuen weitreichend. Ein Großteil des Untersuchungskollektivs wurde mit neurologischen Erkrankungen vorstellig, allerdings wurden auch PatientInnen mit Lungenerkrankungen, Lebererkrankungen und endokrinologischem Erkrankungsspektrum untersucht. Die Betroffenen zeigten sowohl syndromale als auch nicht-syndromale Formen der vermuteten genetischen Erkrankungen. Ein Teil der PatientInnen verstarb bereits im Vorfeld dieser Arbeit.

Tabelle 1.1 und Tabelle 1.2 geben einen Überblick über die Phänotypen und die zugeordneten HPO-Terms (Köhler et al., 2021) der 16 eingeschlossenen Individuen.

Tabelle 1.1: Zusammenfassung der Phänotypen für Individuen mit Organerkrankungen u/o Metabolischen Erkrankungen

ID	K	SS/Geburt	Maße	Organe	Labor/Histologie	NB	HPO-Terms
MS006 Fam 01	♂ †	16 J ✓ -	mit 14 J: 158 cm (11 P) 66 kg (79 P)	Interstitielle Lungenerkrankung fibrosierend, DD: Alveolitis (EM 13J)	Bürstenabstrich: negativ Schweißtest auf CF: negativ	Juv. MD (EM: 8J)	HP:0002206 HP:0006528 HP:0006552
MS007 Fam 01	♂	19 J ✓ -	mit 17 J: 184,4 cm (71 P) 82 kg (86 P)	Interstitielle Lungenerkrankung fibrosierend, DD: Alveolitis (EM 15J)	Bürstenabstrich negativ Schweißtest auf CF: negativ	-	HP:0002206 HP:0006528 HP:0006552
MS010 Fam 02	♂ †	4 T - IUGR Oligohydramnion Sectio: 34+1 ppm	nach Geburt: 43,5 cm (<1 P) 1,7 kg (<1 P) KU: 29 cm (<1 P)	Leberinsuffizienz Bilaterale Nieren- hypoplasie Anurie	Anämie (8,4 g/dl) Thrombopenie (64T/μl) Laktat 3-4 mmol/l Aminosäuren ↑ (v. a. Glycin) Atmungskette relativ ↓ Leber: Eisen ↑ intra- und extrahepatisch	-	HP:0006583 HP:0012584 HP:0001399 HP:0001396
MS011 Fam 02	♂ †	15 T - IUGR Oligohydramnion Sectio: 35+3 ppm	nach Geburt: 47 cm (3 P) 2 kg (<1 P) KU: 32,5 cm (1 P)	Leberinsuffizienz Bilaterale Nieren- hypoplasie Anurie ARDS Hypoglykämien	Anämie (7,3 g/dl) Thrombopenie (65T/μl) Hypoalbuminämie Aminosäuren ↑ (v. a. Tyrosin Phenylalanin, Methionin, Citrullin, Glycin, Lysin) Atmungskette: Komplex I ↓ Leber: Eisen ↑ intra- und extrahepatisch	-	HP:0006583 HP:0012584 HP:0001399 HP:0001396
MS013	♂	15 J - Bei Geburt 61 cm (>99 P)	Kongenitaler Riesnwuchs >2,40 m (>99 P) Makrozephalie	Hypoglykämien	keine messbaren Wachstumshormone, kein Insulin	-	HP:0001943 HP:0000256 HP:0009125 HP:0000098 HP:0040216 HP:0032367

Tabelle 1.2: Zusammenfassung der Phänotypen für Individuen mit Neurologischen Erkrankungen

ID	K Maße Dysmorphien	Anamnese/ Klinische Untersuchung	Epilepsie	Augen	Bildgebung	Labor/Histo	HPO-Terms
MS001	♀ 59 J - 168 cm Makrozephalie (63 cm) große Hände, große Füße (EU 43)	Leichte ID (Hauptschule, keine Berufsausbildung) Ruhetremor linksbetont und Verlangsamung i.S. Parkinsonismus	-	kongenitaler Nystagmus	-	-	HP:0000256 HP:0001300 HP:0001249 HP:0006934
MS002	♀ 4 M † ✓ unauffällig bei Geburt krauses Haar	Koma n. Infektion mit RSV (Respiratorisches Synzytial-Virus)	Generalisiert	Glaukom	ANEC (s. Abschnitt 4.2.1)	Mitochondrio- pathie	HP:0007146 HP:0006965 HP:0000501
MS003	♂ 6 J - progr. Mikrozephalie mit 3J + 9M: 100 cm (30P) 16 kg (46P) KU: 43 cm (<1P)	Schwere globale Entwicklungsverzögerung spastisch-dystone Bewegungsstörung Rumpfhypotonie kaum Kopfkontrolle kein freies Sitzen Verlust von Greiffähigkeit	West-Syndrom (therapieresistent) Myoklonien beim Einschlafen	Spontan- nystagmus Reaktion auf starke Lichtreize	progr. Hirnatrophie, Leukenzephalopathie mit Hypo- myelinisierung	Laktat ↑ GOT, GPT und γGT ↑	HP:0002283 HP:0002352 HP:0006808 HP:0001332 HP:0001257 HP:0001252 HP:0001263 HP:0001250 HP:0000639 HP:0000253
MS005	♀ 5 J - Kleinwuchs Makrozephalie	Entwicklungsverzögerung Stereotypien	-	-	-	-	HP:0001263 HP:0000733 HP:0000256 HP:0001510 HP:0004322

ID	K Maße Dysmorphien	Anamnese/ Klinische Untersuchung	Epilepsie	Augen	Bildgebung	Labor/Histo	HPO-Terms
MS008 ♀ 17 J	-	Koma nach Magen-Darm-Infekt restitutio ad integrum	-	-	Bilaterale Basalganglienläsion	Laktatazidose	HP:0007146 HP:0001298 HP:0003128
MS009 ♀ 6 J	- mit 4 J: 100 cm (21P) 14,5 kg (18P) KU: 50 cm (49P)	Globale Entwicklungsverzögerung keine Sprachentwicklung Hyperaktivität, Autismus keine Risikoeinschätzung Gastroenteritis mit 19T Grippe H1N1 mit 1,5M Chronische Diarrhö	Generalisiert (therapieresistent; Verbesserung unter Gluthation)	-	-	Glycin ↑	HP:0001263 HP:0001344 HP:0000717 HP:0000752
MS012 ♂ 1 J	-	Kombinierte globale Entwicklungsstörung Trinkschwäche mit 6 W Muskelhypotonie + V.a. COX-Defekt (postpartal bei Bruder)	anfallsverdächtige - Zuckungen am 1T (Bruder)	-	Balkenagenesie Marklagerblutung alt (Bruder)	Laktatazidose Liquorlaktat ↑	HP:0008872 HP:0002490
MS014 ♂ 12 J Fam 04	-		-	-	-	-	HP:0000648 HP:0001272 HP:0001251 HP:0000639
MS015 ♀ 9 J Fam 04	-		-	-	-	-	HP:0000648 HP:0001272 HP:0001251 HP:0000639

ID	K Maße Dysmorphien	Anamnese/ Klinische Untersuchung	Epilepsie	Augen	Bildgebung	Labor/Histo	HPO-Terms
MS016 ♂ 4 J Fam 05	-		-	-	-	-	HP:0007146 HP:0001263
MS017 ♂ 4 J Fam 05	-		-	-	-	-	HP:0007146 HP:0001263

Abkürzungen aus den Tabellen 1.1 und 1.2:

ID = Identität, Fam = Familie, † = verstorben, K = Konsanguinität, SS = Schwangerschaft, NB = Nebenbefunde, P = Perzentile, KU = Kopfumfang, CF = Cystische Fibrose, Juv. MD = Juvenile Makuladegeneration, IUGR = *intrauterine growth retardation* (Intrauterine Wachstumsverzögerung), ID = *Intellectual disability* (Geistige Behinderung), COX = Cytochrom-c-Oxidase

Die Exom- und Genomuntersuchungen wurden im Rahmen der Diagnostik durchgeführt, weitere Untersuchungen im Rahmen der „Genom First Approach“ Studie (367/2019BO1). Nach schriftlichem Einverständnis der PatientInnen bzw. deren gesetzlicher VertreterInnen, wurde allen ProbandInnen, sowie teilweise deren Eltern und weiteren betroffenen Geschwistern Blut abgenommen. Zusätzlich wurden Hautbiopsien der Betroffenen durch behandelnde ÄrztInnen entnommen.

Aus den EDTA-Blutproben wurde in den Laboren des NGS Competence Center Tübingen (mit institutsinterner Core Facility NCCT; Leitung: Dr. Nicolas Casadei) genomische DNA extrahiert und für weitere Analysen verwendet (s. Abschnitt 3.1). Aus den Hautbiopsien der PatientInnen wurden Fibroblasten gewonnen und als Primäre Fibroblastenzelllinien für weitere Experimente in den Laboren der Abteilung für Zytogenetik des Instituts für Medizinische Genetik und Angewandte Genomik Tübingen (Leitung: Dr. Ulrike Mau-Holzmann) kultiviert und kryokonserviert (s. Abschnitte 3.1 und 3.5.1). Aus den Fibroblasten wurde wiederum durch die institutsinterne Core Facility (s. oben) mRNA extrahiert und sequenziert (s. Abschnitt 3.1).

Die erhobenen Daten aller 16 ProbandInnen wurden mittels standardisierter Pipelines des Instituts für Medizinische Genetik und Angewandte Genomik Tübingen aufbereitet und für diese Arbeit ausgewertet (s. Abschnitte 3.2 und 3.3).

Eigene molekulargenetische Analysen wurden im institutsinternen Forschungslabor für ausgewählte ProbandInnen angeschlossen (s. Abschnitt 3.4 - 3.6). Die Daten dieser ProbandInnen werden in dieser Arbeit ausführlich vorgestellt und diskutiert.

2 Material

2.1 Primäre Fibroblastenzelllinien

Bezeichnung	Mutation
<i>Patientenzelllinien</i>	
MS001	homozygote Nonsense-Mutation im Gen <i>SEC31B</i> (ENST00000370345.3, c.[1759G>T; 1759G>T]; p.[(Glu587*); (Glu587*)])
MS002	homozygote Missense-Mutation im Gen <i>DBR1</i> (ENST00000260803: c.[461G>A];[461G>A], p.[Arg>154Lys];[Arg>154Lys])
<i>Kontrollzelllinien</i>	
TF004_01	-
TF024_01	-
TF105_01	-

Anmerkung: Die kryokonservierten primäre Fibroblastenzelllinien wurden freundlicherweise von der institutsinternen Abteilung für Cytogenetik (Patientenzelllinien) und Dr. rer. nat. Rebecca Buchert-Lo (Kontrollzelllinien) zur Verfügung gestellt.

2.2 Chemikalien, Größenmarker, Kits

2.2.1 Chemikalien

Bezeichnung	Hersteller
1,4-Dithothreitol für die Biochemie	Merck KGaA, Darmstadt
Acrylamide-Solution (30 %) Mix 29:1 für die Molekularbiologie	AppliChem GmbH, Darmstadt
Ammonium persulfate for molecular biology, for electrophoresis, ≥98 %	Sigma-Aldrich, Steinheim
Ampuwa®	Fresenius Kabi Deutschland, Bad Homburg
Bicin für Pufferlösungen	AppliChem GmbH, Darmstadt
Borsäure (≥99,8 %, p. a., ACS, ISO)	Carl Roth GmbH + Co. KG, Karlsruhe
Bovine Serum Albumin Fraction V	Roche Diagnostics GmbH, Mannheim
cOmplete Mini, EDTA-free Protease inhibitor cocktail tablets	Roche Diagnostics GmbH, Mannheim
Descosept AF	Dr. Schumacher GmbH, Malsfeld
Dimethylsulfoxid (DMSO) für die Zellkultur	AppliChem GmbH, Darmstadt
Essigsäure (Eisessig) 100 %, wasserfrei zur Analyse EMSURE® ACS, ISO, Reag. Ph Eur	Merck KGaA, Darmstadt

Bezeichnung	Hersteller
Ethidiumbromidlösung 1 % (10 mg/ml, Stammlösung)	Carl Roth, Karlsruhe
Ethylendiamin-tetraessigsäure Dinatriumsalz Dihydrat ≥99 %, p.a.; ACS	Carl Roth, Karlsruhe
Gibco™ Trypsin-EDTA (0.25 %), phenol red (1x)	Thermo Fisher Scientific Inc., Waltham (US)
Gibco™ Antibiotic-Antimycotic (100x)	Thermo Fisher Scientific Inc., Waltham (US)
Gibco™ DMEM (1x), Dulbecco's Modified Eagle Medium (+4,5 g L/D-Glucose, L-Glutamine; ohne Pyruvat)	Thermo Fisher Scientific Inc., Waltham (US)
Gibco™ DPBS (1x), Dulbecco's Phosphate Buffered Saline, (ohne CaCl ₂ ; ohne MgCl ₂)	Thermo Fisher Scientific Inc., Waltham (US)
Gibco™ Heat Inactivated FBS (Fetal Bovine Serum)	Thermo Fisher Scientific Inc., Waltham (US)
Gibco™ MEM NEAA (100x), Minimum Essential Medium Non-Essential Amino Acids	Thermo Fisher Scientific Inc., Waltham (US)
Glycerin, ROTIPURAN® ≥86 %, p.a.	Carl Roth, Karlsruhe
Glycin, PUFFERAN® ≥99 %, p.a.	Carl Roth, Karlsruhe
Isopropanol 70 %	SAV Liquid Production GmbH, Flintsbach am Inn
Lithiumdodecylsulfat ≥99 %, für die Biochemie	Carl Roth, Karlsruhe
Methanol, Puriss. p.a., ACS Reagent, Reag. ISO, Reag. Ph. Eur., ≥98 %	Honeywell/Riedel-de Haën, Seelze
MOPS für die Molekularbiologie	AppliChem GmbH, Darmstadt
Deoxycholsäure Natriumsalz, ≥97 % (titration)	Sigma-Aldrich, Steinheim
Natriumazid reinst.	AppliChem GmbH, Darmstadt
Natriumchlorid (ACS, Reag. Ph. Eur.)	VWR International, Darmstadt
Natriumhydroxid-Plättchen (Reag. USP) zur Analyse, ACS, ISO	AppliChem GmbH, Darmstadt
dNTP Mix, PCR Grade (Mix containing 10 mM each of dATP, dCTP, dGTP, and dTTP)	Qiagen, Venlo (NL)
Phenolrot (Reag. USP) zur Analyse, ACS	AppliChem GmbH, Darmstadt
Ponceau S (C.I: 27195) für die Elektrophorese	Sigma-Aldrich, Steinheim
Protein Assay Dye Reagent Concentrate	Bio-Rad Laboratories GmbH, Feldkirchen
RNaseZAP™	Sigma-Aldrich, Steinheim

Bezeichnung	Hersteller
Salzsäure ROTIPURAN® 37 % rauchend, p.a., ACS, ISO	Carl Roth, Karlsruhe
SDS Pellets, ≥99 %, für die Biochemie	Carl Roth, Karlsruhe
SeaKem® LE Agarose	Lonza, Rockland (US)
Skim Milk Powder for microbiology	Sigma-Aldrich, Steinheim
TEMED	AppliChem GmbH, Darmstadt
Triton® X 100 reinst	Carl Roth, Karlsruhe
Trizma base, Primary Standard and Buffer ≥99,9 % (titration), crystalline	Sigma-Aldrich, Steinheim
Tween® 20 für die Molekularbiologie	AppliChem GmbH, Darmstadt

2.2.2 Größenmarker

Bezeichnung	Hersteller
GeneRulerDNA Ladder Mix, ready-to-use	Thermo Fisher Scientific Inc., Waltham (US)
Precision Plus Protein™ Dual Color Standards	Bio-Rad Laboratories GmbH, Feldkirchen
Invitrogen™ SeeBlue™ Pre-stained Protein Standard	Thermo Fisher Scientific Inc., Waltham (US)

2.2.3 Kits

Bezeichnung	Hersteller
Taq DNA Polymerase Kit	Qiagen, Venlo (NL)
QuantiTect® Reverse Transcription Kit	Qiagen, Venlo (NL)

2.3 Puffer und Lösungen

Bezeichnung	Zusammensetzung
10 % APS-Lösung	10 % Ammoniumpersulfat; in ddH ₂ O
3,5x Bis-Tris Puffer	1,25 M Bis-Tris in ddH ₂ O mit 25 % Salzsäure auf pH 6.5-6.8 eingestellt
5 % Blocking-Puffer	5 % Skim milk Powder; in 1x TBS
BSA-Lösung (0,05 mg/mL)	5 % BSA-Lösung (1 mg/ml) in 100 % ddH ₂ O
BSA-Lösung (1 mg/ml)	0,1 % bovine serum albumin fraction V; in ddH ₂ O
1 M DTT Lösung	1 M 1,4-Dithiothreitol in ddH ₂ O
0,5 M EDTA (pH 8.0)	0,5 M EDTA; in ddH ₂ O; mit 5 M Natronlauge auf pH 8,0 eingestellt

Bezeichnung	Zusammensetzung
Extraktionslysepuffer	5 % 1 M Tris (pH 7.5); 150 mM Natriumchlorid; 12 mM Natrium-Deoxycholat; 1 % Triton X 100; in 100 % ddH ₂ O
25 % HCl-Lösung	67,5 ml Salzsäure, rauchend 37 %; 32,4 ml ddH ₂ O
Kulturmedium für Fibroblasten	10 % Fetal Bovine Serum 1 % Antibiotic-Antimycotic 1 % Non-Essential Amino Acids ad 500 ml DMEM
4x LD-Puffer	2 mL 2.5 M Tris-HCl (pH 8.5); 2,5 mL Glycerol 86 %; 0,4 g Lithiumdodecylsulfat; 0,003 g EDTA; 0,125 mL 1 % Phenolrot; 0,375 mL ddH ₂ O
Lysispuffer	10 mM Tris; 150mM Natriumchlorid; 1 % Triton X 100; in ddH ₂ O
Mild Stripping Buffer	25 mM Glycin; 34,7 mM SDS; 1 % Tween 20; in ddH ₂ O mit 25 % Salzsäure auf pH 2.2 eingestellt
1x MOPS-Puffer	10 % 20x MOPS-Puffer; in 100 % ddH ₂ O
20x MOPS-Puffer	1 M MOPS; 1 M Tris; 69,4 mM SDS; 20,5 mM EDTA; in ddH ₂ O
2 % Natriumazid-Lösung	2 % Natriumazid reinst; in ddH ₂ O
5 M Natronlauge	5 M Natriumhydroxid-Plätzchen in ddH ₂ O
1 % Phenolrot-Lösung	1 % Phenolrot; in ddH ₂ O
Ponceau S staining solution	5 % Essigsäure (Eisessig, 100 %); 0,1 % Ponceau S; in 100 % ddH ₂ O
10x Proteaseinhibitor	1 Tablette cOmplete Mini Protease inhibitor cocktail in 1mL ddH ₂ O
10x TBE-Puffer	892 mM Tris; 890 mM Borsäure; 4 % 0,5 M EDTA (pH 8,0); in ddH ₂ O
1x TBE-Puffer	10 % 10x TBE-Puffer; in 100 % ddH ₂ O
10x TBS	10 % 1 M Tris pH 7.5; 30 % 5 M NaCl; in 100 % ddH ₂ O
1x TBS	10 % 10x TBS; in 100 % ddH ₂ O
1x TBS-T	10 % 10x TBS; 1 % Tween 20; in 100 % ddH ₂ O
1x Transfer-Puffer	5 % 20x Transfer-Puffer; 15 % Methanol; in 100% ddH ₂ O
20x Transfer-Puffer	0,5 M Bicine; 0,5 M Bis-Tris; 20 mM EDTA; in ddH ₂ O
1 M Tris (pH 7,5)	1 M Tris in ddH ₂ O mit 25% Salzsäure auf pH 7.5 eingestellt
2,5 M Tris-HCl (pH 8,5)	2,5 M Tris; in ddH ₂ O mit 25 % Salzsäure auf pH 8.5 eingestellt

2.4 Primer

Bezeichnung	Sequenz
U385_OSCP1_RT_F1	5' - GAA ACT ATG TCC CTG CAC CC - 3'
U386_OSCP1_RT_R1	5' - ACT CTT CTG GCC TGG TTA GC - 3'
Hersteller aller Primer: metabion international AG, Planegg/Steinkirchen	

2.5 Antikörper

2.5.1 Primärantikörper

Bezeichnung	Konzentration	Hersteller
DBR1 Rabbit Polyclonal antibody (16019-1-AP)*	1:1000	Proteintech Europe, Manchester (UK)
Anti-SEC31B antibody (ab80053)*	1:2500	Abcam, Cambridge (UK)
SEC31B Antibody - middle region (ARP66671_P050)*	1:500	Aviva Systems Biology Corporation, San Diego (USA)
GAPDH Antikörper (0411): sc-47724**	1:2500	Santa Cruz Biotechnology, Dallas (USA)

Anmerkung: *polyklonal, produziert in Kaninchen (rabbit, rb); **monoklonal, produziert in Maus (mouse, ms)

2.5.2 Sekundärantikörper

Bezeichnung	Konzentration	Hersteller
IRDye® 800CW Goat anti-Mouse IgG Secondary Antibody	1:10000	LI-COR® Biosciences, Lincoln (US)
IRDye® 680RD Goat anti-Rabbit IgG Secondary Antibody	1:10000	LI-COR® Biosciences, Lincoln (US)

2.6 Geräte

Bezeichnung	Hersteller
BioPhotometer® 6131	Eppendorf AG, Hamburg
Centrifuge 5417C	Eppendorf AG, Hamburg
Centrifuge 5418	Eppendorf AG, Hamburg
Centrifuge 5810 (mit Rotor A-4-62 und Adapter für 15 ml/50 ml konische Röhrchen)	Eppendorf AG, Hamburg
CO2-Inkubator CB 210	Binder, Tuttlingen
Corning® CoolCell® LX, Cell Freezing Container, for 12 x 1 mL or 2 mL Cryogenic Vials	Corning Incorporated, Corning (USA)
G-Storm GS1 Thermocycler	AlphaMetrix Biotech, Rödermark

Bezeichnung	Hersteller
Inkubator Function line BB 16	Heraeus Deutschland GmbH & Co. KG, Hanau
LI-COR® Odyssey® Fc Imaging System	LI-COR® Biosciences, Lincoln (US)
Mikroskop Eclipse TS100	Nikon Instruments Europe BV, Amsterdam (NL)
Mini-PROTEAN® Tetra Vertical Electrophoresis Cell, 4-gel, for 1.0 mm thick handcast gels	Bio-Rad Laboratories GmbH, Feldkirchen
Olympus Camedia C-4040, (4.1 Megapixel)	Olympus Europa SE & Co. KG, Hamburg
PerfectBlue™ Gelsystem Mini M	PEQLAB Biotechnologie, Erlangen
Pipetten: Einkanalpipetten Research® plus	Eppendorf AG, Hamburg
SE 300 miniVE Integrated Vertical Electrophoresis and Blotting Unit, complete	Hoefer Inc., Holliston (USA)
Sonopuls UW 22000	Bandelin electronic GmbH & Co. KG, Berlin
Sterilbank HERAsafe® HS12	Heraeus Deutschland GmbH & Co. KG, Hanau
Sterilbank HERAsafe® HS18	Heraeus Deutschland GmbH & Co. KG, Hanau
Synergy™ HT Multi-detection microplate reader	Bio-Tek Instruments, Inc., Bad Friedrichshall
TE 22 Mighty Small Transfertank	Hoefer Inc., Holliston (USA)
ThermoStat Plus	Eppendorf AG, Hamburg
Tischzentrifuge: Mini-Zentrifuge Spectrafuge 3-1810	neoLab Migge GmbH, Heidelberg
Transilluminator Bio View (UV light)	Biostep, Jahnsdorf
VF2 Laborschüttler (mit Aufsatz Vortex Genie 2)	IKA-Werke, Staufen im Breisgau

2.7 Software und Datenbanken

2.7.1 Software

Bezeichnung	Herausgeber
Argus X1 (Version 4.1.10)	Biostep, Jahnsdorf
Gen5, All-in-One Microplate Reader Software (Version 2.01.14)	Bio-Tek Instruments, Bad Friedrichshall
GSVar (Version 2019_08-14-g8b6a17a)	Institutsinterne Software
Image Studio™ (Version 2.1.10)	LI-COR® Biosciences, Lincoln (US)
Integrative Genomics Viewer (IGV) (Version 2.4.19)	Broad Institute and the Regents of the University of California, Oakland (US)
JMP® (Version 13.0.0)	SAS Institute, Cary (US)
Microsoft® Excel für Mac (Version 16.28)	Microsoft Corporation, Redmond (US)
Webversion Primer3 (Version 4.1.0)	Whitehead Institute for Biomedical Research, Cambridge (US)

2.7.2 Datenbanken

Bezeichnung	Referenzpublikation/Herausgeber
BioGPS	Wu et al., 2013
Genome Aggregation Database (gnomAD)	Karczewski et al., 2020
NCBI ClinVar (ClinVar)	Sayers et al., 2024
NCBI Pubmed®	Sayers et al., 2024
NCBI Single Nucleotide Polymorphism Database (dbSNP)	Sayers et al., 2024
Online Mendelian Inheritance in Man® (OMIM®)	McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore (US)
Orphanet	Institut national de la santé et de la recherche médicale, Paris (FR)
The Human Gene Mutation Database (HGMD®)	Stenson et al., 2020
The UniProt Consortium (UniProt)	Rath et al., 2012
UCSC Genome Browser	Kent et al., 2002

2.7.3 *In silico*-Tools

Tool	Referenzpublikation
CADD	Kircher et al., 2014
FATHMM	Shihab et al., 2013
Polyphen-2	Adzhubei et al., 2010
MetaLR	Dong et al., 2015
SIFT	Ng und Henikoff, 2001
MaxEntScan	Yeo und Burge, 2004
GeneSplicer	Pertea et al., 2001
PhyloP	Pollard et al., 2010

2.8 Verbrauchsmaterialien

Bezeichnung	Hersteller
Amersham™ Protran® Premium 0.2 µm NC (Nitrocellulose Blotting Membran), 300 mm x 4 m (1 roll)	GE Healthcare Life Science, Freiburg
Cellstar® Cell Culture Dishes, PS, 35x10 mm	Greiner Bio-One GmbH, Frickenhausen
Cellstar® Cell Culture Flasks, 250 mL, 75 cm ² , PS, red filter screw cap	Greiner Bio-One GmbH, Frickenhausen
Cellstar® Cell Culture Flasks, 50 mL, 25cm ² , PS, red filter screw cap	Greiner Bio-One GmbH, Frickenhausen
Cellstar® PP Röhrchen, 15 und 50 ml (beide konischer Boden mit Schraubverschluss)	Greiner Bio-One GmbH, Frickenhausen
epT.I.P.S.® Standard, 2-200 µl	Eppendorf AG, Hamburg
Falcon® Tissue Culture Dish, 100x20 mm Style	Corning Incorporated, Corning (US)
Filterpapier für Western Blot (0,8 mm stark)	biostep GmbH, Burkhardtsdorf
Invitrogen™ Bolt™ 4-12 % Bis-Tris Plus Gels, 15-well	Thermo Fisher Scientific Inc., Waltham (US)
Kryoröhrchen, 1,5 ml	TPP Techno Plastic Products AG, Trasadingen (CH)
Microplatte, 96 well, PS, U-Boden, Transparent	Greiner Bio-One GmbH, Frickenhausen
PCR Tubes (0,2 ml, Flat Cap)	STARLAB GmbH, Hamburg
Reaktionsgefäße, 1,5 ml und 2 ml (bei beiden mit angehängtem Deckel)	Greiner Bio-One GmbH, Frickenhausen
Sealing-Folie: ThermoSeal und ClearSeal (Durchstechbar)	Thermo Fisher Scientific Inc., Waltham (US)

Bezeichnung	Hersteller
TipOne® Filterspitzen 10 µl, 200 µl, 1000 µl (alle graduiert und steril)	STARLAB GmbH, Hamburg
TipOne® Pipettenspitzen, 10 µl graduiert, 200 µl konisch, 1000 µl graduiert	STARLAB GmbH, Hamburg

3 Methoden

3.1 DNA- und RNA-Sequenzierungen mittels *Next Generation Sequencing*

Zur Identifizierung genetischer Varianten bei Patienten mit vermuteten Erbkrankheiten wurden das komplette Genom (*Whole Genome Sequencing*, WGS) oder das komplette Exom (*Whole Exome Sequencing*, WES) und zusätzlich bei jedem Patienten das proteinkodierende Transkriptom (RNA-Seq) mittels Technologien des *Next Generation Sequencing* (NGS) aufgeschlüsselt. Unter dem Begriff *Next Generation Sequencing* werden Hochdurchsatzverfahren zur DNA-Sequenzierung zusammengefasst, welche nach der Sanger-Sequenzierung (Sanger et al., 1977) entwickelt wurden und eine parallele Sequenzierung von Millionen bis Billionen DNA-Strängen gleichzeitig ermöglichen (Mardis, 2011).

Die Laborabläufe und die Sequenzierläufe der Proben erfolgten im Rahmen des institutsinternen Standard-Procédere für diagnostische Proben durch das NGS Competence Center Tübingen (institutsinterne Core Facility).

Die Laborabläufe sollen daher in diesem Abschnitt nur in ihren Grundzügen skizziert werden.

3.1.1 Vorbereitung einer NGS-Bibliothek (*library preparation*)

Zu Beginn jeder Vorbereitung einer Sequenzierung mittels Technologien des NGS steht die sogenannte *library preparation*, um eine Kompatibilität zwischen der extrahierten DNA oder RNA und dem verwendeten Sequenziersystem herzustellen.

Diese Kompatibilität wird v. a. durch das Einführen spezifischer Adaptersequenzen ermöglicht. Diese Oligonukleotidsequenzen enthalten neben komplementären Sequenzen zu Oligonukleotiden auf der *flow cell* auch sog. Indices zur Wiedererkennung bei parallelem Sequenzieren von multiplen Individuen.

Die Vorbereitung der *library* umfasst i. d. R. folgende essenzielle Hauptschritte (Head et al., 2014):

1. Fragmentierung der DNA und Größenselektion der zu untersuchenden Fragmente
2. Umwandlung der Zielsequenzen in Doppelstrang-DNA (entfällt bei DNA-Proben)
3. Anfügen von Adaptersequenzen
4. Quantifizierung der fertigen NGS-Bibliothek

Weitere Zwischenschritte hängen vom verwendeten Ausgangsmaterial, der Quantität dieses Ausgangsmaterials, der Fragestellung und dem verwendeten Sequenziersystem ab.

Als Ausgangsmaterial für das WGS und WES wurde DNA verwendet, welche aus EDTA-Blutproben von PatientInnen bzw. deren gesunden Eltern mittels des *FlexiGene DNA Kit* (Qiagen) nach Herstellerangaben extrahiert wurde.

Nach ultraschallbasierter Fragmentation (Covaris-Ultraschall) wurde die DNA für das WGS mit dem *TruSeq DNA PCR-Free Kit* der Firma Illumina nach Herstellerangaben vorbereitet. Hierbei erfolgt ein Auffüllen der ‚klebrigen Enden‘ der DNA-Fragmente zu ‚glatten Enden‘ und eine Größenselektion der Fragmente mittels mitgelieferter magnetischer Perlen. Zur Fertigstellung der NGS-Bibliothek erfolgt die Anheftung spezifischer Adaptersequenzen über die Kreierung eines Adenin-Einzelnukleotidüberhangs und anschließende Ligation (Adenin komplementär zu einem Thymin-Überhang an der Adaptersequenz).

Die *library preparation* für das WES erfolgte mittels des *SureSelectXT Human All Exon V6 Kits* der Firma Agilent Technologies. Analog zur Vorbereitung der *library* für das WGS (s. oben) wird die purifizierte DNA ebenfalls mit Ultraschall fragmentiert (Fragmentgröße: 200 – 500 bp), die Enden werden geglättet und die Adaptersequenzen ebenfalls über einen

Einzelnukleotidüberhang gebunden und nachfolgend ligiert (vgl. vorheriger Absatz). Nach Anreicherung der Adapter-ligierten Fragmente durch PCR und mehreren Aufreinigungsschritten, werden Fragmente selektiert, welche proteinkodierende Sequenzen enthalten. Für diese Selektion werden biotinylierte RNA-Sequenzen verwendet. Diese sog. Köder hybridisieren mit den kodierenden Sequenzen (Gnirke et al., 2009). Durch Bindung des Biotins an Streptavidin-markierte magnetische Perlen können die selektierten DNA-Fragmente aufgereinigt werden.

Die RNA, welche für die Transkriptom-Sequenzierung verwendet wurde, stammt aus Fibroblasten von betroffenen PatientInnen. Die primären Fibroblastenzelllinien wurden durch die institutsinterne Abteilung für Cytogenetik kultiviert und geerntet.

Die NGS-Bibliothek wurde mit dem *NEBNext® Ultra™ II RNA Library Prep Kit for Illumina®* der Firma New England Biolabs vorbereitet. Der Arbeitsablauf dieses Kits sieht zunächst eine Aufreinigung der mRNA mittels *Oligo d(T)25 beads* (Bindung des Poly-A-Endes der mRNA) vor. Nach der Aufreinigung erfolgt eine Fragmentierung der mRNA und eine cDNA-Synthese (Umschreiben der RNA in doppelsträngige komplementäre DNA). Die so erhaltene DNA kann anschließend analog zum WGS und WES mit Adaptern versehen werden.

Nach einer Qualitätsanalyse erfolgte die Quantifizierung der fertigen NGS-Bibliotheken mit dem Qubit™-System (Invitrogen) nach Herstellerangaben.

3.1.2 Vorbereitung vor dem Sequenzierlauf (*template preparation*)

Vor der Sequenzierreaktion mittels der Methode *sequencing by synthesis* (s. Abschnitt 3.1.3) muss zunächst die sogenannte Probenvorbereitung (*template preparation*; Metzker, 2010) erfolgen. Hierbei wird die vorbereitete NGS-Bibliothek (s. Abschnitt 3.1.1) durch Bindung an eine Glasoberfläche (*flow cell*) immobilisiert und die Bildung von Clustern durch klonale Amplifikation induziert (Adessi et al., 2000, Fedurco et al., 2006).

Für das *NextSeq 6000-System (Illumina)* erfolgte die Clustergenerierung mittels Brückenamplifikation (*bridge amplification*; Bentley et al., 2008). Hierbei bindet das 3'-Ende des gebundenen DNA-Fragments an einen weiteren oberflächengebundenen Primer – es entsteht eine ‚Brücke‘. Durch erneute Extension am 3'-Ende des zweiten oberflächengebundenen Primers entsteht eine neue Kopie des DNA-Stranges. Durch Wiederholung dieses Zyklus aus Bindung, Extension und Denaturierung entstehen gleichzeitig Millionen von Clustern (alle resultierend aus einem DNA-Fragment) auf der *flow cell*.

Die Bildung von Clustern war notwendig, um detektierbare Fluoreszenzsignale im nachfolgenden Sequenzierlauf zu erreichen.

3.1.3 Sequenzierung durch Synthese (*sequencing by synthesis*)

Die Sequenzierläufe erfolgten mittels der Technik des Sequenzierens durch Synthese (*sequencing by synthesis*; Bentley et al., 2008) mit dem *NovaSeq 6000-System* der Firma Illumina. Bei dieser Technik erfolgt eine zyklische Wiederholung von (i) Nukleotidinkorporation, (ii) Detektion des Fluoreszenzsignals und (iii) Abspaltung des Fluorophors (Bentley et al., 2008).

Der Schlüsselmechanismus des *sequencing by synthesis* liegt in der Verwendung von fluoreszenzbasierten reversiblen Terminatoren (Metzker, 2010). Diese speziell veränderten Nukleotide werden an das gepimte 3'-Ende als komplementäre Basen zum vorliegenden Strang durch eine modifizierte DNA-Polymerase inkorporiert. Anschließend erfolgt die Anregung zur Fluoreszenz mit Lasern unterschiedlicher Wellenlängen und die Detektion des Fluoreszenzsignals mit einer hochauflösenden Kamera. Nach Abspaltung des Fluorophors kann die nächste Base durch die Polymerase inkorporiert werden. Die Anzahl der durchgeführten Zyklen setzt die Länge der abgelesenen Basenabfolge (*read*) fest.

Durch die simultane Sequenzierung von hunderten Millionen Clustern auf der *flow cell* entstehen bei jedem Zyklus ebenfalls mehrere Millionen Lichtsignale. Diese werden durch einen integrierten Algorithmus den

korrespondierenden Basen inklusive eines Qualitätsparameters zugeordnet (sog. *base calling*). Das Ausgabeformat ist eine FASTQ-Datei – eine Textdatei, welche die Basenabfolge der *reads* und den Qualitätsscore (als Phred-Score kodiert in ASCII-Kodierung) des jeweiligen *base calls* enthält. Die beschriebenen Sequenzierreaktionen wurden zunächst für alle DNA-Stränge in einer Leserichtung und für die jeweiligen Indices (zur Zuordnung zum Individuum) durchgeführt. Zur Erzeugung eines Leserasters von beiden Seiten (*paired-end reads*) wurden die DNA-Stränge nach erneuter Brückenbildung in komplementäre DNA-Stränge umgeschrieben. Die Originalstränge wurden entfernt und eine erneute Sequenzierung wurde angeschlossen. Für alle Anwendungen wurden 100 bp *reads* im *paired-end*-Modus generiert.

3.2 Bioinformatische Aufarbeitung der Rohdaten der Sequenzierungsreaktionen

Die bioinformatische Aufarbeitung der durch das *NovaSeq 6000-System* ausgegebenen Rohdaten (FASTQ-Datei) erfolgte im Rahmen des institutsinternen Standard-Procedere für diagnostische Proben durch das NGS Competence Center Tübingen (institutsinterne Core Facility). Ziel dieser Datenanalyse war die Darstellung und Möglichkeit der Filterung von kleinen Varianten und CNVs. In diesem Abschnitt sollen lediglich die Grundlagen erläutert werden.

Die in diesem Abschnitt beschriebenen Datenanalysen für die Genom- und Exom-Sequenzierungen wurden mit der *megSAP pipeline* durchgeführt (Institute of Medical Genetics and Applied Genomics, 2016). Die *megSAP pipeline* nutzt verschiedene Tools und unterliegt den ständigen Weiterentwicklungen der Abteilung für diagnostische Bioinformatik. Mehrere weitere Analysen (z.B. Qualitätskontrollen) wurden mit dem *ngs-bits-Toolset* (Sturm et al., 2018) durchgeführt. Eine Übersicht der verwendeten Tools und deren Versionen zum Zeitpunkt der Datenanalyse kann der Tabelle 3.1 entnommen werden.

Tabelle 3.1: *MegSAP pipeline*: Verwendete Tools und Versionen zum Zeitpunkt der Datenanalyse

Prozess	Verwendetes Tool	Version
Trimmen der Adapter	SeqPurge	2019_09
Read mapping	BWA-MEM	0.7.17
Indel realignment	ABRA2	2.19
Duplicate removal	samblaster	0.1.24
Variant calling	freebayes	1.2.0
Variant annotation	Ensembl Variant Effect Predictor	96.3
Qualitätskontrollen	SeqPurge, MappingQC, VariantQC (ngs-bits)	2019_09
CNV calling	ClinCNV	1.16.1

Die Verarbeitung der RNA-Seq-Rohdaten wurde mit anderen Tools durchgeführt (s. Abschnitt 3.2.5).

3.2.1 Zuordnung zum Referenzgenom (*mapping*)

Als *mapping* bezeichnet man die Zuordnung der im FASTQ-Format gespeicherten *reads* zum Referenzgenom (in dieser Arbeit: GRCh37/hg19). Das Ausgabeformat war eine BAM-Datei. In dieser primären Datenanalyse wurden mehrere Schritte zusammengefasst. Nach einer Qualitätskontrolle der FASTQ-Datei (readQC, Version 2019_09; ngs-bits) wurden zunächst die *reads* auf eine Adapterkontamination mittels des Tools SeqPurge (Version 2019_09; Sturm et al., 2016) überprüft und die aufgrund zu kurzer DNA-Fragmente mitsequenzierten Adaptersequenzen wurden entfernt. Anschließend erfolgte das *mapping* und die Zuordnung eines Qualitätsscores (*mapping quality*) mit der Software BWA-MEM (Version 0.7.17; Li, 2013).

Zuletzt wurden Leseduplikate entfernt (samblaster Version 0.1.24; Faust und Hall, 2014) und die Indels anhand der Referenz neu abgeglichen (ABRA2, Version 2.19; Mose et al., 2019). Eine abschließende Qualitätskontrolle erfolgte mit MappingQC (Version 2019_09; ngs-bits).

3.2.2 Auflistung und Annotation der Varianten (*variant calling und variant annotation*)

Beim *variant calling* werden die zugeordneten *reads* anhand des Referenzgenoms (in dieser Arbeit: GRCh37/hg19) auf Varianten (SNPs und Indels) überprüft. Anschließend erfolgt eine Annotation der Varianten mittels der Abfrage von gängigen Datenbanken. Das Ausgabeformat nach diesen Schritten war eine VCF-Datei. Qualitätskontrollen wurden mit VariantQC (Version 2019_09; ngs-bits) durchgeführt.

Für das *variant calling* wurde das Tool freebayes (Version 1.2.0; Garrison und Marth, 2012) verwendet. Per Konvention werden Varianten, für die es mehrere mögliche Anordnungen in Bezug auf das Referenzgenom gibt, an eine festgelegte Position geschoben. Ausschlaggebend ist hier die Position mit den niedrigsten genomischen Koordinaten ergo die Position, welche sich am weitesten links befindet (sog. *left-alignment*; Tool: VcfLeftNormalize, Version 2019_09; ngs-bits).

Nach erfolgtem *variant calling* wurde die Liste der Varianten mit Informationen aus gängigen Datenbanken mittels des Ensembl Variant Effect Predictor (VEP, Version 96.3; McLaren et al., 2016) ergänzt, sodass eine Auswertung nach bestimmten Filtern und Vorhersagewerten (s. Abschnitt 3.3) möglich wurde.

3.2.3 Kopienzahlvarianten (*Copy number variations*)

Die Detektion vorhandener *Copy number variations* (CNVs) wurde mit der Software ClinCNV (Version 1.16.1; Demidov und Ossowski, 2019) durchgeführt. Analog zur *variant annotation* (s. Abschnitt 3.2.2) wurde auch diese Auflistung mit Informationen aus Datenbanken ergänzt.

3.2.4 Eltern-Kind- und Familienanalysen

Eltern-Kind- und Familienanalysen mit mehreren Betroffenen (sog. Trio- und Multi-Analysen) wurden mittels verschiedener Skripte der *megSAP-pipeline* (Institute of Medical Genetics and Applied Genomics, 2016) durchgeführt. Der Vorteil dieser Analysen liegt in der Erweiterung der

Filterungsschritte, sowie die Eingrenzung der auszuwertenden Varianten durch den Ausschluss von Varianten gesunder Individuen.

3.2.5 Bioinformatische Pipeline für die Rohdaten der RNA-Seq

Um die RNA-Seq-Rohdaten aufzubereiten wurden die *reads* ebenfalls mit dem Tool SeqPurge (Version 2019_09) auf Adapterkontamination überprüft und getrimmt. Anschließend erfolgte eine Zuordnung mit der Software STAR (Version 2,7,0f; Dobin et al., 2013) zum Referenzgenom GRCh37/hg19. Die Qualitätskontrolle des erfolgten *mappings* wurde mit MappingQC (Version 2019_09; ngs-bits) durchgeführt. Es erfolgte die Ausgabe einer BAM-Datei nach abgeschlossener Datenanalyse.

3.3 Auswertung der Genom-, Exom- und Transkriptom-Sequenzierungen

Nach Durchlauf der *megSAP-pipeline* enthält eine VCF-Datei beispielsweise für ein WES ca. 40.000 Varianten. Um diese Masse an Varianten für den Auswerter auf eine Anzahl zu reduzieren bei der jede Variante hinsichtlich ihrer möglichen Pathogenität bewertet werden kann, bedarf es einer Filterung nach verschiedenen Kriterien.

3.3.1 Verwendete Software

Zur Auswertung der Sequenzierungen (vorliegend im VCF- und BAM-Format) wurden die Softwares GSVar (Version 2019_08-14-g8b6a17a; ngs-bits) und Integrative Genomics Viewer (IGV, Version 2.4.19; Robinson et al., 2011) verwendet.

Die institutsinterne Software GSVar ist eine graphische Bedienoberfläche zur Filterung von genetischen Varianten in Genom- und Exomdaten. Nach Import einer VCF-Datei ist eine Filterung anhand folgender Parameter möglich:

- Frequenz in der Allgemeinbevölkerung (1000g, gnomAD)
- Frequenz in der institutsinternen Datenbank
- Art der Mutation (z.B. Missense, Nonsense, etc.)
- Art der Vererbung (z.B. dominant, rezessiv, compound-heterozygot)

- Vorhersage der Pathogenität nach CADD, FATHMM, PolyPhen2, MetaLR, Sift und phyloP
- Phänotyp durch vorherige Zuordnung von HPO-Ziffern
- Bekannte (bereits pathogen klassifizierte) Gene in Medizinischen Datenbanken (OMIM, HGMD, ClinVar)

Der Integrative Genomics Viewer ist eine Software zur graphischen Darstellung der Daten einer BAM-Datei. Die Bedienoberfläche ermöglicht die Exploration der Genomischen Daten, indem Referenzgenom, Abdeckung der Referenz (*coverage*), die einzelnen *reads* (mit farbigen Einstellungsmöglichkeiten) und die Varianten (als Abweichung von der Referenz) dargestellt werden (Thorvaldsdóttir et al., 2012). Des Weiteren bietet der IGV diverse Tools, wie z.B. die Generation eines Sashimi-Plots, um das Spleißen in den RNA-Seq-Daten zu überprüfen (Katz et al., 2013). Es besteht eine Verknüpfung zwischen beiden Softwares, sodass eine Variante über GSVar im IGV aufgerufen werden kann. Dies ermöglicht die direkte visuelle Überprüfung der Variante.

3.3.2 Filterung der vorhandenen Varianten

Die Daten des WGS und WES wurden mit festgelegten Filterungsschritten überprüft.

Die eingestellten Parameter der Filterungsschritte für die kleinen Varianten (SNP und Indels) können Tabelle 3.2 entnommen werden. Nacheinander wurden die einzelnen Filter für rezessiven und dominanten Erbgang durchlaufen und die angezeigten Varianten bewertet (s. Abschnitt 3.3.3). Im Anschluss an diese frequenzbasierte Filterung wurden alle X-chromosomalen und mitochondrialen Varianten erneut gesondert dargestellt.

Bei Eltern-Kind-Analysen erfolgte statt der Filterungen für die einzelnen Erbgänge eine Überprüfung der kleinen Varianten mit den in Tabelle 3.3 dargestellten Parametern. Die Filterungsschritte der Familienanalysen können ebenfalls Tabelle 3.3 entnommen werden.

Abschließend erfolgte bei jedem Individuum eine Filterung der *SNPs* und Indels nach Phänotyp anhand der zugeordneten HPO-Terms (s. Kapitel 1.6).

Für die Filterung der CNVs wurden Qualitätsscores, bereits pathogen annotierte Gene und ebenfalls zugeordnete HPO-Terms zugrunde gelegt. Ein besonderer Fokus wurde auch auf den Vererbungsmechanismus durch Betrachtung der homozygoten und compound-heterozygoten CNVs (auch in Bezug auf vorhandene kleine Varianten) gelegt.

Die Filterschritte in GSVar sind aufeinander aufbauend, das heißt alle Varianten, welche einem einzelnen Filter nicht entsprechen, werden nicht mehr angezeigt. Mit dem „REMOVE“-Befehl verhält es sich gegensätzlich. Eine Ausnahme stellt der „KEEP“-Befehl dar, welcher eine Anzeige der Variante mit dem gegebenen Merkmal forciert. Aus diesem Grund wird nach dem ‚Annotated pathogenic KEEP: HGMD, ClinVar‘-Befehl erneut eine frequenzbasierte Selektion durchgeführt, um die wahrscheinlich benignen Varianten in bekannten Krankheitsgenen herauszunehmen.

Tabelle 3.2: Filterungsschritte für kleine Varianten (Original-Wortlaut aus dem Programm GSVar)

Recessive stringent	Dominant stringent	Recessive relaxed	Dominant relaxed
Allele frequency ≤ 0.50 %	Allele frequency ≤ 0.10 %	Allele frequency ≤ 1.00 %	Allele frequency ≤ 0.10 %
Allele frequency (sub-populations) ≤ 0.50 %	Allele frequency (sub-populations) ≤ 0.10 %	Allele frequency (sub-populations) ≤ 1.00 %	Allele frequency (sub-populations) ≤ 0.10 %
Count NGSD ≤ 15	Count NGSD ≤ 5	Count NGSD ≤ 40	Count NGSD ≤ 5
Impact HIGH, MODERATE	Impact HIGH	Impact HIGH, MODERATE, LOW	Impact HIGH, MODERATE
Filter columns REMOVE: gene_blacklist	Gene constraint o/e ≤ 0.35 (pLI ≥ 0.80)	Filter columns REMOVE: gene_blacklist	Filter columns REMOVE: gene_blacklist
Annotated pathogenic KEEP: HGMD, ClinVar	Filter columns REMOVE: gene_blacklist Annotated pathogenic KEEP: HGMD, ClinVar	Annotated pathogenic KEEP: HGMD, ClinVar	Annotated pathogenic KEEP: HGMD, ClinVar
Allele frequency ≤ 3.00 %	ClinVar	Allele frequency ≤ 3.00 %	Allele frequency ≤ 3.00 %
Classification NGSD REMOVE: 1, 2	Allele frequency ≤ 3.00 %	Classification NGSD REMOVE: 1, 2	Classification NGSD REMOVE: 1, 2
Classification NGSD KEEP: 4, 5	Classification NGSD REMOVE: 1, 2	Classification NGSD KEEP: 4, 5	Classification NGSD KEEP: 4, 5
Genotype affected hom, comp-het	Classification NGSD KEEP: 4, 5	Genotype affected hom, comp-het	

Anmerkung: Die Allelfrequenz bezieht sich auf die Datenbank gnomAD (s. Abschnitt 2.7). NGSD stellt die hausinterne Datenbank dar – hiermit können Pipeline-spezifische Artefakte herausgefiltert werden. Die Klassifikation in NGSD bezieht sich auf die fünf Kategorien der ACMG Klassifikation für genetische Varianten (Richards et al., 2015). Die Vorhersage der Auswirkung (impact) basiert auf der Annotation der Ensembl Datenbank (Yates et al., 2019). Die gene blacklist ist der ENCODE blacklist (Amemiya et al., 2019) entnommen.

Tabelle 3.3: Filterungsschritte für Eltern-Kind- und Familienanalysen (Original-Wortlaut aus dem Programm GSVar)

Trio stringent	Trio relaxed	Multi-sample recessive	Multi-sample comp-het	Multi-sample dominant
Allele frequency ≤ 1.00 %	Allele frequency ≤ 1.00 %	Allele frequency ≤ 1.00 %	Allele frequency ≤ 0.10 %	Allele frequency ≤ 0.10 %
Allele frequency (sub-populations) ≤ 1.00 %	Allele frequency (sub-populations) ≤ 1.00 %	Allele frequency (sub-populations) ≤ 1.00 %	Allele frequency (sub-populations) ≤ 0.10 %	Allele frequency (sub-populations) ≤ 0.10 %
Count NGSD ≤ 20	Count NGSD ≤ 20	Count NGSD ≤ 20	Count NGSD ≤ 20	Count NGSD ≤ 20
Variant quality qual ≥ 100 depth ≥ 20 mapq ≥ 55	Variant quality qual ≥ 30 depth ≥ 10 mapq ≥ 30	Impact HIGH, MODERATE, LOW	Impact HIGH, MODERATE, LOW	Impact HIGH, MODERATE, LOW
Variant type frameshift_variant, splice_acceptor_variant, splice_donor_variant, start_lost, start_retained_variant, stop_gained, stop_lost, inframe_deletion, inframe_insertion, missense_variant, splice_region_variant	Annotated pathogenic KEEP: HGMD, ClinVar (also likely pathogenic)	Annotated pathogenic KEEP: HGMD, ClinVar	Annotated pathogenic KEEP: HGMD, ClinVar	Annotated pathogenic KEEP: HGMD, ClinVar
Allele frequency ≤ 3.00 %	Allele frequency ≤ 3.00 %	Allele frequency ≤ 3.00 %	Allele frequency ≤ 3.00 %	Allele frequency ≤ 3.00 %
Classification NGSD REMOVE: 1, 2	Classification NGSD REMOVE: 1, 2	Classification NGSD REMOVE: 1, 2	Classification NGSD REMOVE: 1, 2	Classification NGSD REMOVE: 1, 2
Classification NGSD KEEP: 4, 5	Classification NGSD KEEP: 4, 5	Classification NGSD KEEP: 4, 5	Classification NGSD KEEP: 4, 5	Classification NGSD KEEP: 4, 5
Trio de-novo, recessive, LOH, comp-het, x-linked	Trio de-novo, recessive, LOH, comp-het, x-linked	Genotype affected hom Genotype control het, wt	Filter columns REMOVE: gene_blacklist Genotype affected comp-het Genotype control wt, het	Genotype affected het, hom Genotype control wt

Anmerkung: Für Spezifizierungen zu den einzelnen Filterschritten s. Anmerkung unter Tabelle 3.2.

3.3.3 Bewertung der gefilterten Varianten

Die gefilterten Varianten wurden zunächst mit der Software IGV auf Qualität kontrolliert. Diesbezüglich wurden die *mapping*-Qualität (*reads* mit der *mapping*-Qualität 0 werden im IGV weiß angezeigt) und die Tiefe der Sequenzierung (*depth* oder *coverage*; $\leq 15-20$ je nach *mapping*-Qualität insuffizient) begutachtet. Nachfolgend wurde der in GSVar ausgegebene Genotyp auf Richtigkeit überprüft.

Bei CNVs wurde zur Qualitätskontrolle auf die *coverage* (Abnahme bei Deletionen und Zunahme bei Insertionen) und das Vorhandensein von sogenannten *split-reads* (in IGV i. d. R. rot oder grün dargestellt) geachtet. *Split-reads* werden durch Abgleichung der erwarteten und der tatsächlich berechneten Größe sog. *inserts* ausgegeben. Als *inserts* werden die unsequenzierten Mittelstücke von DNA-Fragmenten im *paired-end*-Modus bezeichnet, dessen Größe durch Abgleich anhand der Referenz kalkuliert werden kann. Ein Hinweis auf eine Deletion ergibt sich, wenn die erwartete Größe über der tatsächlichen Größe liegt. Bei Hinweisen auf Insertionen verhalten sich die beiden Größen gegensätzlich.

Bei erfolgreicher Qualitätskontrolle wurden die Varianten auf mögliche Pathogenität untersucht. Für diese Bewertung wurden die Richtlinien des *American College of Medical Genetics and Genomics (ACMG)* herangezogen (Richards et al., 2015).

Eine umfangreiche Recherche in gängigen Datenbanken (s. Abschnitt 2.7), sowie eine Literaturrecherche wurde in die Klassifikation der Varianten einbezogen. Gene mit bereits pathogen eingestuft Varianten an anderen Stellen wurden ebenso priorisiert wie Gene mit passenden biochemischen Vorgängen (z.B. Assoziationen in vivo des resultierenden Proteins mit anderen bereits krankheitsassoziierten Proteinen, etc.). Bei Priorisierungen aufgrund in der Fachliteratur etablierten biochemischen Pathways wurde insbesondere auf die Expression des Gens im entsprechenden pathogen veränderten Gewebe geachtet. Dies wurde mittels der Datenbank BioGPS

überprüft (s. Abschnitt 2.7). Im Verlauf der Verfassung dieser Arbeit wurden die gängigen Datenbanken zur Literaturrecherche regelmäßig in Bezug auf mögliche Kandidatengene durchsucht. Neu publizierte Studien wurden entsprechend in die Bewertung miteinbezogen.

Des Weiteren wurden prädiktive Scores zur Analyse des Ausmaßes von Missense-Mutationen (CADD, FATHMM, Polyphen-2, MetaLR und SIFT), des Ausmaßes von Spleiß-Varianten (MaxEntScan, GeneSplicer) und der evolutionären Konservierung des Nukleotids (PhyloP) verwendet. Die einzelnen Scoring-Systeme sind in Abschnitt 2.7.3 aufgelistet.

Eine zusätzliche Ebene der Auswertung wurde durch die Anzeige der RNA-Seq-Daten des entsprechenden Individuums hinzugefügt. Die RNA-Seq-Daten wurden inklusive drei Kontrollen (Betroffene aus dem dargestellten Untersuchungskollektiv mit anderem Phänotyp) parallel zu den WGS- oder WES-Tracks in IGV aufgerufen. Es erfolgte auch hier eine Qualitätskontrolle (s. oben) und anschließend eine Bewertung. Die *coverage* (ergo die Expression), eventuelle präferentielle Expressionen von Varianten (z.B. LOH) und das Spleißmuster (überprüft mittels Sashimi-Plot) waren ausschlaggebend.

LOH bezeichnet bei autosomalen Genen den Verlust eines Allels und nachfolgend homozygote Ausprägung einer parallel auf dem zweiten Allel vorhandenen Variante. Dieses Phänomen konnte sowohl anhand eines visuellen Abgleichs von Varianten im WGS/WES mit den RNA-Seq-Daten (Verlust der Heterozygotie bei der Genexpression) als auch bei Eltern-Kind-Analysen (Verlust der Heterozygotie bei der Ontogenese) detektiert werden.

Zur Auswertung der Sashimi-Plots auf vorhandene Spleiß-Defekte wurden aus derselben Studie jeweils drei zufällig ausgewählte Vergleichsproben hinzugezogen. Durch visuelles Abgleichen der Verbindungsbögen zwischen den einzelnen Exons (Darstellung der exonübergreifenden *reads*) der zu untersuchenden Probe mit den Kontrollen wurden suspekta Spleißmuster identifiziert.

Mittels des OUTRIDER-Protokolls wurde auf aberrante Exprimierung der entsprechenden Transkripte des vorliegenden Kandidatengens geprüft (Methodik s. Brechtmann et al., 2018).

Durch eine Synopsis all dieser Kriterien erfolgte eine Klassifikation der wichtigsten Varianten in fünf Gruppen – pathogen, wahrscheinlich pathogen, Variante von ungewisser Bedeutung, wahrscheinlich benigne, benigne – nach den Richtlinien des ACMG (Richards et al., 2015).

3.3.4 X-Inaktivierungsanalyse

Um eine gestörte X-Inaktivierung bei allen Trios nachzuweisen, wurde die Trio-Analyse mit der quantitativen Expression von heterozygoten Varianten bei Betroffenen in der RNA-Seq kombiniert. Zunächst erfolgte die tabellarische Darstellung aller heterozygoten exonischen Varianten des X-Chromosoms in GSVar. Anschließend wurden die Varianten in IGV visualisiert. Die relative Häufigkeit der heterozygoten Varianten konnte schließlich anhand der Gesamtanzahl der *reads* und der absoluten Anzahl der heterozygoten Variante bestimmt werden. Ausgeschlossen wurden *de novo* Varianten, Varianten bei denen beide Eltern denselben Genotyp aufwiesen und Varianten mit einer *coverage* kleiner/gleich zehn in der RNA-seq.

Als gestörte bzw. veränderte X-Inaktivierung wurde ein Verhältnis von größer gleich 80/20 % für entweder Wildtyp oder heterozygoter Variante angesehen. Eine extrem gestörte X-Inaktivierung wurde bei prozentualen Anteilen von größer 90/10 % angenommen (vgl. Moortgat et al., 2018).

3.4 Aufklärung unklarer Befunde aus den Transkriptom-Daten mittels PCR aus cDNA

Um einen durch niedrige *coverage* bedingten unklaren Befund in den RNA-Seq-Daten für das Kandidatengen *OSCP1* zu validieren, wurde komplementäre DNA (cDNA) aus RNA synthetisiert und einer Polymerkettenreaktion (PCR) zugeführt. Die PCR ist eine Methode zur Amplifikation von Nukleinsäuren (Mullis et al., 1986).

Proben bereits isolierter RNA wurden von der institutsinternen Core Facility (Patientenproben, s. Abschnitt 3.1.1) und Dr. rer. nat. Rebecca Buchert-Lo (Kontrollen) zur Verfügung gestellt. Nach Erhalt dieser Proben wurden diese bis zur weiteren Verwendung bei -20 °C gelagert.

3.4.1 cDNA-Synthese

cDNA wurde unter Verwendung des *QuantiTect® Reverse Transcription Kit* (Qiagen) aus den RNA-Proben (Patientenproben und Kontrollen) synthetisiert, um eine Anreicherung der zu untersuchenden RNA-Abschnitte mittels PCR zu ermöglichen.

Um einen vorzeitigen Abbau der RNA durch ubiquitär vorkommende RNasen zu verhindern, wurden alle Pipetten und Geräte vor Verwendung mit RNaseZAP™ (Sigma-Aldrich) behandelt. Alle Pipettierschritte wurden, soweit möglich, auf Eis durchgeführt.

Zunächst wurde die RNA-Konzentration der verwendeten Proben photometrisch mittels BioPhotometer®6131 (Eppendorf; Parameter: RNA, Lichtweg: 1 mm, Faktor: 1,0 OD₂₆₀ 40 ng µl⁻¹) bestimmt. Die jeweils 500 ng RNA enthaltenden Volumina wurden anschließend einem Ansatz zum Verdau genomischer DNA (*gDNA Wipeout*) zugeführt. Die verwendeten Volumina des *gDNA Wipeout* - Ansatzes können Tabelle 3.4 entnommen werden. Der Verdau wurde anhand der Herstellerangaben durchgeführt und anschließend direkt auf Eis gestellt.

Tabelle 3.4: Ansatz für die Elimination genomischer DNA

Bezeichnung	Volumen [µl]
gDNA Wipeout Buffer, 7x	2
RNA-Probe (500 ng RNA)	x
RNase freies H ₂ O	12-x
Totales Volumen	14

Anmerkung: x entspricht dem Volumen, welches 500 ng RNA enthält.

Zu dem Ansatz des *gDNA Wipeouts* wurden die für die anschließende Retrotranskription benötigten Reagenzien, dessen Volumina Tabelle 3.5 entnommen werden können, hinzugefügt. Die reverse Transkription wurde entsprechend der Herstellerangaben mit folgenden Modifikationen durchgeführt: Inkubation bei 42 °C für 30 min statt 15 min. Bis zur weiteren Verarbeitung wurde die synthetisierte cDNA bei -20 °C eingefroren und gelagert.

Tabelle 3.5: Ansatz für die Retrotranskription

Bezeichnung	Volumen [µl]
Quantitect Reverse Transcriptase	1
Quantitext RT Buffer, 5x	4
RT Primer Mix	1
gDNA Wipeout Ansatz	14
Totales Volumen	20

3.4.2 Design der Primersequenzen

Die Primersequenzen für oben genanntes Ziel (Teil des Gens *OSCP1*) wurden mit der Webversion der Software Primer3 (Version 4.1.0, Whitehead Institute for Biomedical Research) entworfen. Die Standardeinstellungen der Software wurden bis auf folgende Modifikationen übernommen: *Max Poly-X* = 3 (maximal dreimal dasselbe Nukleotid nacheinander in der Sequenz des Primers) und *GC Clamp* = 1 (Guanin oder Cytosin am 3'-Ende des Primers). Tabelle 3.6 gibt einen Überblick über die wichtigsten Auswahlkriterien für das Design der Primersequenzen.

Tabelle 3.6: Überblick über die wichtigsten Auswahlkriterien beim Primerdesign

Auswahlkriterium	Einstellung
Länge	19-25 bp
GC-Gehalt	30-70 %
G- oder C am 3'-Ende	ja
Maximale Wiederholung desselben Nukleotids	3
Bindungstemperatur des Primers	57-62 °C
Maximale Differenz der Bindungstemperaturen des <i>forward</i> - und des <i>reverse</i> -Primers	1 °C
Maximale Länge des PCR-Produktes	1000 bp

Zur Ermittlung der DNA-Sequenz, welche in die Software Primer3 eingegeben werden konnte, wurden die Daten der Transkriptom-Sequenzierungen (s. Abschnitt 3.1) herangezogen. Der zu amplifizierende Bereich wurde zunächst in den Daten mittels IGV (Version 2.4.19) (Robinson et al., 2011) manuell aufgesucht. Durch die entsprechenden Koordinaten wurde eine Quellsequenz mit jeweils deutlichem Abstand zum gewünschten PCR-Produkt mittels *UCSC Genome Browser* (Referenzgenom: GRCh37/hg19) ausgegeben. Intronische Bereiche wurden aus der Quellsequenz entfernt, die Zielsequenz wurde mittels eckiger Klammern (definitiver Einschluss in das resultierende PCR-Produkt) markiert und die vollständige Sequenz wurde schließlich in die Software Primer3 eingegeben.

Alle Primer wurden mit der Funktion *In Silico PCR* des *UCSC Genome Browser* überprüft. Bei Vorkommen von SNPs oder gängigen Polymorphismen in $\geq 2\%$ der Allgemeinbevölkerung in den Primersequenzen wurden die Primer verworfen.

Die Synthese der final ausgewählten Primer (Sequenzen s. Abschnitt 2.4) wurde bei der Firma Metabion International in Auftrag gegeben.

3.4.3 Polymerasekettenreaktion

In Vorexperimenten wurden basierend auf verschiedenen Protokollen PCRs getestet und optimiert. In diesem Abschnitt wird lediglich auf die für die Datenerhebung letztendlich durchgeführte PCR eingegangen.

Für die PCR wurde das *Taq DNA Polymerase Kit* (Qiagen) verwendet.

Alle nachfolgenden Pipettierschritte wurden, sofern möglich, auf Eis durchgeführt.

Nach Verdünnung der Primerlösungen mit Ampuwa®-Wasser auf eine Konzentration von 10 pM, konnten die Ansätze für die PCR (Proben und Kontrolle) entsprechend Tabelle 3.7 zusammengeführt werden.

Tabelle 3.7: PCR-Ansatz

Bezeichnung	Volumen [μ l]
Master-Mix	17,4
Vorwärtsprimer (Primer f)	0,8
Rückwärtsprimer (Primer r)	0,8
cDNA-Proben	1
Totales Volumen	20

Anmerkung: Das Protokoll *PCR Using Taq DNA Polymerase* wurde entsprechend den Herstellerangaben zur Erstellung des Mastermixes (ohne $MgCl_2$) verwendet.

Nach kurzem Herunterzentrifugieren (Tischzentrifuge) erfolgte der Transfer in den bereits vorgeheizten G-Storm GS1 Thermocycler (AlphaMetrix Biotech).

Der Thermocycler wurde entsprechend einem Touchdown-PCR-Protokoll programmiert (s. Tabelle 3.8), um eine höhere Spezifität der erhaltenen PCR-Produkte zu gewährleisten (Don et al., 1991).

Tabelle 3.8: Programmierung des Thermocyclers G-Storm GS1 für die Touchdown-PCR

Anzahl der Zyklen	Reaktionsschritt	Temperatur	Zeit
1x	Heated Lid	105 °C	-
1x	Hot Start	94°C	3 min
10x	Denaturierung	94°C	1 min
	Annealing	68-59°C*	1 min
	Extension	72°C	1 min
30x	Denaturierung	94°C	1 min
	Annealing	59°C	1 min
	Extension	72°C	1 min
1x	Extension	72°C	10 min
1x	Aufbewahrung	15°C	Halten

Anmerkung: *Die Temperatur wurde im angegebenen Bereich beginnend mit der höchsten Temperatur pro Zyklus um 1 °C reduziert.

Nach Abschluss des Programms wurden die PCR-Produkte bis zur weiteren Verwendung bei -20 °C gelagert.

3.4.4 Agarosegelelektrophorese und Detektion der PCR-Produkte

Nukleinsäureproben können durch Agarosegelelektrophorese nach ihrer Größe aufgetrennt (Takahashi et al., 1969) und anschließend durch den DNA-bindenden Farbstoff Ethidiumbromid unter UV-Licht detektiert werden (LePecq und Paoletti, 1967).

Eine Agaroselösung (1 %) wurde zunächst durch Aufkochen von 0,55 g Agarose in 55 ml 1x TBE-Puffer hergestellt und mit 3 µl Ethidiumbromid unter leichtem Schwenken vermischt. Anschließend wurde die Lösung in ein horizontales PerfectBlue™ Gelsystem Mini M (Breite: 9 cm, Laufstrecke: 11 cm; PEQLAB Biotechnologie) gegossen. Ein entsprechender Taschenkamm wurde in die Gellösung gesteckt.

Während der Aushärtung des Gels konnten die Proben der durchgeführten PCR zur Beladung vorbereitet werden. Hierfür wurden je 5 µl Probe mit 1 µl DNA Ladepuffer durch Auf- und Abpipettieren homogenisiert.

Nach Transfer des Gels in die mit 1x TBE-Puffer gefüllte Elektrophoresekammer (PEQLAB Biotechnologie) erfolgte die Beladung der einzelnen Taschen jeweils mit dem kompletten Volumen (6 µl) eines der Probenansätze oder mit 2 µl des DNA-Größenmarker *O'Gene Ruler DNA Ladder Mix* (Thermo Fisher Scientific).

Die Gelelektrophorese wurde anschließend durch Anlage einer Spannung von 120 V gestartet und bei dieser Spannung für 25-30 min laufen gelassen.

Die Detektion der PCR-Produkte erfolgte unter UV-Licht (Transilluminator Bio View, biostep) mit der Kamera Olympus Camedia C-4040 (4,1 Megapixel, Olympus) und der Software Argus X1 (Version 4.1.10, Biostep).

3.5 Zellkultur

Alle Arbeiten mit humanen Zellen wurden unter der Schutzatmosphäre der UV-desinfizierten Sterilbänke HERAsafe® HS12 oder HERAsafe® HS18 (beide Heraeus) durchgeführt. Alle notwendigen unsteril verpackten Verbrauchsmaterialien wurden autoklaviert und vor Einbringen in die Schutzatmosphäre nochmals gründlich desinfiziert. Eine angegebene Erwärmung der benötigten Reagenzien vor Verwendung erfolgte stets bei 37 °C im Wasserbad.

Als Ausgangsmaterial für die Arbeiten mit adhären Zellen dienten kryokonservierte primäre Fibroblastenzelllinien, welche freundlicherweise von Dr. rer. nat. Rebecca Buchert-Lo und der institutsinternen Abteilung für Cytogenetik zur Verfügung gestellt wurden.

3.5.1 Kultivierung primärer Fibroblastenzelllinien

Primäre Fibroblastenzelllinien von Patienten mit nachgewiesenen Mutationen in Kandidatengenen, sowie drei Kontrollzelllinien wurden zur Bereitstellung von Ausgangsmaterial für weitere Experimente und Analysen (s. Abschnitte 2.1) in den Brutschränken CB 210 (Binder) und Function line BB 16 (Heraeus) bei 37 °C und 91 % Luftfeuchtigkeit unter Anreicherung von 5 % CO₂ kultiviert.

Die Kultivierung der adhärenenten Fibroblastenzellen erfolgte je nach Größe der gewählten Zellkulturflasche bzw. -schale in 5-10 ml Fibroblasten-Kulturmedium (5 ml für ein Zellkulturgefäß von 25 cm², 10 ml für ein Zellkulturgefäß von 75 cm²).

3.5.2 Passagieren von adhärenenten Zelllinien

Bei Erreichen einer Konfluenz von nahezu 100 %, erfolgte ein Aufsplitten der Zellen.

Das Kulturmedium wurde abgenommen und Reste des Mediums wurden mit 5-6 ml vorgewärmter *Dulbecco's Phosphate Buffered Saline* (1x, DPBS) unter leichtem Schwenken entfernt. Das DPBS wurde verworfen und die Zellen schließlich durch Zugabe von 1 ml vorgewärmten Trypsin-EDTA (0,25 %) und Inkubation für 2-4 min im Brutschrank vollständig abgelöst. Nach Aufnahme der abgelösten Zellen in vorgewärmtes Fibroblasten-Kulturmedium wurden diese zu gleichen Teilen auf neue Zellkulturflaschen bzw. -schalen aufgeteilt und die Kultivierung (s. Abschnitt 3.5.1) wurde fortgesetzt.

3.5.3 Herstellung und Auftauen kryokonservierter Fibroblasten

Zur Langzeit-Aufbewahrung von frühen Passagen der einzelnen Zelllinien wurden Kryostocks hergestellt. Die adhärenenten Zellen eines 75 cm² mit nahezu 100 % Konfluenz bewachsenen Zellkulturgefäßes (Flasche oder Schale) wurden hierfür entsprechend Abschnitt 3.5.2 abgelöst und in 10 ml vorgewärmtes Fibroblasten-Kulturmedium aufgenommen. Es erfolgte eine Zentrifugation bei 1000 rpm für 4 min (Centrifuge 5810 mit Rotor A-4-62, Eppendorf). Nach Absaugen des Überstandes wurde das Zellpellet in 2,16 ml vorgewärmtem Fibroblasten-Kulturmedium unter Zusatz von 600 µl FBS und 240 µl DMSO resuspendiert. Die Zellsuspension wurde in zwei Kryoröhrchen (TPP; je 1,5 ml) aufgeteilt. Die Kryoröhrchen wurden abschließend schnellstmöglich in den *CoolCell® LX Cell Freezing Container* (Corning) transferiert und mitsamt dem Behälter bei -80 °C eingefroren. Zur Langzeit-Aufbewahrung wurden die Kryoröhrchen

frühestens nach 24 h auf Trockeneis in den Stickstofftank transferiert und dort gelagert.

Zum Auftauen wurden die Kryoröhrchen aus dem Stickstofftank auf Trockeneis entnommen und anschließend für ca. 1 min in ein Wasserbad bei 37 °C gegeben. Nach Auftauen der Zellsuspension in den Kryoröhrchen wurde diese in 10 ml vorgewärmtes Medium überführt und für 8 min bei 1000 rpm zentrifugiert (Centrifuge 5810 mit Rotor A-4-62, Eppendorf). Der Überstand inklusive des in Zellkultur toxischen DMSO wurde abgenommen, die Zellen in 5-10 ml vorgewärmtes Medium (abhängig von der visuellen Größe des Zellpellets) aufgenommen und in ein 25 cm² oder 75 cm² Zellkulturgefäß transferiert. Nach 24 h erfolgte die Überprüfung der Zelldichte unter dem Mikroskop Eclipse TS100 (Nikon Instruments), sowie bei Bedarf ein Wechsel des Kulturmediums. Bedarfskriterien waren eine Verfärbung des Mediums (pH-Umschlag) oder eine erhöhte Dichte an nicht adhärenen (somit vermutlich toten) Zellen.

3.5.4 Ernte von Zellen

Zur Ernte von Fibroblasten wurden diese analog zu Abschnitt 3.5.2 mit Trypsin-EDTA abgelöst und in 9 ml vorgewärmtes Fibroblasten-Kulturmedium aufgenommen. Nach Zentrifugation bei 1000 g für 10 min (Centrifuge 5810) wurde der Überstand abgesaugt. Nach zwei Waschsritten mit je 10 ml vorgewärmter DPBS und ebenfalls Zentrifugation bei 1000 g für 10 min (Centrifuge 5810) wurden die Zellen eines Zellkulturgefäßes in 0,5-2 ml DPBS resuspendiert. Eine bzw. mehrere Zellsuspensionen derselben Zelllinie wurden in ein 1,5 ml oder 2 ml Reaktionsgefäß überführt und erneut für 5 min bei 1000 g zentrifugiert (Centrifuge 5418, Eppendorf). Nach Absaugen des Überstandes wurden die Zellpellets (i. d. R. zwei bis drei 75 cm² Zellkulturgefäße pro Pellet) bei -80 °C bis zur weiteren Verwendung für mind. eine Stunde eingefroren.

3.6 Qualitativer und semiquantitativer Nachweis der Kandidatengenexpression auf Proteinebene

Um für die Kandidatengene *DBR1* und *SEC31B* qualitative und semiquantitative Analysen auf Proteinebene durchführen zu können, wurde die Methode des Western Blots gewählt. Hierbei wurde der Gesamtproteingehalt aus Fibroblasten- Zellpellets (s. Abschnitt 3.5.4) extrahiert und mittels SDS-PAGE nach Molekulargewicht aufgetrennt (Laemmli, 1970). Schließlich konnten die Proteine durch Transfer auf eine Nitrocellulosemembran für die Markierung mittels Primärantikörper und nachfolgende Detektion mittels fluoreszenzmarkierter Sekundärantikörper zugänglich gemacht werden (Towbin et al., 1979).

In Vorexperimenten wurden verschiedene Optimierungsschritte getestet. Diese umfassten vor allem die Wahl des für die Proteinextraktion verwendeten Lysepuffers, die Art und die Konzentration der für die SDS-PAGE benötigten Polyacrylamidgele und die Art des verwendeten Blocking-Puffers. In den nachfolgenden Abschnitten wird lediglich die Methodik dargestellt, welche für die im Ergebnisteil (s. Kapitel 4) dargestellten Daten verwendet wurde.

3.6.1 Proteinextraktion aus Zellpellets

Die bei -80 °C gelagerten Fibroblasten-Zellpellets wurden zunächst für ca. 20 Min auf Eis aufgetaut. Alle weiteren Arbeitsschritte zur Proteinextraktion wurden, soweit möglich, ebenfalls auf Eis durchgeführt.

Abhängig von der visuell abgeschätzten Größe des Zellpellets wurden nach Auftauen der Zellpellets zwischen 10 und 100 µl Lysis- oder Extraktionslysepuffer unter Zusatz von 10x Proteaseinhibitor (Endkonzentration 1 x) hinzugegeben und bei 4 °C für 20 min unter leichtem Schütteln inkubiert. Die in den Zellpellets enthaltene DNA wurde mit dem Ultraschallhomogenisator UW 2200 (BANDELIN electronic GmbH & Co. KG; Parameter: 5 s, 3 Zyklen, Power: 40 %) geschert. Zertrümmerte DNA und enthaltene Zellreste wurden abschließend für 2 min bei 16000 g (Centrifuge 5417C, Eppendorf) abzentrifugiert. Die so erhaltenen

Überstände (Proteinlysate) wurden in den folgenden Arbeitsschritten weiterverwendet oder bis zur weiteren Analyse bei -80 °C gelagert.

Bei einigen Western Blots wurden Kontrolllysate aus anderen Zelltypen verwendet (HEK- und RPE-Zellen, sowie CPT-Blut). Diese zusätzlichen Kontrollen wurden freundlicherweise von Dr. rer. nat. Rebecca Buchert-Lo zur Verfügung gestellt. In diesem Fall entfiel die Proteinextraktion.

3.6.2 Bradford-Test

Um eine standardisierte Menge Proteingemisch aus den Lysaten für die nachfolgende Polyacrylamidgelelektrophorese vorbereiten zu können, erfolgte die Bestimmung der Proteinkonzentrationen der Lysate mittels Bradford-Test (Bradford, 1976). Hierfür war die Erstellung einer Standardkurve bei definierten Proteinkonzentrationen von bovinem Serumalbumin (BSA, *bovine serum albumin*) nötig. Das Pipettierschema, sowie die resultierenden Proteinkonzentrationen der BSA-Standardreihe können Tabelle 3.9 entnommen werden. Die Probenansätze wurden in einer Verdünnung von 1:500 entsprechend Tabelle 3.10 vorbereitet.

Tabelle 3.9: Pipettierschema der BSA-Standardreihe für den Bradford-Test

Bezeichnung	ddH ₂ O [µl]	BSA-Lösung [µl]	Bradford-Reagenz [µl]	Proteinkonzentration [µg/ml]
S1	240	0	60	0
S2	232,5	7,5	60	1,25
S3	225	15	60	2,5
S4	210	30	60	5
S5	195	45	60	7,5
S6	180	60	60	10
S7	150	90	60	15
S8	120	120	60	20

Anmerkung: Die BSA-Lösung lag in einer Konzentration von 0,05mg/ml vor. Bradford-Reagenz bezeichnet das Protein Assay Dye Reagent Concentrate (Bio-Rad Laboratories).

Tabelle 3.10: Pipettierschema der Probenansätze für den Bradford-Test

Bezeichnung	Volumen [µl]
Probe des zu messenden Proteinlysats	0,6
H ₂ O	239,4
Bradford-Reagenz	60

Anmerkung: Bradford-Reagenz bezeichnet das Protein Assay Dye Reagent Concentrate (Bio-Rad Laboratories).

Nach Erstellung aller Ansätze (Standardreihe und Probenansätze) wurden diese durch kurzes vortexen homogenisiert und für 5 min bei Raumtemperatur inkubiert. Die Ansätze wurden anschließend auf eine 96-Well Platte transferiert (je 100 µl, in Triplikaten) und die Extinktionswerte konnten mittels des Plattenlesers Synergy™ HT und der Software Gen5 (Version 2.01.14; beide Bio-Tek Instruments) bei einer Wellenlänge von 595 nm ermittelt werden.

Für die Berechnung der Proteinkonzentration der Proben wurde zunächst aus den ermittelten Werten des dekadischen Absorptionsmaßes der Standardreihe und der Proben jeweils ein Mittelwert (A_{595}) gebildet. Von diesen Mittelwerten wurde anschließend der Mittelwert der Blindprobe (Konzentration 0 µg/ml) abgezogen (A_{595}^*). Anschließend erfolgte auf Basis der berechneten Werte (A_{595}^*) für die Standardreihe die Ermittlung einer Standardkurve mittels linearem Regressionsmodell. Die Determinationskonstante R^2 ist hierbei ein Maß der Güte der Standardreihe. Ein Wert von $R^2 \geq 0,97$ wurde als ausreichend angesehen.

Die Proteinkonzentrationen der Proben wurden schließlich nach nachfolgender Formel berechnet:

$$\text{Proteinkonzentration} \left[\frac{\mu\text{g}}{\mu\text{l}} \right] = \frac{A_{595}^*}{m} * d$$

- A_{595}^* Mittlere dekadische Absorbanz (alt: Extinktion/Optische Dichte) der Proben nach Abzug des Mittelwertes der A_{595} der Blindprobe
- m Steigung der erstellten Standardkurve
- d Dilutionsfaktor, hier: 0,5 (Multiplikation mit dem Faktor 500 aufgrund der Probenverdünnung 1:500 und Division durch den Faktor 1000 aufgrund der Umrechnung von ml in µl)

3.6.3 SDS-Polyacrylamidgelelektrophorese

Mittels SDS-Polyacrylamidgelelektrophorese (SDS-PAGE) können die in den Lysaten (s. Abschnitt 3.6.1) enthaltenen Proteingemische nach ihrem Molekulargewicht aufgetrennt werden. Bei dieser Methode bindet das negativ geladene Detergenz SDS reversibel an die Peptidbindungen der denaturierten Proteine, sodass die Eigenladungen der Proteine vernachlässigbar werden und eine Abhängigkeit der Wanderungsgeschwindigkeit von der Molekulargröße erreicht wird (Laemmli, 1970).

Es erfolgte zunächst die Vorbereitung eines Bis-Tris Polyacrylamidgels in einem Gestell aus Glasplatten (Bio-Rad Laboratories, Dicke: 0,1 cm Breite: 8,3 cm, Trennstrecke: 7,3 cm). Als erster Schritt wurde zunächst eine 8 % oder 10 %ige Trenngellösung je nach gewünschter Trennschärfe anhand Tabelle 3.11 angesetzt und durch mehrmaliges Invertieren homogenisiert. Ca. 4,5 ml der Trenngellösung wurden in den Spalt zwischen den beiden Glasplatten (*short plate* und *spacer plate*) des Gestells gegossen. Es erfolgte eine Überschichtung mit Isopropanol (70 %), um eine gleichmäßige bzw. gerade Kante des Trenngels zu erreichen. Die Auspolymerisierung erfolgte für ca. 20 min bei Raumtemperatur.

Nach Verwerfen des Isopropanols wurde eine 6 %ige Sammelgellösung anhand Tabelle 3.12 angesetzt und ebenfalls durch mehrmaliges Invertieren homogenisiert. Ca. 1,5 ml der Sammelgellösung wurden auf das Trenngel gegossen. Ein entsprechender Taschenkamm wurde abschließend auf das Glasgestell und in das Sammelgel gesteckt. Die Auspolymerisierung erfolgte auch hier für ca. 20 min bei Raumtemperatur.

Tabelle 3.11: Ansätze für eine 8 % oder 10 %ige Trenngellösung

Bezeichnung	8 % Trenngel	10 % Trenngel
	Volumen	Volumen
3,5x Bis-Tris Puffer	1,42 ml	1,42 ml
Acrylamid (30 %)	1,34 ml	1,67 ml
ddH ₂ O	2,24 ml	1,91 ml
APS (10 %)	30 µl	30 µl
TEMED	4 µl	4 µl

Tabelle 3.12: Ansatz für eine 6 %ige Sammelgellösung

Bezeichnung	Volumen [µl]
3,5x Bis-Tris Puffer	1420
Acrylamid (30 %)	1000
ddH ₂ O	2580
APS (10 %)	30
TEMED	4

Bei Verwendung eines vorgefertigten Invitrogen™ Bolt™ 4-12 % Bis-Tris Plus Gels (ThermoFisher Scientific) entfiel die Vorbereitung eines handgefertigten Gels.

Vor Durchführung der Gelelektrophorese war eine Probenvorbereitung zur Denaturierung der in den Lysaten enthaltenen Proteine durch das Detergenz LDS und das reduzierende Thiol DTT nötig. Die folgenden Schritte zur Probenvorbereitung der Lysate wurden, sofern nicht anders spezifiziert, auf Eis durchgeführt.

Hierzu wurden die Volumina der Lysate, welche eine vorher definierte Proteinmenge (je nach Fragestellung 15-50 µg) enthielten, einem Ansatz mit 4x LD-Puffer (Endkonzentration: 1x), 1M DTT-Lösung (Endkonzentration: 0,1M) und ddH₂O zugeführt. Das jeweils verwendete Gesamtvolumen der Ansätze richtete sich nach der Taschengröße des

Polyacrylamidgels (max. 20 µl für 15-Well Gele und max. 30 µl für 10-Well Gele). Alle Ansätze wurden durch Auf- und Abpipettieren homogenisiert, für 10 min bei 90 °C im ThermoStat Plus (Eppendorf) inkubiert und anschließend unverzüglich auf Eis gestellt.

Nach Abschluss der Probenvorbereitung wurde das gegossene Polyacrylamidgel bzw. das vorgefertigte Gel in eine mit 1x MOPS-Puffer gefüllte Elektrophoresekammer (Mini-PROTEAN® Tetra Vertical Electrophoresis Cell, Bio-Rad Laboratories bzw. SE 300 miniVE Integrated Vertical Electrophoresis and Blotting Unit, Hoefer) überführt und anschließend der Taschenkamm vorsichtig entfernt. Um Gelreste aus den Taschen zu entfernen, wurden diese anschließend vorsichtig mit 1x MOPS-Puffer gespült. Die Beladung der einzelnen Taschen erfolgte jeweils mit dem kompletten Volumen eines der Probenansätze, sowie 5 µl unterschiedlicher Proteingrößenstandards (Precision Plus Protein™ Dual Color Standards, Bio-Rad Laboratories und Invitrogen™ SeeBlue™ Pre-stained Protein Standard, Thermo Fisher Scientific). Unbeladene Taschen wurden mit dem für die Probenvorbereitung pro Probe verwendeten Volumen an 4x LD-Puffer befüllt, um eine ungleichmäßige Lauffront zu verhindern.

Nach dem Beladen aller Taschen wurde die Gelelektrophorese durch die Anlage einer Spannung von 80 V gestartet. Sobald die Lauffront sich am unteren Ende des Sammelgels befand, wurde die Spannung von 80 auf 120 V erhöht. Bis die Lauffront das untere Ende des Gels fast erreicht hatte, wurde die Elektrophorese bei dieser Spannung durchgeführt. Sie wurde schließlich durch Entfernen der Spannung beendet.

3.6.4 Western Blot

Um die im Rahmen der SDS-PAGE nach Molekülgröße aufgetrennten Proteine für eine Immundetektion durch fluoreszierende Antikörper zugänglich zu machen, wurden die Proteine zunächst auf eine

Nitrocellulosemembran überführt. Dieses erfolgte mit der Methode des Western Blots.

Zunächst wurde eine 0,2 µm starke Nitrocellulosemembran (Amersham™ Protran® Premium 0,2 µm NC (Nitrocellulose Blotting Membran), GE Healthcare Life Science) auf die Größe des Trenngels zugeschnitten und anschließend für ca. 10 Min in Transfer-Puffer (1x; Zusatz von 15 % Methanol zur Aktivierung der Membran) inkubiert.

Anschließend erfolgte die Vorbereitung der Blotting-Kassette (Hoefer). Ein bereits mit Transfer-Puffer durchtränkter Schwamm (Hoefer), gefolgt von einem durchtränkten Filterpapier (0,8 mm stark, biostep) und der Nitrocellulosemembran wurden auf der einen Seite der Kassette übereinander geschichtet. Nach Abtrennen des Sammelgels wurde das Trenngel vorsichtig aus dem Glasgestell herausgelöst und auf die Nitrocellulosemembran gelegt. Falls sich Luftblasen zwischen Nitrocellulosemembran und Gel gebildet hatten, wurden diese durch vorsichtige Applikation von Transfer-Puffer entfernt. Auf das Gel wurden ebenfalls ein mit Transfer-Puffer durchtränktes Filterpapier und ein durchtränkter Schwamm geschichtet. Abbildung 3.1 zeigt den Aufbau der Blotting-Kassette schematisch.

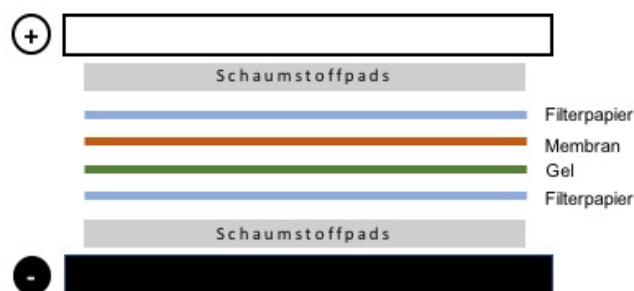


Abbildung 3.1: Aufbau einer Blotting-Kassette

Die Abbildung zeigt den Aufbau einer vorbereiteten Blotting-Kassette und die benötigten Komponenten - Schwamm bzw. Schaumstoffpads, Filterpapiere, Polyacrylamidgel (Gel) und Nitrocellulosemembran (Membran) - sowie die korrekte Orientierung an Kathode (Pluszeichen) und Anode (Minuszeichen).

Nach Einführen der Kassette mit korrekter Orientierung (Nitrocellulosemembran auf Seite der Kathode, s. Abbildung 3.1) in den mit 1x Transfer-Puffer gefüllten Transfertank (TE 22 Mighty Small Transfertank,

Hoefer) wurde der Proteintransfer mit einer Spannung von 80 V je nach Fragestellung für 1,5 - 2 h bei 4 °C unter leichtem Rühren durchgeführt.

Der Erfolg und die Qualität des Proteintransfers wurden vor der Durchführung einer Immundetektion mittels einer reversiblen Ponceau-Färbung kontrolliert. Hierfür wurde die Membran aus der Blotting-Kassette entnommen, in eine Schale mit ddH₂O überführt und unter leichtem Schwenken von eventuell vorhandenen Gelresten gereinigt. Anschließend erfolgte eine Überschichtung mit *Ponceau S staining solution* und eine Inkubation für ca. 5 Min. Die Membran wurde schließlich erneut in eine Schale mit ddH₂O gelegt, um ungebundenen Ponceau S-Farbstoff zu entfernen. Bei der nachfolgenden Qualitätsbeurteilung wurden die Gleichmäßigkeit der nebeneinander vorhandenen Spuren, sowie eventuell vorhandene grobe Unregelmäßigkeiten, beispielsweise durch unvollständigen Transfer bei vorhandenen Luftblasen, bewertet.

Die Membran wurde abschließend auf ihre endgültige Größe zugeschnitten, indem alle Teile ohne sichtbare Ponceau-Proteinfärbung entfernt wurden.

3.6.5 Immundetektion

Nach als erfolgreich beurteiltem Proteintransfer wurde eine Immundetektion mit spezifischen Antikörpern für die zu untersuchenden Zielproteine angeschlossen. Hierbei wurden die Zielproteine zunächst mit einem Primärantikörper markiert, welcher nachfolgend durch einen Sekundärantikörper mit Fluoreszenzmarkierung sichtbar gemacht werden konnte.

Die Membran wurde zunächst in ein 50 ml Probenröhrchen (Greiner) überführt.

Im nächsten Schritt erfolgte die Absättigung der vorhandenen freien Bindungsstellen der Membran durch Inkubation mit 25 ml Blocking-Puffer (5 %, magermilchbasiert) für 1 h bei Raumtemperatur unter leichter Rotation. Dieser Schritt war nötig, um unspezifische Bindungen der nachfolgend verwendeten Antikörper zu verhindern.

Nach kurzem Waschen für ca. 5 min in 10 ml 1x *tris buffered saline with tween* (TBS-T) zur Entfernung von Resten der Magermilchlösung unter leichter Rotation, erfolgte die Inkubation mit 5 ml eines Gemisches aus Primärantikörpern (Antikörper gegen das Zielprotein und Antikörper gegen das Haushaltsgen *GAPDH*) in 1x TBS-T (Zusatz von Natriumacid 2 % in einer Verdünnung 1:100). Die eingesetzte Verdünnung des Primärantikörpers gegen das Haushaltsgen *GAPDH* lag bei allen durchgeführten Immundetektionen bei 1:2500. Die eingesetzten Verdünnungen der Primärantikörper gegen die zu analysierenden Zielproteine variierten je nach Fragestellung von 1:500 – 1:2500. Tabelle 3.13 zeigt beispielhaft den Ansatz eines Primärantikörper-Gemisches für das Zielprotein *DBR1* und einer Verdünnung des entsprechenden Antikörpers auf 1:1000. Die Inkubation mit den Primärantikörpern wurde über Nacht (ca. 14-16 h) bei 4 °C unter leichter Rotation durchgeführt.

Tabelle 3.13: Beispiellansatz eines Primärantikörper-Gemisches

Bezeichnung	Volumen
1x TBS-T	5 ml
Mouse-anti-GAPDH (Santa Cruz), 1:2500	2 µl
Rabbit-anti-DBR1 (Proteintech), 1:1000	5 µl
Natriumacid (NaN ₃), 1:100	50 µl

Anmerkung: Natriumazid dient hier als antimikrobieller Zusatz, um das Primärantikörper-Gemisch haltbarer und wiederverwendbar zu machen.

Am Folgetag wurde die Lösung aus Primärantikörpern zunächst aus dem Probengefäß mit der Membran entfernt und die Membran drei- bis viermal mit jeweils 10 ml 1x TBS-T für 5-10 min bei Raumtemperatur unter leichter Rotation gewaschen.

Parallel hierzu erfolgte der Ansatz einer Lösung mit Sekundärantikörpern (fluoreszenzbasiert), dessen Volumina und Verdünnungen Tabelle 3.14 entnommen werden können. Alle Arbeitsschritte mit fluoreszenzbasierten Antikörpern erfolgten unter Lichtausschluss.

Tabelle 3.14: Ansatz für die Lösung mit Sekundärantikörpern

Bezeichnung	Volumen	Verdünnung
1x TBS-T	5 ml	-
IRDye® 680RD Goat anti-Rabbit (LI-COR®)	0,5 µl	1:10000
IRDye® 800CW Goat anti-Mouse (LI-COR®)	0,5 µl	1:10000

Nach erfolgten Waschschriften wurde die Membran mit dem Sekundärantikörpergemisch (5 ml) für eine Stunde bei Raumtemperatur unter leichter Rotation inkubiert. Die Lösung mit Sekundärantikörpern wurde verworfen. Nach erneutem drei- bis viermaligem Waschen mit 10 ml 1x TBS-T für jeweils 5-10 min bei Raumtemperatur unter leichter Rotation, konnten die gebundenen Sekundärantikörper (fluoreszenzbasiert) mit dem LI-COR® Odyssey® Fc Imaging System und der Software Image Studio™ (Version 2.1.10; beide LI-COR® Biosciences) detektiert werden. Tabelle 3.15 zeigt die Detektionswellenlängen und die Belichtungszeiten.

Tabelle 3.15: Detektionswellenlängen und Belichtungszeiten

Detektionswellenlänge	Belichtungszeit
600 nm	0,5 min
700 nm	10 min
800 nm	10 min

3.6.6 Stripping und Reprobing

Um eine erneute Detektion derselben Nitrocellulosemembran mit einem anderen als den zunächst verwendeten Primärantikörper durchzuführen, wurde die entsprechende Membran gestrippt und erneut mit Primär- und Sekundärantikörpern inkubiert. Hierfür wurde die Membran zunächst zweimal für 10 min mit 10 ml Mild Stripping Buffer (pH 2,2) unter Rollen inkubiert und anschließend zweimal mit 10 ml DPBS für 10 min unter Rollen gewaschen. Nach erneutem zweimaligem Waschen mit 10 ml 1x TBS-T für 5 min, konnte die Membran erneut für 1 h blockiert werden (magermilchbasiert, s. Abschnitt 3.6.5). Die Inkubation mit dem

Primärantikörpergemisch, die Waschschritte und die Inkubation mit dem Sekundärantikörpergemisch, sowie die Detektion der Nitrocellulosemembran erfolgten anschließend analog Abschnitt 3.6.5.

Anmerkung: Das Stripping und Reprobing von Membranen erfolgte lediglich für Membranen bei denen zwei Primärantikörper zur Markierung des Zielproteins zur Verfügung standen. Dies war nur für das Zielprotein SEC31B der Fall (vorliegende Antikörper von abcam und Aviva Systems Biology Corporation).

3.6.7 Messung der Signalintensitäten der detektierten Banden

Um semiquantitative Berechnungen der Expression auf Proteinebene für das Kandidatengen *DBR1* durchführen zu können, wurden die detektierten Signalintensitäten der Banden für das Zielprotein DBR1 und für das Haushaltsprotein GAPDH gemessen. Hierfür wurde die Software Image Studio™ (Version 2.1.10, LI-COR® Biosciences) verwendet.

Als Erstes wurden die Fluoreszenzbanden der gegen die Ladekontrolle (Haushaltsprotein GAPDH) gerichteten Antikörper im 700 nm-Kanal in den einzelnen Spuren der Proben bei ca. 35 kDa (erwartetes Molekulargewicht von GAPDH) anhand des Proteingrößenmarkers aufgesucht und entsprechend den Herstellerangaben mit je einem Messrechteck umfahren. Mittels der Profil-Funktion der Software wurde anschließend eine Feinjustierung der Messrechtecke vorgenommen, sodass die komplette und ausschließliche Erfassung der Signale der Banden von Interesse sichergestellt wurde.

Analog zu der Messung der Signalintensitäten für die Banden bei 35 kDa (Haushaltsprotein GAPDH) wurde ebenfalls für die Fluoreszenzbanden des gegen das Zielprotein DBR1 gerichteten Antikörpers bei ca. 62 kDa (erwartete Molekulargröße von DBR1) im 800 nm-Kanal verfahren.

Abschließend wurde für beide Kanäle eine manuelle Hintergrundsubtraktion durch Positionierung eines Messrechteck über die gesamte Breite der Membran (ohne Erfassung der Spuren des Proteingrößenmarkers) im Bereich nahe der gemessenen

Fluoreszenzbanden (ohne Erfassung von unspezifischen Banden) festgelegt.

3.6.8 Statistische Auswertung der Signalintensitäten

Im Anschluss an die Hintergrundsubtraktion wurden die resultierenden Signalintensitäten statistisch ausgewertet. Dazu wurde zunächst die Signalintensität für das Zielprotein (DBR1) der einzelnen Spuren auf die Signalintensität der jeweiligen Ladekontrolle (GADPH) normalisiert. Anschließend wurden die relativen Signalintensitäten für die Patientenproben und Kontrollen berechnet. Hierfür erfolgte die Kalkulation des Mittelwerts der jeweiligen Proben der unterschiedlichen Zelllinien bezogen auf die Membranen. Die normalisierten Signalintensitäten wurden schließlich durch den Zelllinien-spezifischen Mittelwert der Membran geteilt. Die Berechnung der relativen Signalintensitäten war erforderlich, um die unterschiedlichen Membranen miteinander zu vergleichen.

Nach den beschriebenen Berechnungen wurden die Werte der einzelnen Gruppen (Proben der Zelllinien mit vorhandener Mutation im Gen *DBR1* und die Proben der Kontrollzelllinien) mittels der Software JMP® (Version 13.0.0, SAS Institute) auf Normalverteilung und Varianzgleichheit (Standardtests der Software JMP®: Shapiro-Wilk-Test und Levene-Test, jeweils mit $\alpha = 0,05$) überprüft. Trotz kleinem n ($n = 6$) in den einzelnen Gruppen wurde eine Normalverteilung und eine Varianzgleichheit angenommen, da bei einem nichtparametrischen Test die Power des anschließendem Signifikanztest zu gering wäre und bei steigender Zahl der durchgeführten Western Blots sich unter methodisch einwandfreien Umständen eine Normalverteilung ergeben müsste. Daher wurden Zweistichproben t-Tests für unabhängige Stichproben ($\alpha = 0,05$) für jedes Paar Patient und Kontrolle, sowie für die Kontrollen untereinander (Gesamtzahl: 6) mittels der Software JMP® durchgeführt. Die Bonferroni-Korrektur wurde zur Adjustierung des α -Niveaus bei multiplem Testen angewendet.

4 Ergebnisse

4.1 DNA- und RNA-Sequenzierungen

Bei einem Untersuchungskollektiv von 16 Individuen und teilweise deren gesunden Eltern, sowie weiteren betroffenen Geschwistern wurden umfangreiche genetische Untersuchungen durchgeführt. Sequenzierungen des gesamten Genoms (TruSeq DNA PCR-Free Kit for Illumina®, New England Biolabs) oder des proteinkodierenden Genoms (Exom; SureSelectXT Human All Exon V6 Kit, Agilent Technologies), sowie der Gesamtheit der mRNA aus Fibroblasten (RNA-Seq; NEBNext® Ultra™ II RNA Library Prep Kit for Illumina®, New England Biolabs) wurden auf der NovaSeq 6000-Plattform (Illumina) durch die Core Facility des NGS Competence Center Tübingen (Leitung: Dr. Nicolas Casadei) durchgeführt. Nachfolgend wurde eine bioinformatische Aufarbeitung der Rohdaten durch standardisierte institutsinterne Pipelines (s. Abschnitt 3.2) vollzogen.

In Vorarbeit von Dr. rer. nat. Rebecca Buchert-Lo und Dr. med. Lucia Laugwitz, sowie weiteren MitarbeiterInnen des Instituts für Angewandte Genetik und Medizinische Genomik Tübingen erfolgte eine Auswertung der Exom- und Genomsequenzierungen im diagnostischen Kontext. Hierdurch wurde bei einigen PatientInnen bereits die Priorisierung eines oder mehrerer sog. Kandidatengene (Klasse 3-Variante nach ACMG, vgl. Abschnitt 3.3.3) vorgenommen. Ferner wurde durch die klinisch tätigen KollegInnen und einweisenden ÄrztInnen bei einigen PatientInnen bereits anhand des Phänotyps (s. Tabelle 1.1 und Tabelle 1.2) ein Verdacht bezüglich eines Defektes in einem spezifischen Gen oder einer Gruppe von Genen ausgesprochen. Diese Verdachtsmomente und Priorisierungen wurden durch eine erneute Auswertung im Rahmen dieser Arbeit überprüft und mittels der zusätzlichen Ebene der RNA-Sequenzierungen ergänzt. Zudem wurden ebenfalls PatientInnen untersucht, bei welchen bisher keine Verdachtsmomente bezüglich eines oder mehrerer Kandidatengene bestand.

Nach Durchlaufen der bioinformatischen Pipeline inklusive Qualitätskontrollen (s. Abschnitt 3.2) wurden in den einzelnen Proben für die Genomsequenzierungen der Betroffenen ca. 300.000 – 500.000 Varianten ausgegeben. Bei den Proben für die Exomsequenzierungen ergaben sich Gesamtzahlen von ca. 65.000 – 90.000 Varianten pro Individuum.

Diese Varianten wurden mithilfe standardisierter Filterungsschritte (s. Abschnitt 3.3.2) reduziert, um eine Auswertung zu ermöglichen. In einem ersten Schritt wurden seltene Varianten isoliert, indem solche mit einer Allelfrequenz (*minor allele frequency*, MAF) größer als 0,1 % in der *Genome Aggregation Database* (gnomAD) ausgeblendet wurden.

Anschließend erfolgte eine Filterung nach weiteren Kriterien (s. Abschnitt 3.3.2). Durch Durchsicht der einzelnen Varianten im *Integrative Genomics Viewer* (IGV, Version 2.4.19) konnten die Varianten auf Qualität überprüft und unter anderem technische Artefakte identifiziert werden.

Tabelle 4.1 gibt einen Überblick über die Art der Sequenzierung (WGS oder WES) und die Anzahl der nach den Filterschritten verbleibenden Varianten für die einzelnen Individuen.

Bei den meisten Individuen wurden sog. Trio-Analysen (Eltern-Kind-Analyse) oder Multi-Analysen mit anderen Betroffenen aus der Familie durchgeführt (s. Abschnitt 3.2.4). Die Art der Analysen und die verbleibenden Varianten nach den entsprechenden Filterschritten (s. Abschnitt 3.3.2) können in Tabelle 4.2 und Tabelle 4.3 entnommen werden.

Tabelle 4.1: Anzahl der verbleibenden Varianten nach den beschriebenen Filterungsschritten in Proben der einzelnen Individuen

Individuum	Art	Familie	Gesamt	Allele-frequency based	Recessive stringent	Recessive relaxed	Dominant stringent	Dominant relaxed	Homozygot/ comp.-het. trunkierend
MS001	WES	-	66996	336	12	28	8	141	1
MS002	WGS	-	357787	879	42	179	7	235	3
MS003	WGS	-	305414	386	27	65	14	138	3
MS005	WES	-	67765	321	6	31	13	130	0
MS006	WGS	1	351919	594	76	201	20	208	2
MS007	WGS	1	369733	586	58	145	21	202	2
MS008	WES	-	67763	568	32	90	13	223	1
MS009	WES	-	67170	373	14	45	18	182	0
MS010	WGS	2	363014	533	28	102	19	178	2
MS011	WGS	2	371481	509	25	123	16	153	2
MS012	WGS	3	299142	369	11	52	13	133	0
MS013	WGS	-	337778	379	26	67	17	139	4
MS014	WES	4	66723	419	19	53	17	188	0
MS015	WES	4	66507	398	22	50	15	167	0
MS016	WGS	5	340304	434	26	93	16	166	7
MS017	WGS	5	300321	435	25	94	17	168	7

Tabelle 4.2: Anzahl der verbleibenden Varianten nach den beschriebenen Filterungsschritten für die durchgeführten Multi-Analysen

Familie	Individuen	Art	Allele-frequency based multi	Multi-sample recessive	Multi-sample comp.-het.	Multi-sample dominant	Bemerkung
1	MS006 MS007	WGS	962	30	13	228	Dritter Bruder nicht eingeschlossen (kein WGS vorhanden)
2	MS010 MS011	WGS	896	13	29	244	
3	MS012	WGS	682	4	8	161	Zweiter Bruder nicht eingeschlossen
4	MS014 MS015	WES	603	2	12	152	
5	MS016 MS017	WGS	615	30	38	370	

Tabelle 4.3: Anzahl der verbleibenden Varianten nach den beschriebenen Filterungsschritten für die durchgeführten Trio-Analysen

Individuum	Art	Allele-frequency based trio	Trio stringent	Trio relaxed
MS005	WES	892	10	73
MS008	WES	1215	17	104
MS009	WES	945	12	95

Mithilfe zu den WGS- oder WES-Tracks parallel im IGV angezeigten RNA-Seq-Daten der betroffenen Individuen und jeweils drei zufällig ausgewählten Kontrollen konnten die entsprechenden genomischen Varianten direkt auf ihre Expression und eventuelle Bedeutung für die RNA-Synthese unter anderem auch durch Überprüfen des Spleißens mittels Sashimi-Plot beurteilt werden. Aus diesem Grund konnten eventuell priorisierte Klasse-3-Varianten verworfen oder der Verdacht erhärtet werden. Tabelle 4.4 fasst die Kandidatengene, Gründe für deren Auswahl und die resultierenden Erkenntnisse aus der Analyse der RNA-Seq-Daten zusammen.

Bei einem Großteil der Individuen (10 von 16 [~63 %]) ergaben sich aus dem klinischen Phänotyp (für einen Gesamtüberblick s. Tabelle 1.1 und Tabelle 1.2 in Abschnitt 1.6) in Kombination mit der ersten diagnostischen Auswertung der WGS/WES bereits Verdachtsmomente bezüglich eines oder mehrerer Kandidatengene. Bei 40 % (4 von 10) dieser PatientInnen ergaben sich durch die RNA-Seq Zusatzbefunde, die den Verdacht bezüglich der bereits entdeckten Klasse 3-Varianten erhärteten bzw. in einem Fall zur Klassifizierung einer wahrscheinlich pathogenen Variante führten. Bei zusätzlich drei Individuen (3 von 16 [~18 %]) führte die erneute Auswertung der Genom- bzw. Exom-Daten mit der zusätzlichen Ebene des Transkriptoms zur Entdeckung weiterer Kandidatengene (s. Tabelle 4.4). Insgesamt ergibt sich somit ein Zugewinn an unter Umständen diagnostisch wertvollen Information bei einem Großteil der hier vorgestellten Kohorte (7 von 16 [~44 %]) durch die Sequenzierung der mRNA. Eine Variante im

Gen *HUWE1* konnte zudem als wahrscheinlich pathogen (Klasse 4 nach ACMG) klassifiziert werden.

Daher wurden für vielversprechende Kandidatengene (Individuen in Tabelle 4.4 zusätzlich grau markiert) weitere molekulargenetische Analysen im hausinternen Forschungslabor durchgeführt. Die Phänotypen der entsprechenden Individuen, die identifizierten Varianten und die weiteren Analysen, sowie die abschließende Klassifikation der Mutationen sollen im nachfolgenden Ergebnisteil ausführlich dargestellt werden.

Tabelle 4.4: Übersicht über Kandidatengene, Auswahlgründe und Erkenntnisse aus RNA-Seq-Daten

Familie	Individuum	Kandidatengen vor RNA-Seq	Zusatzbefunde durch RNA-Seq	Kandidatengen nach RNA-Seq	Auswahlgrund	bekanntes Krankheitsgen
	MS001	✓ <i>SEC31B</i>	×	<i>SEC31B</i>	Trunkierende Variante	×
	MS002	✓ <i>DBR1</i>	×	<i>DBR1</i>	Phänotyp	✓ virale Hirnstamminfektionen (OMIM #619441)
	MS003	×	×	×		×
	MS005	✓ <i>HUWE1</i>	✓	<i>HUWE1</i>	X-Inaktivierung gestört Expression herabgesetzt	✓ X-chromosomale Form geistiger Behinderung (OMIM #309590)
1	MS006	×	✓	<i>OSCP1</i>	Spleiß-Defekt + Literaturrecherche	×
	MS007					
	MS008	✓ <i>SQRDL, GANC</i>	×	<i>SQRDL, GANC</i> ¹		¹ s. unten
	MS009	✓ <i>FASTKD3</i>	✓	<i>FASTKD3</i>	präferentielle Expression Missense-Variante; comp.- het. frameshift-Variante	×
2	MS010	×	×	×		×
	MS011					
3	MS012	×	✓	<i>NSUN4</i>	Spleiß-Defekt	×
	MS013	✓ <i>EIF4EBP2</i>	×	<i>EIF4EBP2</i>	Pathway + Phänotyp	×
4	MS014	✓ <i>SDHB</i>	×	<i>SDHB</i>		×
	MS015					
5	MS016	✓ <i>GANC</i>	✓	<i>GANC</i>	Präferentielle Expression Missense-Variante; comp.-het. -Variante	×
	MS017					

Anmerkung: Die graue Markierung kennzeichnet die Individuen für welche tlw. weitere Untersuchungen im Forschungslabor durchgeführt wurden bzw. welche in den folgenden Abschnitten näher beschrieben werden. ¹Durch anderweitige Analysen ohne RNA-seq wurde SQRDL in diesem Fall als pathogen beschrieben (s. Friederich et al., 2020).

4.2 Identifizierung einer Mutation in *DBR1* bei Individuum MS002

4.2.1 MS002: Phänotyp

Individuum MS002 wurde als Kind von Cousins 1. Grades in Neuseeland geboren. Die Eltern sind Angehörige der indigenen Bevölkerung Neuseelands (Māori). Ein Stammbaum lag zum Zeitpunkt der Auswertung dieser Arbeit nicht vor.

Die Schwangerschaft und die Geburt verliefen unauffällig. Das Mädchen wurde zum Termin reif geboren und zeigte Entwicklungsparameter (Größe und Gewicht) im altersgerechten normalen Perzentilenbereich. Es wurde sehr krauses Haar als einzige mögliche syndromale Auffälligkeit beschrieben.

Im Alter von drei Monaten wurde Betroffene MS002 aufgrund von generalisierten tonisch-klonischen Anfällen und komatösem Zustand nach einer milden febrilen Infektion des oberen Respirationstraktes in eine Klinik eingewiesen. Durchgeführte MRT-Aufnahmen zeigten deutliche Schwellungen beidseits im Bereich der Basalganglien und des Hirnstamms (s. Abbildung 4.1). Eine Involvierung des Kleinhirns konnte ebenfalls nachgewiesen werden. Virologische Untersuchungen ergaben einen positiven Test auf das Respiratorische Synzytial-Virus (RSV). Eine Liquoruntersuchung wurde nicht durchgeführt. Blutuntersuchungen ergaben zusätzlich eine Konstellation passend zu einer mitochondrialen Dysfunktion.

Die Diagnose einer akuten nekrotisierenden Enzephalopathie des Kindesalters (*acute necrotizing encephalopathy of childhood*, ANEC; Mizuguchi et al., 1995, Mizuguchi, 1997) wurde gestellt. Individuum MS002 verstarb im Alter von vier Monaten.

Nebenbefundlich wurde ein bilateral vorliegendes Glaukom diagnostiziert.

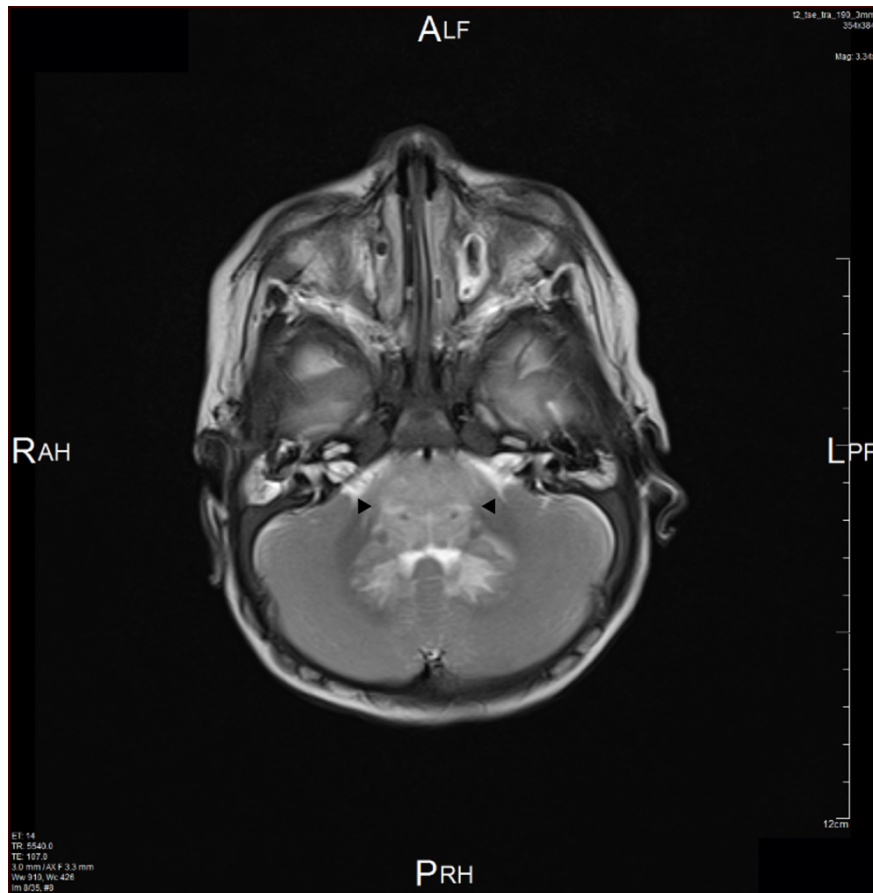


Abbildung 4.1: Exemplarisches MRT im Axialschnitt (T2-gewichtet) von MS002 auf Höhe der Pons/Abgang VIII. Hirnnerv

Die Pfeile markieren den deutlich hyperintens dargestellten Hirnstamm, welches ein typisches radiologisches Merkmal der ANEC darstellt.

4.2.2 Auswertung der DNA- und RNA-Sequenzierungen

Bei Individuum MS002 wurde eine Exom-Sequenzierung in Kombination mit der Sequenzierung der Eltern (Trio-Exom), eine Genom-Sequenzierung und eine RNA-Sequenzierung durchgeführt. Die Daten ohne Trio-Analyse ergaben 879 Varianten mit einer MAF <0,1 % und 179 Varianten für einen potenziell rezessiven Vererbungsmodus. Außerdem wurden 235 Varianten für einen potenziell dominanten Vererbungsweg identifiziert.

Diese Anzahl an Varianten konnte durch die Analyse des Trio-Exoms eingegrenzt werden. Es ergaben sich 33 homozygote Varianten, 41 compound-heterozygote Varianten und keine *de novo* Varianten.

Nach Überprüfen aller Varianten auf technische Artefakte wurden anhand des Phänotyps zwei potenziell interessante Missense-Varianten in homozygotem Status identifiziert (s. Tabelle 4.5).

Tabelle 4.5: Potenzielle Kandidatengene für Individuum MS002

Gen	Art	Zygotität	Position
<i>DBR1</i>	Missense	homozygot	ENST00000260803: c.461G>A, p.Arg>154Lys
<i>PAICS</i>	Missense	homozygot	ENST00000264221: c.665G>C, p.Arg222Pro

Das Gen *PAICS* kodiert für die Phosphoribosylaminoimidazol Carboxylase, ein bifunktionales Enzym, welches mehrere Schritte der *de novo* Purinbiosynthese katalysiert. Defekte in anderen Genen der Purinbiosynthese stellen bereits etablierte Krankheitsentitäten zu einem weiten phänotypischen Spektrum, inkl. verschiedener neurologischer Diagnosen dar (Dewulf et al., 2022). Die Variante ENST00000264221: c.[665G>C];[665G>C], p.[Arg222Pro];[Arg222Pro] erschien aus diesem Grund in Bezug auf Individuum MS002 zunächst interessant. Die Variante ist aktuell zweimal in der gnomAD ausschließlich im heterozygoten Status gelistet (Stand: 09.10.2025). Der PhyloP-Score steht mit 8,01 für eine hochkonservierte Stelle im Genom. *In silico*-Analysen sagen divergente Ergebnisse auf die Proteinfunktion voraus – so ist die Variante nach PolyPhen-2 benigne klassifiziert, während fathmm-MKL- und CADD-Score

einen potenziell schädlichen Effekt vorhersagen. Im gnomAD-Browser steht ein o/e-Score von 0,92 (90 % KI = 0,82 – 1,04) und ein Z-Score von 0,42 für einen niedrigen Selektionsdruck im Gen *PAICS* gegenüber Missense-Varianten. Allerdings finden sich lediglich drei ProbandInnen, welche eine extrem seltene Missense-Variante (MAF <0,01 %) in homozygotem Status aufweisen.

Die Variante hat keinen erkennbaren Effekt auf die RNA-Seq aus Fibroblasten.

Das Gen *PAICS* wurde aus phänotypischen Gründen als Kandidatengen für MS002 verworfen. Kürzlich publizierten Pelet et al., 2019 eine Familie mit zwei Fällen von homozygoten Missense-Mutationen, welche zu einer *PAICS*-Defizienz führten. Der beschriebene Phänotyp bestand in schweren mehrere Organsysteme betreffenden Fehlbildungen, welche bei beiden Betroffenen zum Tod innerhalb weniger Tage führten. Dies trifft keinesfalls den Phänotyp von MS002, welche als zunächst gesundes Kind ohne syndromale Auffälligkeiten geboren wurde.

Durch Literaturrecherche konnte die homozygot vorliegende Missense-Mutation im Gen *DBR1* (ENST00000260803: c.[461G>A];[461G>A], p.[Arg>154Lys];[Arg>154Lys]) favorisiert werden (in Tabelle 4.5 grau hinterlegt). Die Variante findet sich im gnomAD-Browser insgesamt drei Mal ausschließlich in heterozygoter Form (Stand: 09.10.2025) und scheint somit extrem selten zu sein. Das Nukleotid wird mit einem PhyloP-Score von 7,43 als hochkonserviert vorhergesagt.

DBR1 ist ein acht Exons umfassendes Gen, welches für das *RNA lariat debranching enzyme* kodiert. Das Enzym sorgt für die Konversion von Intron-Lassos in lineare Moleküle durch eine Hydrolyse der 2'-5'-Phosphodiesterbindungen und ermöglicht damit deren rapiden Umsatz (Mohanta und Chakrabarti, 2021).

Die Variante von Individuum MS002 liegt am Übergang zweier funktioneller Domänen, der Kerndomäne (Metallophosphoesterase) und des Lariat-Erkennungsloops (Zhang et al., 2018). *In silico*-Tools (SIFT, PolyPhen-2,

fathmm-MKL und CADD) sagen einheitlich einen voraussichtlich schädigenden Effekt der Variante auf das Protein voraus.

Zhang et al., 2018 verdeutlichen in ihrer Publikation, dass *DBR1* ein hochkonserviertes Gen mit Intoleranz gegenüber schädlichen Varianten ist. Ein o/e-Score von 0,87 (90 % KI = 0,81 – 0,94) für Missense-Varianten in der gnomAD spricht zwar für eine höhere Toleranz des Gens gegenüber Missense-Varianten (v. a. im heterozygoten Zustand). Aktuell finden sich jedoch nur sechs homozygot ausgeprägte Missense-SNPs mit einer MAF <0,01 %, wovon keinem ebenfalls in allen *In silico*-Tools ein schädlicher Effekt vorausgesagt wird. In der Datenbank finden sich bis dato keine trunkierenden homozygoten Varianten – dies stützt die These, dass eine komplette *DBR1*-Defizienz vermutlich pränatal tödlich ist (bisher keine Berichte von lebenden *DBR1*^{-/-}-Mäusen, vgl. Zheng et al., 2015).

Auf Ebene des Transkriptoms ist c.461G>A bei MS002 homozygot exprimiert, hat jedoch keinen erkennbaren Effekt auf die mRNA – weder auf die Expression im Vergleich zu Kontrollen noch auf das Muster des Spleißens im Sashimi-Plot.

Das Gen *DBR1* wurde im Jahr 2018 erstmals zu viralen Hirnstamminfektionen assoziiert (OMIM #619441; Zhang et al., 2018). Es handelt sich hierbei um eine Suszeptibilität des Hirnstamms gegenüber der Infektion durch bekannte Krankheitsviren, i. S. einer angeborenen herabgesetzten Hirnstamm-spezifischen Immunabwehr. Dieser Pathomechanismus wurde bereits für die Herpes-simplex-Enzephalitis, welche ebenfalls spezifische Hirnregionen betrifft, implementiert (hier die TLR3-regulierte IFN- α/β -Immunabwehr von Viren, s. u. a. Casanova, 2015). Bemerkenswert in Bezug auf den Phänotyp von MS002 ist das Auftreten von krausem Haar bei einer Probandin in der Kohorte von Zhang et al., 2018.

Aufgrund des Phänotyps (ANEC in Zusammenhang mit einer RSV-Infektion, krauses Haar) passend zur beschriebenen Literaturrecherche wurde *DBR1* in diesem Fall als klares Kandidatengen priorisiert. Die

Variante (s. oben) wurde in die Klasse 3 nach ACMG eingeordnet, da sie bisher nicht als pathogen in Bezug auf das beschriebene Krankheitsbild publiziert wurde.

4.2.3 Qualitative und semiquantitative Analyse der Proteinexpression mittels Western Blot

Western Blot-Analysen mit Proteinextrakten aus Fibroblasten von MS002 und gesunden Kontrollen wurden zum qualitativen und semiquantitativen Nachweis von DBR1 durchgeführt (Methodik s. Abschnitt 3.6). Die Immundetektion wurde hierbei mittels eines Primärantikörpers der Firma proteintech (s. Abschnitt 2.5.1) durchgeführt.

Die vorhergesagte molekulare Masse für das Protein in voller Länge beträgt ~ 62 kDa (544 aa). Zhang et al., 2018 konnten in ihrer Arbeit zusätzlich eine ~ 40 kDa schwere Isoform, insbesondere in transformierten *E. coli*-Lysaten und in geringerer Ausprägung auch in HEK-Zellen, detektieren.

In einem ersten Vorexperiment zeigte sich eine gute Detektion von DBR1 als Bande bei ~ 70 kDa in den Kontrollen, welche sich in deutlich schwächerem Ausmaß bei MS002 präsentierte. Es wurden keine Signale im Bereich der in der Literatur vorbeschriebenen Größe von 40 kDa detektiert.

Nach methodischer Optimierung des Western Blot-Protokolls wurden zwei Membranen mit jeweils drei Proteinlysaten von MS002 und neun Kontrolllysaten angefertigt, um die Reduktion des Proteinlevels des mutierten DBR1 semiquantitativ zu bestimmen (s. Abschnitt 3.6.7 und 3.6.8).

Rein visuell ist bereits die erniedrigte Menge an vorhandenem DBR1 in den Patientenzellen, nicht jedoch den Kontrollen auf den Western Blots zu erkennen (s. Abbildung 4.2A und B).

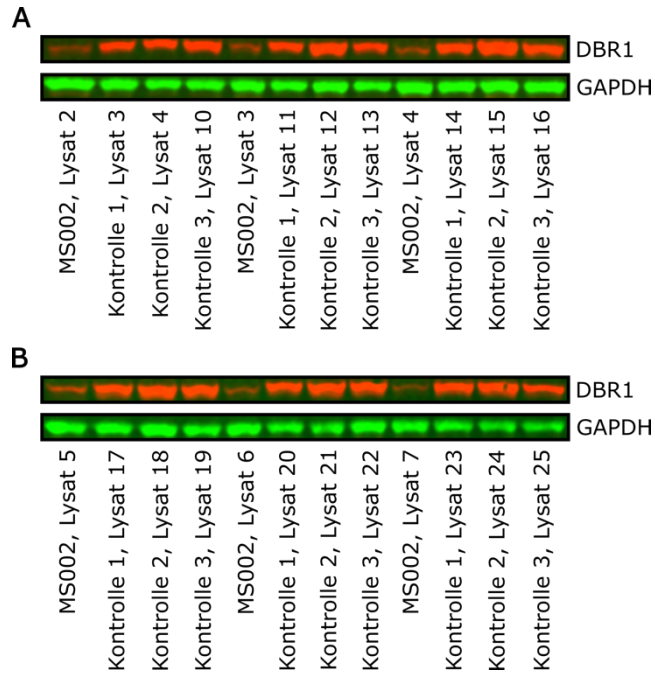


Abbildung 4.2: Western Blots zur semiquantitativen Bestimmung von DBR1 in MS002 und Kontrollen

Rot = Detektion für DBR1, grün = Detektion für GAPDH

A 1. Membran; **B** 2. Membran; bei beiden Membranen wurde der Primärantikörper 16019-1-AP der Firma Proteintech Europe (Konzentration 1:1000) und je 40 µg Proteinlysate verwendet

Nach statistischer Auswertung ergibt sich eine Reduktion von DBR1 in den Patientenzellen gegenüber den drei Kontrollen um durchschnittlich 76 % (0,73 ggü. Kontrolle 1 (95 %-KI = 0,61-0,85); 0,80 ggü. Kontrolle 2 (95 %-KI = 0,61-0,99); 0,74 ggü. Kontrolle 3 (95 %-KI = 0,62-0,86); $p < 0,0001$ für alle drei Kontrollen). Die Unterschiede der Kontrollen zeigten sich nicht signifikant (s. Abbildung 4.3).

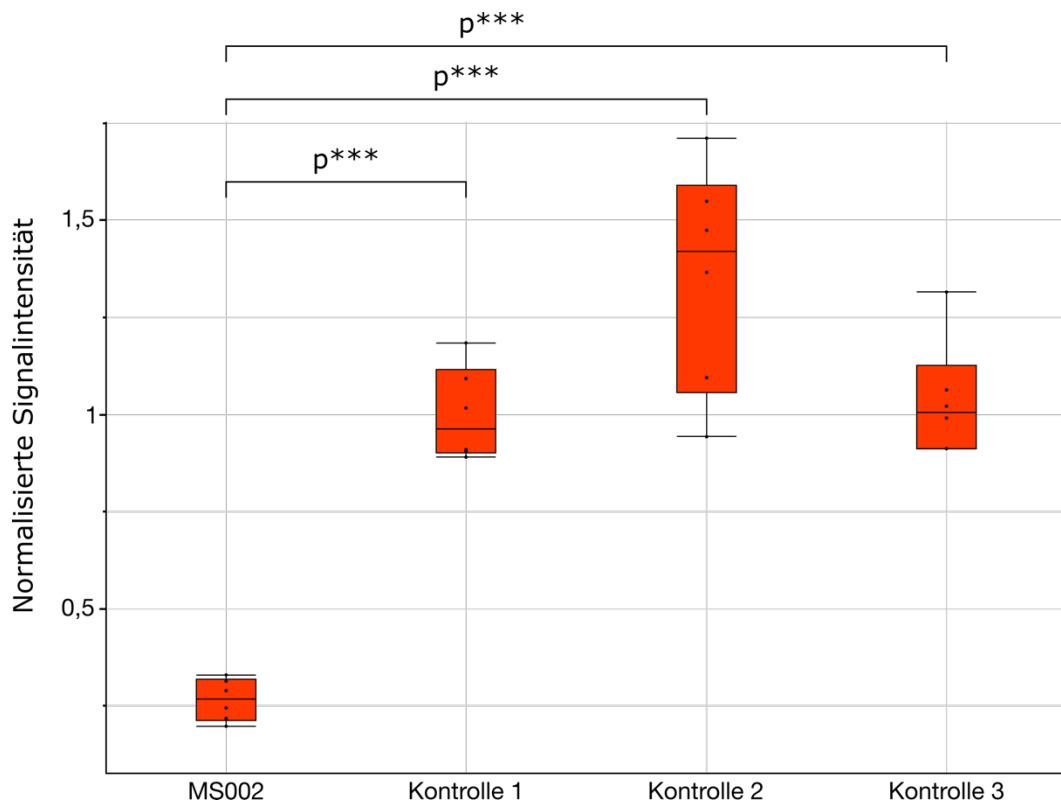


Abbildung 4.3: Boxplots der normalisierten Signalintensitäten für DBR1

4.3 Identifizierung einer Mutation in *SEC31B* bei Individuum MS001

4.3.1 MS001: Phänotyp

Individuum MS001 war bei Vorstellung in der genetischen Sprechstunde 57 Jahre alt und wurde durch KollegInnen der Neurologie überwiesen. Die Patientin litt seit ca. 5 - 6 Jahren unter einer Bewegungsstörung mit Ruhetremor linksbetont und Verlangsamung i. S. eines Parkinsonismus, welche sich seit Einleitung einer Therapie (Medikament unbekannt) durch die neurologischen KollegInnen bereits gebessert hätte.

Seit Geburt besteht ein Nystagmus und die Patientin wurde bereits in ihrer Jugend aufgrund von motorischen Problemen vom Schulsport befreit. Diese motorischen Probleme konnten durch Befragung der Betroffenen nicht weiter eingeordnet werden. Zudem zeigt die Betroffene eine leichte Intelligenzminderung. So konnte Individuum MS001 ein Hauptschuläquivalent in ihrem Heimatland abschließen, jedoch besitzt die Patientin keine abgeschlossene Berufsausbildung. Die Betroffene hat zum Zeitpunkt der Datenerhebung eine Arbeitsstelle in einer Firma in der Montage und als Reinigungskraft.

Syndromale Auffälligkeiten bestehen in einer Makrozephalie von 63 cm und großen Händen, sowie Füßen (Schuhgröße 43 bei einer Körpergröße von 168 cm).

Die Familienanamnese ist unauffällig für genetisch bedingte oder neurologische Erkrankungen. Es gibt keine weiteren betroffenen Geschwister. Eine Schwester hat ebenfalls einen großen Kopf, dieser sei laut Aussage der Familienmitglieder vom Vater vererbt.

4.3.2 Auswertung der Exom-Sequenzierung in Kombination mit den RNA-Seq-Daten

Bei Individuum MS001 wurde ein WES und eine RNA-Seq durchgeführt. Die Exomsequenzierung ergab insgesamt 66996 Varianten und 336 Varianten nach einer frequenzbasierten Filterung (MAF <0,1 %). 28 Varianten waren davon grundsätzlich passend auf einen rezessiven

Erbgang und 141 Varianten kamen für einen dominanten Erbgang in Frage (Filterungsschritte s. Abschnitt 3.3.2).

Nach Durchsicht der gefilterten Varianten und Ausschluss von technischen Artefakten, erschienen zunächst drei Varianten besonders interessant (s. Tabelle 4.6). Diese Varianten wurden durch ausführliche Recherche in genetischen Datenbanken und Pubmed (Literaturrecherche) auf mögliche Pathogenität bewertet. Die Variante im Gen *SEC31B* wurde als Kandidatengen priorisiert (in Tabelle 4.6 grau hinterlegt).

Tabelle 4.6: Potenzielle Kandidatengene von Individuum MS001

Gen	Art	Zygotität	Position
<i>SEC31B</i>	Stopcodon	homozygot	ENST00000370345.3: c.1759G>T; p.Glu587*
<i>NTNG1</i>	Spleißakzeptor	heterozygot	ENST00000370061: c.1088-1G>A
<i>FBXO11</i>	In-frame- Insertion	heterozygot	ENST00000403359: c.191_199dup; p.Pro64_Pro66dup

Das Gen *NTNG1* kodiert für das Protein *netrin-G1*, auch bekannt als *laminet-1*, und ist vorrangig im gesamten Gehirn exprimiert mit prädominanter Expression im dorsalen Thalamus, im Riechkolben und in den Colliculi inferiores. Besonders interessant erschien die Tatsache, dass *NTNG1* eine Rolle in der Entwicklung von neuronalen Verknüpfungen im Gehirn spielt (Yin et al., 2002, Nishimura-Akiyoshi et al., 2007). Es ist außerdem im Nucleus olivaris inferior exprimiert und könnte deshalb im Speziellen in die Formation von funktionellen Schleifen für motorische Aktivitäten involviert sein (Borg et al., 2005).

In der *Genome Aggregation Database* (gnomAD; Version 4.1.0) wird dieser SNP (ENST00000370061: c.1088-1G>A) in 21 Allelen in heterozygoter Form gelistet (Stand: 09.10.2025). Der o/e-Score für pLoF liegt bei 0,29 (90 % CI: 0,2 – 0,44) und der entsprechende pLI-Score bei 1, entsprechend für eine hohe Intoleranz gegenüber LoF-Mutationen in *NTNG1*. Die Variante ist in Datenbanken (ClinVar und HGMD) nicht als pathogen aufgeführt (Stand: 09.10.2025). Der SNP kann in der Datenbank dbSNP unter der

Nummer rs1436065445 eingesehen werden. Der Score PhyloP spricht für eine hochkonservierte Position im Genom.

Auch wenn der Variante als Spleißakzeptor eine hohe Auswirkung vorhergesagt wird (u. a. auch durch *In silico*-Analyse-Tools wie bspw. MaxEntScan), kann diese These in der RNA-Seq nicht eindeutig bestätigt oder widerlegt werden. Vielmehr ist das Exon, welches die Variante enthält, Teil eines alternativen Transkripts und schlecht exprimiert.

Eine balancierte *de novo* Translokation im Gen *NTNG1* wurde durch Borg et al., 2005 zum atypischen Rett-Syndrom (RTT) assoziiert. Weitere Studien konnten jedoch keine weiteren wahrscheinlich pathogenen Varianten in Bezug auf das RTT identifizieren (Archer et al., 2006, Nectoux et al., 2007). *De novo* Missense-Mutationen wurden weiterhin mit Patienten mit Autismus Spektrum Störungen (ASD) in Verbindung gebracht (O'Roak et al., 2012).

Der Phänotyp von Individuum MS001 trifft in keinerlei Hinsicht die Kriterien eines typischen oder atypischen RTT oder einer ASD (vgl. Neul et al., 2010, Lord et al., 2018). Aus phänotypischer Sicht ergab sich somit kein Hinweis auf einen Zusammenhang von der Erkrankung der Betroffenen zu *NTNG1*. Als Konsequenz der Ausführungen in diesem Abschnitt wurde die Variante im Gen *NTNG1* als Kandidatengen aus phänotypischen Gründen nicht weiter priorisiert.

An der Variante in dem Gen *FBXO11* erschien interessant, dass es kürzlich zu einer syndromalen Form der geistigen Behinderung assoziiert wurde (Gregor et al., 2018, Jansen et al., 2019). Der Phänotyp der Betroffenen reichte von milder bis zu schwerer geistiger Behinderung und zeigte inkonstante syndromale Merkmale (hohe breite Stirn, lange Lidspalte, schmale Oberlippe mit breitem Philtrum, Handabnormalitäten i. S. einer Brachydaktylie). Grundsätzlich würde der Phänotyp unter der Abwesenheit der beschriebenen charakteristischen Stigmata zu Individuum MS001 (milde geistige Behinderung) passen. Dennoch zeigt keiner der von Jansen

et al., 2019 eingeschlossenen Betroffenen einen Parkinsonismus, eine Makrozephalie oder große Hände und Füße.

Die Variante (ENST00000403359: c.191_199dup; p.Pro64_Pro66dup) wird insgesamt 55-mal in gnomAD aufgeführt und in der Datenbank dbSNP erfasst (rs764506114). Allerdings ist die Qualität der Daten für diese Position in gnomAD als insuffizient angegeben, sodass die Frequenz in der Allgemeinbevölkerung unsicher ist. Die Variante ist nicht in ClinVar oder HGMD gelistet. Ein Z-Score von 4,1 für Missense-Varianten, welcher *inframe*-Varianten einschließt, spricht eher für eine Intoleranz des Gens gegenüber solchen Mutationen, während der o/e-Score mit 0,62 (90 %-KI = 0,57 – 0,66) nicht signifikant erscheint.

Eine Krankheitsassoziation erscheint in Bezug auf Individuum MS001 vor allem aus phänotypischen Gesichtspunkten unwahrscheinlich, kann jedoch aufgrund der ungenügenden Datenlage nicht abschließend bewertet werden (Variante unklarer Signifikanz). Dennoch wurde *SEC31B* als Kandidatengen gegenüber *FBXO11* bevorzugt (s. folgende Absätze).

Die homozygote Stopp-Variante im Gen *SEC31B* (ENST00000370345.3, c.[1759G>T; 1759G>T]; p.[(Glu587*); (Glu587*)]) wurde nach Recherche in Pubmed und in genetischen Datenbanken als Kandidatengen priorisiert.

Diese Variante ist in der institutsinternen Datenbank 13-mal als heterozygot gelistet. Sie findet sich in 1614206 Allelen des gnomAD-Browser 436-mal ausschließlich in heterozygoter Form (Stand: 09.10.2025). An anderen Positionen des Gens finden sich in der Datenbank lediglich vier weitere trunkierende Varianten (alles *frameshift*-Mutationen) in homozygoter Ausprägung. Der PhyloP-Score spricht mit 2,58 für eine konservierte Stelle im Genom.

Dennoch sagt der kalkulierte o/e-Score in der gnomAD für trunkierende (pLoF) Mutationen mit 0,81 eine hohe Toleranz voraus (pLI = 0).

Die Variante bei Individuum MS001 befindet sich mittig in den meisten Transkripten von *SEC31B* (bspw. Exon 15 von 26 in ENST00000370345.3;

für weitere Transkripte s. *UCSC Genome Browser*) und wird in der langen 26 von 29 Exons umfassenden Isoform von *SEC31B* in allen Transkripten exprimiert (Stankewich et al., 2006). Dies würde folglich zu einem massiv verkürzten Protein oder zu einem Abbau der fehlerhaften mRNA durch *Nonsense mediated decay* (NMD) führen. Das Ergebnis wäre somit eine Null-Mutation. Allerdings berichten Stankewich et al., 2006 von massivem alternativem Spleißen und einer kürzeren Variante von *SEC31B*, welche von der beschriebenen Mutation unbeeinflusst wäre. Die Funktion der kürzeren Variante, welche die für die Bindung von SEC23 wichtige Prolinreiche Domäne nicht enthält, ist bisher nicht untersucht.

Die RNA-Seq war bei einer sehr schlechten *coverage* (<10 reads) in der Probe der Betroffenen, sowie den Kontrollen nur eingeschränkt verwertbar. Erkennbar war jedoch dennoch eine klare Exonstruktur ohne Hinweis auf NMD oder verändertes Spleiß-Pattern.

Das Gen *SEC31B* kodiert für eine Komponente des *coat complex II (COPII)* (Stankewich et al., 2006), welcher für den Cargo-Export vom Endoplasmatischen Retikulum benötigt wird (Bonifacino und Glick, 2004). Varianten in *SEC31B* wurden bisher nicht in Zusammenhang mit Phänotypen im Bereich der geistigen Behinderung beschrieben (Recherche in Pubmed, Stand: 09.10.2025). Allerdings sind Mutationen in multiplen Komponenten der *COPII* Vesikelhülle krankheitsassoziiert worden (vgl. Lu und Kim, 2020). Kürzlich wurde eine *frameshift*-Mutation in *SEC31A*, einem Homolog zu *SEC31B*, mit einem schweren neurologischen Phänotyp bestehend aus spastischer Quadriplegie, Optikusatrophie, Epilepsie und Hirnfehlbildungen in Verbindung gebracht (OMIM #618651; Halperin et al., 2019). Außerdem wurden Missense-Varianten im Gen *SEC24A* zur cranio-lenticulo-suturalen Dysplasie assoziiert (OMIM #607812; Boyadjiev et al., 2006).

Der milde Phänotyp von Individuum MS001 im Vergleich zu den oben beschriebenen Krankheitsbildern könnte hierbei an einer differentiellen Expression von *SEC31B* liegen. Laut Datenbank BioGPS ist *SEC31B* jedoch ubiquitär mit Prädominanz in CD8+- und CD4+-positiven

Lymphozyten exprimiert. Eine weitere Möglichkeit die phänotypischen Abweichungen zu erklären, ist die durch Stankewich et al., 2006 vermutete spezifische Cargo-Selektion durch *SEC31B*.

Zusammenfassend wurde die oben genannte Variante in *SEC31B* als Variante unklarer Signifikanz (Klasse 3 nach ACMG) klassifiziert und als Kandidatengen für Individuum MS001 v. a. aufgrund des trunkierenden Aspektes priorisiert.

4.3.3 Qualitative Analyse der Proteinexpression mittels Western Blot

Für das Kandidatengen *SEC31B* wurde mittels Western Blot die Proteinexpression aus Fibroblasten überprüft (s. Abschnitt 3.6). Die Immundetektion wurde hierbei auf Basis zweier unterschiedlicher Primärantikörper (abcam und Aviva Systems Biology Corporation) umgesetzt. Ziel war die Darstellung des Proteins bei den Kontrollen, jedoch nicht in der Patientenzelllinie, um die Nullmutation resultierend aus der in den vorherigen Absätzen beschriebenen Mutation nachzuweisen. Eine Visualisierung eines verkürzten Proteins bei Individuum MS001 gegenüber des Vollängenproteins bei den Kontrollen wäre ebenfalls denkbar gewesen.

Laut Stankewich et al., 2006 besteht das längste Transkript von *SEC31B* aus 4648 Nukleotiden (inklusive 5'- und 3'-UTRs, offenes Leseraster von 3540 Nukleotiden) und kodiert für ein Protein, welches aus 1180 Aminosäuren besteht. Die Molekularmasse des Proteins wird mit 128.711 Da (\approx 129 kDa) vorhergesagt. Die Arbeitsgruppe um Stankewich et al., 2006 konnte allerdings auch multiple Banden im Western Blot unterschiedlicher Molekülmassen detektieren. Die Molekülmassen dieser Proteine variierten stark zwischen 45 und 129 kDa und passen zum ebenfalls detektierten massiven alternativen Spleißen der prä-mRNA.

Die Darstellung des Proteins gelang trotz der Verwendung zweier Primärantikörper weder in den Patientenzellen noch in den Kontrollen (s. Abbildung 4.4).

In ersten Vorexperimenten wurden mit dem Antikörper der Firma abcam (ab80053) bei allen Proteinlysaten (Patientin und Kontrollen) lediglich Banden bei ca. 230-240 kDa identifiziert (Western Blots nicht in dieser Arbeit gezeigt).

Einer hieraufhin vermuteten mangelnden Lyse der Fibroblasten wurde mit einem stärkeren Lysepuffer (Extraktionslysepuffer, s. Abschnitt 2.3) begegnet, welcher in der Folge für alle gezeigten Western Blots in dieser Arbeit verwendet wurde. Trotz der Verwendung des veränderten Lysepuffers gelang eine Bandendetektion lediglich bei ca. 230-240 kDa. Diese Banden wurden als unspezifisch für SEC31B gewertet (s. Abbildung 4.4A).

Es wurde in der Folge ein weiterer Primärantikörper (Aviva Systems Biology Corporation, ARP-66671_P50) verwendet. Die Darstellung des ~ 129 kDa-Proteins gelang auch bei Inkubation mit steigenden Konzentrationen an Primärantikörper von Aviva Systems Biology Corporation bis zu einer Konzentration von 1:500 nicht (s. Abbildung 4.4B).

Die Expression von *SEC31B* liegt laut der Datenbank BioGPS in der Haut unter dem Durchschnitt, der höchste Wert wird bei CD8-positiven Lymphozyten angegeben. Daher wurde eine mangelnde Expression in Fibroblasten vermutet und mithilfe des parallelen Auftragens von Kontrolllysaten aus CPT-Blut, HEK- und RPE-Zellen überprüft. Es zeigten sich Banden unterschiedlicher Höhe abhängig von den Primärantikörpern und den unterschiedlichen Zellarten. Die Banden konnten allerdings SEC31B nicht eindeutig zugeordnet werden (s. Abbildung 4.5).

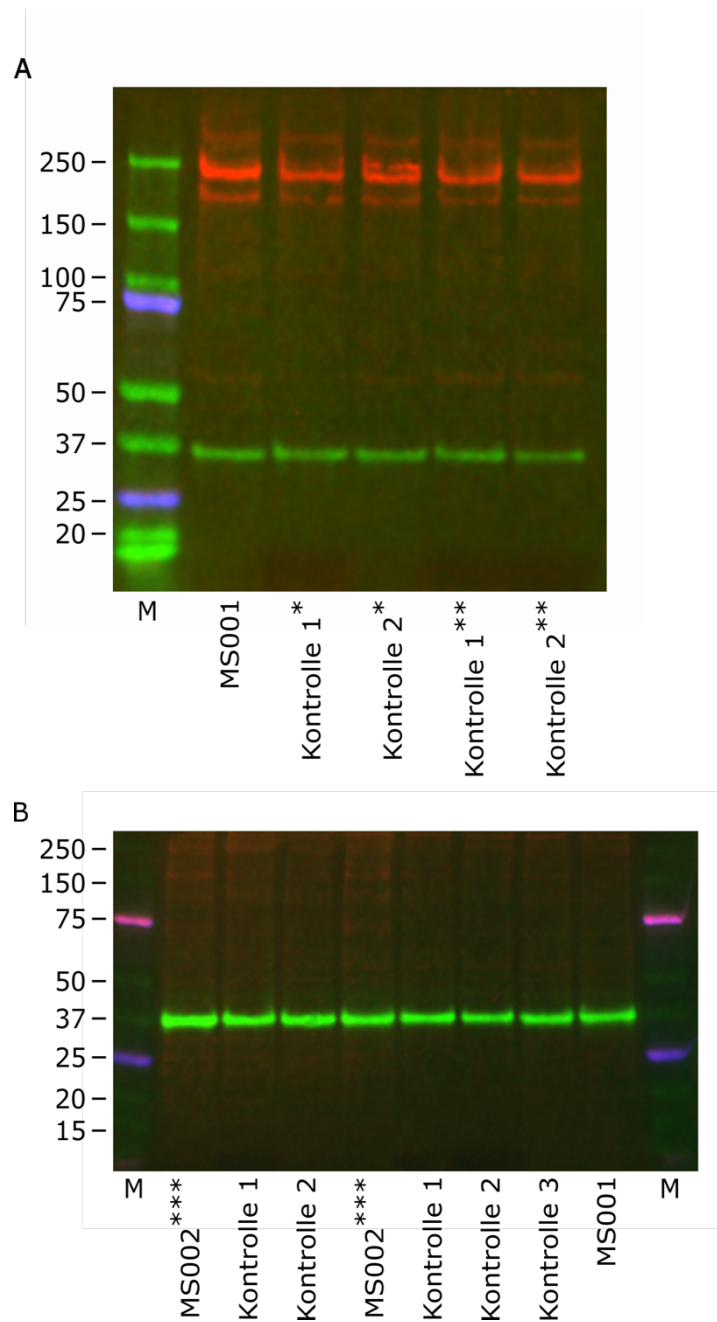


Abbildung 4.4: Western Blots zum Nachweis von SEC31B mit Proteinlysaten aus Fibroblasten unter Verwendung von zwei unterschiedlichen Primärantikörpern

Rot = Detektion für SEC31B, grün = Detektion für GAPDH, M = Marker; Skalierung in kDa;

*Lysate unter Verwendung von Lysispuffer hergestellt, **Lysate unter Verwendung von Extraktionslysepuffer hergestellt, ***MS002 wurde aus organisatorischen Gründen als 4. Kontrolle für diesen Western Blot verwendet.

A Western Blot unter Verwendung des Primärantikörpers ab80053 der Firma abcam (Konzentration 1:2500) und je 50 µg Proteinlysat.

B Western Blot unter Verwendung des Primärantikörpers ARP66671_P050 der Firma Aviva Systems Biology Corporation (Konzentration 1:500) und je 50 µg Proteinlysat.

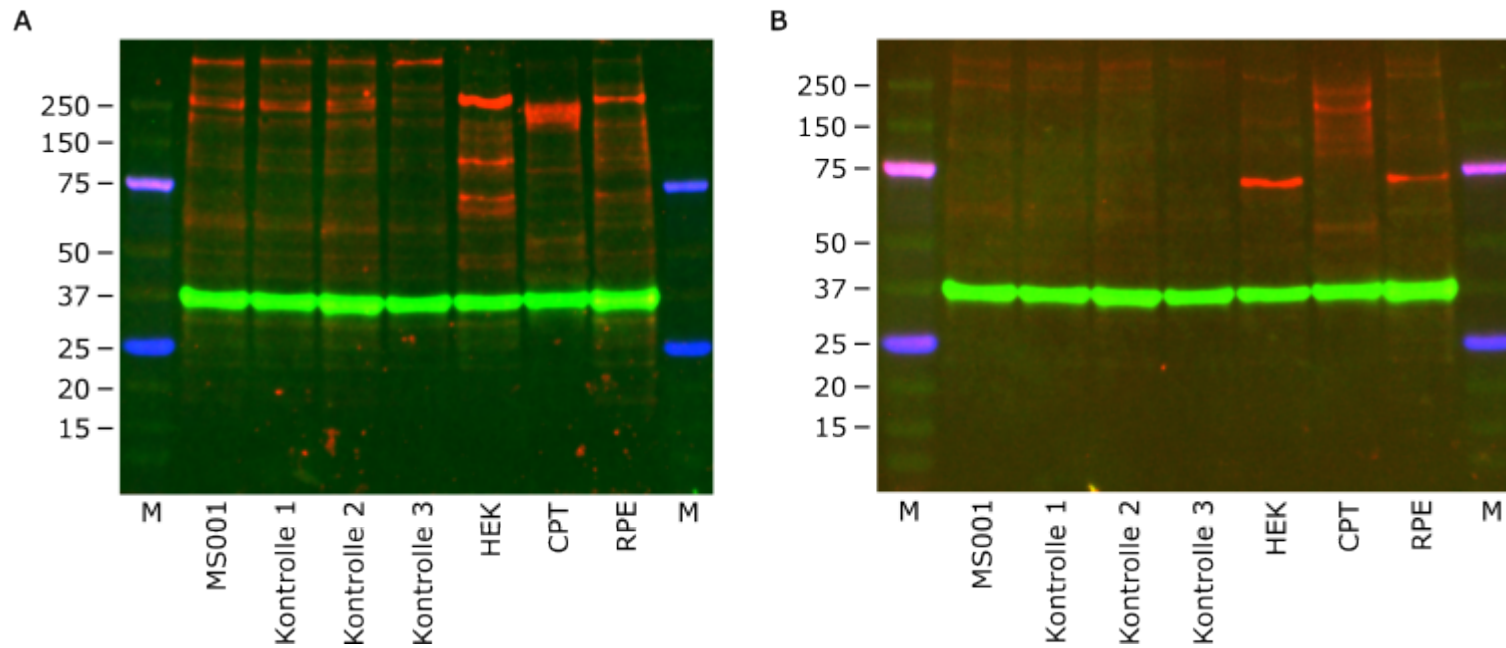


Abbildung 4.5: Western Blots zum Nachweis von SEC31B mit Proteinlysaten aus Fibroblasten (MS001 und Kontrollen 1-3) sowie aus HEK-Zellen, CPT-Blut und RPE-Zellen

Rot = Detektion für SEC31B, grün = Detektion für GAPDH, M = Marker; Skalierung in kDa

A Western Blot unter Verwendung des Primärantikörpers ab80053 der Firma abcam (Konzentration 1:2500) und je 50 µg Proteinlysat.

Es zeigen sich unspezifische Banden bei ca. 240 – 250 kDa bei allen Zelltypen und zusätzlich Banden zwischen 60 – 100 kDa im HEK-Proteinlysat.

B Western Blot unter Verwendung des Primärantikörpers ARP66671_P050 der Firma Aviva Systems Biology Corporation (Konzentration 1:500) und je 50 µg Proteinlysat. Es zeigt sich jeweils jeweils eine deutliche Bande bei ca. 70 kDa für die Lysate aus HEK- und RPE-Zellen. In der Probe aus CPT-Blut kommen erneut nur Banden unspezifischer Art mit Prädominanz bei ca 200 kDa zur Detektion

4.4 Identifizierung einer Mutation in *OSCP1* bei Familie 01

4.4.1 Familie 01: Phänotyp

Familie 01 ist eine konsanguine Familie mit drei männlichen Betroffenen (s. Stammbaum in Abbildung 4.6). Lediglich zwei der Betroffenen dieser Familie (mittleres und jüngstes Kind) wurden in das Untersuchungskollektiv dieser Studie eingeschlossen.

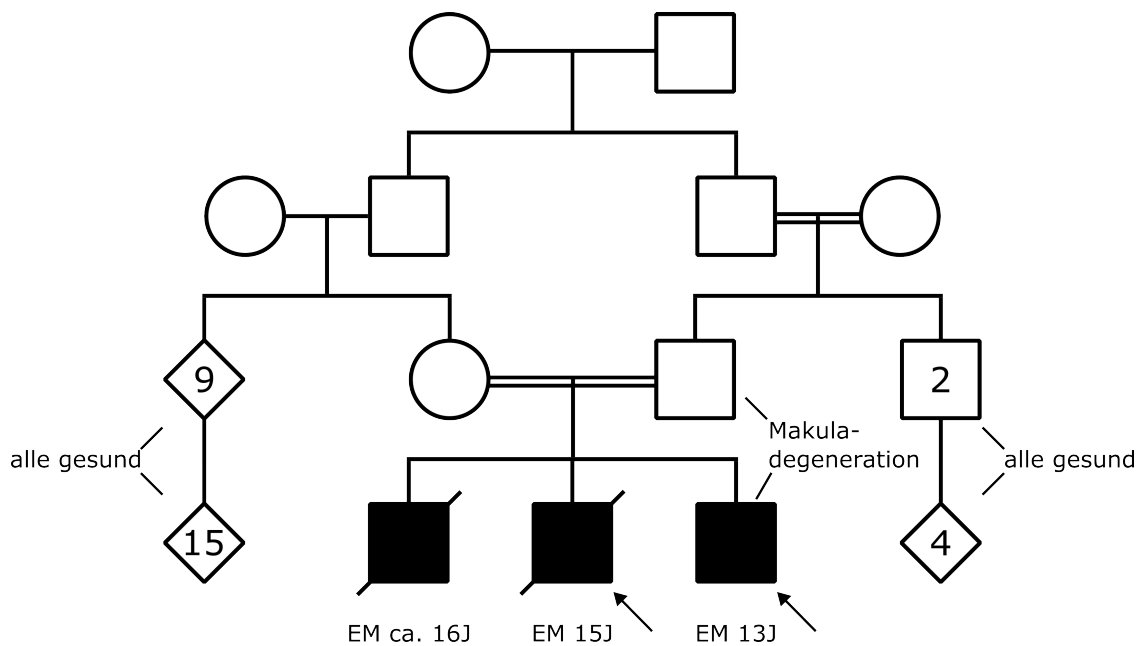


Abbildung 4.6: Stammbaum der Familie 01

EM = Erstmanifestation, J = Jahre

Die Eltern der drei Betroffenen sind Cousin und Cousine 1. Grades.

Alle Brüder leiden bzw. litten an einer bisher ungeklärten oberlappenbetonten fibrosierenden Lungengerüsterkrankung, welche im Lungenversagen mit Todesfolge für bisher zwei von drei Betroffenen mündete (Anm.: zum Zeitpunkt der Auswertung im Oktober 2019). Die Lungenerkrankung manifestierte sich als respiratorische Partialinsuffizienz mit schwerer Hypoxämie. Das Alter der Erstmanifestation lag im vorjugendlichen bis jugendlichen Alter (11 – 16 Jahre).

Im Verlauf der Erkrankung kam es beim verstorbenen ältesten Bruder zu einem Befall mit verschiedenen Mikroorganismen (u. a. Herde einer

chronisch nekrotisierenden Aspergillose). Außerdem wurde eine 4/6 Holosystolikum ohne morphologisches Korrelat (Echokardiographie und Kardio-MRT unauffällig) diagnostiziert.

Die Betroffenen der Familie 01 zeigen keine eindeutigen syndromalen Auffälligkeiten. Beim verstorbenen ältesten Bruder ist ein schmales elfenartiges Gesicht beschrieben, während die anderen Geschwister keine äußeren Stigmata, welche an einen spezifischen Phänotyp denken lassen, aufwiesen. Die Familienhistorie ist leer bezüglich genetischer oder pulmonaler Vorerkrankungen.

Im Vorfeld der genetischen Untersuchungen ergaben durchgeführte Untersuchungen keinen Anhalt für das Vorliegen einer Mukoviszidose (Schweißtest 2x negativ) und keinen Anhalt für das Vorliegen einer primären Ziliendyskinesie (Bürstenabstrich ohne pathologischen Befund). Nebenbefundlich wurde beim dritten Kind der Familie eine Makuladegeneration (EM: 8 Jahre) diagnostiziert. Der Vater ist hiervon ebenfalls betroffen.

4.4.2 Auswertung der Genom-Sequenzierungen in Kombination mit den RNA-Seq-Daten

Die Betroffenen aus Familie 01 erhielten zur genetischen Abklärung jeweils ein WGS in Kombination mit einer RNA-Sequenzierung. Die Genom-Sequenzierungen wurden als Multi-Genomanalyse ausgewertet – hier ergab sich eine Gesamtanzahl von 474261 Varianten. 962 Varianten wiesen eine MAF <0,1 % auf. Nach den Filterungsschritten (vgl. Abschnitt 3.3.2) zeigten sich 30 Varianten passend für einen rezessiven Erbgang, 13 Varianten wurden als compound-heterozygot vorliegend deklariert und 228 Varianten wurden für einen möglichen dominanten Vererbungsweg ausgegeben.

Nach manueller Qualitätskontrolle und ausführlicher Literaturrecherche wurde die Variante im Gen *OSCP1* als Kandidatengen priorisiert (in Tabelle 4.7 grau hinterlegt). Alle interessanten Varianten sind in Tabelle 4.7 aufgelistet. Eine kurze Beschreibung der Bewertung der Varianten und der Ausschlusskriterien als Kandidatengene folgt in den nächsten Absätzen.

Erfahrungsgemäß sind bei einer Multi-Genomanalyse von Kindern von konsanguinen Eltern viele Varianten zu bewerten, daher beschränkt sich diese Auswahl nur auf die vielversprechendsten Kandidatengene (bereits krankheitsassoziiert oder vom biochemischen *pathway* passend erscheinende Gene).

Tabelle 4.7: Interessante Varianten für Familie 01

Gen	Art	Zygotität	Position
<i>OSCP1</i>	Missense	homozygot	ENST00000235532: c.925C>T:p.Arg309Trp
<i>MUC5B</i>	Missense Synonym	compound- heterozygot	ENST00000447027: c.6839C>G:p.Thr2280Arg ENST00000447027: c.6858A>G:p.Thr2286=
<i>DNAH11</i>	Synonym	heterozygot	ENST00000328843: c.12048C>T:p.Ser4016=
<i>DNAH5</i>	Missense	heterozygot	ENST00000265104: c.4564G>T:p.Ala1522Ser
<i>CCDC40</i>	Synonym	heterozygot	ENST00000269318: c.1416G>A:p.Arg472=

Bei Familie 01 wurden compound-heterozygote Varianten in *MUC5B* entdeckt

(ENST00000447027:c.6839C>G:p.Thr2280Arg;c.6858A>G:p.Thr2286=).

Eine Expression in Fibroblasten (RNA-Seq-Daten) fand sich nicht.

MUC5B ist eines von bis dato 20 entdeckten Muzin-Genen, welches in den oberen und unteren Atemwegen mit Prädominanz in den distalen Atemwegen exprimiert wird. Eine Überproduktion, bedingt durch eine Mutation (rs35705950) im Bereich der Promoterregion, wurde durch Seibold et al., 2011 als Risikofaktor für die Entwicklung einer IPF postuliert. Die vorliegenden Varianten in *MUC5B* in Familie 01 wurden aufgrund pathomechanistischer Überlegungen als Kandidatenvarianten verworfen. Beide Varianten scheinen weder an konservierten Stellen im Genom zu liegen (PhyloP-Score -0,65 und -3,36), noch eine potenzielle Überproduktion des Muzins zur Folge haben. Der Effekt der Missense-Mutation auf das Protein wird mit *In silico*-Tools unterschiedlich vorhergesagt (SIFT und CADD potentiell schädlich, PolyPhen, fathmm-MKL und REVEL tolerabel/benigne). Sollte das Protein durch die Missense-Mutation in seiner Funktion alteriert sein, würde dies nicht den bisher

publizierten Erkenntnissen zum möglichen Pathomechanismus von *MUC5B*-assoziierten Varianten entsprechen.

Die Betroffenen zeigen mehrere Varianten in Genen, welche als krankheitsursächlich für die Primäre Ziliendyskinesie (*primary ciliary dyskinesia*, PCD) etabliert sind (in Tabelle 4.7 mit einer gestrichelten Linie umrandet). Interessant erschien hier vor allem die Tatsache, dass ca. 30 % der Patienten mit PCD keine strukturelle Veränderungen der Zilien aufweisen (Zariwala et al., 2007) und somit ein negatives Ergebnis der Bürstenabstriche eine PCD nicht definitiv ausschließt. Klinisch zeigen die Brüder der Familie 01 dennoch keinen klassischen Phänotyp für eine PCD, welcher in einer frühen (häufig neonatalem) Atemstörung, chronischen Atemwegsinfekten (sowohl nasale, als auch pulmonale Infekte) und einer daraus resultierenden obstruktiven Lungenerkrankung besteht (Mullowney et al., 2014, Halbeisen et al., 2018). Typisch wäre außerdem ein *Situs inversus totalis* oder ein *Situs ambiguus*, welcher jedoch nur in ca. 50 % bzw. 10 % der PatientInnen auftritt (Best et al., 2019, Shapiro et al., 2014). Vor allem der fibrosierende Aspekt der Lungenerkrankung und der Zeitpunkt der Erstmanifestation von respiratorischen Symptomen im Alter von 11 – 16 Jahren sorgten für einen Ausschluss der Varianten aus phänotypischen Gründen. Außerdem ist der Vererbungsmodus bei allen hier beschriebenen Genen autosomal-rezessiv – ein weiteres Ausschlusskriterium stellte insofern die Heterozygotie dieser Varianten dar. Bei allen heterozygot vorliegenden Varianten wurde explizit auf compound-heterozygot vorliegende CNVs, Indels und SNPs geprüft, welche in keinem Fall vorlagen. Die Analyse der RNA-Seq ergab bei all diesen Varianten keinen wegweisenden Befund.

Nach kompletter Auswertung der Multi-Genom-Analyse und ausführlicher Literaturrecherche wurde die homozygot vorliegende Missense-Variante im Gen *OSCP1* (ENST00000235532: c.[925C>T];[925C>T], p.[Arg309Trp];[Arg309Trp]) als Kandidatengen priorisiert. Die Variante ist

unter der Nummer rs1406005863 in der Datenbank dbSNP gelistet. Im gnomAD-Browser findet sich dieser SNP in 1613948 Exomen lediglich 16-mal in heterozygoter Form (Stand: 09.10.2025). Der prädiktive Score für Konservierung (PhyloP) liegt bei 7,42 und spricht somit für ein hochkonserviertes Nukleotid. *In silico*-Analysen sagen durchgängig einen schädlichen Einfluss der Missense-Variante auf das Protein voraus (SIFT: D(0,00) \triangleq *deleterious*; PolyPhen: D(1,00) \triangleq *deleterious*; fathmm-MKL: 0,99; CADD: 34,00).

In der RNA-Seq zeigt sich zusätzlich ein alteriertes Spleißen von Exon 8 von 10 i. S. eines Herausspleißens des Exons bei ca. 73% der Transkripte (49 *reads* mit entferntem und 18 *reads* mit vorhandenem Exon 8 bei einem der beiden Individuen, s. Abbildung 4.7A). Der andere Bruder zeigte bei *low-coverage* keine verwertbaren Daten, es konnte jedoch anhand weniger *reads* ein ebenfalls vorhandener Spleißdefekt angenommen werden (Validierung s. nächster Abschnitt). Das Phänomen des alternativen Spleißens zeigte sich in keiner der angezeigten Kontrollen und ist somit ein Hinweis auf eine direkte Auswirkung der vorhandenen Missense-Variante in Exon 8. Abbildung 4.8A zeigt die Auswirkungen der Missense-Variante und des Spleißdefekt schematisch im Vergleich zum Wildtyp.

Exon 8 besteht aus insgesamt 140 bp, sodass ein Herausspleißen zu einem *frameshift* um zwei Basenpaare führt. In Abbildung 4.8B ist gezeigt, wie es durch diesen *frameshift* zu einer Bildung eines Stopcodons 36 bp nach dem Spleißdefekt mittig in Exon 9 kommt. Es würde also hypothetisch ein um 47 Aminosäuren verkürztes Protein resultieren (p.Glu320ThrfsTer12).

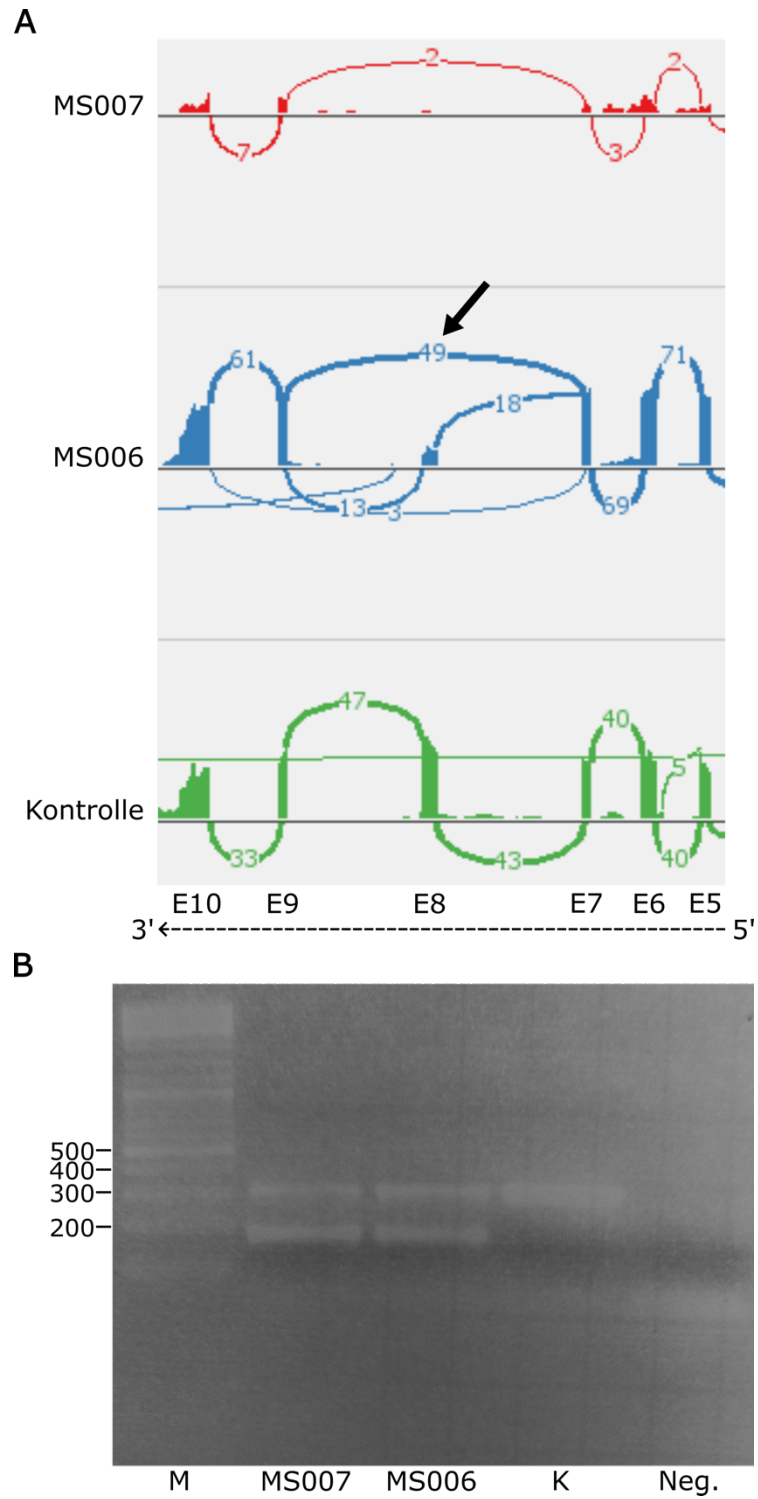


Abbildung 4.7: Spleißdefekt in *OSCP1* bei Familie 01

A Sashimi-Plot von *OSCP1* für MS006, MS007 und eine Kontrolle

E = Exon; Pfeilmarkierung zeigt den Spleißdefekt bei MS006

B cDNA-Befund für MS007, MS006 und eine Kontrolle; Skalierung in bp

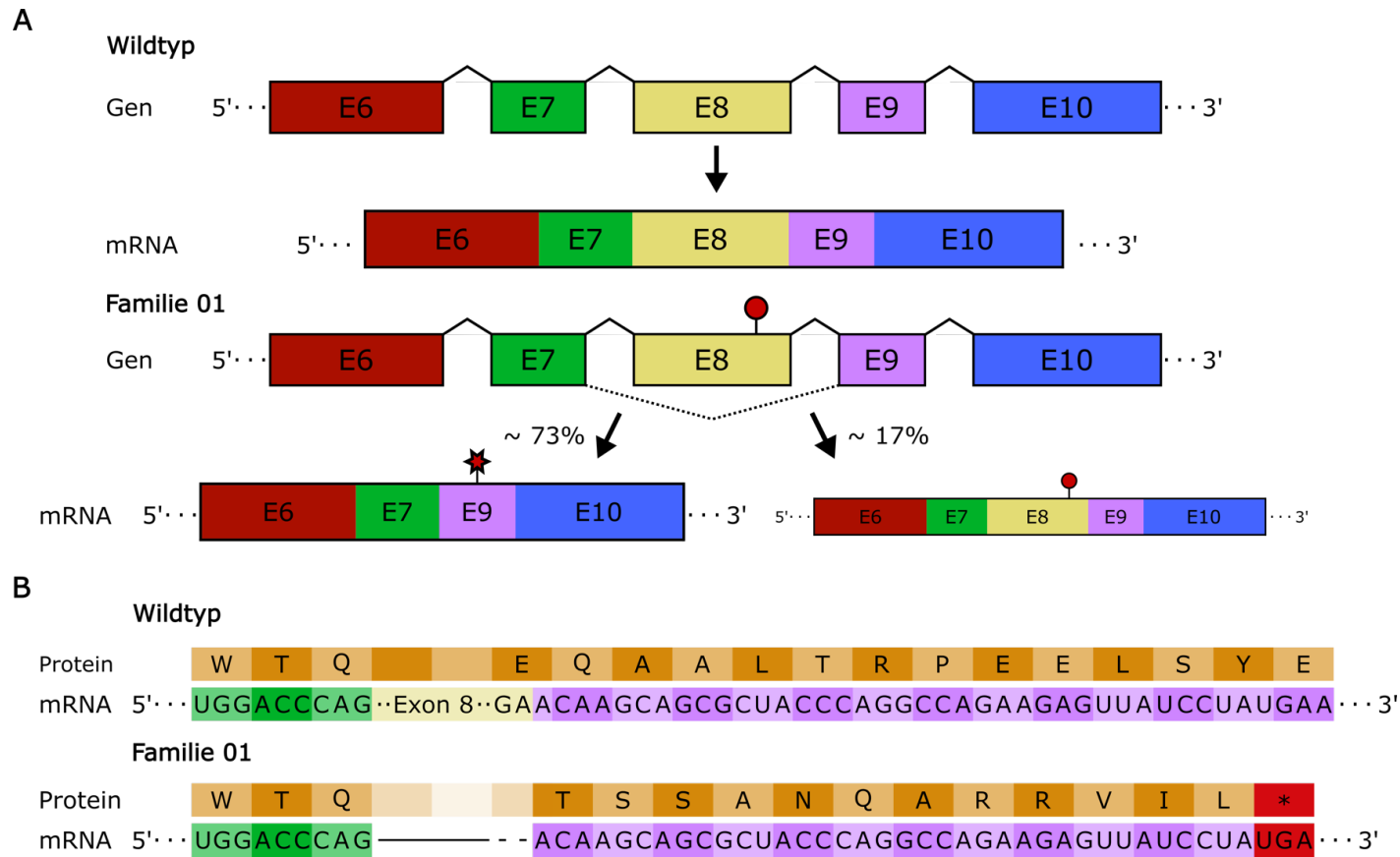


Abbildung 4.8: Schematische Darstellung des Spleißdefekts bei Familie 01 im Vergleich zum Wildtyp

A Ausprägung des Spleißdefekts bei Familie 01 im Vergleich zum Wildtyp; **B** Frameshift mit Einführung eines Stopcodons durch Spleißen von Exon 8

E = Exon; Rote Punktmarkierung = Mutation bei Familie 01; Rote Sternmarkierung = Eingeführtes Stopcodon durch Spleißen von Exon 8

OSCP 1 besteht aus elf Exons und kodiert für das *organic solute carrier protein 1*, auch bekannt als *NOR1 (oxidored-nitro domain-containing protein 1)*. Durch alternatives Spleißen von Exon 2 entstehen zwei Isoformen aus 379 bzw. 389 Aminosäuren.

Laut der Datenbank BioGPS wird *OSCP1* prädominant in Zellen des Atemweges (v. a. epithelialen Zellen des oberen und unteren Atemweges; z-score >12 (Median: 1,52)) exprimiert. Eine deutliche Expression findet sich außerdem im Hoden und den Spermien (z-Score >7), sowie im zentralen Nervensystem (z-Scores variabel). Bemerkenswert war in bisherigen Studien, dass die längere Isoform des Proteins nur im Gehirn exprimiert wird (Xiang et al., 2011).

OSCP1 erscheint als Kandidatengen interessant, da es kürzlich in einer Studie als relevant für die Entwicklung von Zilien im Zebrafischmodell publiziert wurde (van Dam et al., 2019). Die Arbeitsgruppe konnte *OSCP1* bei verschiedenen Zelltypen an der Zilienbasis und am Axonem nachweisen. Außerdem zeigte sich, dass bei *OSCP1*-depletierten Zebrafischen die Zilien desorganisiert und kürzer waren. Die Embryos wiesen phänotypische Auffälligkeiten auf – eine kurze, gebogene Körperachse, kleine Augen, kleiner Kopf, Zysten im *Ductus pronephricus* und Ödeme des Herzens. Besonders interessant im Zusammenhang mit Familie 01 erschien zusätzlich der dosisabhängige Phänotyp der Zebrafische.

Weitere diverse Funktionen von *OSCP1* wurden bereits in anderen Studien postuliert, u. a. Regulation von Apoptose, Inflammation, Proliferation und Tumorsuppression (Yi et al., 2017).

Die homozygote Missense-Variante im Gen *OSCP1* wurde abschließend als Variante unklarer Signifikanz (Klasse 3 nach ACMG) klassifiziert, da der Pathomechanismus der fibrosierenden Lungenerkrankung bei Familie 01 durch eine Ziliopathie oder auch andere bereits für *OSCP1* publizierte *pathways* abschließend nicht geklärt ist (vgl. auch die Ausführungen zur PCD oben in diesem Abschnitt).

Als Zusatzbefund zeigte sich bei dem jüngeren Bruder, welcher zusätzlich an einer Makuladegeneration leidet, eine homozygote Deletion im Gen *ABCA4*

(ENST00000370225:c.[4640del];[4640del],p.[Lys1547ArgfsTer34];[Lys1547ArgfsTer34]). Das Gen *ABCA4* kodiert für den Retina-spezifischen *ATP-binding cassette*-Transporter der Unterfamilie A. Die Variante ist in der Literatur in Bezug auf die autosomal-rezessiv vererbte Makuladegeneration bereits als pathogen beschrieben (Schulz et al., 2017) und erklärt dieses zusätzliche phänotypische Merkmal bei dem jüngeren Bruder.

4.4.3 Aufklärung des unklaren Befundes aus der RNA-Seq mittels PCR aus cDNA

Der unklare Befund im Rahmen der RNA-Seq-Auswertung (anhand weniger reads konnte bei Individuum MS007 ebenfalls ein Spleißdefekt angenommen werden; s. vorheriger Abschnitt) wurde mithilfe einer RT-PCR aus cDNA validiert (Methodik s. Abschnitt 3.4).

Für beide Individuen wurde eine Doppelbande in der PCR nachgewiesen (s. Abbildung 4.7B) Die Kontrollprobe zeigt nur eine Einzelbande. Die amplifizierten Fragmente passen zu der erwarteten Größe (angezeigt mit dem Tool *In Silico* PCR) von 315 bp für das Vollängentranskript inklusive Exon 8 und dem entsprechend um 140 bp verkürzten Transkript ohne Exon 8 (175 bp). Somit konnte der vermutete Spleißdefekt bei beiden Brüdern mithilfe dieser Methode validiert werden.

4.5 Identifizierung einer Mutation in *HUWE1* bei Individuum MS005

4.5.1 MS005: Phänotyp

Individuum MS005 wurde aufgrund einer unklaren Entwicklungsstörung an die Humangenetische Sprechstunde verwiesen. Die Tochter zweier nicht-konsanguiner Eltern war zum Zeitpunkt der NGS-Auswertungen fünf Jahre alt. MS005 zeigt eine geistige Entwicklungsverzögerung und stereotypische Verhaltensauffälligkeiten. Syndromale Auffälligkeiten bestehen in einem Kleinwuchs kombiniert mit einer Makrozephalie. Eine genauere Phänotypisierung, v. a. der Schwere der Entwicklungsstörung und der Stereotypien, als auch ein Stammbaum, war für diese Arbeit nicht vorhanden.

Die Anamnese ergab keine Hinweise auf familiäre genetische oder neurologische Erkrankungen. Es gibt, soweit zum Zeitpunkt dieser Arbeit bekannt, keine betroffenen Geschwister.

4.5.2 Auswertung der Trio-Exom-Sequenzierung in Kombination mit den RNA-Seq-Daten

Betroffene MS005 erhielt eine Abklärung mittels WES in Kombination mit der Sequenzierung der Exome der Eltern (sog. Trio-Exom, s. Abschnitt 3.2.4) und außerdem eine RNA-Sequenzierung aus Fibroblasten.

In der Exomsequenzierung (ohne Trio-Analyse) konnte eine Gesamtzahl von 67765 Varianten detektiert werden, von denen 321 eine MAF $<0,1\%$ aufwiesen. Für einen rezessiven Erbgang passend zeigten sich 31 Varianten, während sich für einen dominanten Erbgang 130 Varianten überprüft wurden. In der Trio-Exom-Auswertung kamen sich 91611 Varianten zustande, 892 hiervon mit einer MAF $<0,1\%$. Durch den Filter *Trio relaxed* konnte die Anzahl auf 73 Varianten eingegrenzt werden. Der Filter *Trio stringent* belief sich auf 10 Varianten (für alle hier erwähnten Filterschritte s. Abschnitt 3.3.2).

Nach ausführlicher Durchsicht und Literaturrecherche, sowie Integration der RNA-Seq-Daten ergaben sich die in Tabelle 4.8 aufgelisteten Gene, welche in Bezug auf den Phänotyp von MS005 interessant erschienen.

Die Variante im Gen *HUWE1* wurde abschließend als Kandidatengen priorisiert (in Tabelle 4.8 grau hinterlegt). Erläuterungen zu den Gründen für diese Wahl werden in den nächsten Absätzen erläutert.

Tabelle 4.8: Potenzielle Kandidatengene für MS005

Gen	Art	Zygotität	Position
<i>HUWE1</i>	Speißregion <i>De novo</i>	heterozygot; X-chromosomal	ENST00000218328: c.2496+3A>G
<i>PRR12</i>	Missense Synonym	compound- heterozygot	ENST00000418929: c.6100C>T;p.Arg2034Cys ENST00000418929: c.1947G>A;p.Ala649=

Mutationen im Gen *PRR12* wurden kürzlich zu einer syndromalen Form der geistigen Behinderung (*Neuroocular syndrome*, NOC; OMIM # 619539) verknüpft. Syndromale Auffälligkeiten bestehen bei den Betroffenen in strukturellen Augen- und Visusdefekten, einer Mikrozephalie, einem Kleinwuchs und Skelettanomalien (u. a. Brachy- oder Syndaktylien; Leduc et al., 2018, Chowdhury et al., 2021, Reis et al., 2021). NOC wird nach aktuellem Stand der Literatur autosomal-dominant vererbt, molekulare Analysen zeigten bisher ausschließlich *de novo*-Mutationen.

MS005 zeigt weder Augenauffälligkeiten noch sonst eines der phänotypischen Stigmata aus den oben erwähnten Studien. Des Weiteren handelt es sich bei den beiden Varianten ENST00000418929:c.[1947G>A];[6100C>T], p.[Ala649=];[Arg2034Cys] nicht um *de novo* Varianten. Daher wurde *PRR12* als Kandidatengen aufgrund des Pathomechanismus und des Phänotyps verworfen.

Eine heterozygote *de novo* Spleißstellen-Variante im Gen *HUWE1* (ENST00000218328: c.2496+3A>G) konnte mittels Trio-Exom-Analyse ermittelt werden. Die Variante wird im gnomAD-Browser nicht gelistet und scheint daher exklusiv für MS005 zu sein (Stand: 09.10.2025). Ein PhyloP-

Score von 2,19 spricht für eine Konservierung des Nukleotids an dieser Stelle.

Das Programm MaxEntScan sagt einen Spleißdefekt und eine verkürzte mRNA um -28,94 % voraus. Ein o/e-Score für pLoF von 0,03 (90 % - CI: 0,01 – 0,06) im gnomAD-Browser spricht für eine hohe Intoleranz gegenüber LoF-Mutationen im heterozygoten Zustand (aus gnomAD Version 2.1.1, in aktueller Version 4.1.0 nicht vorhanden). Es zeigen sich keine compound-heterozygot vorliegenden Varianten – explizit geachtet wurde auch auf CNVs.

Der vorhergesagte Spleißeffekt konnte in den RNA-Seq-Daten nicht bestätigt werden (s. Abbildung 4.9; aus Übersichtsgründen sind keine Kontrollen dargestellt). Die Expression von HUWE1 ist stattdessen mit einem log₂FC von - 0,9 und einem p-Wert von 0,041 statistisch signifikant geringer exprimiert, was auf einen Abbau der mRNA durch NMD hindeuten könnte.

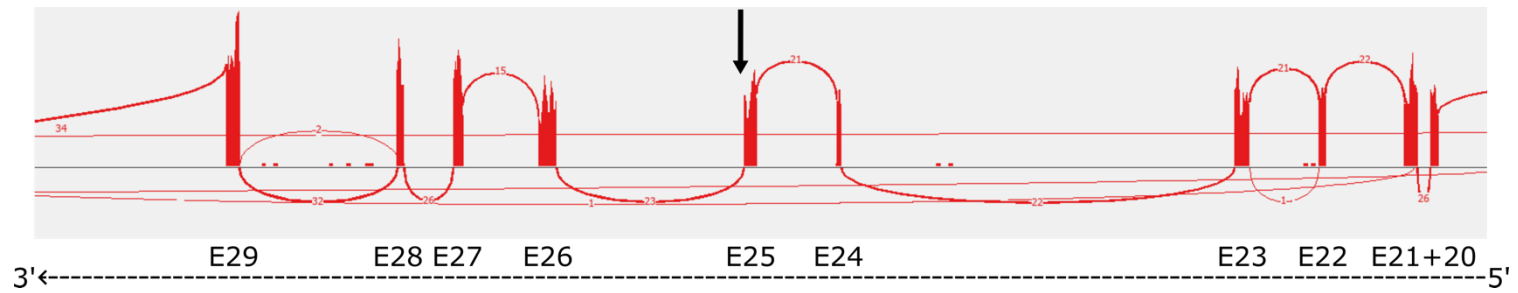


Abbildung 4.9: Sashimi-Plot für *HUWE1* bei Individuum MS005
 E = Exon; Pfeilmarkierung an der Stelle der vorhandenen Mutation

Die Variante erschien als Kandidatengen hochinteressant, da Mutationen in *HUWE1* in der Literatur im Rahmen einer X-chromosomalen Form geistiger Behinderung (OMIM #309590) auch bei weiblichen Individuen als pathogen beschrieben wurden (Moortgat et al., 2018). Moortgat et al. beschreiben bei weiblichen Betroffenen ein variables X-Inaktivierungsmuster, welches in unterschiedlichen Geweben (bspw. Fibroblasten und Blut) schwächer oder deutlicher auftritt. Interessant erschien an dieser Studie, dass die Allele von *HUWE1* (sowohl das Wildtyp, als auch das mutierte Allel) je nach Untersuchungsmaterial unterschiedlich präferentiell oder ausschließlich exprimiert wurden.

4.5.3 X-Inaktivierungsanalyse

Die bei Individuum MS005 heterozygot vorliegenden X-chromosomalen Varianten wurden auf eine präferentielle Expression in der RNA-Seq überprüft. Durch Kombination mit dem Genotyp der Eltern aus der Trio-Exom-Analyse konnte auf eine gestörte X-Inaktivierung geschlossen werden (s. Abschnitt 3.3.4).

Es zeigte sich eine deutlich präferenzielle Expression der durch den Vater vererbten Allele (im Durchschnitt 98% Expression des väterlichen Allels und 2% Expression des mütterlichen Allels) und somit eine gestörte X-Inaktivierung in den Fibroblasten (s. Tabelle 4.9). Lediglich bei Varianten in genomischen Regionen, welche der X-Inaktivierung entgehen und ein pseudoautosomales Verhalten haben, zeigte sich eine Expression von mütterlichen und väterlichen Varianten. Ob die Variante in *HUWE1* auf dem väterlichen oder mütterlichen Allel liegt, konnte nicht abschließend geklärt werden. Eine Möglichkeit der Sequenzierung von cDNA mit einer väterlichen Variante in der Nähe der Mutation in *HUWE1* wurde geprüft, zeigte sich aber durch Limitierungen der Sanger-Sequenzierung als unmöglich. Tabelle 4.9 zeigt zusammenfassend aufgrund der besseren Lesbarkeit dieser Arbeit lediglich die untersuchten Varianten mit einer coverage > 50 in der RNA-Seq.

In Zusammenschau der hier vorliegenden Ergebnisse, insbesondere der reduzierten Expression und dem verschobenen X-Inaktivierungsmuster, wurde die Variante neu als wahrscheinlich pathogen klassifiziert.

Tabelle 4.9: X-Inaktivierungsanalyse bei Individuum MS005 (für Varianten mit coverage >50 in der RNA-Seq)

Trio-Exom								RNA-Seq								
N°	Position	ref	obs	Mut	MS005	Vat	Gen	cov	obs	reads	%	obs	reads	%	X-	Expression
									ref		reads	reads	other	reads	In-	
															aktivierung	
1	chrX: 229493	G	A	het	het	wt	GTPBP6	58	G	27	53%	A	31	47%	pseudo	Vat+Mut
2	chrX: 229542	G	A	het	het	wt	GTPBP6	53	G	30	57%	A	23	43%	pseudo	Vat+Mut
3	chrX: 1531648	C	T	hom	het	wt	ASMTL	65	C	25	38%	T	40	62%	pseudo	Vat+Mut
4	chrX: 2632482	C	T	hom	het	wt	CD99	2406	C	1506	62%	T	897	37%	pseudo	Vat+Mut
5	chrX: 2656258	G	A	het	het	wt	CD99	3684	G	2050	56%	A	1632	44%	pseudo	Vat+Mut
6	chrX: 2700157	G	A	het	het	wt	XG	254	G	188	74%	A	66	26%	pseudo	Vat+Mut
7	chrX: 3228096	C	T	het	het	wt	MXRA5	331	C	323	98%	T	7	2%	abnorm	Vat wt
8	chrX: 3228144	C	G	het	het	wt	MXRA5	248	C	245	99%	G	2	1%	abnorm	Vat wt
9	chrX: 3228378	A	C	het	het	wt	MXRA5	215	A	215	100%	-	-	-	abnorm	Vat wt
10	chrX: 3228382	G	A	het	het	wt	MXRA5	216	G	216	100%	-	-	-	abnorm	Vat wt
11	chrX: 3228420	C	T	het	het	wt	MXRA5	198	C	197	99%	T	1	1%	abnorm	Vat wt

Trio-Exom								RNA-Seq								
N°	Position	ref	obs	Mut	MS005	Vat	Gen	cov	obs ref	reads	% reads	obs other	reads	% reads	X-Inaktivierung	Expression
12	chrX: 3228675	C	T	hom	het	wt	MXRA5	185	C	179	97%	T	6	3%	abnorm	Vat wt
13	chrX: 3228711	G	A	hom	het	wt	MXRA5	340	G	328	96%	A	12	4%	abnorm	Vat wt
14	chrX: 6995417	C	T	het	het	wt	HDHD1	62	C	45	73%	T	17	27%	pseudo	Vat+Mut
15	chrX: 6995438	C	T	het	het	wt	HDHD1	67	C	42	63%	T	25	37%	pseudo	Vat+Mut
16	chrX: 7023678	G	A	het	het	wt	HDHD1	51	G	34	67%	A	17	33%	pseudo	Vat+Mut
17	chrX: 34674970	C	G	wt	het	hom	TMEM47	99	C	3	3%	G	96	97%	abnorm	Vat mut
18	chrX: 47444663	C	T	wt	het	hom	TIMP1, SYN1	2825	C	80	3%	T	2742	97%	abnorm	Vat mut
19	chrX: 48382174	G	T	het	het	wt	EBP	50	G	47	94%	T	3	6%	abnorm	Vat wt
20	chrX: 54841737	G	A	het	het	wt	MAGED 2	481	G	478	99%	-	-	-	abnorm	Vat wt
21	chrX: 114879399	T	C	wt	het	hom	PLS3	254	T	9	4%	C	245	96%	abnorm	Vat mut
22	chrX: 114880423	T	C	wt	het	hom	PLS3	127	T	4	3%	C	123	97%	abnorm	Vat mut
23	chrX:	G	A	wt	het	hom	GPC4	111	G	2	2%	A	109	98%	abnorm	Vat mut

Trio-Exom								RNA-Seq								
N°	Position	ref	obs	Mut	MS005	Vat	Gen	cov	obs ref	reads	% reads	obs other	reads	% reads	X-Inaktivierung	Expression
	132437337															
24	chrX: 132438872	C	A	wt	het	hom	GPC4	53	C	2	4%	A	51	96%	abnorm	Vat mut
25	chrX: 135310785	T	C	het	het	wt	MAP7D3	64	T	62	97%	C	2	3%	abnorm	Vat wt
26	chrX: 148582549	G	A	het	het	wt	IDS	141	G	139	99%	A	2	1%	abnorm	Vat wt
27	chrX: 152771509	C	T	het	het	wt	BGN	191	C	187	98%	T	4	2%	abnorm	Vat wt
28	chrX: 153278550	G	A	het	het	wt	IRAK1	60	G	60	100%	-	-	-	abnorm	Vat wt

ref = Referenzbase, obs = Observierte Base, Mut = Mutter, Vat = Vater, cov = coverage, pseudo = pseudoautosomal, entgeht X-Inaktivierung, wt = Wildtyp, mut = mutiert

5 Diskussion

Seit der Einführung der Technologie des *Next generation sequencing* wurde die Diagnostik von genetisch bedingten Erkrankungen revolutioniert.

Tatsächlich ist es im letzten Jahrzehnt zu einer überschießenden Publikation von jährlich ca. 200 – 300 neuen krankheitsursächlichen Varianten, welche monogenetischen Erbkrankheiten zugeordnet werden können, gekommen (Chong et al., 2015, Boycott et al., 2017). Das Tempo an Aufklärungen im Vergleich zu *konventionellen* genetischen Methoden hat somit seit Einführung des NGS enorm zugenommen – von 2013 bis 2017 wurde durch NGS-Technologien (v.a. WES und WGS) im Vergleich zu klassischen molekulargenetischen Methoden die dreifache Anzahl an Genen in Bezug auf RGDs entdeckt (Boycott et al., 2017).

Dennoch bleibt ein Großteil der vermuteten seltenen genetischen Erkrankungen auch mit den zunehmend verbesserten Technologien im Bereich der Diagnostik weiterhin ungelöst. In älteren Arbeiten findet sich häufig eine Aufklärungsrate von ~ 30 – 50 % (s. bspw. Shashi et al., 2014, Smith et al., 2019). Aktuellere Studien bestätigen jedoch nach wie vor diese Aufklärungsraten. Tenorio et al., 2023 fanden bspw. in ihrem systematischen Review eine diagnostische Ausbeute von 46 % für Fälle von Hereditärer Ataxie und Britten-Jones et al., 2023 konnten positive genetische Diagnosen via NGS bei 47,6 % bis 61,3 % für Fälle von vererbten retinalen Erkrankungen zuordnen (abhängig von der genauen Subklassifikation der Netzhauterkrankung), um nur zwei Gruppen von Erkrankungen mit genetischem Hintergrund zu nennen.

In dieser Arbeit werden insgesamt 16 ungelöste Fälle mit Verdacht auf eine seltene genetisch bedingte Erkrankung mit unterschiedlichem phänotypischem Spektrum vorgestellt. Ziel war die Evaluation von molekulargenetischen Methoden in Bezug auf die Klassifikation von Kandidatengenen, um eine Aussage über den Nutzen zusätzlicher Methoden (in dieser Arbeit RNA-Seq, Western Blots, Validierungen durch

klassische molekulargenetische Verfahren) im Bereich der genetischen Diagnostik treffen zu können (s. Abschnitt 1.5).

Insgesamt konnte bei 63 % der Individuen ein Kandidatengen durch die hier beschriebene Methodik etabliert werden. Bei 44 % der Betroffenen erbrachte die RNA-Seq bei der Etablierung des Kandidatengens weitere Erkenntnisse (vgl. Tabelle 4.4). Eine Variante konnte zudem nach Durchführung der RNA-Seq als wahrscheinlich pathogen bewertet werden. Dennoch bleiben die meisten Fälle mit der in dieser Arbeit vorgestellten Methodik weiterhin ungelöst.

Die Ursachen für diese sowohl in der Literatur und auch in dieser Arbeit weiterhin unbefriedigende Rate an ungelösten Fällen sind in vielfältigen Themenfeldern zu suchen.

5.1 Diskussion der Beispiele aus dem Ergebnisteil

5.1.1 Identifizierung einer Mutation in *DBR1* bei Individuum MS002

Bei Individuum MS002 wurde durch die NGS-Auswertung eine homozygote Missense-Variante (c.461G>A; p.Arg>154Lys) im Gen *DBR1* als Kandidatengen ausgewählt. Die Mutation liegt in einem bekannten Krankheitsgen.

Die Auswahl von *DBR1* gelang vor allem aufgrund des spezifischen Phänotyps in Kombination mit einer passenden Publikation. In Bezug auf MS002 waren zunächst zwei homozygote Varianten in unterschiedlichen Genen (*DBR1* und *PAICS*, s. Tabelle 4.5) interessant, welche beide in der Vergangenheit bereits krankheitsassoziiert wurden. Diese Tatsache ist in der Kandidatengensuche nicht ungewöhnlich – in vielen Studien werden gelegentlich mehrere Varianten in unterschiedlichen Genen als Kandidaten für ein Individuum bzw. eine Familie entdeckt (vgl. bspw. Kahrizi et al., 2019, Carneiro et al., 2018, Monies et al., 2019). Im Zeitalter des NGS ist folglich weniger die Detektion von Varianten ein Problem, als die Interpretation und Validierung von Varianten unklarer Signifikanz (VUS), um diese für die klinische Praxis aufzubereiten und falsch-positive Ergebnisse zu limitieren (Lohmann und Klein, 2014). Literatur und Informationen in Datenbanken

fehlen jedoch bei VUS häufig per definitionem und die Interpretation gestaltet sich daher besonders schwierig (Lohmann und Klein, 2014, Gburek-Augustat et al., 2021). In diesen Fällen ist eine Zusammenschau aller Befunde und eine präzise Beschreibung des Phänotyps essenziell, um valide Aussagen über die Relation zwischen Geno- und Phänotyp treffen zu können und ggf. bei inkongruenten Befunden die genetische Diagnose oder Kandidatengenauswahl in Frage zu stellen (Robinson, 2012, Zschocke et al., 2023).

Der Phänotyp von MS002 scheint auf das Kandidatengen *PAICS* nicht zu passen (vgl. Abschnitt 4.2.2). Dennoch bleibt eine gewisse diagnostische Unsicherheit, da das phänotypische Spektrum von Defekten in der Purinbiosynthese weit ist (Dewulf et al., 2022). Ein milder (klinisch unauffälliger) Phänotyp bei MS002 durch c.665G>C ist jedoch vor dem Hintergrund, dass LoF-Varianten vermutlich pränatal tödlich sind, Missense-Varianten in der bisher einzigen publizierten Familie mutmaßlich für eine hohe Anzahl habitueller Aborte verantwortlich waren und MS002 keinerlei phänotypische Auffälligkeiten bei Geburt aufwies als unwahrscheinlich zu betrachten (Dewulf et al., 2022, Pelet et al., 2019).

Die eindeutige klinische Diagnose einer ANEC in Kombination mit der kürzlich vor Auswertung der Daten publizierten Studie von Zhang et al., 2018 sorgten schließlich für die Entdeckung von c.461G>A, p.Arg>154Lys in *DBR1*. Eine präzise Phänotypisierung ist zwar in jeder genetischen Abklärung essenziell, jedoch im Fall von MS002 nicht ausreichend, um c.461G>A als wahrscheinlich pathogen zu klassifizieren. Die Variante wurde daher als Klasse 3 nach ACMG (VUS) bewertet (vgl. Abschnitt 4.2.2). Erwähnenswert ist in diesem Zusammenhang die Publikation von einer weiteren homozygoten Missense-VUS in *DBR1*, welche ebenfalls zu einer viralen Hirnstammenzephalitis assoziiert wurde (French et al., 2022).

Die Unsicherheit bei der Bewertung der vorliegenden Variante ergibt sich unter anderem aus den unklaren Auswirkungen der Variante auf das Protein. Unterschiedliche Szenarien bei Missense-Varianten sind

grundsätzlich denkbar. Zum einen könnte sich die Variante durch einen rein quantitativen Effekt bemerkbar machen (z.B. über den Abbau der fehlerhaften mRNA, den Abbau des fehlerhaften Proteins oder ein komplett defektes Protein durch Störung der funktionalen Domäne). Diese Art der Auswirkungen einer Missense-Variante wäre analog von LoF-Varianten zu bewerten (vgl. Absatz 5.1.2). Des Weiteren könnte es zur Expression eines stabilen Proteins kommen, welches eine veränderte Funktion besitzt (qualitativer Effekt). Am schwierigsten wird die Interpretation, wenn eine Mischform aus quantitativen und qualitativen Effekten vorliegt (Zschocke et al., 2023).

Ein kompletter LoF von *DBR1* scheint pränatal tödlich zu sein (vgl. Absatz 4.2.2; Zheng et al., 2015), daher kommt für den Pathomechanismus bei einem quantitativen Effekt der vorliegenden Variante lediglich eine partielle Insuffizienz in Frage (sog. hypomorphe Variante). Des Weiteren liegt bei in *normalem* quantitativem Ausmaß exprimierter mRNA ebenfalls die Interpretation nahe, dass der Pathomechanismus von c.461G>A sich über eine Wirkung auf Proteinebene entfaltet.

Mit dem Verdacht einer auf Proteinebene hervorgerufenen partiellen Insuffizienz von *DBR1* bei MS002 wurden Western Blots mit semiquantitativen Bestimmungen von DBR1 aus Proteinlysaten von kultivierten Fibroblasten durchgeführt. Die Detektion von spezifischen Proteinen und die semiquantitative Messung desselbigen in einer Probe ist die häufigste Anwendung von Western Blots in der Wissenschaft (Meftahi et al., 2021). Im Vergleich zu den drei Kontrollen zeigte sich die Proteinmenge von DBR1 bei MS002 um durchschnittlich 76 % reduziert ($p < 0,0001$ ggü. allen Kontrollen). Über den Pathomechanismus dieser signifikanten Reduktion kann anhand der in dieser Arbeit vorgestellten Daten nur spekuliert werden. Ein Abbau der mRNA durch NMD müsste in der Regel in einer verminderten Expression in der RNA-Seq sichtbar sein und scheint somit unwahrscheinlich. Überlegungen zu einem Abbau des fehlerhaften Proteins durch zelluläre Mechanismen müssten zur weiteren Einordnung der Variante in Bezug auf MS002 nachgegangen werden und

kann den in dieser Arbeit präsentierten Daten leider nicht entnommen werden.

Jedoch scheint es sich um eine deutlich hypomorphe Variante zu handeln. Die absolute Zahl der Reduktion an DBR1-Protein deckt sich mit den Ergebnissen der Studie von Zhang et al., 2018, welche für vier unabhängige Mutationen in HEK293T-Zellen ~ 25 % der Menge an Vollängenprotein im Vergleich zum Wildtyp detektieren konnten. Des Weiteren konnten die KollegInnen mittels RT-qPCR gleiche Expressionslevel von WT- und mutierten DBR1-transfizierten Zellen beobachten (Zhang et al., 2018). Dieser Umstand passt sehr gut zu der unauffälligen RNA-Seq im vorliegenden Fall.

Methodisch lässt sich bei der vorliegenden Arbeit die Wahl des *housekeeping*-Gens *GAPDH* in Bezug auf die möglicherweise nicht vorhandene Linearität der Konzentration desselbigen in Bezug auf die totale Proteinkonzentration bemängeln (Gorr und Vogel, 2015). Dennoch zeigt sich das optische Signal von DBR1 für die Patientenzellen bereits in der rein qualitativen Beurteilung der Membranen erniedrigt, sodass in jedem Fall ein quantitativer Effekt von c.461G>A angenommen werden kann. Somit gelingt mithilfe der durchgeführten Western Blots die Zuordnung in die pathophysiologische Gruppe der hypomorphen Mutationen.

Eine weitere Unsicherheit liegt bei MS002 in der infektiologischen Anamnese. Bisher wurden Hirnstammenzephalitiden in Verbindung mit einer partiellen *DBR1*-Insuffizienz lediglich mit Herpes-simplex-Virus-1- (HSV1), Influenza-B-Virus- (IBV), Norovirus- und SARS-Cov2-Infektionen in Zusammenhang gebracht (Zhang et al., 2018, Chan et al., 2024). In den meisten Fällen ist hier der Hirnstamm die einzige betroffene Hirnregion, sodass aktuell von einer Hirnstamm-spezifischen Immundefizienz ausgegangen wird (Zhang, 2020, Zhang et al., 2018). Die Inzidenz neurologischer Manifestationen (inkl. Epilepsien, Status epilepticus, zentrale Apnoe und Enzephalopathie/-itis) in Folge einer RSV-Infektion wird mit 1,2 – 6,5 % angegeben und kann für Kinder unter zwei Jahren auf bis zu 40 % ansteigen (Bohmwald et al., 2016, Saravanos et al., 2021). Die

zugrundeliegende Pathophysiologie und somit, ob das Virus hierbei tatsächlich Neuronen oder andere Zellen aktiv befällt, ist trotz aller wissenschaftlichen Bemühungen weiter unklar (Bohmwald et al., 2016). Ebenso ist die Rolle von *DBR1* bei der Hirnstamm-spezifischen Immunabwehr aufgrund unklarer molekularer Mechanismen ebenso Gegenstand von aktuellen wissenschaftlichen Bemühungen und wirft die Frage auf, ob Immunantwort im Zentralen Nervensystem nach anatomischer Region aufgeteilt ist oder es sich um spezifische Mechanismen gegen dezidierte Viren handelt (Zhang, 2020). Vor dem Hintergrund der RSV-Infektion lässt sich die Kandidatenvariante aufgrund der Ausführungen in diesem Absatz daher lediglich als VUS klassifizieren. Der Fall MS002 verdeutlicht eindrücklich die Problematik der Interpretation einiger genetischen Varianten im Kontext von äußeren Einflüssen (z.B. Infektionen, Umwelteinflüsse, Ernährung, etc.), vor allem wenn sich die Symptomatik ähnlich einer häufig vorkommenden Erkrankung präsentiert. Nichtsdestotrotz bleibt c.461G>A abschließend eine sehr interessante VUS, welche durch die Durchführung von funktionellen Analysen als wahrscheinlich pathogen validiert werden könnte. Diese funktionellen Analysen sollten die Messung der Rest-Enzymaktivität durch Bestimmung der Intron-Lariat-Level in Zellen (z.B. Fibroblasten) von MS002 beinhalten, um die Minderung der Funktion von *DBR1* durch c.461G>A zu erfassen. Eine Etablierung der Messung der Intronlariat-Level in Zusammenarbeit mit Dr. med. Lucia Laugwitz war jedoch bisher nicht erfolgreich. Zusätzlich wären Experimente mit Infektionen in vitro von bspw. Fibroblasten durch RSV und erneutem Assessment der Funktion von *DBR1* sicherlich sinnvoll und notwendig (für beide Methoden vgl. Zhang et al., 2018).

5.1.2 Identifizierung einer Mutation in *SEC31B* bei Individuum MS001

Bei Individuum MS001 wurde im Rahmen der NGS-Auswertung eine homozygote Stopp-Mutation (c.1759G>T; p.Glu587*) in *SEC31B* als Kandidatengen favorisiert. Die Variante war aufgrund einer niedrigen *coverage* im Transkriptom weder bei MS001 noch den Kontrollen bewertbar. Varianten in *SEC31B* wurden bisher zu keiner monogenetischen Erkrankung assoziiert (Recherche in pubmed, Stand: 08.05.2024).

Die oben genannte homozygote Variante wurde unter anderem aufgrund ihres trunkierenden Aspektes ausgewählt. Hierbei wird c.1759G>T; p.Glu587* pathophysiologisch als Mutation betrachtet, die mit einem *loss-of-function* (LoF) einhergehen müsste. Bei LoF-Mutationen handelt es sich um Varianten, welche einen quantitativen Effekt auf die Gendosis durch kompletten Verlust des Proteins aufgrund verschiedener möglicher Mechanismen hervorrufen (Zschocke et al., 2023). Aufgrund dieses Mechanismus sind homozygot vorliegende LoF-Varianten in der Kandidatengensuche besonders interessant.

Bei der Bewertung dieser Varianten ist es von enormer Bedeutung die Gene nach ihrer Haplosuffizienz zu unterscheiden. Dies bedeutet, dass die Reduktion der Gendosis durch heterozygote LoF-Varianten durch das andere Allel kompensiert werden und es zu keinen phänotypischen Ausprägungen im Vergleich zum homozygoten Wildtyp kommt (Zschocke et al., 2023). Ein pLI-Score = 0 beschreibt in diesem Fall die voraussichtliche Haplosuffizienz von *SEC31B* und könnte erklären, warum bis dato > 2000 pLoF-Varianten heterozygot in der gnomAD gelistet sind, *SEC31B* jedoch bisher nicht ursächlich für monogenetische Erkrankungen publiziert wurde. In diesem Kontext kann spekuliert werden, dass es bei *SEC31B* bei Heterozygotie einer trunkierenden Variante zu keiner phänotypischen Ausprägung kommt. Diese Annahme passt mit einem möglichen rezessiven Vererbungsmuster bei MS001 überein, in dessen Familie es bisher keine weiteren betroffenen Geschwister oder Familienmitglieder gibt. Denkbar wäre jedoch auch ein autosomal-dominanter Vererbungsweg in Kombination mit einer *de novo* Mutation

aufgrund dessen es wünschenswert gewesen wäre, eine Segregationsanalyse mit den Eltern und ggf. den gesunden Geschwistern durchzuführen. Dies kann beispielsweise via klassischer Sanger-Sequenzierung oder Trio-Sequenzierung erfolgen.

Im Themenfeld der *Intellectual Disability* (ID) wird der Sequenzierung von Eltern-Kind-Trios zusätzlicher Nutzen in Bezug auf die Aufklärungsrate zugeschrieben (Kahrizi et al., 2019, Wieczorek, 2018, Brea-Fernández et al., 2022). Besonders sporadische Fälle von ID stellen die genetische Diagnostik vor eine besondere Herausforderung, welche sich in einer niedrigen Aufklärungsrate von ca. 15 – 30 % bei Nutzung von WES widerspiegelt (Chérot et al., 2018, Monroe et al., 2016, Volk et al., 2015). Vor allem die allgemein hohe genetische Heterogenität und die höhere Rate an *de novo* Varianten im Vergleich zu autosomal-rezessiven Formen sind hier als Gründe hervorzuheben (Wieczorek, 2018, Kahrizi et al., 2019). Carneiro et al., 2018 verdeutlichen in diesem Kontext in ihrer Arbeit mit insgesamt acht Fällen von sporadischer ID den Nutzen von Trio-Exom-Sequenzierungen bei einer Aufklärungsrate von 62 % und konkludieren, dass vor allem zur Detektion von *de novo* Varianten das Trio ausschlaggebend war (insgesamt drei ProbandInnen). Die homozygoten Mutationen (folglich rezessiv vererbten Erkrankungen) wären nach Einschätzung der AutorInnen auch ohne elterliche Daten detektiert worden. Einschränkend ist zu nennen, dass die beschriebenen Fälle sich auf syndromale Formen der ID beziehen, welche bei MS001 lediglich in milder Form (große Hände, große Füße, vgl. Abschnitt 4.3.1) vorliegt.

Ob eine Trio-WES oder eine Sequenzierung von gesunden Geschwistern für den Fall MS001 zusätzlichen diagnostischen Nutzen erbracht hätte, bleibt zusammenfassend unklar. Eine Segregationsanalyse (auch mittels Sanger-Sequenzierung) war aufgrund fehlender Proben der Angehörigen nicht möglich.

Des Weiteren bleibt ebenfalls offen, ob eine Abklärung mittels WGS zielführend gewesen wäre. Häufigste Ursache von ID sind in ca. 12 – 15 % nach wie vor chromosomale Ursachen (inkl. Trisomie 21, CNVs,

Translokation, etc.; Gupta, 2023, Maia et al., 2021). Ein erneutes WGS von 50 ungelösten Fällen (im Vorfeld abgeklärt mit *genomic microarray* und WES) erbrachte bereits vor 10 Jahren zusätzliche Diagnosen bei 42 % der Kohorte, davon acht *de novo* Strukturvarianten (Gilissen et al., 2014). Eine relevante Strukturvariante kann bei MS001 mit dem vorliegenden WES nicht ausgeschlossen werden. Nichtsdestotrotz stellt sich bei allen wissenschaftlichen Bemühungen, um eine korrekte Diagnosefindung zu ermöglichen, immer auch die Frage der Kosten. Diese wären bei MS001 vermutlich mit zusätzlicher Trio-/Angehörigen-Sequenzierung und/oder Genomsequenzierung vor dem Hintergrund der bereits vorhandenen Detektion eines Kandidatengens im WES zu hoch gewesen.

Der quantitative Effekt auf die Gendosis entfaltet sich bei LoF-Varianten über unterschiedliche Mechanismen. Denkbar bei Vorliegen von c.1759G>T; p.Glu587* wären Instabilität der mRNA mit konsekutivem NMD, Instabilität des Proteins, vorzeitiger Proteinabbau oder Inaktivität der Proteins (Ma et al., 2024, Zschocke et al., 2023). Bis auf die Inaktivität des Proteins führen alle hier aufgeführten Mechanismen den beschriebenen Gendosisverlust über eine Nicht-Expression des Proteins hervor.

Eine zusätzliche Evidenz für *SEC31B* als Kandidatengens sollte daher in Form von Western Blots realisiert werden. Ziel war im Idealfall die Darstellung des Proteins in den Kontrollen nicht jedoch bei MS001, um die resultierende Nullmutation aus c.1759G>T darzustellen.

Leider gelang es in den Western Blot-Analysen aus Fibroblasten trotz Verwendung zweier Primärantikörper nicht, das Protein SEC31B eindeutig zu detektieren. Der Antikörper der Firma abcam wies lediglich unspezifische Banden nach und der Antikörper der Firma Aviva Systems Biology Corporation erbrachte keine Proteindetektion (s. Abschnitt 4.3.3).

Gründe für diese Negativresultate können in vielfältigen methodischen Ursachen liegen. Beginnend mit Artefakten bei der Proteinisolation (bspw. Proteolyse oder Aggregation), über Probleme beim Blotting (bspw. inkompletter Proteintransfer) bis hin zur Immundetektion (bspw.

unspezifische Bindung) liegen viele Fallstricke auf dem Weg zur finalen Detektion des Zielproteins (vgl. Gorr und Vogel, 2015).

Da jedoch das Haushaltsprotein GAPDH bei allen Blots problemlos dargestellt wurde und die Ponceau-Färbung zur Überprüfung des Blotting-Ergebnisses bei allen Membranen positiv ausfiel, scheint es unwahrscheinlich das technische Probleme hier ursächlich waren. Vielmehr liegt auch mit erneutem Blick auf die RNA-Seq-Ergebnisse (niedrige *coverage*) die Spekulation nahe, dass die kultivierten Fibroblasten *SEC31B* nicht oder nicht in ausreichender Menge exprimieren.

Voruntersuchungen mit Proteinlysaten aus anderen Zellreihen (HEK- und RPE-Zellen, sowie CPT-Blut) scheiterten ebenfalls mit der Detektion von unspezifischen Banden, welche dem Zielprotein nicht eindeutig zugeordnet werden konnten (vgl. Abschnitt 4.3.3). Eine alternative Analyse-Pipeline konnte daher für *SEC31B* nicht etabliert werden.

Die Ergebnisse der Western Blots sorgten jedoch nicht für ein Verwerfen der Variante als Kandidatengen für MS001. Interessant in Kombination mit dem Phänotyp könnte hier die hohe Expression von *SEC31B* im Cerebellum (Störung der Motorik), die vermutete Cargo-Selektion und die bisher ungeklärte Funktion der kürzeren Isoform sein (Stankewich et al., 2006). Schlussendlich wären weitere Analysen (u.a. Etablierung eines funktionalen Assays in-vitro) notwendig, um die Variante in *SEC31B* eindeutig als pathogen zu validieren. Diese weiterführenden Untersuchungen müssten ggf. auch aus anderen Geweben oder entsprechend induzierten Zelltypen durchgeführt werden – dies hätte jedoch leider den Rahmen der vorliegenden Arbeit gesprengt.

5.1.3 Identifizierung einer Mutation in *OSCP1* bei Familie 01

Bei Familie 01 wurde eine homozygote Missense-Mutation im Gen *OSCP1* (c.925C>T; p.Arg309Trp) als Kandidatengen ausgewählt. *OSCP1* scheint in dieser Arbeit bedingt durch fehlende Evidenz in der Literatur und die Wahl der Auswertungstechnik ein sehr unsicheres Kandidatengen zu sein.

Bei Konsanguinität der Eltern steigt das Risiko für autosomal-rezessive Erkrankungen, dementsprechend steigt die Anzahl der zu bewertenden Varianten an (Temaj et al., 2022, Mezzi et al., 2024, Khayat et al., 2024). Außerdem konnte in den vor dieser Studie bereits erfolgten diagnostischen Auswertungen kein Kandidatengen anhand der genomischen Daten gefunden werden. Unter diesen Umständen ist eine rein manuelle Auswertung der RNA-Seq anhand der detektierten genomischen Varianten als unglücklich anzusehen. Die schiere Anzahl der zu bewertenden Varianten ist zeitaufwendig, anfällig für menschliches Versagen (ergo Übersehen von Effekten) und die exakte Auswertung der RNA-Seq erfordert Kandidatenvarianten in der initialen genomischen Analyse (Murdock et al., 2021). Außerdem kommen falsch-negative Resultate im WES/WGS zwar selten vor, können aber dennoch zusätzlich eine Fehlerquelle bei dieser Art der Auswertung darstellen und für falsch-negative Resultate sorgen (Goldfeder et al., 2016, Sun et al., 2015). Daher wäre es sinnvoll gewesen, automatisierte bioinformatische Berechnungen zur Detektion abnormer Transkripte zu etablieren. Vor diesem Hintergrund richtet sich besondere Aufmerksamkeit zunehmend der Detektion von abnormalen Spleißevents, da fehlerhaftes Spleißen in bis zu 15 – 30 % als krankheitsverursachend für Erbkrankheiten vermutet wird (López-Bigas et al., 2005, Wang und Cooper, 2007, Rowlands et al., 2021, Mertes et al., 2021). Die Entdeckung von aberranten *de novo* Spleißevents hat bereits einige molekulargenetische Diagnosen in der Vergangenheit ermöglicht (Cummings et al., 2017, Kremer et al., 2017). Mehrere Algorithmen wurden in der Folge zur transkriptomweiten Detektion ebendieser Events entwickelt bzw. weiterentwickelt (bspw. LeafCutterMD, SPOT, FRASER; Jenkinson et al., 2020, Ferraro et al., 2020, Mertes et al., 2021). Selbstverständlich wäre

es wünschenswert gewesen, die RNA-Seq-Daten der vorliegenden Kohorte auch mit einem der genannten Algorithmen auszuwerten. Mit Blick auf Familie 01 würden die oben genannten Berechnungen anhand der Transkriptom-Daten parallel zum Multi-WGS sicherlich die Sensitivität bei der Kandidatengensuche und die Sicherheit in der Bewertung der Kandidatenvarianten erhöhen. Dies gilt jedoch ebenso in Bezug auf die restliche Kohorte.

Trotz der methodischen Kritikpunkte erschien *OSCP1* in Bezug auf Familie 01 aufgrund der kürzlich erfolgten Zuordnung zu Zilien hochinteressant (van Dam et al., 2019). Problematisch in der Interpretation ist jedoch der fehlende Phänotyp einer klassischen Ziliopathie (s. Abschnitt 4.4.2). Interessanterweise scheint *OSCP1* jedoch in Würmern und Zebrafischen unterschiedliche Funktionen zu erfüllen – *OSCP1*^{-/-}-Zebrafische zeigen einen spezifischen zilienassozierten Phänotyp, wohingegen bei *OSCP1*^{-/-}-*C. elegans* sich intakte Zilien finden (van Dam et al., 2019). Somit scheint eine Involvierung von *OSCP1* als Ursache für den Phänotyp der Brüder trotz bisheriger negativer Diagnostik in Bezug auf Ziliopathien (Bürstenabstrich) nicht ausgeschlossen. Dennoch bleiben viele Fragen ungeklärt, z.B. wie der Phänotyp einer Lungenfibrose durch eine Ziliopathie ausgelöst wird oder ob der Phänotyp doch auf eine der anderen bis dato publizierte Funktionen von *OSCP1* zurückzuführen ist (vgl. Ausführungen im Abschnitt 4.4.2). Hier kann letztendlich anhand der aktuellen Datenlage nur spekuliert werden. Zur Validierung von *OSCP1* als Ursache der fibrosierenden Lungengerüsterkrankung wäre weitere Studien bspw. in einem Modellorganismus notwendig.

Schlussendlich bleibt *OSCP1* in dieser Arbeit ein sehr unsicheres Kandidatengen. Vor diesem Hintergrund ist von enormer Relevanz, die Daten von ungelösten Fällen regelmäßig zu reevaluieren und erneut auszuwerten, da neue wissenschaftliche Erkenntnisse oder neue Methoden in einer erneuten Auswertung für eine molekulare Diagnose sorgen könnten (Matalonga et al., 2021, Zurek et al., 2021, Laurie et al., 2025).

5.1.4 Identifizierung einer Mutation in *HUWE1* bei Individuum MS005

Bei MS005 konnte eine heterozygot vorliegende X-chromosomale Spleißstellenmutation in *HUWE1* (c.2496+3A>G) als wahrscheinlich pathogene Variante postuliert werden.

HUWE1 ist ein bekanntes Krankheitsgen. Sowohl Mikroduplikationen, Missense- und Spleißstellen-Varianten wurden in der Vergangenheit zur X-chromosomalen geistigen Behinderung (*X-linked intellectual disability*, XLID) assoziiert (Froyen et al., 2008, Froyen et al., 2012, Moortgat et al., 2018). Das klinische Spektrum ist weit und reicht von milder XLID bis hin zu schwer betroffenen Kindern (Froyen et al., 2008, Taylor et al., 2015, Moortgat et al., 2018). Viele der Betroffenen zeigen syndromale Auffälligkeiten, wobei faziale Dysmorphien mit Augenanomalien die markantesten Merkmale darzustellen scheinen (Moortgat et al., 2018). Selbstverständlich ist die Beschreibung des Phänotyps in dieser Arbeit als insuffizient anzusehen, da die Beschreibung wichtiger klinischer/syndromaler Merkmale fehlt (vgl. Absatz 4.5.1). Auf das Thema der Bedeutung der exakten Phänotypisierung wurde jedoch in Abschnitt 5.1.1 (Bedeutung der Phänotypisierung für die Differenzierung zwischen unterschiedlichen Kandidatengenen im Fall von MS002) bereits ausführlich eingegangen. Der klinische Phänotyp von MS005 könnte grundsätzlich zu den bisher publizierten Merkmalen in Verbindung mit *HUWE1*-Mutationen passen (Globale Entwicklungsverzögerung und Stereotypien, vgl. Moortgat et al., 2018) und es konnten im Rahmen dieser Arbeit keine *besser passenden* Kandidatenvarianten detektiert werden. Falsch negative Resultate kommen zwar beim WES sehr selten vor, könnten aber natürlich in diesem Design der Auswertung eine Rolle spielen (Sun et al., 2015).

Die *de novo* Variante c.2496+3A>G, welche bei MS005 detektiert werden konnte, ist ein Beispiel für den Vorteil von Trio-basiertem WES. Gerade im Bereich der Forschung zum Thema ID konnten in der Vergangenheit durch Trio-Exom-Sequenzierungen höhere Diagnoseraten als mit WES allein erzielt werden und deutlich häufiger *de novo* Varianten als pathogen

validiert werden (vgl. auch Ausführungen zu MS001 in Abschnitt 5.1.2; Farwell et al., 2015, Carneiro et al., 2018, Brea-Fernández et al., 2022).

Pathophysiologisch wurden weibliche Individuen lange Zeit in Bezug auf die *HUWE1*-assoziierte ID als asymptomatische oder niedrig symptomatische Carrier aufgrund einer möglicherweise vorteilhaften X-Inaktivierung zugunsten des gesunden Allels verstanden (Froyen et al., 2012). Diese Annahme wurde durch die Publikation der ersten weiblichen Betroffenen mit Lernschwierigkeiten und Kraniosynostose in Kombination mit einer *de novo* Missense-Variante in *HUWE1* stark in Frage gestellt (Taylor et al., 2015). Tatsächlich konnte bis dato bei den schwer betroffenen weiblichen Individuen in nahezu allen Fällen eine gestörte X-Inaktivierung nachgewiesen werden (Taylor et al., 2015, Moortgat et al., 2018). In der Kohorte von Moortgat et al. zeigten die Probandinnen wechselnde Inaktivierungsmuster, sowohl in Bezug auf die vorwiegende Expression der Mutation selbst oder des Wildtyp-Allels, als auch eine unterschiedlich stark gestörte Inaktivierung in unterschiedlichen Geweben (Moortgat et al., 2018).

Dieses Pattern deckt sich mit den Befunden bei MS005, welche ebenfalls eine hoch präferentielle Expression (98/2 %) zugunsten des väterlich vererbten X-Chromosoms in der RNA-Seq aus Fibroblasten zeigt (vgl. Abschnitt 4.5.3). Problematisch in der Interpretation dieser gestörten X-Inaktivierung ist jedoch die nicht-vorhandene Expression der Spleißstellenvariante. Fraglich ist ebenso, ob die Variante auf dem väterlich oder auf dem mütterlich vererbten Allel liegt. In Fibroblasten wird lediglich das gesunde Allel exprimiert – ein Spleißdefekt konnte nicht nachgewiesen werden (vgl. Abschnitt 4.5.2). Stattdessen zeigte sich eine signifikant reduzierte Expression von *HUWE1*, was auf den Abbau eines fehlerhaft gespleißten Transkripts hinweist. Aufgrund der reduzierten Expression und der verschobenen X-Inaktivierung wurde die Variante neu als wahrscheinlich pathogen klassifiziert. Eine Transkriptomsequenzierung in einem anderen Gewebe könnte ggf. einen Spleißdefekt nachweisen und

verdeutlicht die Bedeutsamkeit der Wahl der Gewebeart für etwaige Transkriptom-Analysen (Frésard et al., 2019, Murdock et al., 2021).

5.2 Gibt es eine ideale Pipeline für ungelöste Fälle?

Die Aufklärung der molekulargenetischen Ursachen für ungelöste Fälle im Bereich der RGDs stellt ForscherInnen trotz sich stetig verbessernden Methoden vor eine große Herausforderung. Unterschiedlichste Pipelines zur Analyse von großen Kohorten wurden in den letzten Jahren angewendet und mit unterschiedlichen Erfolgsraten publiziert. Anhand der vorherigen Fallbeispiele mit den spezifischen Problemen in der Etablierung von Kandidatengenen in dieser Arbeit, sollen nun essenzielle Entscheidungen in der Wahl der Pipeline abschließend diskutiert werden.

5.2.1 Phänotypisierung

Der erste Schritt jeder medizinischen Konsultation stellt die eingehende Anamnese und körperliche Untersuchung dar. Anschließend erfolgen bei Bedarf weiterführende Untersuchungen wie bspw. Blutentnahme und Labortests, Bildgebung und psychologische Testung, um letztendlich eine Diagnosestellung zu ermöglichen. Diesen Prozess, die Abweichungen von *normaler* Physiologie, Morphologie oder Verhalten im Kontext einer genetischen Erkrankung zu beschreiben, fasst man als Phänotypisierung zusammen (Robinson, 2012). So wichtig robuste phänotypische Daten sind, um Geno-Phänotyp-Korrelationen zu etablieren, so schwierig kann in der klinischen Praxis die Differentialdiagnose sein (Hennekam und Biesecker, 2012). Bei RGDs wird dies zudem durch die Seltenheit der Erkrankungen und die phänotypische Variabilität erschwert (Chong et al., 2015).

Deep Phenotyping ist die sorgfältige Aufarbeitung, präzise Beschreibung und oft technisch gestützte Charakterisierung des Phänotyps (Delude, 2015). Hierbei wird i. d. R. als Standard auf das Vokabular des *Human Phenotype Ontology* (sog. HPO-Terms) zurückgegriffen, um Phänotypen untereinander vergleichbar zu machen (Weng et al., 2020).

Die Bedeutung von *Deep Phenotyping* wurde erst kürzlich in einer Studie mit 172 pädiatrischen PatientInnen untermauert, welche mittels WGS (tlw. auch mit Trio-WGS) abgeklärt wurden (Akgun-Dogan et al., 2024). Durch extensive Konsultationen (ca. 4x pro Patient, ca. 1h pro Termin) vor und nach Sequenzierung mit einem einzelnen klinisch tätigen Genetiker konnten viele VUS durch *Deep* und *Reverse Phenotyping* (mehr als 10 HPO-Terms pro PatientIn) aufgeklärt werden und eine diagnostische Ausbeute von 61 % erreicht werden (Akgun-Dogan et al., 2024).

Auch in dieser Studie konnte anhand des Beispiels von MS002 (vgl. Abschnitt 5.1.1) gezeigt werden, dass eine korrekte Phänotypisierung die Grundlage für die molekulargenetische Diagnose bildet und die Kandidatengensuche im Falle von mehreren interessanten Varianten deutlich erleichtert.

Letztendlich sind genomische Daten ohne präzisen Phänotyp allenfalls von limitierter Aussagekraft (Boycott et al., 2017). Dies wird in dieser Arbeit durch insuffiziente Daten bezüglich der Krankheitsgeschichte bei einigen Fällen (vgl. Phänotypentabellen in Abschnitt 1.6 und Diskussion zum Phänotyp von Individuum MS005 in Abschnitt 5.1.4) verdeutlicht und sollte daher bei der Reanalyse der Daten (s. Abschnitt 5.2.5) verbessert werden (sog. *reverse phenotyping*).

Zusammenfassend stellen die ausführliche Beschreibung und korrekte Darstellung des Phänotyps die Grundlage jeder genetischen Diagnose dar. Im Idealfall sollte daher *Deep phenotyping* in keiner Studie zu ungelösten Fällen mit vermuteten RGDs fehlen und sollte von erfahrenen AnwenderInnen durchgeführt werden (Akgun-Dogan et al., 2024). Dennoch kann die interdisziplinäre Zusammenarbeit (v.a. auch mit auswärtig tätigen KollegInnen von externen Einsendern) die verfügbaren Informationen über den klinischen Status der/des Betroffenen sehr limitieren. Die Etablierung eines allgemeingültigen Standards zur Definition und zum Austausch phänotypischer Daten wird daher vom *International Rare Diseases Research Consortium* als kritisch für die weitere Forschung im Bereich der RGDs gesehen (Boycott et al., 2017).

5.2.2 Wahl des molekulargenetischen Verfahren zur Detektion genomischer Varianten

Die Mehrzahl der krankheitsauslösenden Varianten von monogenetischen Erkrankungen liegen vermutlich in proteinkodierenden Sequenzen (Precone et al., 2015, Wortmann et al., 2022). Daher wurde lange Zeit das WES als Hochdurchsatzsequenzierungsverfahren der ersten Wahl bei vermuteter monogenetischer Erbkrankheit empfohlen. Je nach Krankheitsspektrum hat das WES im Vergleich zu *klassischen* zytogenetischen und molekulargenetischen Methoden (3 % für Karyogramm mit Giemsa-Färbung und 15 – 20 % für hochauflösende *Array-CGH (Array-based Comparative Genomic Hybridization; komparative genomische Hybridisierung)* mit ca. 24 – 52 % im Bereich der undiagnostizierten RGDs eine deutlich höhere Aufklärungsrate (Miller et al., 2010, Zhong et al., 2021). Bei klarer Anamnese und klinischem Verdacht auf eine genetische Erkrankung, welche mit einem der oben genannten genetischen Verfahren diagnostiziert werden kann, besitzen diese Verfahren natürlich weiterhin ihren Stellenwert (z.B. die Karyotypisierung in der Diagnose des Down-Syndrom) und sind wesentlich kostengünstiger (Miller et al., 2010). Da diese Arbeit sich jedoch insbesondere mit der Methodik des NGS beschäftigt, soll in den folgenden Absätzen lediglich auf die Vor- und Nachteile von Panel-Untersuchungen, WGS und WES eingegangen werden.

Lange Zeit aufgrund der höheren Kosten nur speziellen Fragestellungen vorbehalten, rückt in den letzten Jahren zunehmend das WGS in den Fokus des wissenschaftlichen Interesses. Die Vorteile des WGS sind nicht von der Hand zu weisen und bestehen in der Erfassung von intronisch liegenden Varianten, der besseren Detektion von Strukturvarianten und der gleichmäßigeren Abbildung des Exoms (Sun et al., 2015). In der Vergangenheit konnte außerdem bereits gezeigt werden, dass Varianten durch WGS detektiert werden können, welche bei alleiniger Sequenzierung des Exoms verpasst worden wären (Belkadi et al., 2015). Problematisch bei der Wahl des WGS als Erstlinien-Sequenzierung wird die tlw. als schwierig

beschriebene Interpretation vor allem von tief intronisch liegenden Varianten beschrieben, sodass die Anzahl der Detektionen von VUS potenziell steigt. Das WGS sollte aus diesem Grund vor allem Instituten und Forschungsgruppen vorbehalten sein, welche eine große Expertise in der Interpretation von VUS aufweisen oder die Möglichkeit besitzen, entsprechende Validierungsstudien durchzuführen oder weitere molekulargenetische Methoden (z.B. RNA-Seq) anzuschließen. In diesen Fällen kann jedoch mit einer höheren Erfolgsrate der Aufklärungen im Vergleich zum WES gerechnet werden (Zhong et al., 2021).

Die Wahl des *Exome Sequencing* (ES) resultiert in repräsentativen Studien in einer Aufklärungsrate von ca. 24-52 % für undiagnostizierte Fälle (Zhong et al., 2021). Hierbei muss unterschieden werden, ob das ES als WES oder als Panel-Untersuchung durchgeführt wird. Letztere kann aufgrund der niedrigeren Kosten eine valide Wahl bei typischem Phänotyp darstellen, die UntersucherInnen sollten sich jedoch im Klaren sein, dass Varianten in bis dato nicht beschriebenen Genen hierdurch unter Umständen verpasst werden. Für ungelöste Fälle v. a. bei heterogener Krankheitspräsentation stellt daher das Exom-Panel in der Regel keine geeignete Untersuchungsmethode dar (Zhong et al., 2021).

Wie bereits in Abschnitt 5.1.2 für den Fall von MS001 diskutiert, stellt sich die Frage, ob die Detektion der kausalen Variante das Problem darstellt oder das Verständnis der pathophysiologischen Auswirkung auf den Organismus insuffizient ist und somit, ob ein Wissenszuwachs durch das WGS generiert werden kann oder ein WES eventuell ausreichend ist. Das WGS hat einen klaren Vorteil bei negativem ES oder rezessiv vererbten Erkrankungen mit Detektion lediglich einer Mutation und Verdacht auf eine kausale intronisch liegende compound-heterozygote Variante oder eine compound-heterozygote Strukturvariante aufgrund der Möglichkeit der Detektion dieser Varianten.

Im Bereich der ungelösten Fälle kann dem WGS ebenfalls eine Stellung als Verfahren mit höherer diagnostischer Ausbeute als bei alleinigem WES

eingesäumt werden (Wortmann et al., 2022, Zhong et al., 2021). Sicherlich wird daher der Trend hin zum WGS in den nächsten Jahren fortgeführt werden, um ein umfassenderes Verständnis im Bereich der RGDs zu entwickeln.

Dennoch kann die Frage, ob WES oder direkt WGS als erstes Verfahren gewählt werden sollte, vor dem Hintergrund der aktuellen Literatur und der Daten dieser Arbeit nicht abschließend beantwortet werden. Vor allem in Ländern mit schlechterer medizinischer Versorgung und somit der Notwendigkeit kosteneffektiv zu arbeiten, scheint nach wie vor das WES eine gute Möglichkeit der Erstlinien-Abklärung zu sein. Diese Kontroversität hält aktuell auch Einzug in entsprechende Empfehlungen der ACMG – Exom- oder Genomsequenzierung werden hier als first- oder second-line Verfahren (nach *chromosomal microarray analysis* oder fokussierter Testung) bei kongenitalen Anomalien im Kindesalter in Anlehnung an die Expertenmeinung der KlinikerIn empfohlen (Manickam et al., 2021).

Sehr sinnvoll kann die Trio- oder Multi-Sequenzierung mit Eltern und/oder weiteren betroffenen Geschwistern/Familienangehörigen sein, um die große Anzahl der zu interpretierenden Varianten weiter einzugrenzen (Carneiro et al., 2018, Brea-Fernández et al., 2022). Eindrucksvoll konnten die Vorteile dieser Methodik anhand des Falles von MS005 (Detektion einer *de novo* Variante mit pathologischem X-Inaktivierungsmuster (Kombination mit RNA-Seq erforderlich)) und Familie 01 (Eingrenzung der vielen Varianten im Bereich der Konsanguinität durch Sequenzierung von mehreren betroffenen Geschwistern) demonstriert werden. Bei dieser Technik ist es nicht nur möglich *de novo* und compound-heterozygote Varianten zu erkennen, sondern auch die atypische Inaktivierung oder epigenetischen Vererbungsmuster in Kombination mit einer RNA-Seq nachzuvollziehen. Automatisierte bioinformatische Berechnungen wären in diesem Bereich, wie bereits ausführlich in Bezug auf die Auswertung bei Familie 01 diskutiert, von Vorteil (vgl. Abschnitt 5.1.3). Trio-

Sequenzierungen können die diagnostische Effektivität steigern und sind vor allem im Erkrankungsspektrum mit hohem Detektionsanteil an *de novo* Varianten etabliert.

Angemerkt sei an dieser Stelle, dass in jedem Fall bei vermuteter dominanter Vererbung oder Verdacht auf Compound-Heterozygotie eine Segregationsanalyse der Eltern (sofern Untersuchungsmaterial verfügbar) angestrebt werden sollte.

5.2.3 RNA-Seq

Aktuelle Studien konnten einen eindeutigen Zuwachs der molekulargenetischen Diagnosestellungen von 10 – 36 % durch die zusätzliche Anwendung von RNA-Seq verzeichnen (Kremer et al., 2017, Cummings et al., 2017, Gonorazky et al., 2019, Murdock et al., 2021). Die Vorteile der Transkriptom-Analyse in Zusammenschau mit den genomischen Befunden liegt in der direkten Interpretation der Varianten auf Expression und Spleißveränderungen.

Auch in dieser Arbeit können mehrere Fälle (MS005 und Familie 01) als Beispiele für unerwartete Auswirkungen von Varianten auf das Transkriptom herangezogen werden. Besonders bei Spleißveränderungen aufgrund tief exonisch liegenden Varianten und der X-Inaktivierungsanalyse entfaltet die RNA-Seq in den genannten Beispielen ihre Vorteile. In aktuellen Arbeiten finden sich außerdem gute Daten zur Interpretation von tief intronisch liegenden Varianten, welche zu Exon-Skipping, Einführung von kryptischen Exons oder Intron-Retention führen (Bournazos et al., 2022). Als allgemeiner Konsens wird vermutet, dass 15 – 60 % der genetischen Varianten ihre desaströse Wirkung über Veränderung des Spleißens entfalten (Wang und Cooper, 2007, Shiraishi et al., 2022). Somit könnte demnach die RNA-Seq ein guter ergänzender Baustein zum WGS sein, um die bereits in Absatz 5.2.2 angesprochene schwierige Interpretation der vielen Varianten des WGS zu erleichtern.

Ausdruck der fortschreitenden Entwicklungen in diesem Bereich der molekulargenetischen Diagnostik sind die mittlerweile verfügbaren

Algorithmen zur bioinformatischen Auswertung der Transkriptom-Daten (bspw. auf Expressions- und Spleißunterschiede oder epigenetische Alterationen). In dieser Arbeit wurden diese Analyse-Tools nicht verwendet und dennoch konnte eine Rate von 44 % an Zusatzerkenntnissen in Bezug auf Kandidatengene erzielt werden. Diese Rate kann selbstverständlich nicht als Diagnoserate verstanden werden und ist damit nicht mit aktuellen Studien vergleichbar, sondern bezieht sich größtenteils auf VUS. Wünschenswert wäre jedoch die Verwendung von entsprechenden Tools in Pipelines für ungelöste monogenetische Erkrankungen.

Zusammenfassend wirft die Transkriptom-Sequenzierung jedoch auch viele weiterführende Fragen auf und ist genauso wie WES/WGS nicht trivial in der Interpretation. Vor allem die Expression unterschiedlichster Gene in diversen Gewebearten, nicht-vergleichbare Konditionen in der Kultur von Zellen für entsprechende Analysen und Befunde, welche klassischen genetischen Theorien widersprechen, stellen ForscherInnen vor Herausforderungen und bedingen die oben erwähnte Diagnoserate von *nur* 10 – 36 %. Jedoch hat die RNA-Seq im Laufe der letzten Jahre ihren Stellenwert unter Beweis gestellt und sollte in einer Pipeline für ungelöste Fälle auf jeden Fall in Betracht gezogen werden.

5.2.4 Multiomics Approach

Neben WES/WGS und Transkriptom-Analysen erfreuen sich in letzter Zeit auch sogenannte Multiomics-Ansätze zunehmender Beliebtheit in der genetischen Forschung, um die Interpretation von VUS weiter zu verbessern.

Vor diesem Hintergrund konnte durch die Arbeitsgruppe um Kopajtich und KollegInnen eine Diagnose bei 22 % WES/WGS-negativen Fällen (n = 121) durch eine Kombination aus *Genomics*, *Transcriptomics* und *Proteomics* erreicht werden (Kopajtich et al., 2021). Die Daten der Studie verdeutlichen vor allem den Vorteil von Proteinanalysen im Bereich der Missense-VUS, welche keinen Effekt auf die RNA haben und somit bei einem alleinigem

Einsatz von molekulargenetischen Sequenziermethoden nicht eingeordnet werden können.

Ebenso konnte kürzlich eine Studie den Einsatz von *Metabolomics* als hilfreich in der Interpretation von 43,5 % der vorgestellten Kohorte hervorheben (Alaimo et al., 2020). Die AutorInnen machen in ihrer Studie zusätzlich deutlich, dass es bei der Mehrzahl der Kandidatengene bisher keine validierten funktionellen Tests gibt, sodass ein ungezielter Ansatz von *Metabolomics* zur näheren Einordnung von Varianten sinnvoll sein kann (Alaimo et al., 2020).

Diese zwei Arbeiten verdeutlichen, dass Multiomics-Ansätze im Spektrum der ungelösten genetischen Fälle für zusätzliche Evidenz für oder gegen Kandidatengene sorgen können. Andere Omics-Technologien (z.B. Lipidomics, Glycomics, etc.) stecken jedoch noch in den Kinderschuhen und benötigen weitere Forschung, um im klinisch-wissenschaftlichen Setting Diagnosestellungen zu ermöglichen (Wortmann et al., 2022).

In Bezug auf die in dieser Arbeit vorgestellten Fälle erscheinen die Omics-Technologien wie die fehlende zusätzliche Ebene, welche eine Diagnosestellung unter Umständen ermöglichen und ggf. auch die unklaren Befunde aus den RNA-Seq-Daten aufklären könnten.

Schlussendlich sollten diese Technologien auf jeden Fall in Betracht gezogen werden, sofern keine funktionale Testung (sowohl in vitro als auch in vivo eines Modellorganismus) möglich ist oder ein Test nicht etabliert werden kann. Des Weiteren könnten Omics in großen Kohorten mit vermuteten monogenetischen Erkrankungen vor allem im Bereich der Missense und heterozygoten Varianten Evidenz für oder gegen Kandidatengene bereitstellen und sollten somit im Baukasten der Pipeline für ungelöste Fälle nicht fehlen.

5.2.5 Reanalyse der Daten

Regelmäßige Reanalyse von ungelösten Fällen konnte in der Vergangenheit in einigen Studien zur Auflösung von Fällen beitragen. Die Aufklärungsrate bei systematischer Herangehensweise beträgt je nach

Arbeit bis zu 15 % (Nambot et al., 2018, Matalonga et al., 2021, Laurie et al., 2025). Aktuell wird von Experten eine Reanalyse alle 1 – 3 Jahre empfohlen und sollte gemeinsam mit aktualisierten phänotypischen Daten (sog. *reverse phenotyping*) erfolgen (Wortmann et al., 2022).

Bei dieser Reanalyse muss die Datenqualität insbesondere von älteren WES beachtet werden, da die verwendeten Kits zur Anreicherung des proteinkodierenden Genoms deutlich weniger annotierte Gene abdecken als die entsprechenden Nachfolgeversionen (Wortmann et al., 2022).

Als Paradebeispiel kann in diesem Zusammenhang das *Solve-rare disease Consortium* (SOLVE-RD) genannt werden, welches in einer groß angelegten europaweiten Studie Datensets von 6.449 Individuen mit einer seltenen Erkrankung mit einem Erfolg in 12,6 % der Familien reanalysiert hat (Laurie et al., 2025).

Zusammenfassend sollte eine Reanalyse vor jedem weiteren molekulargenetischen Verfahren erfolgen, um sicherzustellen, dass einerseits nicht doch eine molekulare Diagnose mittels des ursprünglichen Ansatzes erfolgen kann und zweitens sich der Aufwand des neuen angestrebten Verfahrens überhaupt lohnt. Eine Reanalyse der Daten war mit 42,8 Min je Individuum in der Studie von Solve-RD bewältigbar (Laurie et al., 2025).

5.3 Schlussfolgerung und Ausblick

In Anbetracht der in dieser Arbeit vorgestellten Resultate bleibt festzustellen, dass die vorgestellte Pipeline sich sehr gut eignet, um Kandidatengene zu finden, jedoch deren Interpretation nur teilweise vereinfacht hat.

Zwar konnte für einen Großteil der beschriebenen Kohorte eine Kandidatenvariante priorisiert werden, jedoch konnte nur in einem Fall eine molekulare Diagnose gestellt werden. Dennoch stellt die Etablierung von VUS, welche durch nachfolgende Studien validiert werden können, die erste Maßnahme in der molekulargenetischen Diagnosefindung dar. Der letzte Schritt, die Validierung der Pathogenität einer Variante in Bezug auf

einen bestimmten Phänotyp, erscheint in dieser Kohorte als der entscheidende Engpass. Dies stützt somit die Aussage von Lohmann und Klein, dass nicht die Detektion von Varianten in der Ära des NGS das Problem darstellt, sondern deren Interpretation problematisch sein kann (Lohmann und Klein, 2014).

Diese Schwierigkeiten in der Etablierung einer Pipeline für ungelöste Fälle zur Kandidatengensuche wurden ausführlich diskutiert. Angefangen bei der Phänotypenbeschreibung, über die Wahl des molekulargenetischen Verfahrens mit Vor- und Nachteilen bis hin zu möglichen funktionellen Analysen, um die Konsequenzen der detektierten Varianten zu bestimmen, sind viele Entscheidungen mit entsprechenden Fallstricken auf dem Weg zu einer molekulargenetischen Diagnose aller bis dato ungelösten Fälle zu bedenken.

Die Frage, ob es eine ideale Pipeline für diese Konstellationen gibt, kann nach aktuellem Stand der Wissenschaft verneint werden. Derzeit werden unterschiedlichste Ansätze mit variablen Erfolgsraten parallel publiziert, sodass bspw. selbst die aktuellen Empfehlungen der ACMG für pädiatrische PatientInnen sich zwar für die Nutzung von NGS als genomisches Sequenzierverfahren ausspricht, jedoch kein klares Statement für ES oder GS, noch für den Zeitpunkt (Erstlinien- vs. Zweitlinien-Sequenzierung) abgibt (Manickam et al., 2021). Die Basis für diese Form der Empfehlung liegen in der Natur der RGDs begründet – die genetische Heterogenität und die Seltenheit sind herausfordernd für die Diagnostik und sorgen gleichzeitig dafür, dass Evidenz lediglich in Fallberichten oder -serien vorliegt und RCTs aufgrund fehlender Vergleichskohorten häufig fehlen. Nichtsdestotrotz werden evidenzbasierte Guidelines wiederholt als entscheidender Baustein für die Weiterentwicklung des Forschungsfeldes gesehen, um für eine einheitlichere Nutzung von verfügbarer Diagnostik und Interpretation derselbigen zu sorgen. Das übergeordnete Ziel – evidenzbasierte Medizin zu leisten, im Idealfall Therapieoptionen anbieten zu können und die Last einer genetischen Erkrankung für die Familie, aber auch im

sozioökonomischen Kontext zu reduzieren – steht bei allen wissenschaftlichen Bemühungen an erster Stelle.

Um diesem Ziel näher nachzukommen, hat *The International Rare Diseases Research Consortium* (IRDiRC) bei seiner Gründung 2011 die Intention ausgerufen, die Ursache aller RGDs aufzuklären und damit den Grundstein, dieses Wissen in der Patientenversorgung zu integrieren, gelegt (Boycott et al., 2017). Hierfür muss die Forschungscommunity in den nächsten Jahren weiterhin einen gewaltigen Kraftakt leisten und ihre Methodik und Interpretation von Forschungsergebnissen laufend verbessern, um der Komplexität des menschlichen Organismus nachzukommen.

Als aktuelle kritische Entwicklungsfelder gelten die in dieser Arbeit angesprochenen Themen (Klinische Daten/*deep phenotyping*, Standardisierte Aufbereitung und Interpretation genomischer Daten, Aufbau von funktioneller Evidenz, Ausbau des Verständnisses von neuen Krankheitsmechanismen, etc). Des Weiteren wird der internationalen Kooperation und damit dem systematischen Austausch von entsprechend standardisierten Daten eine kritische Rolle zugeschrieben (Laurie et al., 2025).

Die nächsten Schritte für die in dieser Arbeit vorgestellten Fälle sollten somit in der Reanalyse der Daten vor dem jeweils aktuellen wissenschaftlichen Hintergrund, der Evaluation von funktionellen Analysen und dem Datenaustausch mit anderen genetischen Instituten oder Forschungsgruppen liegen.

6 Zusammenfassung

Seltene Erkrankungen, welche in Europa durch eine Prävalenz kleiner als 1 : 2.000 definiert sind, stellen in ihrer Gesamtheit sowohl eine große sozioökonomische Belastung, als auch eine individuelle Belastung der Betroffenen dar. Seltene Erkrankungen mit genetischem Ursprung sind hierbei ein bedeutende Untergruppe und machen ca. 70 % aus. Die Diagnostikraten der RGDs haben sich im letzten Jahrzehnt seit Einführung der Technologie des *Next generation sequencing* enorm verbessert, sind aber mit einer Erfolgsrate von ca. 50 % je nach Studiendesign nach wie vor als insuffizient einzustufen. Diese niedrige Diagnoserate äußert sich häufig in einer diagnostischen Odyssee für die Betroffenen und sorgt für eine signifikant schlechtere Gesundheitsversorgung im Bereich der RGDs.

Ein häufiges Problem in der molekulargenetischen Diagnosestellung via NGS stellt nicht die Detektion, sondern die Bewertung von genomischen Varianten auf Pathogenität dar, welches sich in einer hohen Rate an sog. *Variants of unknown significance* (VUS) äußert. Eine Möglichkeit die VUS pathomechanistisch genauer einzuordnen, stellt die Untersuchung auf Expressions- und Prozessierungsunterschiede im Bereich der prä-mRNA mittels Transkriptom-Analyse dar.

Vor diesem Hintergrund wurde in der vorliegenden Arbeit der diagnostische Zusatznutzen der Analyse der Gesamtheit der mRNA zusätzlich zu einer Reanalyse der NGS-Daten untersucht und durch weiterführende Untersuchungen im Forschungslabor ergänzt. Ziel war im Idealfall eine Klassifizierung der Kandidatengene als pathogen oder wahrscheinlich pathogen. Vor- und Nachteile dieser diagnostischen Pipeline wurden herausgearbeitet.

16 bisher molekulargenetisch ungelöste Fälle wurden im Kontext dieser Arbeit erneut qualitativ ausgewertet. Die NGS-Sequenzierungen und bioinformatischen Aufarbeitungen der Rohdaten wurden nach institutsinternem Standard durchgeführt. Anschließend wurden diese Daten mittels strukturierter Filterungsschritte auf potenzielle Kandidatengene ausgewertet. Eine Validierung eines Spleißdefektes mittels cDNA-Synthese

und anschließender PCR, sowie der qualitative und semiquantitative Nachweis von Kandidatengenens mittels Western Blots wurden in Bezug auf spezifische Kandidatengene im hauseigenen Forschungslabor durchgeführt. Hierfür wurden Primäre Fibroblastenzelllinien kultiviert. Kandidatengene konnten in 10 von 16 Fällen der hier vorgestellten Kohorte etabliert werden. In 40 % ergaben sich durch die RNA-Seq Befunde, welche den Verdacht bezüglich eines Kandidatengens erhärteten. Bei drei Individuen sorgte die RNA-Seq für die Etablierung eines neuen Kandidatengens. Für zwei Individuen wurden Western Blots durchgeführt. Es gelang der Nachweis einer deutlich hypomorphen Mutation in Patientenzellen passend zum Kandidatengen *DBR1*. Des Weiteren konnte ein Spleißdefekt im Gen *OSCP1* validiert werden und bei einer weiblichen Betroffenen mittels Kombination aus Trio-WES und RNA-Seq eine gestörte X-Inaktivierung sowie reduzierte Expression vor dem Hintergrund einer Mutation in *HUWE1* nachgewiesen werden. Diese Variante wurde schließlich neu als wahrscheinlich pathogen klassifiziert.

Vor diesem Hintergrund bleibt festzustellen, dass die vorgestellte Pipeline sich nach aktuellem Wissensstand sehr gut eignet, um Kandidatengene zu finden und weitere Evidenz für eine mögliche Pathogenität zu ergänzen. Problematisch scheint nach wie vor die Konsequenzen auf den menschlichen Organismus einer VUS vorherzusagen, vor allem wenn für ein Kandidatengen bisher keine diagnostischen Tests etabliert sind. Nächste Schritte für die hier vorgestellten Fälle könnten daher in einer Prüfung auf mögliche funktionelle Analysen, regelmäßige Reanalysen vor dem aktuellen wissenschaftlichen Kontext und Datenaustausch mit anderen humangenetischen Instituten liegen.

Als kritisch in Bezug auf die Forschung zu RGDs werden die in dieser Arbeit diskutierten Themen rund um Phänotypisierung, Standard-Pipelines, Aufbau von funktioneller Evidenz und somit Ausbau des Verständnisses von neuen Krankheitsmechanismen gesehen, um in Zukunft die medizinische Versorgung von PatientInnen mit RGDs nachhaltig zu verbessern.

7 Literaturverzeichnis

- Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J. J., Mayer, P. & Kawashima, E. 2000. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res*, 28, E87.
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. & Sunyaev, S. R. 2010. A method and server for predicting damaging missense mutations. *Nat Methods*, 7, 248-9.
- Akgun-Dogan, O., Tuc Bengur, E., Ay, B., Ozkose, G. S., Kar, E., Bengur, F. B., Bulut, A. S., Yigit, A., Aydin, E., Esen, F. N., Ozdemir, O., Yesilyurt, A. & Alanay, Y. 2024. Impact of deep phenotyping: high diagnostic yield in a diverse pediatric population of 172 patients through clinical whole-genome sequencing at a single center. *Front Genet*, 15, 1347474.
- Alaimo, J. T., Ginton, K. E., Liu, N., Xiao, J., Yang, Y., Reid Sutton, V. & Elsea, S. H. 2020. Integrated analysis of metabolomic profiling and exome data supplements sequence variant interpretation, classification, and diagnosis. *Genet Med*, 22, 1560-1566.
- Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. 2015. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res*, 43, D789-98.
- Amemiya, H., Haley M., Kundaje, Anshul & Boyle, Alan P. 2019. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports*, 9, 9354.
- Archer, H. L., Evans, J. C., Millar, D. S., Thompson, P. W., Kerr, A. M., Leonard, H., Christodoulou, J., Ravine, D., Lazarou, L., Grove, L., Verity, C., Whatley, S. D., Pilz, D. T., Sampson, J. R. & Clarke, A. J. 2006. NTNG1 mutations are a rare cause of Rett syndrome. *Am J Med Genet A*, 140, 691-4.
- Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., Antipenko, A., Shang, L., Boisson, B., Casanova, J. L. & Abel, L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*, 112, 5473-8.
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E., Catenazzi M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-9.
- Best, S., Shoemark, A., Rubbo, B., Patel, M. P., Fassad, M. R., Dixon, M., Rogers, A. V., Hirst, R. A., Rutman, A., Olsson, S., Jackson, C. L., Goggin, P., Thomas, S., Pengelly, R., Cullup, T., Pissaridou, E., Hayward, J., Onoufriadis, A., O'Callaghan, C., Loebinger, M. R., Wilson, R., Chung, E. M., Kenia, P., Doughty, V. L., Carvalho, J. S., Lucas, J. S., Mitchison, H. M. & Hogg, C. 2019. Risk factors for situs defects and congenital heart disease in primary ciliary dyskinesia. *Thorax*, 74, 203-205.
- Bohmwald, K., Espinoza, J. A., Rey-Jurado, E., Gómez, R. S., González, P. A., Bueno, S. M., Riedel, C. A. & Kalergis, A. M. 2016. Human Respiratory Syncytial Virus: Infection and Pathology. *Semin Respir Crit Care Med*, 37, 522-37.
- Bonifacino, J. S. & Glick, B. S. 2004. The mechanisms of vesicle budding and fusion. *Cell*, 116, 153-66.
- Borg, I., Freude, K., Kübart, S., Hoffmann, K., Menzel, C., Laccone, F., Firth, H., Ferguson-Smith, M. A., Tommerup, N., Ropers, H. H., Sargan, D. & Kalscheuer, V. M. 2005. Disruption of Netrin G1 by a balanced chromosome translocation in a girl with Rett syndrome. *Eur J Hum Genet*, 13, 921-7.
- Bournazos, A. M., Riley, L. G., Bommireddipalli, S., Ades, L., Akesson, L. S., Al-Shinnag, M., Alexander, S. I., Archibald, A. D., Balasubramanian, S., Berman, Y., Beshay, V., Boggs, K., Bojadzieva, J., Brown, N. J., Bryen, S. J., Buckley, M. F., Chong, B., Davis, M. R.,

- Dawes, R., Delatycki, M., Donaldson, L., Downie, L., Edwards, C., Edwards, M., Engel, A., Ewans, L. J., Faiz, F., Fennell, A., Field, M., Freckmann, M. L., Gallacher, L., Gear, R., Goel, H., Goh, S., Goodwin, L., Hanna, B., Harraway, J., Higgins, M., Ho, G., Hopper, B. K., Horton, A. E., Hunter, M. F., Huq, A. J., Josephi-Taylor, S., Joshi, H., Kirk, E., Krzesinski, E., Kumar, K. R., Lemckert, F., Leventer, R. J., Lindsey-Temple, S. E., Lunke, S., Ma, A., Macaskill, S., Mallawaarachchi, A., Marty, M., Marum, J. E., McCarthy, H. J., Menezes, M. P., McLean, A., Milnes, D., Mohammad, S., Mowat, D., Niaz, A., Palmer, E. E., Patel, C., Patel, S. G., Phelan, D., Pinner, J. R., Rajagopalan, S., Regan, M., Rodgers, J., Rodrigues, M., Roxburgh, R. H., Sachdev, R., Roscioli, T., Samarasekera, R., Sandaradura, S. A., Savva, E., Schindler, T., Shah, M., Sinnerbrink, I. B., Smith, J. M., Smith, R. J., Springer, A., Stark, Z., Strom, S. P., Sue, C. M., Tan, K., Tan, T. Y., Tantsis, E., Tchan, M. C., Thompson, B. A., Trainer, A. H., van Spaendonck-Zwarts, K., Walsh, R., Warwick, L., White, S., White, S. M., Williams, M. G., et al. 2022. Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genet Med*, 24, 130-145.
- Boyadjiev, S. A., Fromme, J. C., Ben, J., Chong, S. S., Nauta, C., Hur, D. J., Zhang, G., Hamamoto, S., Schekman, R., Ravazzola, M., Orci, L. & Eyaid, W. 2006. Cranio-lenticulo-sutural dysplasia is caused by a SEC23A mutation leading to abnormal endoplasmic-reticulum-to-Golgi trafficking. *Nat Genet*, 38, 1192-7.
- Boycott, K. M., Rath, A., Chong, J. X., Hartley, T., Alkuraya, F. S., Baynam, G., Brookes, A. J., Brudno, M., Carracedo, A., den Dunnen, J. T., Dyke, S. O. M., Estivill, X., Goldblatt, J., Gonthier, C., Groft, S. C., Gut, I., Hamosh, A., Hieter, P., Höhn, S., Hurles, M. E., Kaufmann, P., Knoppers, B. M., Krischer, J. P., Macek, M., Jr., Matthijs, G., Olry, A., Parker, S., Paschall, J., Philippakis, A. A., Rehm, H. L., Robinson, P. N., Sham, P. C., Stefanov, R., Taruscio, D., Unni, D., Vanstone, M. R., Zhang, F., Brunner, H., Bamshad, M. J. & Lochmüller, H. 2017. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet*, 100, 695-705.
- Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72, 248-54.
- Brea-Fernández, Alejandro J., Álvarez-Barona, Miriam, Amigo, Jorge, Tubío-Fungueiriño, María, Caamaño, Pilar, Fernández-Prieto, Montserrat, Barros, Francisco, De Rubeis, Silvia, Buxbaum, Joseph & Carracedo, Ángel 2022. Trio-based exome sequencing reveals a high rate of the de novo variants in intellectual disability. *European Journal of Human Genetics*, 30, 938-945.
- Brechtmann, F., Mertes, C., Matusevičiūtė, A., Yépez, V. A., Avsec, Ž, Herzog, M., Bader, D. M., Prokisch, H. & Gagneur, J. 2018. OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data. *Am J Hum Genet*, 103, 907-917.
- Britten-Jones, A. C., Gocuk, S. A., Goh, K. L., Huq, A., Edwards, T. L. & Ayton, L. N. 2023. The Diagnostic Yield of Next Generation Sequencing in Inherited Retinal Diseases: A Systematic Review and Meta-analysis. *Am J Ophthalmol*, 249, 57-73.
- Bundesministerium für Gesundheit 2023. *Seltene Erkrankungen* [Online]. URL: <https://www.bundesgesundheitsministerium.de/themen/praevention/gesundheitsgefahren/seltene-erkrankungen.html> [Letzter Zugriff: 19.07.2023].
- Carneiro, T. N., Krepischi, A. C., Costa, S. S., Tojal da Silva, I., Vianna-Morgante, A. M., Valieris, R., Ezquina, S. A., Bertola, D. R., Otto, P. A. & Rosenberg, C. 2018. Utility of trio-based exome sequencing in the elucidation of the genetic basis of isolated syndromic intellectual disability: illustrative cases. *Appl Clin Genet*, 11, 93-98.
- Casanova, J. L. 2015. Severe infectious diseases of childhood as monogenic inborn errors of immunity. *Proc Natl Acad Sci U S A*, 112, E7128-37.
- Chan, Y. H., Lundberg, V., Le Pen, J., Yuan, J., Lee, D., Pinci, F., Volpi, S., Nakajima, K., Bondet, V., Åkesson, S., Khobreakar, N. V., Bodansky, A., Du, L., Melander, T., Mariaggi, A. A., Seeleuthner, Y., Saleh, T. S., Chakravarty, D., Marits, P., Dobbs, K., Vonlanthen, S., Hennings, V., Thörn, K., Rinchai, D., Bizien, L., Chaldebas, M., Sobh, A., Özçelik, T., Keles, S., AlKhater, S. A., Prando, C., Meyts, I., Wilson, M. R., Rosain, J., Jouanguy, E., Aubart, M., Abel, L., Mogensen, T. H., Pan-Hammarström, Q., Gao, D., Duffy, D., Cobat, A., Berg, S., Notarangelo, L. D., Harschnitz, O., Rice, C. M., Studer, L., Casanova, J. L., Ekwall, O. & Zhang, S. Y. 2024. SARS-CoV-2 brainstem encephalitis in human inherited DBR1 deficiency. *J Exp Med*, 221.
- Chérot, E., Keren, B., Dubourg, C., Carré, W., Fradin, M., Lavillaureix, A., Afenjar, A., Burglen, L., Whalen, S., Charles, P., Marey, I., Heide, S., Jacqueline, A., Heron, D., Doummar, D., Rodriguez, D., Billette de Villemeur, T., Moutard, M. L., Guët, A., Xavier, J., Périsset, D., Cohen, D., Demurger, F., Quélin, C., Depienne, C., Odent, S., Nava, C., David, V., Pasquier, L. & Mignot, C. 2018. Using medical exome sequencing to identify the causes of

- neurodevelopmental disorders: Experience of 2 clinical units and 216 patients. *Clin Genet*, 93, 567-576.
- Chong, J. X., Buckingham, K. J., Jhangiani, S. N., Boehm, C., Sobreira, N., Smith, J. D., Harrell, T. M., McMillin, M. J., Wiszniewski, W., Gambin, T., Coban Akdemir, Z. H., Doheny, K., Scott, A. F., Avramopoulos, D., Chakravarti, A., Hoover-Fong, J., Mathews, D., Witmer, P. D., Ling, H., Hetrick, K., Watkins, L., Patterson, K. E., Reinier, F., Blue, E., Muzny, D., Kircher, M., Bilguvar, K., López-Giráldez, F., Sutton, V. R., Tabor, H. K., Leal, S. M., Gunel, M., Mane, S., Gibbs, R. A., Boerwinkle, E., Hamosh, A., Shendure, J., Lupski, J. R., Lifton, R. P., Valle, D., Nickerson, D. A. & Bamshad, M. J. 2015. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet*, 97, 199-215.
- Chowdhury, F., Wang, L., Al-Raqad, M., Amor, D. J., Baxová, A., Bendová, Š, Biamino, E., Brusco, A., Caluseriu, O., Cox, N. J., Froukh, T., Gunay-Aygun, M., Hancárová, M., Haynes, D., Heide, S., Hoganson, G., Kaname, T., Keren, B., Kosaki, K., Kubota, K., Lemons, J. M., Magriña, M. A., Mark, P. R., McDonald, M. T., Montgomery, S., Morley, G. M., Ohnishi, H., Okamoto, N., Rodriguez-Burítica, D., Rump, P., Sedláček, Z., Schatz, K., Streff, H., Uehara, T., Walia, J. S., Wheeler, P. G., Wiesener, A., Zweier, C., Kawakami, K., Wentzensen, I. M., Lalani, S. R., Siu, V. M., Bi, W. & Balci, T. B. 2021. Haploinsufficiency of PRR12 causes a spectrum of neurodevelopmental, eye, and multisystem abnormalities. *Genet Med*, 23, 1234-1245.
- Cummings, Beryl B., Marshall, Jamie L., Tukiainen, Taru, Lek, Monkol, Donkervoort, Sandra, Foley, A. Reghan, Bolduc, Veronique, Waddell, Leigh B., Sandaradura, Sarah A., O'Grady, Gina L., Estrella, Elicia, Reddy, Hemakumar M., Zhao, Fengmei, Weisburd, Ben, Karczewski, Konrad J., O'Donnell-Luria, Anne H., Birnbaum, Daniel, Sarkozy, Anna, Hu, Ying, Gonorazky, Hernan, Claeys, Kristl, Joshi, Himanshu, Bournazos, Adam, Oates, Emily C., Ghaoui, Roula, Davis, Mark R., Laing, Nigel G., Topf, Ana, Consortium, Genotype-Tissue Expression, Kang, Peter B., Beggs, Alan H., North, Kathryn N., Straub, Volker, Dowling, James J., Muntoni, Francesco, Clarke, Nigel F., Cooper, Sandra T., Bönnemann, Carsten G. & MacArthur, Daniel G. 2017. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Science Translational Medicine*, 9, eaal5209.
- Delude, Cathryn M. 2015. Deep phenotyping: The details of disease. *Nature*, 527, S14-S15.
- Demidov, German & Ossowski, Stephan 2019. ClinCNV: novel method for allele-specific somatic copy-number alterations detection. *bioRxiv*, 837971.
- den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., Roux, A. F., Smith, T., Antonarakis, S. E. & Taschner, P. E. 2016. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat*, 37, 564-9.
- Dewulf, J. P., Marie, S. & Nassogne, M. C. 2022. Disorders of purine biosynthesis metabolism. *Mol Genet Metab*, 136, 190-198.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15-21.
- Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res*, 19, 4008.
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. & Liu, X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*, 24, 2125-37.
- Europäische Union (Das Europäische Parlament und der Rat der Europäischen Union), 1999. Verordnung (EG) Nr. 1411/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32000R0141> [Letzter Zugriff: 19.07.2023].
- Farwell, K. D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Tippin Davis, B., Baxter, R. M., Zeng, W., Mroske, C., Parra, M. C., Gandomi, S. K., Lu, I., Li, X., Lu, H., Lu, H. M., Salvador, D., Ruble, D., Lao, M., Fischbach, S., Wen, J., Lee, S., Elliott, A., Dunlop, C. L. & Tang, S. 2015. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med*, 17, 578-86.
- Faust, G. G. & Hall, I. M. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics*, 30, 2503-5.
- Faye, Fatoumata, Crocione, Claudia, Anido de Peña, Roberta, Bellagambi, Simona, Escati Peñaloza, Luciana, Hunter, Amy, Jensen, Lene, Oosterwijk, Cor, Schoeters, Eva, de Vicente, Daniel, Faivre, Laurence, Wilbur, Michael, Le Cam, Yann & Dubief, Jessie 2024. Time to diagnosis and determinants of diagnostic delays of people living with a rare

- disease: results of a Rare Barometer retrospective patient survey. *European Journal of Human Genetics*.
- Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G. 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*, 34, e22.
- Ferraro, N. M., Strober, B. J., Einson, J., Abell, N. S., Aguet, F., Barbeira, A. N., Brandt, M., Bucan, M., Castel, S. E., Davis, J. R., Greenwald, E., Hess, G. T., Hilliard, A. T., Kember, R. L., Kotis, B., Park, Y., Peloso, G., Ramdas, S., Scott, A. J., Smail, C., Tsang, E. K., Zekavat, S. M., Ziosi, M., Aradhana, Ardlie, K. G., Assimes, T. L., Bassik, M. C., Brown, C. D., Correa, A., Hall, I., Im, H. K., Li, X., Natarajan, P., Lappalainen, T., Mohammadi, P., Montgomery, S. B. & Battle, A. 2020. Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science*, 369.
- Ferreira, C. R. 2019. The burden of rare diseases. *Am J Med Genet A*, 179, 885-892.
- French, C. E., Dolling, H., Mégy, K., Sanchis-Juan, A., Kumar, A., Delon, I., Wakeling, M., Mallin, L., Agrawal, S., Austin, T., Walston, F., Park, S. M., Parker, A., Piyasena, C., Bradbury, K., Ellard, S., Rowitch, D. H. & Raymond, F. L. 2022. Refinements and considerations for trio whole-genome sequence analysis when investigating Mendelian diseases presenting in early childhood. *HGG Adv*, 3, 100113.
- Frésard, Laure, Smail, Craig, Ferraro, Nicole M., Teran, Nicole A., Li, Xin, Smith, Kevin S., Bonner, Devon, Kernohan, Kristin D., Marwaha, Shruti, Zappala, Zachary, Balliu, Brunilda, Davis, Joe R., Liu, Boxiang, Prybol, Cameron J., Kohler, Jennefer N., Zastrow, Diane B., Reuter, Chloe M., Fisk, Dianna G., Grove, Megan E., Davidson, Jean M., Hartley, Taila, Joshi, Ruchi, Strober, Benjamin J., Utiramerur, Sowmithri, Adams, David R., Aday, Aaron, Alejandro, Mercedes E., Allard, Patrick, Ashley, Euan A., Azamian, Mahshid S., Bacino, Carlos A., Baker, Eva, Balasubramanyam, Ashok, Barseghyan, Hayk, Batzli, Gabriel F., Beggs, Alan H., Behnam, Babak, Bellen, Hugo J., Bernstein, Jonathan A., Berry, Gerard T., Bican, Anna, Bick, David P., Birch, Camille L., Bonner, Devon, Boone, Braden E., Bostwick, Bret L., Briere, Lauren C., Brokamp, Elly, Brown, Donna M., Brush, Matthew, Burke, Elizabeth A., Burrage, Lindsay C., Butte, Manish J., Chen, Shan, Clark, Gary D., Coakley, Terra R., Cogan, Joy D., Colley, Heather A., Cooper, Cynthia M., Cope, Heidi, Craigen, William J., D'Souza, Precilla, Davids, Mariska, Davidson, Jean M., Dayal, Jyoti G., Dell'Angelica, Esteban C., Dhar, Shweta U., Dipple, Katrina M., Donnell-Fink, Laurel A., Dorrani, Naghmeh, Dorset, Daniel C., Douine, Emilie D., Draper, David D., Dries, Annika M., Duncan, Laura, Eckstein, David J., Emrick, Lisa T., Eng, Christine M., Enns, Gregory M., Eskin, Ascia, Esteves, Cecilia, Estwick, Tyra, Fernandez, Liliana, Ferreira, Carlos, Fieg, Elizabeth L., Fisher, Paul G., Fogel, Brent L., Friedman, Noah D., Gahl, William A., Glanton, Emily, Godfrey, Rena A., Goldman, Alica M., Goldstein, David B., Gould, Sarah E., Gouridine, Jean-Philippe F., Groden, Catherine A., Gropman, Andrea L., Haendel, Melissa, Hamid, Rizwan, Hanchard, Neil A., et al. 2019. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nature Medicine*, 25, 911-919.
- Friederich, M. W., Elias, A. F., Kuster, A., Laugwitz, L., Larson, A. A., Landry, A. P., Ellwood-Digel, L., Mirsky, D. M., Dimmock, D., Haven, J., Jiang, H., MacLean, K. N., Styren, K., Schoof, J., Goujon, L., Lefrancois, T., Friederich, M., Coughlin, C. R., 2nd, Banerjee, R., Haack, T. B. & Van Hove, J. L. K. 2020. Pathogenic variants in SQOR encoding sulfide:quinone oxidoreductase are a potentially treatable cause of Leigh disease. *J Inherit Metab Dis*, 43, 1024-1036.
- Froyen, G., Belet, S., Martinez, F., Santos-Rebouças, C. B., Declercq, M., Verbeeck, J., Donckers, L., Berland, S., Mayo, S., Rosello, M., Pimentel, M. M., Fintelman-Rodrigues, N., Hovland, R., Rodrigues dos Santos, S., Raymond, F. L., Bose, T., Corbett, M. A., Sheffield, L., van Ravenswaaij-Arts, C. M., Dijkhuizen, T., Coutton, C., Satre, V., Siu, V. & Marynen, P. 2012. Copy-number gains of HUWE1 due to replication- and recombination-based rearrangements. *Am J Hum Genet*, 91, 252-64.
- Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H., Chelly, J., Sanlaville, D., van Bokhoven, H., Ropers, H. H., Laumonnier, F., Ranieri, E., Schwartz, C. E., Abidi, F., Tarpey, P. S., Futreal, P. A., Whibley, A., Raymond, F. L., Stratton, M. R., Fryns, J. P., Scott, R., Peippo, M., Sipponen, M., Partington, M., Mowat, D., Field, M., Hackett, A., Marynen, P., Turner, G. & Géczy, J. 2008. Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. *Am J Hum Genet*, 82, 432-43.
- Garrison, Erik & Marth, Gabor 2012. Haplotype-based variant detection from short-read sequencing. *arXiv*, 1207.

- Gburek-Augustat, J., Schoene-Bake, J. C., Bültmann, E., Haack, T., Buchert, R., Synofzik, M., Biskup, S., Feuerhake, F., Sorge, I. & Hartmann, H. 2021. Pitfalls in Genetic Diagnostics: Why Phenotyping is Essential. *Neuropediatrics*, 52, 274-283.
- Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W., Willemsen, M. H., Kwint, M., Janssen, I. M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H. G., de Vries, B. B., Kleefstra, T., Brunner, H. G., Vissers, L. E. & Veltman, J. A. 2014. Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 511, 344-7.
- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S. & Nusbaum, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol*, 27, 182-9.
- Goldfeder, R. L., Priest, J. R., Zook, J. M., Grove, M. E., Waggott, D., Wheeler, M. T., Salit, M. & Ashley, E. A. 2016. Medical implications of technical accuracy in genome sequencing. *Genome Med*, 8, 24.
- Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., Kao, D., Ohri, K., Viththiyapaskaran, S., Tarnopolsky, M. A., Mathews, K. D., Moore, S. A., Osorio, A. N., Villanova, D., Kemaladewi, D. U., Cohn, R. D., Brudno, M. & Dowling, J. J. 2019. Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *Am J Hum Genet*, 104, 466-483.
- Gorr, T. A. & Vogel, J. 2015. Western blotting revisited: critical perusal of underappreciated technical issues. *Proteomics Clin Appl*, 9, 396-405.
- Gregor, A., Sadleir, L. G., Asadollahi, R., Azzarello-Burri, S., Battaglia, A., Ousager, L. B., Boonsawat, P., Bruel, A. L., Buchert, R., Calpena, E., Cogné, B., Dallapiccola, B., Distelmaier, F., Elmslie, F., Faivre, L., Haack, T. B., Harrison, V., Henderson, A., Hunt, D., Isidor, B., Joset, P., Kumada, S., Lachmeijer, A. M. A., Lees, M., Lynch, S. A., Martinez, F., Matsumoto, N., McDougall, C., Mefford, H. C., Miyake, N., Myers, C. T., Moutton, S., Nesbitt, A., Novelli, A., Orellana, C., Rauch, A., Rosello, M., Saida, K., Santani, A. B., Sarkar, A., Scheffer, I. E., Shinawi, M., Steindl, K., Symonds, J. D., Zackai, E. H., Reis, A., Sticht, H. & Zweier, C. 2018. De Novo Variants in the F-Box Protein FBXO11 in 20 Individuals with a Variable Neurodevelopmental Disorder. *Am J Hum Genet*, 103, 305-316.
- Gupta, N. 2023. Deciphering Intellectual Disability. *Indian J Pediatr*, 90, 160-167.
- Haendel, M., Vasilevsky, N., Unni, D., Bologa, C., Harris, N., Rehm, H., Hamosh, A., Baynam, G., Groza, T., McMurry, J., Dawkins, H., Rath, A., Thaxton, C., Bocci, G., Joachimiak, M. P., Köhler, S., Robinson, P. N., Mungall, C. & Oprea, T. I. 2020. How many rare diseases are there? *Nat Rev Drug Discov*, 19, 77-78.
- Halbeisen, F. S., Goutaki, M., Spycher, B. D., Amirav, I., Behan, L., Boon, M., Hogg, C., Casaulta, C., Crowley, S., Haarman, E. G., Karadag, B., Koerner-Rettberg, C., Loebinger, M. R., Mazurek, H., Morgan, L., Nielsen, K. G., Omran, H., Santamaria, F., Schwerk, N., Thouvenin, G., Yiallourou, P., Lucas, J. S., Latzin, P. & Kuehni, C. E. 2018. Lung function in patients with primary ciliary dyskinesia: an iPCD Cohort study. *Eur Respir J*, 52.
- Halperin, D., Kadir, R., Perez, Y., Drabkin, M., Yogev, Y., Wormser, O., Berman, E. M., Eremenko, E., Rotblat, B., Shorer, Z., Gradstein, L., Shelef, I., Birk, R., Abdu, U., Flusser, H. & Birk, O. S. 2019. SEC31A mutation affects ER homeostasis, causing a neurological syndrome. *J Med Genet*, 56, 139-148.
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R. & Ordoukhanian, P. 2014. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*, 56, 61-4, 66, 68, passim.
- Hennekam, R. C. & Biesecker, L. G. 2012. Next-generation sequencing demands next-generation phenotyping. *Hum Mutat*, 33, 884-6.
- Institute of Medical Genetics and Applied Genomics 2016. *megSAP - a Medical Genetics Sequence Analysis Pipeline* [Online]. URL: <https://github.com/imgag/megSAP> [Letzter Zugriff: 07.03.2021].
- Jansen, S., van der Werf, I. M., Innes, A. M., Afenjar, A., Agrawal, P. B., Anderson, I. J., Atwal, P. S., van Binsbergen, E., van den Boogaard, M. J., Castiglia, L., Coban-Akdemir, Z. H., van Dijck, A., Doummar, D., van Eerde, A. M., van Essen, A. J., van Gassen, K. L., Guillen Sacoto, M. J., van Haelst, M. M., Iossifov, I., Jackson, J. L., Judd, E., Kaiwar, C., Keren, B., Klee, E. W., Klein Wassink-Ruiter, J. S., Meuwissen, M. E., Monaghan, K. G., de Munnik, S. A., Nava, C., Ockeloen, C. W., Pettinato, R., Racher, H., Rinne, T., Romano, C., Sanders, V. R., Schnur, R. E., Smeets, E. J., Stegmann, A. P. A., Stray-Pedersen, A., Sweetser, D. A., Terhal, P. A., Tveten, K., VanNoy, G. E., de Vries, P. F., Waxler, J. L., Willing, M., Pfundt, R., Veltman, J. A., Kooy, R. F., Vissers, L. E. M. & de Vries, B. B. A.

2019. De novo variants in FBXO11 cause a syndromic form of intellectual disability with behavioral problems and dysmorphisms. *Eur J Hum Genet*, 27, 738-746.
- Jenkinson, G., Li, Y. I., Basu, S., Cousin, M. A., Oliver, G. R. & Klee, E. W. 2020. LeafCutterMD: an algorithm for outlier splicing detection in rare diseases. *Bioinformatics*, 36, 4609-4615.
- Kahrizi, K., Hu, H., Hosseini, M., Kalscheuer, V. M., Fattahi, Z., Beheshtian, M., Suckow, V., Mohseni, M., Lipkowitz, B., Mehvari, S., Mehrjoo, Z., Akhtarkhavari, T., Ghaderi, Z., Rahimi, M., Arzhang, S., Jamali, P., Falahat Chian, M., Nikuei, P., Sabbagh Kermani, F., Sadeghinia, F., Jazayeri, R., Tonekaboni, S. H., Khoshaeen, A., Habibi, H., Pourfatemi, F., Mojahedi, F., Khodaie-Ardakani, M. R., Najafipour, R., Wienker, T. F., Najmabadi, H. & Ropers, H. H. 2019. Effect of inbreeding on intellectual disability revisited by trio sequencing. *Clin Genet*, 95, 151-159.
- Karczewski, Konrad J., Francioli, Laurent C., Tiao, Grace, Cummings, Beryl B., Alföldi, Jessica, Wang, Qingbo, Collins, Ryan L., Laricchia, Kristen M., Ganna, Andrea, Birnbaum, Daniel P., Gauthier, Laura D., Brand, Harrison, Solomonson, Matthew, Watts, Nicholas A., Rhodes, Daniel, Singer-Berk, Moriel, England, Eleina M., Seaby, Eleanor G., Kosmicki, Jack A., Walters, Raymond K., Tashman, Katherine, Farjoun, Yossi, Banks, Eric, Poterba, Timothy, Wang, Arcturus, Seed, Cotton, Whiffin, Nicola, Chong, Jessica X., Samocha, Kaitlin E., Pierce-Hoffman, Emma, Zappala, Zachary, O'Donnell-Luria, Anne H., Minikel, Eric Vallabh, Weisburd, Ben, Lek, Monkol, Ware, James S., Vittal, Christopher, Armean, Irina M., Bergelson, Louis, Cibulskis, Kristian, Connolly, Kristen M., Covarrubias, Miguel, Donnelly, Stacey, Ferreira, Steven, Gabriel, Stacey, Gentry, Jeff, Gupta, Namrata, Jeandet, Thibault, Kaplan, Diane, Llanwarne, Christopher, Munshi, Ruchi, Novod, Sam, Petrillo, Nikelle, Roazen, David, Ruano-Rubio, Valentin, Saltzman, Andrea, Schleicher, Molly, Soto, Jose, Tibbetts, Kathleen, Tolonen, Charlotte, Wade, Gordon, Talkowski, Michael E., Aguilar Salinas, Carlos A., Ahmad, Tariq, Albert, Christine M., Ardissino, Diego, Atzmon, Gil, Barnard, John, Beaugerie, Laurent, Benjamin, Emelia J., Boehnke, Michael, Bonnycastle, Lori L., Bottinger, Erwin P., Bowden, Donald W., Bow, Matthew J., Chambers, John C., Chan, Juliana C., Chasman, Daniel, Cho, Judy, Chung, Mina K., Cohen, Bruce, Correa, Adolfo, Dabelea, Dana, Daly, Mark J., Darbar, Dawood, Duggirala, Ravindranath, Dupuis, Josée, Ellinor, Patrick T., Elosua, Roberto, Erdmann, Jeanette, Esko, Tõnu, Färkkilä, Martti, Florez, Jose, Franke, Andre, Getz, Gad, Glaser, Benjamin, Glatt, Stephen J., Goldstein, David, Gonzalez, Clicerio, Groop, Leif, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581, 434-443.
- Kato, G. J., Piel, F. B., Reid, C. D., Gaston, M. H., Ohene-Frempong, K., Krishnamurti, L., Smith, W. R., Panepinto, J. A., Weatherall, D. J., Costa, F. F. & Vichinsky, E. P. 2018. Sick cell disease. *Nat Rev Dis Primers*, 4, 18010.
- Katz, Yarden, Wang, Eric T., Silterra, Jacob, Schwartz, Schraga, Wong, Bang, Mesirov, Jill P., Airoidi, Edoardo M. & Burge, Christopher B. 2013. Sashimi plots: Quantitative visualization of RNA sequencing read alignments. URL: <https://ui.adsabs.harvard.edu/abs/2013arXiv1306.3466K> [Accessed June 01, 2013].
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. 2002. The human genome browser at UCSC. *Genome Res*, 12, 996-1006.
- Khayat, A. M., Alshareef, B. G., Alharbi, S. F., AlZahrani, M. M., Alshangity, B. A. & Tashkandi, N. F. 2024. Consanguineous Marriage and Its Association With Genetic Disorders in Saudi Arabia: A Review. *Cureus*, 16, e53888.
- Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M. & Shendure, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, 46, 310-5.
- Köhler, S., Gargano, M., Matentzoglou, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., Danis, D., Balagura, G., Baynam, G., Brower, A. M., Callahan, T. J., Chute, C. G., Est, J. L., Galer, P. D., Ganesan, S., Griese, M., Haimel, M., Pazmandi, J., Hanauer, M., Harris, N. L., Hartnett, M. J., Hastreiter, M., Hauck, F., He, Y., Jeske, T., Kearney, H., Kindle, G., Klein, C., Knoflach, K., Krause, R., Lagorce, D., McMurry, J. A., Miller, J. A., Munoz-Torres, M. C., Peters, R. L., Rapp, C. K., Rath, A. M., Rind, S. A., Rosenberg, A. Z., Segal, M. M., Seidel, M. G., Smedley, D., Talmy, T., Thomas, Y., Wiafe, S. A., Xian, J., Yüksel, Z., Helbig, I., Mungall, C. J., Haendel, M. A. & Robinson, P. N. 2021. The Human Phenotype Ontology in 2021. *Nucleic Acids Res*, 49, D1207-d1217.
- Kole, Anna & Faurisson, François. The Voice of 12,000 Patients. Experiences and Expectations of Rare Disease Patients on Diagnosis and Care in Europe. Paris, France, 2009. Verfügbar unter: <https://www.eurordis.org/publications/the-voice-of-12000-patients/> [Letzter Zugriff: 20.06.2024 Access 2009].
- Kopajtich, Robert, Smirnov, Dmitrii, Stenton, Sarah L., Loipfinger, Stefan, Meng, Chen, Scheller, Ines F., Freisinger, Peter, Baski, Robert, Berutti, Riccardo, Behr, Jürgen, Bucher, Martina,

- Distelmaier, Felix, Graf, Elisabeth, Gusic, Mirjana, Hempel, Maja, Kulterer, Lea, Mayr, Johannes, Meitinger, Thomas, Mertes, Christian, Metodiev, Metodii D., Nadel, Agnieszka, Nasca, Alessia, Ohtake, Akira, Okazaki, Yasushi, Olsen, Rikke, Piekutowska-Abramczuk, Dorota, Rötig, Agnès, Santer, René, Schindler, Detlev, Slama, Abdelhamid, Staufner, Christian, Strom, Tim, Verloo, Patrick, von Kleist-Retzow, Jürgen-Christoph, Wortmann, Saskia B., Yépez, Vicente A., Lamperti, Costanza, Ghezzi, Daniele, Murayama, Kei, Ludwig, Christina, Gagneur, Julien & Prokisch, Holger 2021. Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders. *medRxiv*, 2021.03.09.21253187.
- Kremer, Laura S., Bader, Daniel M., Mertes, Christian, Kopajtich, Robert, Pichler, Garwin, Iuso, Arcangela, Haack, Tobias B., Graf, Elisabeth, Schwarzmayr, Thomas, Terrile, Caterina, Koňářiková, Eliška, Repp, Birgit, Kastenmüller, Gabi, Adamski, Jerzy, Lichtner, Peter, Leonhardt, Christoph, Funalot, Benoit, Donati, Alice, Tiranti, Valeria, Lombes, Anne, Jardel, Claude, Gläser, Dieter, Taylor, Robert W., Ghezzi, Daniele, Mayr, Johannes A., Rötig, Agnes, Freisinger, Peter, Distelmaier, Felix, Strom, Tim M., Meitinger, Thomas, Gagneur, Julien & Prokisch, Holger 2017. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nature Communications*, 8, 15824.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680-5.
- Laurie, S., Steyaert, W., de Boer, E., Polavarapu, K., Schuermans, N., Sommer, A. K., Demidov, G., Ellwanger, K., Paramonov, I., Thomas, C., Aretz, S., Baets, J., Benetti, E., Bullich, G., Chinnery, P. F., Clayton-Smith, J., Cohen, E., Danis, D., de Sainte Agathe, J. M., Denommé-Pichon, A. S., Diaz-Manera, J., Efthymiou, S., Faivre, L., Fernandez-Callejo, M., Freeberg, M., Garcia-Pelaez, J., Guillot-Noel, L., Haack, T. B., Hanna, M., Hengel, H., Horvath, R., Houlden, H., Jackson, A., Johansson, L., Johari, M., Kamsteeg, E. J., Kellner, M., Kleefstra, T., Lacombe, D., Lochmüller, H., López-Martín, E., Macaya, A., Marcé-Grau, A., Maver, A., Morsy, H., Muntoni, F., Musacchia, F., Nelson, I., Nigro, V., Olimpio, C., Oliveira, C., Paulasová Schwabová, J., Pauly, M. G., Peterlin, B., Peters, S., Pfundt, R., Piluso, G., Piscia, D., Posada, M., Reich, S., Renieri, A., Ryba, L., Šablauskas, K., Savarese, M., Schöls, L., Schütz, L., Steinke-Lange, V., Stevanin, G., Straub, V., Sturm, M., Swertz, M. A., Tartaglia, M., Te Paske, Ibaw, Thompson, R., Torella, A., Trainor, C., Udd, B., Van de Vondel, L., van de Warrenburg, B., van Reeuwijk, J., Vandrovцова, J., Vitobello, A., Vos, J., Vyhňáková, E., Wijngaard, R., Wilke, C., William, D., Xu, J., Yaldiz, B., Zalatnai, L., Zurek, B., Brookes, A. J., Evangelista, T., Gilissen, C., Graessner, H., Hoogerbrugge, N., Ossowski, S., Riess, O., Schüle, R., Synofzik, M., et al. 2025. Genomic reanalysis of a pan-European rare-disease resource yields new diagnoses. *Nat Med*, 31, 478-489.
- Leduc, M. S., McGuire, M., Madan-Khetarpal, S., Ortiz, D., Hayflick, S., Keller, K., Eng, C. M., Yang, Y. & Bi, W. 2018. De novo apparent loss-of-function mutations in PRR12 in three patients with intellectual disability and iris abnormalities. *Hum Genet*, 137, 257-264.
- LePecq, J. B. & Paoletti, C. 1967. A fluorescent complex between ethidium bromide and nucleic acids. Physical-chemical characterization. *J Mol Biol*, 27, 87-106.
- Li, Heng 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv*, 1303.
- Lohmann, K. & Klein, C. 2014. Next generation sequencing and the future of genetic diagnosis. *Neurotherapeutics*, 11, 699-707.
- López-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigó, R. 2005. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett*, 579, 1900-3.
- Lord, C., Elsabbagh, M., Baird, G. & Veenstra-Vanderweele, J. 2018. Autism spectrum disorder. *Lancet*, 392, 508-520.
- Lu, C. L. & Kim, J. 2020. Consequences of mutations in the genes of the ER export machinery COPII in vertebrates. *Cell Stress Chaperones*, 25, 199-209.
- Ma, Z., Sharma, R. & Rogers, A. N. 2024. Physiological Consequences of Nonsense-Mediated Decay and Its Role in Adaptive Responses. *Biomedicines*, 12.
- Maia, N., Nabais Sá, M. J., Melo-Pires, M., de Brouwer, A. P. M. & Jorge, P. 2021. Intellectual disability genomics: current state, pitfalls and future challenges. *BMC Genomics*, 22, 909.
- Manickam, K., McClain, M. R., Demmer, L. A., Biswas, S., Kearney, H. M., Malinowski, J., Masingham, L. J., Miller, D., Yu, T. W. & Hisama, F. M. 2021. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*, 23, 2029-2037.
- Mardis, E. R. 2011. A decade's perspective on DNA sequencing technology. *Nature*, 470, 198-203.
- Matalonga, L., Hernández-Ferrer, C., Piscia, D., Schüle, R., Synofzik, M., Töpf, A., Vissers, Lelm, de Voer, R., Tonda, R., Laurie, S., Fernandez-Callejo, M., Picó, D., Garcia-Linares, C.,

- Papakonstantinou, A., Corvó, A., Joshi, R., Diez, H., Gut, I., Hoischen, A., Graessner, H. & Beltran, S. 2021. Solving patients with rare diseases through programmatic reanalysis of genome-phenome data. *Eur J Hum Genet*, 29, 1337-1347.
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., Flicek, P. & Cunningham, F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol*, 17, 122.
- Meftahi, G. H., Bahari, Z., Zarei Mahmoudabadi, A., Iman, M. & Jangravi, Z. 2021. Applications of western blot technique: From bench to bedside. *Biochem Mol Biol Educ*, 49, 509-517.
- Mertes, C., Scheller, I. F., Yépez, V. A., Çelik, M. H., Liang, Y., Kremer, L. S., Gusic, M., Prokisch, H. & Gagneur, J. 2021. Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun*, 12, 529.
- Metzker, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet*, 11, 31-46.
- Mezzi, N., Abassi, N., Fatnassi, F., Abdelhak, S. & Romdhane, L. 2024. Consanguinity and its impact on health and genome dynamic: An example from Tunisia. *Tunis Med*, 102, 256-265.
- Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., Church, D. M., Crolla, J. A., Eichler, E. E., Epstein, C. J., Faucett, W. A., Feuk, L., Friedman, J. M., Hamosh, A., Jackson, L., Kaminsky, E. B., Kok, K., Krantz, I. D., Kuhn, R. M., Lee, C., Ostell, J. M., Rosenberg, C., Scherer, S. W., Spinner, N. B., Stavropoulos, D. J., Tepperberg, J. H., Thorland, E. C., Vermeesch, J. R., Waggoner, D. J., Watson, M. S., Martin, C. L. & Ledbetter, D. H. 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet*, 86, 749-64.
- Mizuguchi, M. 1997. Acute necrotizing encephalopathy of childhood: a novel form of acute encephalopathy prevalent in Japan and Taiwan. *Brain Dev*, 19, 81-92.
- Mizuguchi, M., Abe, J., Mikkaichi, K., Noma, S., Yoshida, K., Yamanaka, T. & Kamoshita, S. 1995. Acute necrotizing encephalopathy of childhood: a new syndrome presenting with multifocal, symmetric brain lesions. *J Neurol Neurosurg Psychiatry*, 58, 555-61.
- Mohanta, A. & Chakrabarti, K. 2021. Dbr1 functions in mRNA processing, intron turnover and human diseases. *Biochimie*, 180, 134-142.
- Monies, D., Abouelhoda, M., Assoum, M., Moghrabi, N., Rafiullah, R., Almontashiri, N., Alowain, M., Alzaidan, H., Alsayed, M., Subhani, S., Cupler, E., Faden, M., Alhashem, A., Qari, A., Chedrawi, A., Aldhalaan, H., Kurdi, W., Khan, S., Rahbeeni, Z., Alotaibi, M., Goljan, E., Elbardisy, H., ElKalioby, M., Shah, Z., Alruwaili, H., Jaafar, A., Albar, R., Akilan, A., Tayeb, H., Tahir, A., Fawzy, M., Nasr, M., Makki, S., Alfaifi, A., Akleh, H., Yamani, S., Bubshait, D., Mahnashi, M., Basha, T., Alsagheir, A., Abu Khaled, M., Alsaleem, K., Almugbel, M., Badawi, M., Bashiri, F., Bohlega, S., Sulaiman, R., Tous, E., Ahmed, S., Algoufi, T., Al-Mousa, H., Alaki, E., Alhumaidi, S., Alghamdi, H., Alghamdi, M., Sahly, A., Nahrir, S., Al-Ahmari, A., Alkuraya, H., Almehaidib, A., Abanemai, M., Alsohaibaini, F., Alsaud, B., Arnaout, R., Abdel-Salam, G. M. H., Aldhekri, H., AlKhater, S., Alqadi, K., Alsabban, E., Alshareef, T., Awartani, K., Banjar, H., Alsahan, N., Abosoudah, I., Alashwal, A., Aldekhail, W., Alhajjar, S., Al-Mayouf, S., Alsemari, A., Alshuaibi, W., Altala, S., Altalhi, A., Baz, S., Hamad, M., Abalkhail, T., Alenazi, B., Alkaff, A., Almohareb, F., Al Mutairi, F., Alsaleh, M., Alsonbul, A., Alzelaye, S., Bahzad, S., Manee, A. B., Jarrad, O., Meriki, N., Albeirouti, B., Alqasmi, A., AlBalwi, M., Makhseed, N., et al. 2019. Lessons Learned from Large-Scale, First-Tier Clinical Exome Sequencing in a Highly Consanguineous Population. *Am J Hum Genet*, 104, 1182-1201.
- Monroe, G. R., Frederix, G. W., Savelberg, S. M., de Vries, T. I., Duran, K. J., van der Smagt, J. J., Terhal, P. A., van Hasselt, P. M., Kroes, H. Y., Verhoeven-Duif, N. M., Nijman, I. J., Carbo, E. C., van Gassen, K. L., Knoers, N. V., Hövels, A. M., van Haelst, M. M., Visser, G. & van Haften, G. 2016. Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability. *Genet Med*, 18, 949-56.
- Moortgat, S., Berland, S., Aukrust, I., Maystadt, I., Baker, L., Benoit, V., Caro-Llopis, A., Cooper, N. S., Debray, F. G., Faivre, L., Gardeitchik, T., Haukanes, B. I., Houge, G., Kivuva, E., Martinez, F., Mehta, S. G., Nassogne, M. C., Powell-Hamilton, N., Pfundt, R., Rosello, M., Prescott, T., Vasudevan, P., van Loon, B., Verellen-Dumoulin, C., Verloes, A., Lippe, C. V., Wakeling, E., Wilkie, A. O. M., Wilson, L., Yuen, A., Study, D., Low, K. J. & Newbury-Ecob, R. A. 2018. HUWE1 variants cause dominant X-linked intellectual disability: a clinical study of 21 patients. *Eur J Hum Genet*, 26, 64-74.
- Mose, L. E., Perou, C. M. & Parker, J. S. 2019. Improved indel detection in DNA and RNA via realignment with ABRA2. *Bioinformatics*, 35, 2966-2973.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, 51 Pt 1, 263-73.

- Muldowney, T., Manson, D., Kim, R., Stephens, D., Shah, V. & Dell, S. 2014. Primary ciliary dyskinesia and neonatal respiratory distress. *Pediatrics*, 134, 1160-6.
- Murdock, David R., Dai, Hongzheng, Burrage, Lindsay C., Rosenfeld, Jill A., Ketkar, Shamika, Müller, Michaela F., Yépez, Vicente A., Gagneur, Julien, Liu, Pengfei, Chen, Shan, Jain, Mahim, Zapata, Gladys, Bacino, Carlos A., Chao, Hsiao-Tuan, Moretti, Paolo, Craigen, William J., Hanchard, Neil A. & Lee, Brendan 2021. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *The Journal of Clinical Investigation*, 131.
- Nambot, S., Thevenon, J., Kuentz, P., Duffourd, Y., Tisserant, E., Bruel, A. L., Mosca-Boidron, A. L., Masurel-Paulet, A., Lehalle, D., Jean-Marçais, N., Lefebvre, M., Vabres, P., El Chehadeh-Djebbar, S., Philippe, C., Tran Mau-Them, F., St-Onge, J., Jouan, T., Chevarin, M., Poé, C., Carmignac, V., Vitobello, A., Callier, P., Rivière, J. B., Faivre, L. & Thauvin-Robinet, C. 2018. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genet Med*, 20, 645-654.
- Nectoux, J., Girard, B., Bahi-Buisson, N., Prieur, F., Afenjar, A., Rosas-Vargas, H., Chelly, J. & Bienvendu, T. 2007. Netrin G1 mutations are an uncommon cause of atypical Rett syndrome with or without epilepsy. *Pediatr Neurol*, 37, 270-4.
- Neul, J. L., Kaufmann, W. E., Glaze, D. G., Christodoulou, J., Clarke, A. J., Bahi-Buisson, N., Leonard, H., Bailey, M. E., Schanen, N. C., Zappella, M., Renieri, A., Huppke, P. & Percy, A. K. 2010. Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol*, 68, 944-50.
- Ng, P. C. & Henikoff, S. 2001. Predicting deleterious amino acid substitutions. *Genome Res*, 11, 863-74.
- Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y. & Rath, A. 2020. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet*, 28, 165-173.
- Nishimura-Akiyoshi, S., Niimi, K., Nakashiba, T. & Itohara, S. 2007. Axonal netrin-Gs transneuronally determine lamina-specific subdendritic segments. *Proc Natl Acad Sci U S A*, 104, 14801-6.
- O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., Nickerson, D. A., Bernier, R., Shendure, J. & Eichler, E. E. 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, 485, 246-50.
- Pelet, A., Skopova, V., Steuerwald, U., Baresova, V., Zarhrate, M., Plaza, J. M., Hnizda, A., Krijt, M., Souckova, O., Wibrand, F., Andorsdóttir, G., Joensen, F., Sedlak, D., Bleyer, A. J., Kmoch, S., Lyonnet, S. & Zikanova, M. 2019. PAICS deficiency, a new defect of de novo purine synthesis resulting in multiple congenital anomalies and fatal outcome. *Hum Mol Genet*, 28, 3805-3814.
- Pertea, M., Lin, X. & Salzberg, S. L. 2001. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res*, 29, 1185-90.
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*, 20, 110-21.
- Precone, V., Del Monaco, V., Esposito, M. V., De Palma, F. D., Ruocco, A., Salvatore, F. & D'Argenio, V. 2015. Cracking the Code of Human Diseases Using Next-Generation Sequencing: Applications, Challenges, and Perspectives. *Biomed Res Int*, 2015, 161648.
- Rath, A., Olry, A., Dhombres, F., Brandt, M. M., Urbero, B. & Ayme, S. 2012. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Hum Mutat*, 33, 803-8.
- Reis, L. M., Costakos, D., Wheeler, P. G., Bardakjian, T., Schneider, A., Fung, S. S. M. & Semina, E. V. 2021. Dominant variants in PRR12 result in unilateral or bilateral complex microphthalmia. *Clin Genet*, 99, 437-442.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K. & Rehm, H. L. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 17, 405-24.
- Richter, T., Nestler-Parr, S., Babela, R., Khan, Z. M., Tesoro, T., Molsen, E. & Hughes, D. A. 2015. Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. *Value Health*, 18, 906-14.
- Robinson, James T., Thorvaldsdóttir, Helga, Winckler, Wendy, Guttman, Mitchell, Lander, Eric S., Getz, Gad & Mesirov, Jill P. 2011. Integrative genomics viewer. *Nature Biotechnology*, 29, 24-26.

- Robinson, P. N. 2012. Deep phenotyping for precision medicine. *Hum Mutat*, 33, 777-80.
- Rowlands, C., Thomas, H. B., Lord, J., Wai, H. A., Arno, G., Beaman, G., Sergouniotis, P., Gomes-Silva, B., Campbell, C., Gossan, N., Hardcastle, C., Webb, K., O'Callaghan, C., Hirst, R. A., Ramsden, S., Jones, E., Clayton-Smith, J., Webster, A. R., Douglas, A. G. L., O'Keefe, R. T., Newman, W. G., Baralle, D., Black, G. C. M. & Ellingford, J. M. 2021. Comparison of in silico strategies to prioritize rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. *Sci Rep*, 11, 20607.
- Sanger, F., Nicklen, S. & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74, 5463-7.
- Saravanos, G. L., King, C. L., Deng, L., Dinsmore, N., Ramos, I., Takashima, M., Crawford, N., Clark, J. E., Dale, R. C., Jones, C. A., Wood, N. J. & Britton, P. N. 2021. Respiratory Syncytial Virus-Associated Neurologic Complications in Children: A Systematic Review and Aggregated Case Series. *J Pediatr*, 239, 39-49.e9.
- Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Comeau, D. C., Connor, R., DiCuccio, M., Farrell, C. M., Feldgarden, M., Fine, A. M., Funk, K., Hatcher, E., Hoepfner, M., Kane, M., Kannan, S., Katz, K. S., Kelly, C., Klimke, W., Kim, S., Kimchi, A., Landrum, M., Lathrop, S., Lu, Z., Malheiro, A., Marchler-Bauer, A., Murphy, T. D., Phan, L., Prasad, A. B., Pujar, S., Sawyer, A., Schmieder, E., Schneider, V. A., Schoch, C. L., Sharma, S., Thibaud-Nissen, F., Trawick, B. W., Venkatapathi, T., Wang, J., Pruitt, K. D. & Sherry, S. T. 2024. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, 52, D33-d43.
- Schulz, H. L., Grassmann, F., Kellner, U., Spital, G., Ruther, K., Jagle, H., Hufendiek, K., Rating, P., Huchzermeyer, C., Baier, M. J., Weber, B. H. & Stohr, H. 2017. Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort-Impact of Selected Deep Intronic Variants and Common SNPs. *Invest Ophthalmol Vis Sci*, 58, 394-403.
- Seibold, M. A., Wise, A. L., Speer, M. C., Steele, M. P., Brown, K. K., Loyd, J. E., Fingerlin, T. E., Zhang, W., Gudmundsson, G., Groshong, S. D., Evans, C. M., Garantziotis, S., Adler, K. B., Dickey, B. F., du Bois, R. M., Yang, I. V., Herron, A., Kervitsky, D., Talbert, J. L., Markin, C., Park, J., Crews, A. L., Slifer, S. H., Auerbach, S., Roy, M. G., Lin, J., Hennessy, C. E., Schwarz, M. I. & Schwartz, D. A. 2011. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med*, 364, 1503-12.
- Shapiro, A. J., Davis, S. D., Ferkol, T., Dell, S. D., Rosenfeld, M., Olivier, K. N., Sagel, S. D., Milla, C., Zariwala, M. A., Wolf, W., Carson, J. L., Hazucha, M. J., Burns, K., Robinson, B., Knowles, M. R. & Leigh, M. W. 2014. Laterality defects other than situs inversus totalis in primary ciliary dyskinesia: insights into situs ambiguus and heterotaxy. *Chest*, 146, 1176-1186.
- Shashi, V., McConkie-Rosell, A., Rosell, B., Schoch, K., Vellore, K., McDonald, M., Jiang, Y. H., Xie, P., Need, A. & Goldstein, D. B. 2014. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet Med*, 16, 176-82.
- Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., Day, I. N. & Gaunt, T. R. 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*, 34, 57-65.
- Shiraishi, Yuichi, Okada, Ai, Chiba, Kenichi, Kawachi, Asuka, Omori, Ikuko, Mateos, Raúl Nicolás, Iida, Naoko, Yamauchi, Hirofumi, Kosaki, Kenjiro & Yoshimi, Akihide 2022. Systematic identification of intron retention associated variants from massive publicly available transcriptome sequencing data. *Nature Communications*, 13, 5357.
- Smith, H. S., Swint, J. M., Lalani, S. R., Yamal, J. M., de Oliveira Otto, M. C., Castellanos, S., Taylor, A., Lee, B. H. & Russell, H. V. 2019. Clinical Application of Genome and Exome Sequencing as a Diagnostic Tool for Pediatric Patients: a Scoping Review of the Literature. *Genet Med*, 21, 3-16.
- Stankewich, M. C., Stabach, P. R. & Morrow, J. S. 2006. Human Sec31B: a family of new mammalian orthologues of yeast Sec31p that associate with the COPII coat. *J Cell Sci*, 119, 958-69.
- Stenson, P. D., Mort, M., Ball, E. V., Chapman, M., Evans, K., Azevedo, L., Hayden, M., Heywood, S., Millar, D. S., Phillips, A. D. & Cooper, D. N. 2020. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet*, 139, 1197-1207.
- Sturm, M., Schroeder, C. & Bauer, P. 2016. SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics*, 17, 208.
- Sturm, M., Schroeder, C., Mattes, J. & Ossowski, S. 2018. *ngs-bits - Short-read sequencing tools* [Online]. URL: <https://github.com/imgag/ngs-bits> [Letzter Zugriff: 07.03.2021].

- Sun, Y., Ruivenkamp, C. A., Hoffer, M. J., Vrijenhoek, T., Kriek, M., van Asperen, C. J., den Dunnen, J. T. & Santen, G. W. 2015. Next-generation diagnostics: gene panel, exome, or whole genome? *Hum Mutat*, 36, 648-55.
- Takahashi, M., Ogino, T. & Baba, K. 1969. Estimation of relative molecular length of DNA by electrophoresis in agarose gel. *Biochim Biophys Acta*, 174, 183-7.
- Taylor, J. C., Martin, H. C., Lise, S., Broxholme, J., Cazier, J. B., Rimmer, A., Kanapin, A., Lunter, G., Fiddy, S., Allan, C., Aricescu, A. R., Attar, M., Babbs, C., Becq, J., Beeson, D., Bento, C., Bignell, P., Blair, E., Buckle, V. J., Bull, K., Cais, O., Cario, H., Chapel, H., Copley, R. R., Cornall, R., Craft, J., Dahan, K., Davenport, E. E., Dendrou, C., Devuyst, O., Fenwick, A. L., Flint, J., Fugger, L., Gilbert, R. D., Goriely, A., Green, A., Greger, I. H., Grocock, R., Gruszczyk, A. V., Hastings, R., Hatton, E., Higgs, D., Hill, A., Holmes, C., Howard, M., Hughes, L., Humburg, P., Johnson, D., Karpe, F., Kingsbury, Z., Kini, U., Knight, J. C., Krohn, J., Lamble, S., Langman, C., Lonie, L., Luck, J., McCarthy, D., McGowan, S. J., McMullin, M. F., Miller, K. A., Murray, L., Németh, A. H., Nesbit, M. A., Nutt, D., Ormondroyd, E., Oturai, A. B., Pagnamenta, A., Patel, S. Y., Percy, M., Petousi, N., Piazza, P., Piret, S. E., Polanco-Echeverry, G., Popitsch, N., Powrie, F., Pugh, C., Quek, L., Robbins, P. A., Robson, K., Russo, A., Sahgal, N., van Schouwenburg, P. A., Schuh, A., Silverman, E., Simmons, A., Sørensen, P. S., Sweeney, E., Taylor, J., Thakker, R. V., Tomlinson, I., Trebes, A., Twigg, S. R., Uhlig, H. H., Vyas, P., Vyse, T., Wall, S. A., Watkins, H., Whyte, M. P., Witty, L., et al. 2015. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet*, 47, 717-726.
- Temaj, G., Nuhii, N. & Sayer, J. A. 2022. The impact of consanguinity on human health and disease with an emphasis on rare diseases. *Journal of Rare Diseases*, 1, 2.
- Tenorio, R. B., Camargo, C. H. F., Donis, K. C., Almeida, C. C. B. & Teive, H. A. G. 2023. Diagnostic Yield of NGS Tests for Hereditary Ataxia: a Systematic Review. *Cerebellum*.
- Thorvaldsdóttir, Helga, Robinson, James T. & Mesirov, Jill P. 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14, 178-192.
- Towbin, H., Staehelin, T. & Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A*, 76, 4350-4.
- United States Congress 2002. Rare Diseases Act of 2002. URL: <https://www.govinfo.gov/content/pkg/PLAW-107publ280/html/PLAW-107publ280.htm> [Letzter Zugriff].
- van Dam, T. J. P., Kennedy, J., van der Lee, R., de Vrieze, E., Wunderlich, K. A., Rix, S., Dougherty, G. W., Lambacher, N. J., Li, C., Jensen, V. L., Leroux, M. R., Hjeij, R., Horn, N., Texier, Y., Wissinger, Y., van Reeuwijk, J., Wheway, G., Knapp, B., Scheel, J. F., Franco, B., Mans, D. A., van Wijk, E., Képès, F., Slaats, G. G., Toedt, G., Kremer, H., Omran, H., Szymanska, K., Koutroumpas, K., Ueffing, M., Nguyen, T. T., Letteboer, S. J. F., Oud, M. M., van Beersum, S. E. C., Schmidts, M., Beales, P. L., Lu, Q., Giles, R. H., Szklarczyk, R., Russell, R. B., Gibson, T. J., Johnson, C. A., Blacque, O. E., Wolfrum, U., Boldt, K., Roepman, R., Hernandez-Hernandez, V. & Huynen, M. A. 2019. CiliaCarta: An integrated and validated compendium of ciliary genes. *PLoS One*, 14, e0216705.
- Volk, A., Conboy, E., Wical, B., Patterson, M. & Kirmani, S. 2015. Whole-Exome Sequencing in the Clinic: Lessons from Six Consecutive Cases from the Clinician's Perspective. *Mol Syndromol*, 6, 23-31.
- Wang, G. S. & Cooper, T. A. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet*, 8, 749-61.
- Weng, C., Shah, N. H. & Hripacsak, G. 2020. Deep phenotyping: Embracing complexity and temporality-Towards scalability, portability, and interoperability. *J Biomed Inform*, 105, 103433.
- Wieczorek, D. 2018. Autosomal dominant intellectual disability. *Med Genet*, 30, 318-322.
- Wortmann, S. B., Oud, M. M., Alders, M., Coene, K. L. M., van der Crabben, S. N., Feichtinger, R. G., Garanto, A., Hoischen, A., Langeveld, M., Lefeber, D., Mayr, J. A., Ockeloen, C. W., Prokisch, H., Rodenburg, R., Waterham, H. R., Wevers, R. A., van de Warrenburg, B. P. C., Willemsen, Maap, Wolf, N. I., Vissers, Lelm & van Karnebeek, C. D. M. 2022. How to proceed after "negative" exome: A review on genetic diagnostics, limitations, challenges, and emerging new multiomics techniques. *J Inherit Metab Dis*, 45, 663-681.
- Wu, C., Macleod, I. & Su, A. I. 2013. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res*, 41, D561-5.
- Xiang, B., Wang, W., Yi, M., Li, W., Zhou, M., Li, X. & Li, G. 2011. Distribution and expression of alternative splice isoforms of NOR1 in human tissues and cell lines. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*, 36, 597-603.

- Yates, Andrew D, Achuthan, Premanand, Akanni, Wasiu, Allen, James, Allen, Jamie, Alvarez-Jarreta, Jorge, Amode, M Ridwan, Armean, Irina M, Azov, Andrey G, Bennett, Ruth, Bhai, Jyothish, Billis, Konstantinos, Boddu, Sanjay, Marugán, José Carlos, Cummins, Carla, Davidson, Claire, Dodiya, Kamalkumar, Fatima, Reham, Gall, Astrid, Giron, Carlos Garcia, Gil, Laurent, Grego, Tiago, Haggerty, Leanne, Haskell, Erin, Hourlier, Thibaut, Izuogu, Osagie G, Janacek, Sophie H, Juettemann, Thomas, Kay, Mike, Lavidas, Ilias, Le, Tuan, Lemos, Diana, Martinez, Jose Gonzalez, Maurel, Thomas, McDowall, Mark, McMahon, Aoife, Mohanan, Shamika, Moore, Benjamin, Nuhn, Michael, Oheh, Denye N, Parker, Anne, Parton, Andrew, Patricio, Mateus, Sakthivel, Manoj Pandian, Abdul Salam, Ahamed Imran, Schmitt, Bianca M, Schuilenburg, Helen, Sheppard, Dan, Sycheva, Mira, Szuba, Marek, Taylor, Kieron, Thormann, Anja, Threadgold, Glen, Vullo, Alessandro, Walts, Brandon, Winterbottom, Andrea, Zadissa, Amonida, Chakiachvili, Marc, Flint, Bethany, Frankish, Adam, Hunt, Sarah E, Ilsley, Garth, Kostadima, Myrto, Langridge, Nick, Loveland, Jane E, Martin, Fergal J, Morales, Joannella, Mudge, Jonathan M, Muffato, Matthieu, Perry, Emily, Ruffier, Magali, Trevanion, Stephen J, Cunningham, Fiona, Howe, Kevin L, Zerbino, Daniel R & Flicek, Paul 2019. Ensembl 2020. *Nucleic Acids Research*, 48, D682-D688.
- Yeo, G. & Burge, C. B. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*, 11, 377-94.
- Yi, M., Yang, J., Li, W., Li, X., Xiong, W., McCarthy, J. B., Li, G. & Xiang, B. 2017. The NOR1/OSCP1 proteins in cancer: from epigenetic silencing to functional characterization of a novel tumor suppressor. *J Cancer*, 8, 626-635.
- Yin, Y., Miner, J. H. & Sanes, J. R. 2002. Laminets: laminin- and netrin-related genes expressed in distinct neuronal subsets. *Mol Cell Neurosci*, 19, 344-58.
- Zariwala, M. A., Desportes, K. A. & Davis, S. D. 2007. Primary Ciliary Dyskinesia. In: Adam, M. P., Feldman, J., Mirzaa, G. M., Pagon, R. A., Wallace, S. E. & Amemiya, A. (eds.) *GeneReviews*® [Online]. Seattle (WA): University of Washington, 1993-2025. URL: <https://pubmed.ncbi.nlm.nih.gov/20301301/> [Letzter Zugriff: 09.10.2025].
- Zhang, S. Y. 2020. Herpes simplex virus encephalitis of childhood: inborn errors of central nervous system cell-intrinsic immunity. *Hum Genet*, 139, 911-918.
- Zhang, S. Y., Clark, N. E., Freije, C. A., Pauwels, E., Taggart, A. J., Okada, S., Mandel, H., Garcia, P., Ciancanelli, M. J., Biran, A., Lafaille, F. G., Tsumura, M., Cobat, A., Luo, J., Volpi, S., Zimmer, B., Sakata, S., Dinis, A., Ohara, O., Garcia Reino, E. J., Dobbs, K., Hasek, M., Holloway, S. P., McCammon, K., Hussong, S. A., DeRosa, N., Van Skike, C. E., Katolik, A., Lorenzo, L., Hyodo, M., Faria, E., Halwani, R., Fukuhara, R., Smith, G. A., Galvan, V., Damha, M. J., Al-Muhsen, S., Itan, Y., Boeke, J. D., Notarangelo, L. D., Studer, L., Kobayashi, M., Diogo, L., Fairbrother, W. G., Abel, L., Rosenberg, B. R., Hart, P. J., Etzioni, A. & Casanova, J. L. 2018. Inborn Errors of RNA Lariat Metabolism in Humans with Brainstem Viral Infection. *Cell*, 172, 952-965.e18.
- Zheng, S., Vuong, B. Q., Vaidyanathan, B., Lin, J. Y., Huang, F. T. & Chaudhuri, J. 2015. Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA. *Cell*, 161, 762-73.
- Zhong, Y., Xu, F., Wu, J., Schubert, J. & Li, M. M. 2021. Application of Next Generation Sequencing in Laboratory Medicine. *Ann Lab Med*, 41, 25-43.
- Zschocke, J., Byers, P. H. & Wilkie, A. O. M. 2023. Mendelian inheritance revisited: dominance and recessiveness in medical genetics. *Nat Rev Genet*, 24, 442-463.
- Zurek, Birte, Ellwanger, Kornelia, Vissers, Lisenka E. L. M., Schüle, Rebecca, Synofzik, Matthis, Töpf, Ana, de Voer, Richarda M., Laurie, Steven, Matalonga, Leslie, Gilissen, Christian, Ossowski, Stephan, 't Hoen, Peter A. C., Vitobello, Antonio, Schulze-Hentrich, Julia M., Riess, Olaf, Brunner, Han G., Brookes, Anthony J., Rath, Ana, Bonne, Gisèle, Gumus, Gulcin, Verloes, Alain, Hoogerbrugge, Nicoline, Evangelista, Teresinha, Harmuth, Tina, Swertz, Morris, Spalding, Dylan, Hoischen, Alexander, Beltran, Sergi, Graessner, Holm, Haack, Tobias B., Zurek, Birte, Ellwanger, Kornelia, Demidov, German, Sturm, Marc, Kessler, Christoph, Wayand, Melanie, Wilke, Carlo, Traschütz, Andreas, Schöls, Ludger, Hengel, Holger, Heutink, Peter, Brunner, Han, Scheffer, Hans, Steyaert, Wouter, Sablauskas, Karolis, de Voer, Richarda M., Kamsteeg, Erik-Jan, van de Warrenburg, Bart, van Os, Nienke, te Paske, Iris, Janssen, Erik, de Boer, Elke, Steehouwer, Marloes, Yaldiz, Burcu, Kleefstra, Tjitske, Veal, Colin, Gibson, Spencer, Wadsley, Marc, Mehtarizadeh, Mehdi, Riaz, Umar, Warren, Greg, Dizjikan, Farid Yavari, Shorter, Thomas, Straub, Volker, Bettolo, Chiara Marini, Specht, Sabine, Clayton-Smith, Jill, Banka, Siddharth, Alexander, Elizabeth, Jackson, Adam, Faivre, Laurence, Thauvin, Christel, Vitobello, Antonio, Denommé-Pichon, Anne-Sophie, Duffourd, Yannis, Tisserant, Emilie, Bruel, Ange-Line, Peyron, Christine, Pélissier, Aurore, Beltran, Sergi, Gut, Ivo Glynne, Laurie, Steven, Piscia, Davide, Matalonga, Leslie, Papakonstantinou, Anastasios, Bullich,

Gemma, Corvo, Alberto, Garcia, Carles, Fernandez-Callejo, Marcos, Hernández, Carles, Picó, Daniel, Paramonov, Ida, Lochmüller, Hanns, Gumus, Gulcin, Bros-Facer, Virginie, Hanauer, Marc, Oly, Annie, Lagorce, David, Havrylenko, Svitlana, Izem, Katia, et al. 2021. Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases. *European Journal of Human Genetics*, 29, 1325-1331.

Erklärungen zum Eigenanteil

Die Arbeit wurde am Institut für Medizinische Genetik und Angewandte Genomik Tübingen unter Betreuung von Prof. Dr. med. Olaf Rieß durchgeführt.

Die Konzeption der Studie erfolgte in Zusammenarbeit mit Prof. Dr. med. Olaf Rieß, Dr. rer. nat. Rebecca Buchert-Lo, PD Dr. med. Tobias Haack und Dr. med. Lucia Laugwitz.

Alle Daten in Zusammenhang mit *Next generation sequencing* wurden im Vorfeld meiner Studie durch die *Core facility* des Instituts für Medizinische Genetik und Angewandte Genomik generiert (Extraktion der DNA/RNA aus Patientenproben, NGS-Sequenzierläufe). Die bioinformatische Aufarbeitung der Rohdaten erfolgte durch die Standard-Pipelines des Instituts (s. Methodik-Teil dieser Arbeit) unterstützt durch wissenschaftliche Mitarbeiter des Instituts im Bereich der Bioinformatik.

Die Auswertung der NGS-Daten (WES, WGS und RNA-seq) erfolgte selbstständig durch mich nach Einarbeitung durch Dr. rer. nat. Rebecca Buchert-Lo.

Alle weiteren Versuche mit Fibroblasten (Zellkultur, Extraktion von Proteinen), Proteinanalysen (Alle Versuchsschritte des Western Blots, inkl. Detektion und Auswertung) und DNA-Analysen (PCR, Sanger Sequenzierung, Herstellung von cDNA) wurden nach Einarbeitung durch Dr. rer. nat. Rebecca Buchert-Lo von mir eigenständig durchgeführt.

Die statistische Auswertung der Western Blot-Daten erfolgte eigenständig durch mich. Dr. rer. nat. Rebecca Buchert-Lo stand bei Fragen zu Laborabläufen und Auswertungen beratend zur Verfügung.

Ich versichere, das Manuskript selbständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen verwendet zu haben.

Es wurde in keine Künstliche Intelligenz für diese Forschungsarbeit verwendet.

Freiburg, den 10.10.2025

Anna Sophia Ehrhardt

Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die maßgeblich daran beteiligt waren, dass Sie alle nun dieses fertige Manuskript in den Händen halten.

Zuallererst möchte ich Prof. Dr. Olaf Rieß meinen besonderen Dank aussprechen, dass er mich als Doktorvater während der letzten sechs Jahre begleitet hat. Dieses Projekt wäre ohne dein Zutun nicht zustande gekommen. Ich möchte diese spannende Zeit auf keinen Fall missen. Vielen Dank, Olaf!

Ein riesiges Dankeschön gilt Dr. Rebecca Buchert-Lo, die sich bereit erklärt hat mich während aller Phasen der Doktorarbeit zu betreuen. Sei es durch zügige und strukturierte Einarbeitung in diverse Methoden im Forschungslabor, jederzeit ein offenes Ohr bei Forschungsfragen oder die wertvollen Anmerkungen in der Schreibphase, ohne dich wäre ich nie so weit gekommen. Dankeschön, Rebecca!

Dr. Lucia Laugwitz danke ich für ihre tatkräftige Unterstützung bei einigen Fällen dieser Doktorarbeit. Es hat immer Spaß gemacht sich mit dir auszutauschen!

Danke an PD Dr. med. Tobias Haack, dass du bei Fragen jederzeit zur Verfügung standest und dieses Projekt unterstützt hast.

Außerdem danke an Chrissie, Dorothea, Jonasz und alle anderen Labormitglieder, die immer ein offenes Ohr und einige Anmerkungen bei der Kaffeepause für mich parat hatten.

Von Herzen danke ich meiner Familie, die immer für mich da war, mich aufgefangen, ermutigt und an mich geglaubt hat – auch dann, wenn ich selbst gezweifelt habe. Ich hab euch unglaublich lieb!

Mein letzter und ganz besonderer Dank gilt meiner Partnerin Tesse, die mich T-Rex-mäßig unterstützt und jederzeit für mich da ist. Dein Rückhalt bedeutet mir sehr viel! ♥

Anhang

Qualitätsparameter der DNA-Sequenzierungen

ID	Art	Anzahl reads	Reads mit Q-Score >20 [%]	GC-Gehalt [%]	Reads Zielregion [%]	Duplikate [%]	Read depth Zielregion	Anteil Zielregion mit 20x coverage [%]	Anzahl Varianten	Bekannte Varianten [%]	Ti/Tv
MS001	WES	104204030	99,87	51,14	72,94	5,32	130,23	95,32	38828	99,8	2,76
MS002	WES	123957470	99,98	53,07	72,08	15,75	112,06	94,36	38718	99,72	2,71
MS003	WGS	1317555798	99,63	40,92	97,83	27,57	42,82	99,71	4480831	99,72	2,02
MS005	WES	192528348	99,96	52,57	64,07	27,1	130,65	96,45	45857	99,57	2,57
MS006	WES	87232936	99,83	51,39	73,86	6,9	108,6	94,27	36881	99,69	2,73
MS007	WES	92133066	99,89	51,25	74,35	9,49	112,02	94,19	38170	99,66	2,7
MS008	WES	144985332	99,97	53,12	73,88	22,41	127,53	95,04	39203	99,57	2,72
MS009	WES	153064788	99,98	52,87	73,1	23,31	128,94	95	39135	99,79	2,72
MS010	WGS	1204399030	99,91	40,87	97,97	8,49	50,43	99,76	4529086	99,78	2,03
MS011	WGS	1197573178	99,9	40,98	97,86	8,19	50,16	99,77	4596596	99,64	2,02
MS012	WES	116836818	99,81	51,4	73,03	7,31	142,82	95,7	38801	99,83	2,76
MS013	WGS	1388978280	99,9	41,27	98,2	15,71	53,42	99,81	4492352	99,68	2,01
MS014	WES	130916278	99,98	53,61	77,17	15,68	132,24	94,72	38597	99,79	2,71
MS015	WES	139239882	99,98	53,98	78,2	17,36	140,2	94,38	38868	99,77	2,73
MS016	WES	83050146	99,87	50,48	71,59	4,85	101,56	94,69	39027	99,8	2,74
MS017	WES	119631756	99,87	50,65	71,59	9,87	138,14	95,87	39117	99,8	2,74

Anmerkung: Der Q-Score von 20 entspricht einer Fehlerwahrscheinlichkeit beim *Base calling* von 1%. Ti = Transition, Tv = Transversion.

Qualitätsparameter RNA-Sequenzierungen

ID	Anzahl reads	Reads mit Q-Score >20 [%]	GC-Gehalt [%]	Reads Zielregion [%]	Duplikate [%]	Read depth Zielregion	Anteil Zielregion mit 20x coverage [%]
MS001	139850984	99,94	51,81	64,46	35,84	156,02	40,38
MS002	133281318	99,93	51,23	69,2	26,51	187,28	43,33
MS003	147616340	99,96	51,83	70,72	33,82	190,26	44,37
MS005	88990242	99,96	51,22	65,84	27,29	41,69	26,7
MS006	132173538	99,93	51,57	68,42	28,86	180,56	45,2
MS007	41459454	99,9	48,78	65,79	11,83	34,71	25,99
MS008	140885840	99,95	52,17	67,59	48,25	135,61	39,34
MS009	125955836	99,93	51,01	68,83	29,62	166,15	42,01
MS010	139482850	99,93	51,73	71,78	32,08	184,74	43,65
MS011	42698930	99,9	50,04	66,62	14,71	35,27	26,57
MS012	148507190	99,96	52,22	73,34	36,14	188,93	43,66
MS013	93171798	99,96	50,26	64,33	24,19	45,15	30,24
MS014	126814718	99,96	51,73	66,18	25,92	171,51	43,03
MS015	141741368	99,95	52,21	67,96	27,25	191,37	43,88
MS016	125049674	99,93	52,3	66,91	27,3	161,95	41,13
MS017	127512568	99,94	52,21	66,33	26,67	168,09	41,63