

Aus dem Institut für Medizinische Genetik und Humangenetik  
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Untersuchung des diagnostischen Potenzials der  
Ganzgenomsequenzierung am Beispiel von 69 Fällen mit  
angeborenen Extremitätenfehlbildungen

Investigation of the diagnostic potential of  
Whole-Genome-Sequencing by the example of 69 cases with  
congenital limb malformations

zur Erlangung des akademischen Grades  
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Jonas Elsner

Datum der Promotion: 30.11.2023



## Inhaltsverzeichnis

Tabellenverzeichnis .....	vi
Abbildungsverzeichnis .....	vii
Abkürzungsverzeichnis .....	viii
Zusammenfassung .....	1
1. Einleitung .....	4
2. Methodik .....	10
2.1 Kohortenrekrutierung und Studiendesign .....	10
2.2 Phänotypisierung der Patient*innen .....	11
2.3 Whole-Genome-Sequencing .....	11
2.3.1 Datenverarbeitung in SODAR und VarFish .....	11
2.4 Analyse der Single-Nucleotide-Variants .....	12
2.4.1 Trio und Singleton Filtration in VarFish .....	12
2.4.2 Analyse durch MutationDistiller und Exomiser .....	13
2.4.3 Einzel-Analyse der Index-Patient*innen .....	14
2.4.4 Manuelle Auswertung der Kandidaten-Varianten .....	15
2.5 Analyse der Strukturvarianten .....	16
2.5.1 Heterozygote Varianten .....	16
2.5.2 Homozygote Varianten .....	18
2.6 Analyse der nicht-kodierenden Varianten .....	18
2.7 Genotyp-Phänotyp-Korrelationsanalyse von <i>GLI3</i> .....	19
2.8 Analyse des Genlocus um <i>HMGB1</i> .....	20
2.9 Analyse von <i>HMGB1</i> Patient*innen mittels Face2Gene .....	21
3. Ergebnisse .....	22
3.1 Auswertung und Filterprozess der SNV .....	22
3.2 Auswertung und Filterprozess der SVs .....	23
3.3 Varianten in bekannten Krankheits-Lokalisationen .....	25

---

3.3.1 SNVs in bekannten Lokalisationen .....	25
3.3.2 Repeat-Expansionen in <i>HOXD13</i> .....	27
3.3.3 SVs in bekannten Lokalisationen .....	27
3.4 Mutationen in kodierenden Bereichen von Kandidatengen .....	32
3.4.1 Loss of Function Mutationen in <i>UBA2</i> lösen Ektrodaktylie aus.....	32
3.4.2 <i>HMGB1</i> -Mutation als Ursache von BPTAS .....	33
3.4.3 <i>SEMA3D</i> als Kandidatengen für Short-Stature-Limb-Abnormality-Syndrome	34
3.4.4 Varianten in <i>ALDH1A2</i> als potenzieller Auslöser für häutige Syndaktylien....	34
3.5 Varianten unklarer Signifikanz .....	35
3.6 Analyse des nicht-kodierenden Bereichs .....	37
3.6.1 <i>De novo</i> Varianten in nicht-kodierenden Bereichen.....	39
3.6.2 Nicht-kodierende Varianten in VISTA Elementen .....	40
3.6.3 Analyse von räumlich benachbarten Varianten .....	40
3.7 Genotyp-Phänotyp-Korrelationsanalyse von <i>GLI3</i> .....	41
3.7.1 Analyse der hauseigenen Fälle .....	41
3.7.2 Analyse bereits publizierter Fälle.....	44
3.8 Validierung der Pathogenität von <i>HMGB1</i> -Mutationen.....	46
3.8.1 <i>HMGB1</i> -Varianten bei Patient*innen mit BPTAS.....	46
3.8.2 Face2Gene Analyse von BPTAS Individuen .....	47
3.8.3 Phasentrennung abhängig von <i>HMGB1</i> .....	48
4. Diskussion.....	50
4.1 Identifizierte Varianten .....	50
4.1.1 <i>UBA2</i> als neues Krankheitsgen.....	52
4.1.2 Varianten unklarer Signifikanz.....	52
4.2 Analyse des nicht-kodierenden Bereichs .....	53
4.3 Qualität der Analyse.....	54

---

4.4 Limitationen von WGS bei der Untersuchung von 69 IPs mit Extremitätenfehlbildungen.....	56
4.5 Genotyp-Phänotyp-Korrelation von <i>GLI3</i> .....	57
4.6 Kleine intragenische Deletion von <i>GLI3</i> .....	59
4.7 Final-exon frameshift Varianten in <i>HMGB1</i> lösen BPTAS aus.....	61
4.8 Genotyp-Phänotyp-Korrelationen von <i>HMGB1</i> .....	63
4.9 Veränderte Proteinfunktion von HMGB1 .....	64
4.10 WGS ermöglicht Auffinden neuer Varianten .....	64
5. Schlussfolgerungen.....	66
Literaturverzeichnis .....	67
Eidesstattliche Versicherung .....	78
Anteilerklärung an den erfolgten Publikationen.....	79
Auszug aus der Journal Summary List Publikation 1.....	82
Druckexemplar der Publikation.....	84
Auszug aus der Journal Summary List Publikation 2.....	95
Druckexemplar der Publikation.....	98
Auszug aus der Journal Summary List Publikation 3.....	106
Druckexemplar der Publikation.....	107
Lebenslauf.....	115
Komplette Publikationsliste.....	116
Danksagung .....	117

**Tabellenverzeichnis**

Tabelle 1: VUS-Klassifikation.....	36
Tabelle 2: GLI3 Varianten.....	42

## Abbildungsverzeichnis

Abbildung 1: Anzahl der kodierenden SNVs pro Patient nach dem Filtern in VarFish...	22
Abbildung 2: Identifizierte SVs von 68 Patient*innen nach dem Filter in VarFish.....	24
Abbildung 3: Visualisierung Strukturvariante und Phänotyp von I11.....	29
Abbildung 4: Analyse der Strukturvariante von I11.....	31
Abbildung 5: Visualisierung der Analyse des nicht-kodierenden Bereichs.....	38
Abbildung 6: Lokalisation von 15 Varianten in <i>GLI3</i> .....	43
Abbildung 7: Übersicht der bereits publizierten <i>GLI3</i> Varianten und Phänotypen.....	45
Abbildung 8: Workflow und Ergebnisse der Genotyp-Analyse der HMGB1-I1 bis -I5.....	47
Abbildung 9: Analyse des Proteins HMGB1.....	49
Abbildung 10: Sanger-Sequenzierung von <i>GLI3</i> Exon 2-4 des Trios von IP GLI3-I17.....	60
Abbildung 11: Struktur und Ladung von HMGB1.....	62

## Abkürzungsverzeichnis

aCGH	array Comparative Genomic Hybridization
ACGS	Association for Clinical Genomic Science
ACMG	American College of Medical Genetics
AER	apical ectodermal ridge
Asp	Aspartat
bp	Basenpaare
BPES	Blepharophimosis-ptosis-epicanthus inversus Syndrom
BPTAS	Brachyphalangy-Polydactyly-and-Tibial-Aplasia-Syndrom
CNV	Copy Number Variation
Dup	Duplikation
ExAC	Exome Aggregation Consortium
F2GC	Face2Gene CLINIC
F2GR	Face2Gene RESEARCH
GCPS	Greig-Zephalopolysyndaktylie-Syndrom
GERP	Genomic Evolutionary Rate Profiling
Glu	Glutamat
gnomAD	Genome Aggregation Database
GWAS	genomweite Assoziationsstudien
HMG	high mobility group
HPO	Human Phenotype Ontology
IDR	intrinsically disordered region
IMG	Instituts für Medizinische Genetik und Humangenetik
I	Individuum
IP	Index-Patient*in
IPD	Isolierte Polysyndaktylie
kb	Kilobasen
LoF	Loss of Function
LSS	Laurin-Sandrow-Syndrom
M	Male
MGI	Mouse Genome Informatics
NMD	nonsense mediated decay
OMIM	Online Mendelian Inheritance in Man
PAP-A	postaxiale Polydaktylie Typ A
PAP-B	postaxiale Polydaktylie Typ B
PHS	Pallister-Hall-Syndrom
PPD-IV	präaxiale Polydaktylie Typ IV
qPCR	quantitative real-time PCR
SCT	Spondylocarpotarsal-Synostose-Syndrom
SHFM	Spalthand-Spaltfuß-Malformation
SNV	Single Nucleotide Variant
SODAR	System for Omics Data Analysis and Retrieval
SV	Strukturvariante
TAD	topologisch assoziierte Domäne
VCF	Variant Calling File
VUS	Variante unklarer Signifikanz
WGS	Whole-Genome-Sequencing

## Zusammenfassung

Seit der Sequenzierung des ersten menschlichen Genoms 2001 hat die Ganz-Genom-Analyse (Whole-Genome-Sequencing, WGS) große Fortschritte gemacht. Mittlerweile ist es möglich, die Methode im klinisch-diagnostischen Alltag anzuwenden. Allerdings existieren etablierte Verfahren zur Identifikation von Mutationen, welche durch jahrelange Anwendung einen hohen Stellenwert in der genetischen Diagnostik einnehmen. Inwiefern WGS eine Ergänzung oder gar Alternative zu diesen Verfahren darstellt, und welches diagnostische Potential die Methode besitzt, soll in der vorliegenden Arbeit untersucht werden.

In 69 Fällen mit angeborenen Fehlbildungen von mindestens einer Extremität, bei denen die genetische Routinediagnostik keine pathogene Variante identifizieren konnte, wurde ein WGS durchgeführt. Die Daten aller Patient\*innen wurden mit Hilfe von VarFish (Version v0.17.2) und SODAR (System for Omics Data Analysis and Retrieval) prozessiert und danach manuell ausgewertet. Dabei erfolgte die Betrachtung von Einzelnukleotidvarianten (Single Nucleotide Variants, SNVs) und Strukturvarianten (SVs) im proteinkodierenden wie nicht-proteinkodierenden Bereich des Genoms. Im Anschluss wurden *GLI3* und *HMGB1*, hinsichtlich möglicher Genotyp-Phänotyp-Korrelationen untersucht, wobei Erkenntnisse aus dem WGS als Grundlage genutzt wurden.

Insgesamt konnten pathogene Varianten in 12 Fällen (17.4%) identifiziert werden. SNVs zeigten sich in vier Patienten jeweils in den Genen *FGFR1*, *FGFR2*, *GLI3* und *BHLHA9*. Die Variante in *FGFR1* war bereits in der Literatur beschrieben worden, die anderen drei wurden erstmals nachgewiesen. In einem weiteren Fall konnte eine Variante in *HMGB1* als pathogen identifiziert werden. Weiterhin konnten SNVs in zwei potenziell krankheitsauslösenden Genen (*SEMA3D*, *ALDH1A*) aufgefunden werden. Darüber hinaus fanden sich *UBA2*-Varianten in zwei Ektrodaktylie-Fällen der Kohorte sowie einem weiteren, dem Institut für Medizinische Genetik und Humangenetik bekannten, Fall. Dies unterstützt die aktuelle These von *UBA2* als Krankheitsgen. Neben den SNVs konnten zwei komplexe SVs gefunden werden. Weiterhin konnte eine pathogene Alanin-Repeat-Expansion in *HOXD13* identifiziert werden. Es konnten keine pathogenen Varianten im nicht-proteinkodierenden Bereich des Genoms gefunden werden. Bei *GLI3* und *HMGB1* wurden neue Erkenntnisse bezüglich der Genotyp-Phänotyp-Korrelationen an den entsprechenden Loci gewonnen.

Insgesamt konnte bei 12 (17,4%) von 69 Fällen, bei denen herkömmliche Verfahren bisher keine molekular-genetische Diagnose erbrachten, durch das WGS die Etablierung einer Diagnose ermöglicht werden. Darüber hinaus, lieferte WGS die Grundlage für die genauere Betrachtung der Loci von *GLI3* und *HMGB1*, wobei final-exon frameshifts von *HMGB1* als Ursache des Brachyphalangie-Polydaktylie-Tibia-Hypo/Aplasie-Syndroms identifiziert werden konnten.

Durch künftige Verbesserung von Sequenzierungsmethoden sowie prozessierender und auswertender Software ist zu erwarten, dass die diagnostische Rate des WGS weiter ansteigen wird. Allerdings zeigt es sich bereits heute als potentes Verfahren für den klinisch-genetischen Alltag sowie verschiedene Forschungsansätze. Es ermöglicht als ein einziger umfassender Test die Identifikation von Varianten unterschiedlichen Typs, inklusiver kleiner, aber komplexer Strukturvarianten, die durch herkömmliche Verfahren nicht erfasst werden.

## Abstract

Since the sequencing of the first human genome in 2001, Whole-Genome-Sequencing (WGS) has made great progress. It is now possible to apply the method in everyday clinical diagnostics. However, there are established approaches for the identification of mutations, which, due to years of application, have a high status in genetic diagnostics. To what extent WGS represents a complement or even an alternative to these methods, and which diagnostic results are possible by considering the genome in its completeness, will be investigated in the present work.

Sixty-nine cases with malformations of at least one limb received WGS, excluding those who had already received a molecularly confirmed genetic diagnosis. The data of all were processed using VarFish (version v0.17.2) and SODAR (System for Omics Data Analysis and Retrieval) and then analyzed manually. Single nucleotide variants (SNVs) and structural variants (SVs) in the protein-coding and non-protein-coding regions of the genome were examined. Subsequently, *GLI3* and *HMGB1* were investigated for possible genotype-phenotype correlations using findings from the WGS.

Overall, we identified pathogenic variants in twelve cases (17.4%). SNVs appeared in four patients in each of the genes *FGFR1*, *FGFR2*, *GLI3*, and *BHLHA9*. The variant in

*FGFR1* had already been described in the literature, the other three were detected for the first time. In one case, a variant in *HMGB1* was identified as pathogenic. Furthermore, SNVs were detected in two potentially disease-causing genes (*SEMA3D*, *ALDH1A*). In addition, *UBA2* variants were found in two ectrodactyly cases in the cohort, as well as another case known to the Institute of Medical Genetics and Human Genetics. This supports the current hypothesis of *UBA2* as a disease gene. In addition to SNVs, two complex SVs were detected. One inversion, which is flanked by a deletion, was found in a patient with ectrodactyly at the locus of split hand foot malformation (SHFM) 3. The other variant is a complex translocation between 7q36.3 and 9p24.1 in a patient with bilateral mirror-image polydactyly of the hands and feet. Furthermore, a pathogenic alanine repeat expansion in *HOXD13* was identified in a patient with synpolydactyly. No pathogenic variants could be found in the non-protein-coding region of the genome. In *GLI3* and *HMGB1*, new findings regarding the genotype-phenotype correlations at the corresponding loci were obtained.

Thus, of sixty-nine cases in which conventional methods did not previously yield a genetic diagnosis, WGS enabled the establishment of a diagnosis in twelve cases (17.4%). In addition, WGS provided the basis for a closer look at the loci of *GLI3* and *HMGB1*, with final-exon frameshifts of *HMGB1* being identified as the cause of brachyphalangia-polydactyly-tibia-hypo/aplasia syndrome. Future improvements in sequencing methods, as well as in processing and evaluation software, are expected to further increase the diagnostic yield of WGS. However, it is already proving to be a potent tool for everyday clinical genetics, as well as various research approaches. As a single comprehensive test, it allows the identification of multiple variants of several types, including small but complex structural variants that are not detected by conventional methods.

## 1. Einleitung

Extremitätenfehlbildungen gehören zu den häufigsten kongenitalen Erkrankungen des Menschen. In europäischen Populationen beträgt die Inzidenz 1,4 bis 30 Fälle auf 10.000 Geburten (Shi et al., 2018), wobei circa 18% aufgrund zusätzlicher Organdysfunktionen oder -malformationen in den ersten 6 Lebensjahren sterben. Ursachen für Extremitätenfehlbildungen können in exogenen Umweltfaktoren sowie in spontan aufgetretenen oder vererbten genetischen Läsionen gefunden werden, welche die Entwicklung der Extremitäten *in utero* beeinflussen (Zuniga et al., 2012).

Physiologischerweise beginnt die Entwicklung der Gliedmaßen bereits am Ende der vierten Embryonalwoche. Zuerst erscheinen circa am 25. Tag faltenförmige Knospen an der seitlichen Rumpfwand, welche später die oberen Extremitäten bilden. Ungefähr zwei Tage später sind die Extremitätenknospen der unteren Extremitäten zu erkennen. Diese Knospen bestehen aus einem mesenchymalen Kern mit einem Überzug aus Ektoderm. Durch das Einwirken von regulatorischen Genen verdickt sich das distale Ektoderm zur apical ectodermal ridge (apikale Randleiste - AER). Vor allem der fibroblastic growth factor 8 reguliert die initiierten Prozesse des Extremitätenwachstums, welche von der AER ausgehen (Newman et al., 2018). Zwischen dem mesenchymalen Kern und der Randleiste bilden sich daraufhin Regulationskreisläufe, welche die weitere Entwicklung der Extremität steuern (Tanaka, 2013; Mundlos & Horn, 2014). In der sechsten Woche bilden sich durch Einschnürungen Hand- und Fußplatten, in welchen sich Finger bzw. Fußstrahlen entwickeln. Durch Apoptose der ekto- und mesodermalen Zellen in den Finger- und Zehenzwischenräumen kommt es zur Ausbildung der getrennten Digiti. Dieser Prozess wird unter anderem von sogenannten patterning networks reguliert. Dabei handelt es sich um komplexe Molekül-Netzwerke, welche durch ihre räumliche und zeitliche Varianz die Entwicklung der Extremität steuern (Newman et al., 2018). Postnatal erfolgt die Bildung von Knochenkernen sowie das Längenwachstum (Ulfig, 2009).

Diese Bildungs- und Entwicklungsprozesse der Extremitäten sind sowohl zeitlich als auch räumlich hoch reguliert. Dabei spielen diverse Gene eine entscheidende Rolle, welche in hochdynamischen, selbst regulierenden Netzwerken das unabhängige Wachstum der Extremität steuern (Tanaka, 2013; Zeller et al., 2009). Durch molekular-genetische Diagnostik, wie das Sequenzieren von menschlichen Genomen und dem Genom der Maus

konnten in den letzten Jahren zahlreiche Erkenntnisse über den Entwicklungsprozess der Extremitäten gewonnen werden. Man begann Zusammenhänge zwischen einzelnen genetischen, regulatorischen Pfaden zu entdecken, was zur Identifizierung von sogenannten Krankheitsgenen beitrug. Sollte es durch Mutationen dieser Gene zu einer Fehlregulation kommen, können bestimmte genetische Erkrankungen entstehen. Im Mausmodell konnte man diese Fehlregulationen durch gezielte genetische Veränderungen simulieren, was zu einem erweiterten Verständnis der Extremitätenentwicklung führte. Dies konnte dann wiederum auf pathogenetische Prozesse der menschlichen *in utero* Entwicklung übertragen werden (Zuniga et al., 2012).

Ein Beispiel hierfür ist die Spalthand-Spaltfuß-Malformation 3 (SHFM3; MIM 246560). Bei dieser seltenen Fehlbildung sind die zentralen Strahlen von Händen und Füßen betroffen. Diese zeigen dabei Syndaktylien, mediale Spalten und Aplasien bzw. Hypoplasien der Phalangen, Metacarpalknochen und Metatarsalknochen. Weiterhin können mentale Retardierung sowie ektodermale oder kraniofaziale Fehlbildungen auftreten (Elliott & Evans, 2006). Als Ursache hierfür wird eine Duplikation (Dup) im Bereich 10q24 vermutet. De Mollerat *et al.* zeigten, dass in sieben SHFM3 Patient\*innen Bereiche dieses Locus dupliziert waren. Alle trugen dabei mindestens partielle Duplikationen des Gens *Dactylin* sowie vollständige Duplikationen der Gene *LBX1*, *BTRC*, und *POLL* (Mollerat & de Mollerat, 2003). Im Mausmodell führte eine strukturelle Veränderung im *Dactylin*-Gen zu einer verringerten Transkription und zur Dactylaplasie, dem Korrelat der menschlichen SHFM3. Auch in anderen Individuen mit SHFM3 fand sich eine verminderte Transkription von *Dactylin* (Basel et al., 2003). Neben diesen scheinbar eindeutigen Genotyp-Phänotyp-Korrelationen existieren allerdings auch Gene, deren Varianten je nach Art und Lokalisation ein ganzes Spektrum von Phänotypen auslösen können. Bei *GLI3* beispielsweise, lösen verschiedene Varianten diverse Phänotypen wie das Greig-Zephalopolysyndaktylie-Syndrom (GCPS; MIM 175700), die isolierte Polysyndaktylie (IPD) oder das Palister-Hall-Syndrom (PHS; MIM 146510) aus (Johnston et al., 2010).

Um diese Veränderungen der DNA zu analysieren, gibt es diverse Möglichkeiten der Sequenzierung, die sich in ihrem Ergebnis sowie der diagnostischen Funktion unterscheiden. Sequenzfehler, die durch Veränderungen einzelner Basenpaare entstehen, können unter anderem durch Sanger-Sequenzierung gefunden werden (McGinn & Gut, 2013). Handelt es sich um kleine Kopiezahlpolymerismen (Copy Number Variations, CNVs),

bietet sich eine array Comparative Genomic Hybridization (aCGH) an (Cheung & Bi, 2018). Auch mikroskopische Anwendungen, wie die Fluoreszenz *in-situ* Hybridisierung kommen bei der Untersuchung von großen strukturellen Läsionen sowie chromosomalen Aberrationen zum Einsatz (Rodriguez-Mateos et al., 2020). Nicht selten sind komplexe strukturelle Veränderungen ursächlich, zu deren Identifizierung mehrere Verfahren kombiniert werden müssen (Berisha et al., 2020).

Diese und weitere Arten der DNA-Sequenzierung bilden für diverse wissenschaftliche Ansätze eine wichtige Grundlage. Ihren Anfang nahmen sie allerdings erst vor kurzer Zeit. 1968 berichtete Wu die Sequenzierung von 12 Basen der cohesive ends des Escherichia-Virus Lambda, 1973 folgten Gilbert und Maxam mit der lactose-repressor binding site des Virus, welche 24 Basen umfasste. Die Sequenzierung dieser Basen erfolgte über einen Zeitraum von 24 Monaten, also eine Base pro Monat (Shendure et al., 2019). In den darauffolgenden Jahren entwickelte sich die Sequenzierung deutlich. Maschinen und Methoden wurden leistungsfähiger und günstiger, sodass immer größere Abschnitte der DNA in absehbarer Zeit sequenziert werden konnten. Weiterhin waren die Methoden nun nicht mehr nur ausgewählten Wissenschaftlern zugänglich, sondern konnten von der breiten wissenschaftlichen Gemeinde verwendet werden. Durch diesen Fortschritt war es möglich, größere Bereiche des Genoms zu analysieren, oder sogar ganze Genome von Lebewesen darzustellen. So konnte zum Beispiel 1982 das Genom des Escherichia-Virus Lambda sequenziert werden ("Nucleotide Sequence of Bacteriophage  $\lambda$  DNA," 1982). Auf Grundlage dieser Untersuchungen und der durch sie gewonnenen Erkenntnisse über die Vorgänge auf der Ebene des Genoms wurde in den späten 1980er Jahren das Human-Genome-Project ins Leben gerufen. Dieses sollte die Methoden und Erkenntnisse nutzen, um die DNA-Sequenzierung auf die Größe des menschlichen Genoms von circa drei Milliarden Basen zu skalieren. Nachdem 1988 durch ein Komitee des US National Research Council erste Konzepte formuliert wurden, begann nach einem 1995 beginnenden Pilotprojekt, im Jahr 1999 die Sequenzierung. An dieser ersten Sequenzierung arbeiteten wissenschaftliche Gruppen aus über 20 Ländern. Innerhalb eines Jahres konnte eine Skizze des menschlichen Genoms geschaffen werden, welche 90% der genetischen Information abdeckt (Consortium & International Human Genome Sequencing Consortium, 2001). Bis 2004 wurde diese Skizze auf 99% vervollständigt (Gyapay et al., 2004).

Dieser Prozess kostete 2,7 Milliarden US-Dollar, forderte die Mitarbeit diverser wissenschaftlicher Gruppen und dauerte vier Jahre, wenn man die Zeit der Pilotprojekte nicht einrechnet.

20 Jahre später sieht die Realität anders aus. 2010 wurde das 1000 Genomes Project fertig gestellt, mittlerweile liefert die Genome Aggregation Database (gnomAD) mit circa 15.000 humanen Referenzgenomen die größte Datenbank zur Abgleichung von Patient\*innengenomen (Q. Wang et al., 2021). Die Genomsequenzierung eines Menschen kostet aktuell weniger als 1.000 US-Dollar und kann in circa 12 Stunden durchgeführt werden (Schwarze et al., 2019; van Dijk et al., 2014). Durch diese geringen Kosten sowie die schnelle Verfügbarkeit, ist Whole-Genome-Sequencing in immer mehr klinischen Anwendungen zu finden.

Das Whole-Genome-Sequencing liefert einige Vorteile gegenüber anderen Sequenziermethoden. Neben der Analyse von missens, nonsense und splice-site Varianten, bietet WGS die Möglichkeit zur Detektion von CNVs. Wie bereits beschrieben, kann hierfür auch eine aCGH genutzt werden. Bei dieser Methode wird eine vorher festgelegte Auswahl an Kandidatenloci auf CNVs getestet. Moderne Ansätze erreichen dabei bis zu 1945 Kandidatengene, mit einer Auflösung von 15 Kilobasen (kb) (Cheung & Bi, 2018). Obwohl aCGH eine günstigere, weitaus zugänglichere Methode darstellt, konnte in diversen Studien gezeigt werden, dass WGS zum Auffinden von CNVs identische, wenn nicht sogar bessere Ergebnisse liefert (Petersen et al., 2017; Russo et al., 2014). Neben dem Auffinden von klassischen CNVs im proteinkodierenden Bereich (Exom), bietet WGS auch die Möglichkeit, CNVs im nicht-kodierenden Bereich des Genoms zu identifizieren (Hehir-Kwa et al., 2015).

Weiterhin kann eine unvoreingenommene Betrachtung aller bekannten Gene erfolgen, die Fixierung auf vorher festgelegte Kandidatengene bleibt aus. Dadurch können genomweite Assoziationsstudien (GWAS) durchgeführt werden, welche das Verständnis von komplexen Krankheiten ermöglichen. Einen großen Beitrag zum Verständnis lieferten GWAS zum Beispiel im Bereich der chronisch entzündlichen Darmerkrankungen. Diese phänotypisch komplexen Krankheiten besitzen neben unbekanntem Umwelteinfluss eine starke genetische Komponente. Dabei ist allerdings nicht von einer monogenen Ur-

sache auszugehen. Es handelt sich eher um verschiedene Loci, welche die Wahrscheinlichkeit des Krankheitsausbruchs modifizieren. Durch die unvoreingenommene Analyse des Genoms sowie das Abgleichen mit gesunden Referenzgenomen, konnten mittlerweile 200 verschiedene Loci identifiziert werden. WGS bietet also neben der Diagnose von Mendel'schen Krankheiten, auch Möglichkeiten zur Analyse komplexer genetischer Fragestellungen sowie monogener oder polygener Erkrankungen mit variabler Penetranz (Petersen et al., 2017). Darüber hinaus besitzt die unvoreingenommene Betrachtung des gesamten Genoms das Potential, neue Varianten und damit neue Erkenntnisse im Bereich genetischer Regulations- und Pathomechanismen zu gewinnen.

Zusammenfassend lässt sich sagen, dass sich die molekular-genetische Diagnostik in den letzten Jahren deutlich gewandelt hat. Mittlerweile verfügt man über die technischen Möglichkeiten, ganze Genome in einer wirtschaftlichen Art und Weise zu sequenzieren, was die Methode des WGS als molekular-genetische Standarddiagnostik attraktiv erscheinen lässt. Außerdem existieren diverse Projekte auf Forschungsebene, welche immer mehr Genome von Menschen mit zunehmender Vollständigkeit sequenzieren. Dabei rücken neben dem detaillierten Verständnis genetischer Funktionen, auch regulatorische Regionen und der nicht-kodierende Bereich zunehmend in den Fokus der Forschung (Prokop et al., 2018).

Patienten mit Fehlbildungen der Extremitäten bilden für das WGS eine interessante Kohorte, da in ihrer Pathophysiologie unterschiedlichste Gene und deren Interaktion eine Rolle spielen. Weiterhin treten in diesem Spektrum von Krankheiten neben unvollständiger Penetranz auch verschiedene Mutationsmechanismen auf, sodass sie sich als Modell für andere erbliche Erkrankung eignen.

Das WGS, welches im Vergleich zu herkömmlichen Sequenzierungsmethoden mindestens gleichwertige Ergebnisse liefert, bietet hier durch die holistische Betrachtung des Genoms eine Chance für neue Erkenntnisse. Vor allem durch die gleichzeitige Betrachtung von SNVs, CNVs und des nicht-kodierenden Bereichs eines\*r Patient\*in, könnten sich neue Möglichkeiten ergeben, unbekannte Krankheits-Loci zu identifizieren oder bestehende Erkenntnisse zu unterstützen. Auch das Auffinden von komplexen genetischen Varianten könnte ermöglicht, beziehungsweise vereinfacht werden.

Welchen Wert das WGS als diagnostisches Standardverfahren der molekularen Genetik bietet, und ob es als eine one-for-all Teststrategie eingesetzt werden kann, soll in der vorliegenden Arbeit untersucht und erörtert werden.

## 2. Methodik

### 2.1 Kohortenrekrutierung und Studiendesign

Die Rekrutierung der Kohorte erfolgte über einen Zeitraum von 17 Jahren. Sie setzt sich zusammen aus Patient\*innen der Abteilung für Kinderhandchirurgie des Katholischen Kinderkrankenhauses Wilhelmstift in Hamburg sowie Patient\*innen des Instituts für Medizinische Genetik und Humangenetik (IMG) der Charité, deren Proben von externen sowie hauseigenen Ärztinnen und Ärzten zu diagnostischen Zwecken eingesendet wurden. Eingeschlossen wurden nur Proben solcher Individuen, die eine Fehlbildung an mindestens zwei Extremitäten aufwiesen oder in deren Familie mindestens zwei Mitglieder von Fehlbildungen der Extremitäten betroffen waren. Ausschlusskriterien waren eine etablierte molekular-genetische Diagnose, ein Amniotisches-Band-Syndrom (MIM 217100) als vermutete Ursache oder eine isolierte Klinodaktylie des fünften Fingers. Der Einschluss erfolgte vor der Durchführung von WGS.

Insgesamt konnten 69 Patient\*innen eingeschlossen werden. Bei 64 davon handelte es sich um komplette Trios, bei denen neben dem\*der Index-Patient\*in (IP) auch beide Eltern sequenziert werden konnten. In einem Fall konnte nur ein Elternteil sequenziert werden, in vier weiteren Fällen erfolgte nur die Sequenzierung des\*der Index-Patient\*in. Weiterhin konnte in einem Fall zusätzlich zum bestehenden Trio ein Bruder der IP sequenziert werden.

Alle Patient\*innen erhielten vor ihrem Einschluss diagnostische molekular-genetische Untersuchungen, abhängig vom präsentierten Phänotyp. Dabei handelte es sich um eine aCGH sowie, in manchen Fällen, um eine gezielte Sequenzierung von bekannten Genen, welche den entsprechenden Phänotyp des Patienten nachgewiesenermaßen auslösen können. Hierfür bediente man sich bestehender Standardprotokolle (Flöttmann et al., 2018). Die Auswahl dieser Gene erfolgte durch qualifizierte Genetiker\*innen. Für das WGS wurden nur solche Proben verwendet, bei denen im Rahmen dieser Methoden keine krankheitsauslösende Mutation gefunden werden konnte.

## 2.2 Phänotypisierung der Patient\*innen

Zu Beginn erfolgte eine Sichtung der zur Verfügung stehenden Informationen über die Patient\*innen. Man bediente sich dabei mehrerer digitaler sowie analoger hausinterner Datenbanken, welche dem IMG zur Verfügung stehen. Dabei wurde neben der Annotation der Phänotypen der Patient\*innen auch eine Informationszusammenführung und -komprimierung vorgenommen. Inhalte, die bisher auf viele Datenbanken aufgeteilt oder nur in analogen Archiv-Akten zu finden waren, wurden gesichtet, bewertet und einheitlich zusammengeführt. Neben dem Phänotyp lag das Augenmerk hierbei auf Verwandtschaftsbeziehungen für die korrekte Zuordnung der Trios sowie auf Informationen aus bereits erfolgten diagnostischen Untersuchungen. Die Phänotypen wurden mittels der Terminologie der Human Phenotype Ontology (HPO) erfasst (Köhler et al., 2017).

In 60 Fällen wiesen lediglich die jeweiligen IPs eine Extremitätenfehlbildung auf. In fünf weiteren zeigte auch eines der beiden Elternteile eine vergleichbare Extremitätenfehlbildung. In einem Fall präsentierten sich Eltern und Großeltern der IP ohne Extremitätenfehlbildung, jedoch hatte ein Großonkel mütterlicherseits eine vergleichbare Malformation. Dieser wurde in die Analyse eingeschlossen und zusätzlich sequenziert. Darüber hinaus lagen in vier Fällen lediglich geno- und phänotypische Informationen der IPs vor.

## 2.3 Whole-Genome-Sequencing

Die Genome der Kohorte wurden von der Firma MacroGen inc. (Süd-Korea) sequenziert. Hierbei bediente man sich der HiSeq X-Ten Plattform, welche mit einem PCR-free Protokoll lief. DNA-Vorbereitung, Sequenzierung und Auswertung der Sequenzierungsdaten erfolgte anhand der Standardprotokolle von MacroGen (coverage 30X und read length 150 base pairs).

### 2.3.1 Datenverarbeitung in SODAR und VarFish

Für die Analyse der Daten und das Auffinden von Varianten wurden FASTA Dateien an die Core-Unit Bio-Informatik des Berlin Institute of Health weitergereicht. Diese wurden weiter prozessiert und gespeichert, wobei man sich des System for Omics Data Analysis and Retrieval (SODAR; Nieminen et al., 2020) bediente. Um SNVs aufzufinden, wurde der GATK HC Algorithmus verwendet. Diesem ist es möglich SNVs sowie InDels (kleine

Inversionen und Deletionen) durch die lokale Neuankordnungen von Haplotypen zu identifizieren (<https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller>).

Strukturvarianten wurden mit Hilfe von Delly2 (Rausch et al., 2012), PopDel (Niehus et al., 2021) und ERDS (Zhu et al., 2012) bzw. SV2 (Antaki et al., 2018) erfasst.

Nach der Verarbeitung wurden die Daten durch die VarFish Software (Holtgrewe et al., 2020) prozessiert und annotiert.

## 2.4 Analyse der Single-Nucleotide-Variants

Während der Analyse der SNVs wurden drei Gruppen unterschieden. 60 der 69 Fälle wurden hierbei als nicht betroffenes Trio eingeordnet. Dies bedeutet, dass der\*die IP den Phänotyp einer Extremitätenfehlbildung aufweist, beide Elternteile allerdings nicht betroffen sind. Im Gegensatz hierzu finden sich in der Kohorte fünf IPs, bei denen zusätzlich ein Elternteil eine Extremitätenfehlbildung aufweist, im Nachfolgenden als betroffenes Trio bezeichnet. Neben diesen beiden Gruppen existieren vier weitere IPs, bei denen keine Informationen über die Eltern vorliegen, welche im Nachfolgenden als Singleton bezeichnet werden.

### 2.4.1 Trio und Singleton Filtration in VarFish

Für die Durchführung der Filtervorgänge wurde die VarFish Plattform (Version v0.17.2; Holtgrewe et al., 2020) verwendet. Gefiltert wurden dabei nicht-synonyme exonische sowie splice-site Varianten, wobei alle die voreingestellten Qualitätskriterien bezüglich read depth, allelic balance und read quality erfüllen mussten. Weiterhin wurden die Varianten abhängig vom zu untersuchenden Erbgang nach Art und Frequenz gefiltert. Hierbei wurden für die unterschiedlichen Index-Gruppen unterschiedliche Filter verwendet. Die Ergebnisse des Filters wurden in Tabellenform dokumentiert und die Variant Calling Files (VCF) für weitere Analysen aus VarFish exportiert.

Die nicht betroffenen Trios wurden auf einen compound heterozygoten, homozygoten, *de novo* und bei männlichem IP X-chromosomal rezessiven Erbgang untersucht.

Für betroffene Trios wurden zwei verschiedene potenzielle Erbgänge analysiert. Einerseits wurden Varianten gefiltert, welche heterozygot vorlagen. Hierbei würde die Fehlbildung autosomal dominant vererbt werden. Andererseits wurden auch rezessive Erbgänge betrachtet, wobei die Variante hier homo- oder hemizygot vorliegen würde. Alle betroffenen Trios wurden zuerst nach diesen Standarderbgängen betrachtet.

Im nächsten Schritt erfolgte die Einstellung des Filters auf mögliche gonosomale Erbgänge. Bei den Index-Patientinnen mit kranken Vätern wurden Varianten gefiltert, welche vom Vater hemizygot und von der Tochter heterozygot getragen wurden, was auf einen X-chromosomal dominanten Erbgang schließen lässt. Männliche IP mit kranken Vätern bedurften keiner weiteren Betrachtung, da hier X-gonosomale Geschehen ausgeschlossen werden können. Y-chromosomale Varianten wurden aufgrund der Struktur von VarFish bereits vom Filtervorgang für heterozygote Varianten abgedeckt.

Ein Sonderfall, bei dem die Mutter des männlichen IP betroffen ist, wurde separat analysiert. Hierbei wurden zusätzlich zu den oben beschriebenen Filtern Varianten betrachtet, die heterozygot bei der Mutter und hemizygot bei dem Index-Patienten vorliegen könnten.

Für Index-Patient\*innen, bei denen keine Informationen über die Elternteile vorlagen, wurde ein für Genomanalysen üblicher Filtervorgang verwendet. Man betrachtet hierbei zuerst alle heterozygot vorliegenden und danach alle homozygot vorliegenden SNVs. Durch diese Filtervorgänge wurde auch die Möglichkeit einer compound heterozygoten Vererbung abgedeckt.

Bei der Analyse der IPs wurden die Filter für die maximalen Frequenzen in Vergleichsdatenbanken entsprechend der supplementary table 1 (Elsner et al., 2021) gesetzt. Alle weiteren Filtereinstellungen wurden standardmäßig belassen.

#### 2.4.2 Analyse durch MutationDistiller und Exomiser

Bei MutationDistiller handelt es sich um eine webbasierte Software, welche die Vorhersage der Pathogenität einer Variante mit dem für ein betroffenes Gen bekannten Phänotyp und den Symptomen des\*der Patient\*in verknüpft. Dabei wird eine Liste potenziell krankheitsauslösender Varianten erstellt, welche in absteigender Reihenfolge der Wahrscheinlichkeit geordnet ist, dass das betroffene Gen den Phänotyp des\*der Patient\*in auslösen könnte (Hombach et al., 2019). Hierfür wurden die VCFs, welche durch die oben

beschriebenen Filter in VarFish erstellt wurden, mit Hilfe des MutationDistillers analysiert ([mutationdistiller.org](http://mutationdistiller.org)) und mit entsprechenden Informationen zum jeweiligen Phänotyp des\*der Patient\*in versehen. Die ersten zehn ausgegebenen Varianten wurden als potenzielle Kandidaten gewertet und im Weiteren näher betrachtet. Varianten, bei deren zufälligem Fund laut der Liste des American College of Medical Genetics (ACMG) SF v2.0 (Kalia et al., 2017) Patient\*innen bei entsprechendem Einverständnis informiert werden sollten, konnten durch diesen Vorgang zusätzlich identifiziert werden.

Der Exomiser ähnelt dem MutationDistiller in einigen Punkten. So handelt es sich auch hier um eine webbasierte Software, welche Varianten nach Pathogenität und der Wahrscheinlichkeit, dass das betroffene Gen den Phänotyp des\*der Patient\*in auslösen könnte, ordnet. Der entscheidende Unterschied liegt jedoch darin, dass die Analyse nicht nur anhand von menschlichen Datensätzen, sondern auch speziesübergreifend durchgeführt wird. So können krankheitsauslösende Mutationen in Mäusen mit entsprechenden Varianten bei Menschen über den Phänotyp verknüpft werden (P. N. Robinson et al., 2014). Für die Analyse wurden die im vorherigen Schritt erstellten VCFs hochgeladen und - nach der Ergänzung von Phänotyp und zu erwartendem Erbgang - mit Hilfe des Exomisers analysiert. Die Software erstellte daraufhin eine Liste, wobei erneut die ersten zehn vorgeschlagenen Varianten als potenzielle Kandidaten gewertet wurden.

#### 2.4.3 Einzel-Analyse der Index-Patient\*innen

Neben den bisher beschriebenen Analysen, wurden für jede\*n Patient\*in, unabhängig von der vorliegenden Trio-Information, VCFs erstellt, welche sowohl alle seltenen heterozygoten als auch alle seltenen homozygoten SNVs enthielten. Als seltene heterozygote Varianten wurden dabei solche gewertet, die nicht häufiger als zehnmals heterozygot oder homozygot in Vergleichspopulationen auftraten. Weiterhin wurden solche ausgeschlossen, die mehr als 20-mal heterozygot oder zehnmals homozygot in der hauseigenen Datenbank auftraten. Für die Erstellung der homozygoten VCFs wurden lediglich Varianten ausgeschlossen, die mehr als zehnmals homozygot in Vergleichspopulationen vorkamen. Die so gefilterten Varianten wurden mit Hilfe eines Python-Skripts analysiert. Dabei wurden trunkierende oder potenzielle Loss of Function (LoF) Mutationen (CADD >20; Rentzsch et al., 2019) hervorgehoben, die bei mindestens zwei Patient\*innen im gleichen Gen

auftraten. Das entsprechende Gen musste weiterhin einen pLI-Score  $>0,9$  (Lek et al., 2016) aufweisen. Die so gefilterten Varianten wurden der weiteren Analyse zugeführt.

#### 2.4.4 Manuelle Auswertung der Kandidaten-Varianten

Nachdem durch die Webtools eine grobe Auswahl, basierend auf potenzieller Pathogenität der Mutationen, Verknüpfung mit dem Phänotyp des\*der Patient\*in sowie bekannter krankheitsauslösender Mutationen im Mausmodell bzw. deren Interaktion erfolgt war, begann die manuelle Auswertung der Kandidaten-Varianten. Dabei wurde auf diverse Informationsquellen zugegriffen (vgl. supplementary fig. 2 (Elsner et al., 2021)).

Als erstes wurden die Daten des WGS im Integrative Genomics Viewer (J. T. Robinson et al., 2017) betrachtet. Hier konnte festgestellt werden, ob es sich um eine echte Variante handelte und nicht um ein Artefakt der Software, welche zur Datenverarbeitung vor der Analyse genutzt wurde. Auch Varianten, welche in rapid-Repeat Sequenzen lagen, konnten so genauer betrachtet werden. Nur im Fall einer echten Variante, erfolgte die weitere Analyse.

Dafür wurde zuerst die Pathogenität einer jeden Variante mit Hilfe des MutationTasters überprüft (Schwarz et al., 2010). Dabei konnten außerdem weitere Informationen über die Variante und den Locus gewonnen werden. Ein Beispiel hierfür ist die Konservierung. Eine hohe Konservierung tritt in funktionellen Bereichen auf, die vermutlich einen evolutionären Vorteil darstellen und deshalb speziesübergreifend ähnlich strukturiert, also erhalten sind (Cooper & Brown, 2008). Befindet sich eine Variante in einem solchen konservierten Bereich, kann von einer höheren Pathogenitätswahrscheinlichkeit ausgegangen werden. Bei Varianten, welche von MutationTaster als Polymorphismen klassifiziert wurden, erfolgte keine weitere Betrachtung. Da im MutationTaster nur exonische bzw. splice-site Varianten betrachtet werden können, erfolgte die Analyse von intronischen Varianten mit Hilfe des RegulationSpotters (Schwarz et al., 2019). Splice-site Varianten wurden zusätzlich mit webbasierten Anwendungen wie HumanSplicingFinger (Desmet et al., 2009) und VarSeak (<https://varseak.bio/>) betrachtet.

Für missense Varianten wurde außerdem PolyPhen2 (Adzhubei et al., 2010) zur weiteren Überprüfung der Pathogenität genutzt.

Varianten, welche eine hohe Konservierung aufwiesen und/oder als wahrscheinlich krankheitsauslösend klassifiziert wurden, erhielten eine genauere Analyse.

So erfolgte unter anderem eine Betrachtung der Variante im University of California, Santa Cruz-Browser (UCSC-Browser; Kent, 2002), welcher durch einen benutzerdefinierten Track erweitert war (<http://genome-euro.ucsc.edu/s/mensahm/limb%203%20color%20score>). Hier konnte ein Überblick über die Lage der Variante im Gen gewonnen werden. Weiterhin erhielt man einen Einblick in die Häufigkeit der Varianten in gesunden und erkrankten Vergleichspopulationen. Dabei wurde mit hauseigenen Datenbanken und der Database of Genomic Variants (MacDonald et al., 2014) gearbeitet, welche Bestandteile eines IMG-eigenen, auf die Analyse von Fälle mit Extremitätenfehlbildungen ausgelegten, UCSC-Tracks sind.

Parallel dazu erfolgte eine Analyse des betroffenen Gens. Diese bestand zum einen aus einer Betrachtung des Gens im Mausmodell. Dabei wurden die Datenbanken des Mouse Genome Informatics (MGI; Bult et al., 2019), des International Mouse Phenotyping Consortiums (Dickinson et al., 2017) und der Mouse Organogenesis Cell Atlas (Cao et al., 2019) genutzt.

Zum Anderen wurden die Online Mendelian Inheritance in Man (OMIM; Hamosh et al., 2005), PubMed (<https://pubmed.ncbi.nlm.nih.gov>) und Genecards (Safran et al., 2010) sowie die Human Gene Mutation Database (Stenson et al., 2003) genutzt um bereits beschriebene Funktionen oder krankheitsauslösende Varianten des Gens aufzufinden. Varianten, welche im Verdacht standen, die Struktur des kodierten Proteins zu verändern, wurden durch das Simple Modular Architecture Research Tool (Letunic et al., 2021) genauer betrachtet.

Schließlich wurden die Varianten in einer wöchentlichen Konferenz evaluiert.

## **2.5 Analyse der Strukturvarianten**

### **2.5.1 Heterozygote Varianten**

Bei nicht betroffenen Trios konnten lediglich *de novo* Varianten betrachtet werden, betroffene Trios wurden unter der Annahme eines dominanten Erbgangs analysiert.

Alle Varianten-Calls mussten mindestens von einem split-read und einem paired-read oder von zwei split-reads beziehungsweise von zwei paired-reads abgedeckt werden. Weiterhin wurden Varianten ausgeschlossen, welche kleiner als 1.500 Basen waren oder mehr als zehnmal in gesunden Vergleichskohorten auftraten. Dabei zählt eine Variante als ähnlich oder gleich einer Variante in der Vergleichskohorte, wenn sie die entsprechende Vergleichsvariante zu 75% überlappt oder zu 75% von ihr überlappt wird. Des Weiteren wurden nur Varianten betrachtet, die nicht mehr als zwei Mal in der Kohorte vorkamen, um Artefakte auszuschließen. Auch hier wurde eine 75%ige Überlappung in beide Richtungen verwendet.

Sollte das Filter initial mehr als 30 Varianten ausgeben, wurden modifizierte Filtereinstellungen verwendet. Um das Filter zu passieren, musste eine Variante nun von mindestens fünf split- und/oder paired-reads abgedeckt werden, anstatt von zwei. Dieses Procedere wurde in neun Fällen angewandt, um eine detaillierte manuelle Auswertung der Daten zu ermöglichen.

Beim Filtern der Singletons wurden strengere Frequenz- und Qualitätseinstellungen verwendet. Varianten durften nur noch fünf Mal in Vergleichskohorten vorkommen und mussten von mindestens fünf split- und/oder paired-reads abgedeckt sein. Ansonsten wurden die gleichen Einstellungen wie bei nicht betroffenen Trios verwendet.

Nachdem die Varianten nach den entsprechenden Kriterien (vgl. supplementary tab. 2 (Elsner et al., 2021)) gefiltert wurden, durchliefen sie, ähnlich wie die SNVs, einen Analyseprozess. Zuerst wurden sie anhand der WGS Rohdaten im Integrative Genomics Viewer auf Echtheit überprüft. Danach erfolgte eine detaillierte Analyse in einer IMG-eigenen Session des UCSC-Browser, der durch Betrachtung der Variante bzw. des betroffenen Gens in MGI, OMIM, PubMed und Genecards ergänzt wurde (vgl. supplementary fig. 3 (Elsner et al., 2021)).

Anschließend wurden alle Varianten in wöchentlichen Meetings evaluiert.

### 2.5.2 Homozygote Varianten

Die Analyse von homozygoten Varianten erfolgte anhand der VCF Daten, welche durch VarFish zur Verfügung standen. Varianten mussten mindestens von drei paired-reads oder drei split-reads abgedeckt sein. Weiterhin durften sie nicht kleiner als 1.500 Basen sein und höchstens zweimal in der Kohorte vorkommen. Hier wurde die bereits beschriebene 75%ige Überlappung verwendet. Trio und Singleton Patient\*innen wurden gleichermaßen analysiert.

Die so erhaltenen Varianten wurden gleich den heterozygoten SVs analysiert und in wöchentlichen Meetings evaluiert.

## 2.6 Analyse der nicht-kodierenden Varianten

Für die Analyse der nicht-kodierenden Varianten wurden spezifische regulatorische Bereiche des Genoms definiert. Hierfür bediente man sich mehrerer selbstentwickelter und externer Datenbanken. Als Resultat entstand eine Liste aus 1.719 Genen, welche nachgewiesenermaßen in der embryonalen Extremitätenentwicklung eine Rolle spielen. Diese Informationen resultierten aus Datensätzen des MGI und OMIM.

Die regulatorischen Bereiche wurden wie folgt definiert: 1. alle konservierten (phyloP>1,3; Pollard et al., 2010) Varianten, die 2. in der gleichen topologisch assoziierten Domäne (TAD; Dixon et al., 2012) wie eines der vorher definierten Gene anzutreffen sind und 3. einen H3K27 Acetylierungs-Peak in menschlichen Extremitätenknospen besitzen (Cotney et al., 2013). Weiterhin wurden validierte Enhancer Elemente des VISTA Enhancer Browsers eingeschlossen. Dies erfolgte unabhängig davon, ob sie spezifische Expressionsmuster im Mausmodell gezeigt hatten (Visel et al., 2007). Regionen, welche häufige single nucleotide polymorphisms beinhalten, wurden entfernt.

Im Anschluss wurden sowohl kodierende als auch nicht-kodierende SNVs, die mit einer Frequenz von <0,1% in der Kohorte sowie in vorher beschriebenen Vergleichskohorten auftraten, gefiltert. Diese Varianten wurden als VCF aus SODAR exportiert und der weiteren Analyse via Python zugeführt. Sollten zwei oder mehr Varianten einen Abstand von 300 Basenpaaren (bp) oder weniger voneinander haben und in dem vorher definierten regulatorischen Bereich liegen, wurden sie vom Skript ausgegeben. Unter der Annahme,

dass sie die gleiche regulatorische Domäne beeinflussen könnten, folgte eine semi-automatische Evaluation von Kandidatenvarianten.

Zusätzlich wurden alle seltenen, potenziell pathogenen und heterozygot kodierenden Varianten, welche Gene betrafen, die im Zusammenhang mit rezessiven Extremitätenfehlbildungen stehen, genauer betrachtet. Bei einer partiellen Übereinstimmung des Patienten-Phänotyps mit der durch eine Variante des entsprechenden Gens ausgelösten Fehlbildung wurde nach konservierten, nicht-kodierenden, Varianten *in trans*, welche dieselbe TAD beeinträchtigten, gesucht. Die minor allele frequency dieser Varianten durfte dabei 3% nicht überschreiten.

## 2.7 Genotyp-Phänotyp-Korrelationsanalyse von *GLI3*

Parallel zu den bisher beschriebenen Prozessen, erfolgte eine genauere Analyse des *GLI3* Locus. Bei 94 Patient\*innen mit syndromaler und nicht-syndromaler Polysyndaktylie, deren Phänotyp auf eine *GLI3*-assoziierte Polysyndaktylie schließen ließ, erfolgte eine Sanger-Sequenzierung anhand von genomischer DNA aus Vollblut. Zusätzlich wurde, falls möglich, eine Segregationsanalyse durch Sequenzierung der Elternteile durchgeführt. Die Sequenzdaten wurden mit SeqPilot (<https://www.jsi-medisys.de/products/sequence-pilot/>) evaluiert und auffällige Varianten anschließend zur klinischen Interpretation mit den Datenbanken ClinVar (Landrum et al, 2018), HGMD und gnomAD abgeglichen und durch bioinformatische Software, wie MutationTaster, Polyphen2 und SIFT (Ng & Henikoff, 2001) hinsichtlich ihrer Pathogenität bewertet.

Im Anschluss erfolgte eine retrospektive Analyse bereits publizierter Fälle, in denen eine Variante in *GLI3* als krankheitsauslösend klassifiziert worden war. Hierbei erfolgte neben der lokalen und funktionellen Einordnung der Variante auch eine Bewertung des Phänotyps sowie die Erhebung weiterführender Patienteninformation, welche für die spätere Auswertung von Nutzen waren.

Im weiteren Verlauf stieß man auf ein Individuum (I), welches sich mit dem Phänotyp *GLI3*-assoziierte Polydaktylie präsentierte. Dieses war dem IMG im Rahmen der molekular-genetischen Standarddiagnostik zugesandt worden. Es erfolgte ein WGS mit automa-

tisierter und manueller Auswertung analog des in dieser Arbeit beschriebenen Vorgehens. Die gefundenen Mutation wurden per Sanger-Sequenzierung überprüft, wobei man sich der Primer *GLI3*-fwd:ACAGGCATTCTCAGGAAGC, *GLI3*-rev:GCTTTGTGAATCGGAATGCTGT bediente.

## 2.8 Analyse des Genlocus um *HMGB1*

Das Auffinden einer Variante in *HMGB1* sowie das Vorhandensein eines charakteristischen Phänotyps des variantentragenden Individuums (Brachyphalangy-Polydactyly-and-Tibial-Aplasia-Syndrome; BPTAS), zog eine genauere Untersuchung des Gens nach sich. Zu Beginn dieser Analyse wurden zusätzlich vier weitere Patient\*innen eingeschlossen, die den Phänotyp BPTAS aufwiesen (vgl. extended data fig. 1; (Mensah et al., 2023)). Zusätzlich wurde ein weiterer Patient, bei dem der Verdacht auf ein 13q12.3 Deletions-Syndrom bestand, analysiert. Nach detaillierter Phänotypisierung, erfolgte eine WGS sowie eine Sanger-Sequenzierung und quantitative real-time PCR (qPCR), wobei letztere auf einem 3730 DNA-analyzer (Thermo Fisher Scientific, MA, USA) durchgeführt wurden. Die Methodik des WGS orientierte sich am bisher in dieser Arbeit beschriebenen Vorgehen.

Für die Sanger-Sequenzierung der gDNA wurden folgende Primer verwendet: Ex2-3F AAATAACCAGTACTTTGGTTTTTCATTC; EX2-3R GCTTGCTAAATTTCTCTCAGATTCTAC; Ex4F CACAGAAATTCACAGGCAGC; Ex4R CTAAGTACAATCATACATCTGGCG; Ex5F TTCTGGAGTTGCTGTGGATCT; Ex5R GGGGGTTAAATGCTTTATAGACAA. Im Falle der cDNA fanden die folgenden Primer Anwendung: fwd CACTGGGCGACTCTGTG und rev AAGGAGTGAGTTGTGTACAGG, fwd-seq GACTCTGTGCCTCGCTGA und rev-seq AACTGCGCTAGAACCAACTT. Zuletzt wurden die folgenden Primer für die qPCR verwendet: *HMGB1*-fwd GACTCTGTGCCTCGCTGA und *HMGB1*-rev TCTTATGCTCCTCCCGACAAG, *GAPDH*-fwd ATGCCTCCTGCACCACCAAC und *GAPDH*-rev GGCCATCCACAGTCTTCTGG.

Hiernach erfolgte eine aCGH, analog zu der bisher beschriebenen Methodik. Alle Prozeduren wurden mittels der Protokolle durchgeführt, die durch die herstellenden Firmen empfohlen wurden. Zuletzt wurden die Varianten am *HMGB1*-Transkript NM\_002128.7 im Referenz-Genom hg19 annotiert.

Darüber hinaus konnten in der vorhandenen Literatur acht weitere Individuen mit dem Phänotyp BPTAS identifiziert werden. Diese wurden in die weitere Analyse und Bewertung des Gens *HMGB1* eingeschlossen (vgl. supplementary tab. 1; (Mensah et al., 2023)).

## **2.9 Analyse von *HMGB1* Patient\*innen mittels Face2Gene**

Nach der Geno- und Phänotyp-Analyse wurden frontale Aufnahmen der Gesichter der Patient\*innen mittels Face2Gene (v.20.1.4, [www.face2gene.com](http://www.face2gene.com)) analysiert. Hierfür wurden sieben Bilder der bereits in der Literatur beschriebenen BPTAS Fälle sowie Aufnahmen von *HMGB1*-I1 und -I2 verwendet. Es erfolgte eine Phänotypisierung mittels Face2Gene CLINIC (F2GC) sowie das Erstellen eines durchschnittlichen fazialen Phänotyps mittels Face2Gene RESEARCH (F2GR). Um letzteres zu ermöglichen, wurden alle Bilder der neun ausgewählten IPs doppelt eingespeist. Waren mehrere Bilder des Gesichts vorhanden, so wurde stets das aktuellste verwendet.

### 3. Ergebnisse

#### 3.1 Auswertung und Filterprozess der SNV

Insgesamt konnten in den Genom-Rohdaten 333.163.643 SNVs aufgefunden werden. Davon lagen 326.142.877 SNVs im nicht-kodierenden Bereich, wovon 19.362 als seltene Varianten (gnomAD Allel-Frequenz < 0,01) eingestuft worden. Mit Hilfe einer Analyse durch Jannovar (Jäger et al., 2014) konnten 21.369 der kodierenden SNVs als mindestens moderat relevant (missense und truncating Mutationen) und 1.429 sogar als hoch relevant (nur truncating) eingeordnet werden.

Die verbleibenden 7.020.766 SNVs befanden sich im kodierenden Bereich oder flankierten diesen im Abstand von maximal zehn Basenpaaren. Nach dem Filtervorgang in VarFish fanden sich 5.761 Varianten. Dies entspricht im Schnitt 83,49 SNVs pro Patient, wobei der Median bei 22 SNVs liegt (siehe Abbildung 1).

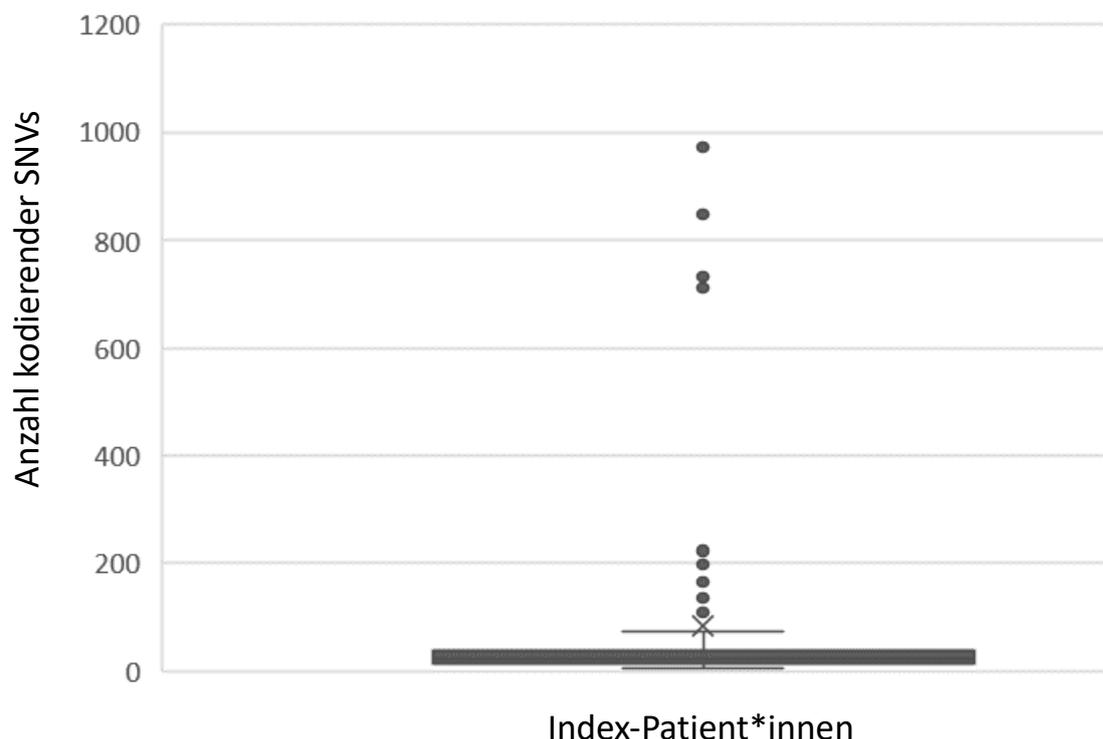


Abbildung 1: Anzahl der kodierenden SNVs pro Patient nach dem Filtern in VarFish. Der Median liegt bei 22 SNVs, die Länge der Whisker wurde bei 1,5-fachem Interquartilsabstand gesetzt. Insgesamt sind neun Ausreißer zu erkennen. *eigene Darstellung*

Von diesen 5.761 wurden 433 Varianten von MutationDistiller und Exomiser als potenziell pathogen bewertet, woraufhin eine genauere manuelle Auswertung dieser erfolgte. Diese 433 SNVs setzten sich aus mehreren Typen von Varianten zusammen. Bei nicht betroffenen-Trios konnten 55 compound heterozygote, 50 *de novo*, 73 homozygote und 53 X-linked hemizygoten Varianten identifiziert werden. Bei betroffenen Trios und Singletons wurden 162 heterozygote und 40 homozygote Varianten festgestellt. Durch die manuelle Auswertung der 433 Varianten, ergaben sich 174, welche in wöchentlichen Meetings evaluiert worden.

Parallel zu diesem Prozess lieferte die Einzelbetrachtung der Index-Patient\*innen VCFs in Hinblick auf CADD- und pLI-Score der Gene sowie die Betrachtung der Ähnlichkeiten der Patient\*innen-Phänotypen weitere 49 Varianten. Auch diese wurden in wöchentlichen Meetings beurteilt.

Insgesamt konnten hierdurch 22 Kandidaten-SNVs identifiziert werden, welche genauer differenziert wurden. Fünf Varianten befanden sich in bereits beschriebenen krankheitsauslösenden Genen, vier weitere wurden auf Grundlage von Informationen aus diversen Quellen (vgl. supplementary fig. 2 (Elsner et al., 2021)) als höchstwahrscheinlich krankheitsauslösend bewertet. Die restlichen 13 wurden als Varianten unklarer Signifikanz (VUS) beurteilt. All diese Varianten werden im Verlauf weiter ausgeführt.

### 3.2 Auswertung und Filterprozess der SVs

Strukturvarianten konnten in 68 von 69 Fällen ausgewertet werden. Insgesamt ergaben sich aus den Genom-Rohdaten 1.638.500 potenzielle SVs, deren Anzahl durch das Filtern in VarFish auf 633 heterozygote reduziert wurde. Diese setzen sich zusammen aus 222 Inversionen, 288 Deletionen, 76 Duplikationen und 47 Break-Ends. Für alle diese Varianten erfolgte eine manuelle Auswertung (vgl. supplementary fig. 3 (Elsner et al., 2021)). 417 der ursprünglich 633 Calls konnten als Artefakte identifiziert werden. Hiervon waren 100 SVs doppelt und 52 fälschlicherweise gecallt worden. 165 Calls besaßen keine ausreichende Qualität und wurden deshalb von der weiteren Analyse ausgeschlossen. 186 der initial 633 SVs lagen in nicht-kodierenden Regionen oder beeinflussten Gene, die in keiner der Quellen, die uns zur Verfügung standen, einen Zusammenhang mit der Entwicklung von Extremitäten zeigten.

Die 30 verbleibenden SVs zeigten sich als potenziell echte Varianten, die Gene, welche im Zusammenhang mit der Entwicklung von Extremitäten stehen, direkt oder indirekt beeinflussen. Es folgten Segregationsanalysen via quantitativer PCR sowie eine Recherche bezüglich der Gen-Phänotyp Assoziation und eine Analyse des entsprechenden Locus, wobei man sich diverser Quellen bediente. Hierzu zählte die HPO, speziesübergreifende Phänotypdatenbanken, Maus-Modelle, Gen-Expressions Daten (Cao et al., 2019), limb Enhancer Elemente (Visel et al., 2009) sowie die lokale TAD-Architektur (Cao et al., 2019; Dixon et al., 2012). Alle 30 Varianten zeigten sich in dieser Analyse als vielversprechend.

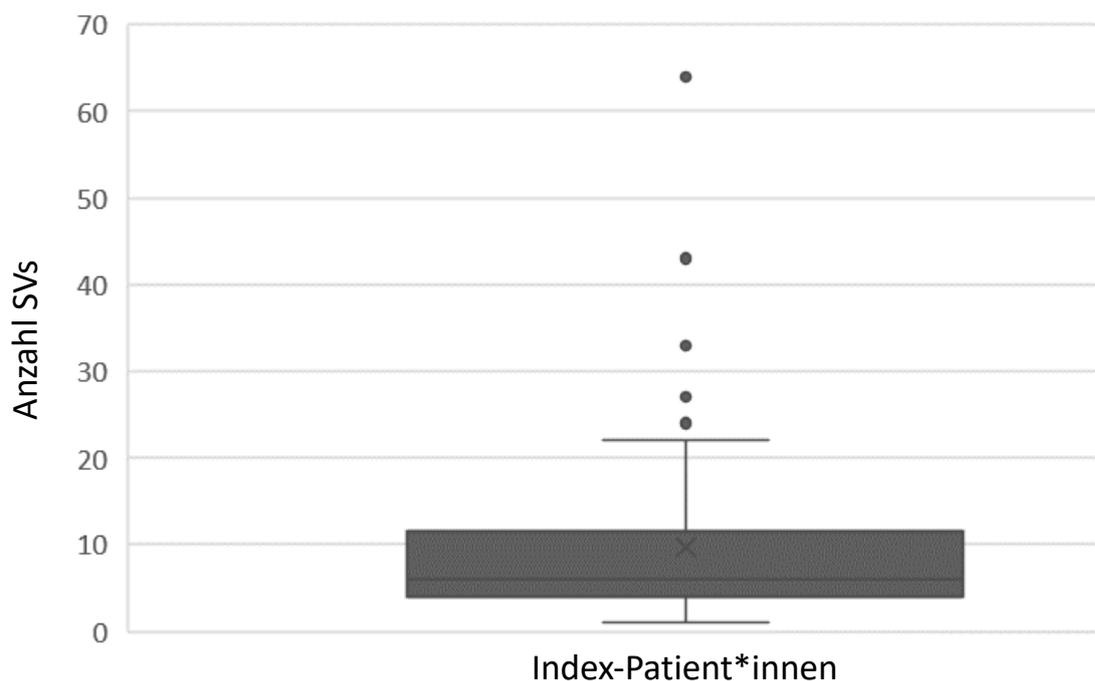


Abbildung 2: Identifizierte SVs von 68 Patient\*innen nach dem Filtern in VarFish. Der Median liegt bei 6, die Länge der Whisker wurde bei 1,5-fachem Interquartilsabstand gesetzt. Insgesamt sind fünf Ausreißer zu erkennen. *eigene Darstellung*

Die Fälle Individuum 1, I2, I3 lieferten keine Ergebnisse nach dem Filtern in VarFish. Weiterhin musste I4 von der Analyse der SVs ausgeschlossen werden. VarFish lieferte im Falle dieses Patienten selbst unter strengsten Filterkriterien 6.788 Strukturvarianten, was auf eine schlechte Datenqualität des Falles, zum Beispiel aufgrund eines Fehlers des Varianten-Callings, schließen ließ. Durch diese hohe Anzahl an SVs war die manuelle Auswertung nicht möglich.

Darüber hinaus konnten 22 homozygote SVs durch die Analyse der VCF-Daten identifiziert werden. Die weitere Analyse ließ allerdings in keinem Fall eine wahrscheinliche Pathogenität vermuten.

### 3.3 Varianten in bekannten Krankheits-Lokalisationen

#### 3.3.1 SNVs in bekannten Lokalisationen

In vier Fällen konnten pathologische SNVs in bereits bekannten krankheitsassoziierten Genen identifiziert werden. Eine Variante im Gen *FGFR1* befand sich an einer Position, welche bereits als krankheitsauslösend beschrieben worden war. Die anderen drei Varianten in den Genen *FGFR2*, *GLI3* und *BHLHA9* befanden sich an Positionen, an denen noch keine pathogenen Varianten beschrieben worden sind. Alle Varianten wurden via Sanger Sequenzierung validiert.

In der Familie I6 konnte eine heterozygote missense Variante in *FGFR1* (NM\_001174063: c.755C>G; P252R) identifiziert werden. Diese ist sowohl im Index-Patienten (I6) als auch dessen ebenfalls betroffenen Vater (Male6; M6) zu finden. I6 weist den Phänotyp Distal symphalangism (hands); Finger syndactyly; 2-3 toe syndactyly und Deviation of the hallux auf. Dieser stimmt größtenteils mit dem des Vaters überein. Die Variante wurde bereits in der Literatur im Zusammenhang mit dem Pfeiffer-Syndrom (MIM 101600) als pathogen beschrieben (Chokdeemboon et al., 2013), welches sich mit dem Phänotyp der Betroffenen deckt. Wir bewerteten diese Variante als krankheitsauslösend.

In einer anderen Familie (I7) konnte eine heterozygote nonsense Variante in *GLI3* (NM\_000168: c.650C>G; S217\*) identifiziert werden. Hier führt eine Punktmutation dazu, dass die transkribierte mRNA ein vorzeitiges Stop-Codon innerhalb des fünften Exons aufweist. Dadurch würde das Ribosom die Translation vorzeitig beenden und es könnte ein verkürztes Protein entstehen. Durch die Lage der Variante in einem der vorderen Exone (5/15) von *GLI3*, scheint allerdings die Aktivierung des nonsense mediated decays deutlich wahrscheinlicher, sodass es nicht zur Translation eines Proteins kommt (Karousis & Mühlemann, 2019).

Im vorliegenden Fall tragen sowohl die Index-Patientin (I7), als auch der betroffene Vater (M7) die besagte Variante in *GLI3*. I7 weist eine Kraniosynostose sowie Polydaktylie und

Syndaktylie der Hände und Füße auf. Beim Vater sind ebenfalls Oligosyndaktylien bekannt.

Nonsense Varianten in *GLI3* sind bekanntermaßen krankheitsauslösend für verschiedene Extremitätenfehlbildungen. Hierzu zählen unter anderem das Pallister-Hall-Syndrom (MIM 146510) oder das Greig-Zephalopolysyndaktylie-Syndrom (MIM 175700). Weiterhin weisen betroffene Patienten postaxiale sowie präaxiale Polydaktylien auf (Sczakiel et al., 2021). Auf Grundlage dieser Erkenntnisse bewerteten wir die Variante als pathogen.

Die Patientin I5 trägt eine homozygote missense Variante in *BHLHA9* (NM\_001164405:c.307A>G; K103E). Diese war von den beiden Eltern, welche heterozygote Träger der Variante sind, vererbt worden. I5's Phänotyp ist 2-3 toe cutaneous syndactyly, 3-4 finger syndactyly, Brachydactyly syndrome und Broad fingertip.

Die Variante ist weder in gnomAD noch in den Daten des Exome Aggregation Consortium (ExAC) aufgeführt. Des Weiteren sind homozygote Varianten in *BHLHA9* dafür bekannt, komplexe Kamptopolysyndaktylien (Phadke et al., 2016) sowie Syndaktylien mit mesoaxialer Synostose und phalangealen Reduktionen auszulösen. Dies lässt auf eine essenzielle Rolle von *BHLHA9* im Entwicklungsprozess von Extremitäten schließen (Malik et al., 2014). Aufgrund dieser Tatsachen bewerteten wir die Variante als pathogen.

Die letzte Variante, welche in einem bereits beschriebenen krankheitsauslösenden Gen identifiziert werden konnte, befand sich in *FGFR2*. Die Patientin I8 weist hier eine heterozygote missense Variante (NM\_001144919:c.451A>G; I151V) auf, welche auch vom Vater getragen wird. Allerdings zeigt in diesem Fall nur die Index-Patientin den pathologischen Phänotyp Brachydactyly syndrome, Aplasia/Hypoplasia of the distal phalanges of the toes, Aplasia/Hypoplasia of the distal phalanges of the hand. Der Vater ist phänotypisch unauffällig.

Die Variante der Patientin findet man lediglich je einmal in den Vergleichskohorten von gnomAD und ExAC. Die Domäne 'Ig-like C2-type 2' des transkribierten Proteins *FGFR2*, welche auch durch die Variante der Patientin beeinflusst wird, weist drei non-VUS, non-synonymous Varianten auf, welche alle als pathogen gewertet wurden. Insgesamt sind 184 missense non-VUS des Gens *FGFR2* bekannt, von denen 95,7% als pathogen gewertet wurden. Mutationen von *FGFR2* sind ursächlich für diverse Fehlbildungen, welche alle einen autosomal dominanten Erbgang besitzen. Hierbei weist vor allem das Saethre-

Chotzen-Syndrom (MIM 101400) phänotypische Überlappungen mit der Patientin auf. Dieses Syndrom wurde in der Literatur bereits mit weniger ausgeprägten und/oder variablen Phänotypen assoziiert (Chun et al., 2002). Auf Grundlage dieser Erkenntnisse wurde die Variante als pathogen bewertet.

### 3.3.2 Repeat-Expansionen in *HOXD13*

Bei Patient I9 konnten zwei *HOXD13* Repeat-Expansionen identifiziert werden. Der Patient, welcher eine Brachy-, Poly- und Syndaktylie hat, erbte eine dieser Expansionen von seiner Mutter, welche den gleichen Phänotyp wie der Patient aufweist. Es handelt sich hierbei um eine Erweiterung um acht Alanin-Repeats, welche bereits in der Fachliteratur als pathogen beschrieben wurde (Brison et al., 2014). Weiterhin trägt der Patient eine benigne Erweiterung um ein Alanin-Repeat auf dem paternalen Allel. Diese Ergebnisse wurden mit Hilfe von herkömmlicher *HOXD13* Mikrosatelliten-Analyse validiert und als pathogen eingeschätzt.

### 3.3.3 SVs in bekannten Lokalisationen

In einem Fall konnte eine Strukturvariante in einem bereits als pathogen beschriebenen Bereich identifiziert werden. Bei Patientin I10, welche den Phänotyp Ektrodaktylie bzw. SHFM aufweist, konnte eine circa 100kb große Inversion im SHFM3 Locus 10q24 nachgewiesen werden. Diese wurde von der phänotypisch unauffälligen Mutter vererbt. Weiterhin konnte dieselbe Variante im Genom des Großonkels mütterlicherseits, welcher ebenfalls eine Ektrodaktylie aufweist, aufgefunden werden. Hierbei könnte es sich um einen Fall von unvollständiger Penetranz handeln.

Der Locus, welcher in I10 als pathogen identifiziert wurde, liegt in anderen SHFM3 Patienten dupliziert vor (de Mollerat et al., 2003). Interessanterweise zeigte I10 keine Duplikation des Locus, sondern eine Inversion, welche von einer Deletion flankiert wird. Punktmutationen im SHFM3-Locus lösen nach aktuellen Erkenntnissen keine Ektrodaktylie aus (Holder-Espinasse et al., 2019; de Mollerat et al., 2003).

Aus den SHFM3-Locus betreffenden Publikationen ergibt sich eine minimale kritische Region, welche die Gene *BTRC*, *POLL* und *DPCD* teilweise oder komplett überlappt. Daraus lässt sich schließen, dass eines dieser Gene oder deren Zusammenspiel, durch Veränderungen der Kopienanzahl so beeinflusst wird, dass sich eine SHFM im Patienten

etabliert. Die von uns identifizierte Inversion hat jedoch keinen Einfluss auf die Zahl der Kopien. Allerdings beeinflusst sie eine TAD, welche von Holder-Espinasse *et al.* im SHFM-Locus identifiziert wurde (Holder-Espinasse *et al.*, 2019). Diese Ergebnisse lassen vermuten, dass nicht nur Änderungen der Kopienanzahl, sondern auch Positionsveränderungen bestimmter Genomabschnitte die Expression und Regulation von *BTRC*, *POLL* und *DPCD* so verändern, dass der Phänotyp Ektrodaktylie entsteht.

Weiterhin konnte in I11 eine SV auf chr7q identifiziert werden (siehe Abbildung 3). Hier war bereits im aCGH eine circa 300kb umfassende Duplikation auf chr7q36.1 aufgefallen. Zum Zeitpunkt der Analyse wurde diese SV jedoch als VUS eingeordnet, da der Phänotyp der Patientin (bilaterale mirror-image Polydaktylien der Hände und Füße) nicht mit einem bereits beschriebenen Individuum, welches eine ähnliche lokalisierte Duplikation der gleichen Größenordnung besaß, übereinstimmte (Kroeldrup *et al.*, 2012). Aufgrund dieser Tatsache wurde die Patientin als nicht diagnostiziert betrachtet und in die vorliegende Arbeit eingeschlossen.

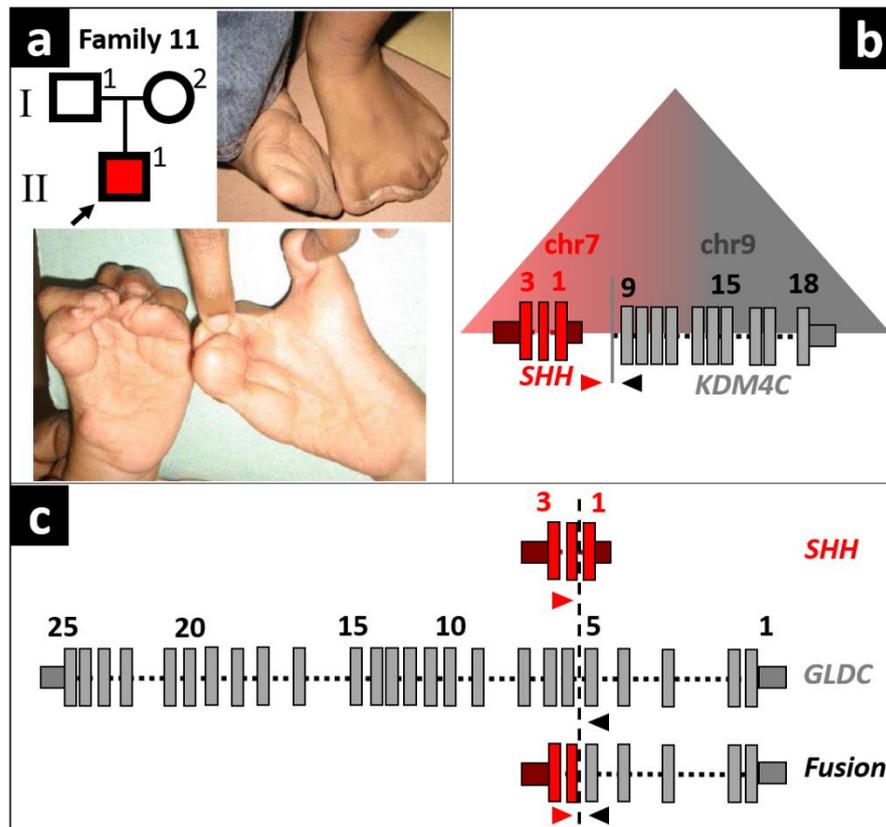


Abbildung 3: Visualisierung von Strukturvariante und Phänotyp von I11. a: Stammbaum und Phänotyp von I11. b: Potenzielle neo-TAD c: Bruchpunkte und Verbindungspunkte zwischen den Loci auf Chromosom 7 und 9. Diese Abbildung wurde bereits in dem publizierten Artikel Elsner et al., 2021 verwendet.

Bei der Analyse mittels WGS konnte die oben beschriebene Duplikation erneut identifiziert werden. Darüber hinaus konnte man erkennen, dass sie in manchen reads von sogenannten Break-Ends (chimeric read-pairs und split-reads) flankiert wurde. Dies sind Hinweise auf Translokationen, weshalb der entsprechende Bereich Sanger-sequenziert wurde. Hierbei zeigte sich eine komplexe SV, welche sich aus zwei überlappenden Duplikationen zusammensetzt (Dup1 und Dup2, siehe Abbildung 4). Der zentrale Bruchpunkt wird dabei von beiden Duplikationen geteilt, wobei der distale Bruchpunkt von Dup2 in Dup1 und damit innerhalb des ersten Introns von *SHH* (chr7:155,603,964) liegt. Durch die Analyse der WGS-Daten konnte weiterhin gezeigt werden, dass die Duplikationen nicht *in tandem* vorliegen, sondern beide an Sequenzen geknüpft sind, die zu einem entsprechenden Locus auf Chromosom 9 gehören (chr9p24.1). Der distale Bruchpunkt von Dup2 befand sich also im fünften Intron von *GLDC*, während der Bruchpunkt von Dup1 im achten Intron von *KDM4C* lokalisiert werden konnte (siehe Abbildung 4 und Abbildung

3c). Durch die Verfügbarkeit der Trio-WGS-Daten konnte die Variante als *de novo* identifiziert werden.

In wöchentlichen Meetings, in denen jede Kandidatenvariante mit mehreren erfahrenen klinischen Genetiker\*innen diskutiert wurde, konnte eine starke Ähnlichkeit des Phänotyps von I11 mit Laurin-Sandrow-Syndrom (LSS, MIM 135750) festgestellt werden. Dieses Syndrom entsteht durch Duplikationen des *SHH* Regulators ZRS, welches im fünften Intron von *LMBR1* auf chr7p36.3 liegt und die Expression von *SHH* pathologisch verändern kann (Lohan et al., 2014). Obwohl weder Dup1 noch Dup2 in der entsprechenden Region des ZRS liegen und bisher keine Duplikationen von *SHH* als krankheitsauslösend für LSS beschrieben wurden, konnte eine Pathogenität der zuvor beschriebenen SV vermutet werden. Diese Annahme stützt sich auf die Struktur der Variante: Ein dupliziertes Fragment von *SHH*, welches in eine andere Region inseriert wird und *de novo* auftritt, macht eine ektopische Expression von SHH sehr wahrscheinlich. Es ist anzunehmen, dass durch die SV eine *SHH-KDM4C* neo-TAD entsteht, wodurch regulatorische Elemente von *SHH* (wie ZRS) als auch von *KDM4C* im Mesenchym der Extremitäten während der embryonalen Entwicklung beeinflusst werden. Weiterhin besitzt auch *KDM4C* eine verstärkte Expression in den embryonalen Extremitätenknospen, was eine Bedeutung in der Extremitätenentwicklung vermuten lässt. Diese Zusammenhänge lassen die Strukturvariante des Patienten als wahrscheinlichste Ursache für eine ektopische SHH-Expression und damit für den beobachteten Phänotyp erscheinen (siehe Abbildung 3b).

Deshalb reklassifizierten wir die SV in I11, welche vor der WGS-Analyse als VUS angenommen wurde, als pathogen.

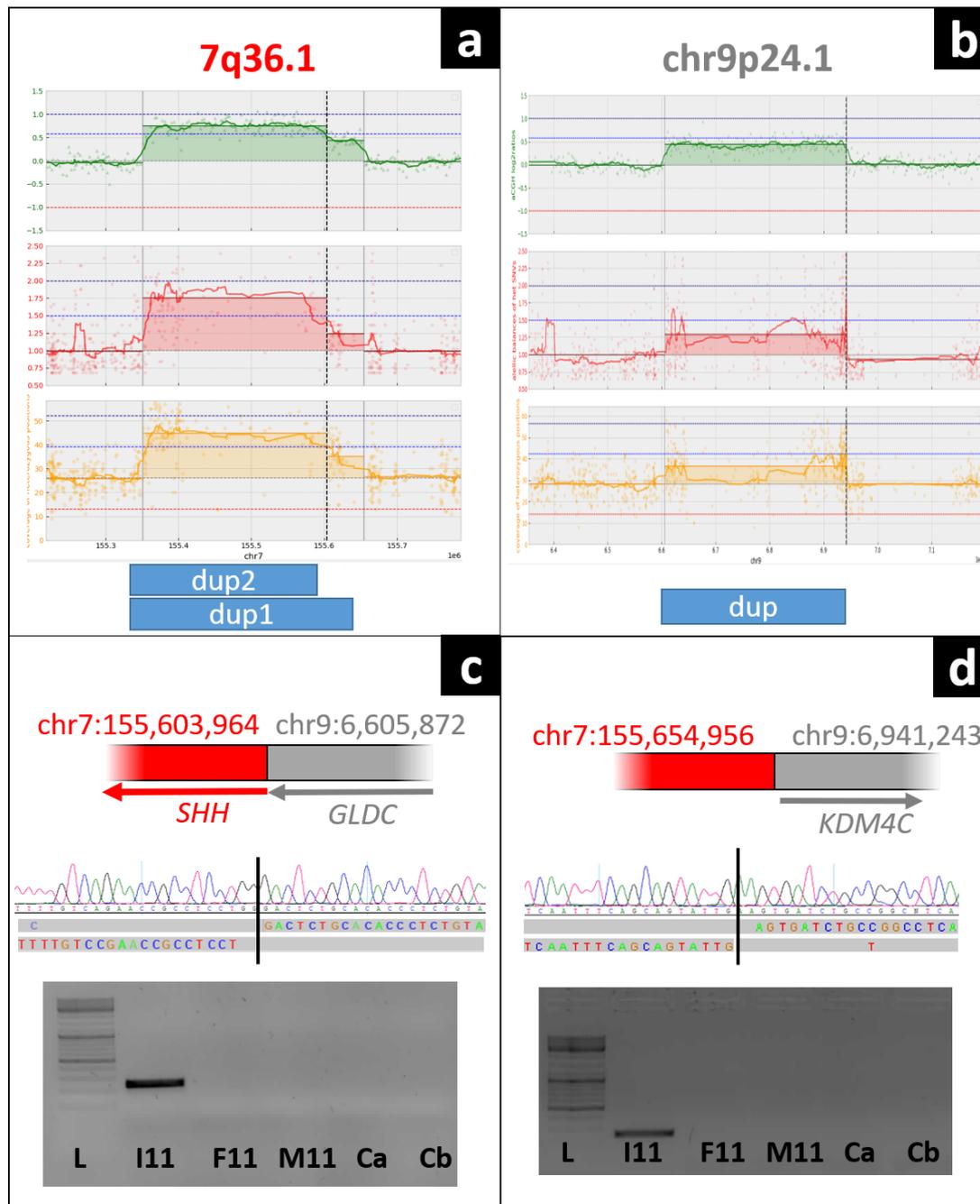


Abbildung 4: Analyse der Strukturvariante von I11: a-b: Duplikationen am Locus ch7q36.1 und chr9p24.1. Grün: aCGH Daten, Rot: Genom-Sequenz allelic balance Daten, Gelb: Genom-Sequenz coverage Daten; c-d: Bestätigung der Verbindungspunkte via Sanger-Sequenzierung, L = ladder, I = index, F = father, M = mother, Ca = control a, Cb = control b. Diese Abbildung wurde bereits in dem publizierten Artikel *Elsner et al., 2021* verwendet.

### 3.4 Mutationen in kodierenden Bereichen von Kandidatengen

#### 3.4.1 Loss of Function Mutationen in *UBA2* lösen Ektrodaktylie aus

In zwei Patienten der Kohorte konnten Varianten im *ubiquitin-like modifier-activating enzyme 2 (UBA2)* identifiziert werden. Bei diesen Varianten handelte es sich um unterschiedliche, heterozygote vorliegende frameshift Varianten. Im Fall von I12 konnte eine *de novo* Variante identifiziert werden (NM\_005499.3:c.1355\_1356delTG; p.Val452Alafs\*6), während I13 eine Variante (NM\_005499.3:c.34\_37delGCTG; p.Ala12Argfs\*34) trug, welche von seiner vermutlich phänotypisch unauffälligen Mutter vererbt wurde. Der genaue Phänotyp der Mutter konnte nicht in Erfahrung gebracht werden, da weder Lichtbild- noch radiologische Aufnahmen vorlagen.

Die Varianten wurden unter Berücksichtigung der ACMG-Richtlinien (Richards et al., 2015) klassifiziert. In beiden Fällen handelt es sich um amorphe Varianten von *UBA2*, welches einen pLI-Score von 1 aufweist (PSV1). Die Variante, welche von I12 getragen wird, trat *de novo* auf (PS2). Weiterhin sind beide Varianten weder im 1000 Genomes Project noch in ExAC gelistet (PM2) und wurden von MutationTaster als pathogen eingeschätzt (PP3). Zusätzlich hierzu berichteten Abe *et al.* 2018 von zwei Patienten, welche Microdeletionen des Locus chr19q13.11, in welchem auch *UBA2* befindlich ist, trugen (Abe et al., 2018). Yamoto *et al.* und Aerden *et al.* beschrieben verschiedene Patienten, welche eine *de novo* frameshift Mutation in *UBA2* aufweisen (Aerden et al., 2020; Yamoto et al., 2019). Weiterhin berichteten Schnur *et al.* kürzlich von diversen Individuen in mehreren, nicht miteinander verwandten Familien, welche *UBA2* Varianten aufwiesen (Schnur et al., 2021). Alle diese Fälle zeigten phänotypisch, unter anderem eine Ektrodaktylie.

Aufgrund der Faktenlage in der entsprechenden Literatur, betrachten wir *UBA2* als krankheitsauslösend und diese Varianten als pathogen (1PVS(+1PS)+1PM+1PP).

Da während der vorliegenden Arbeit neue Erkenntnisse über *UBA2* gewonnen wurden und in der Kohorte zwei Fälle von Ektrodaktylie eine Mutation in diesem Gen aufwiesen, fiel der Entschluss *UBA2* bei allen Patient\*innen des IMGs, welche eine Ektrodaktylie aufweisen, zu sequenzieren. Dies wurde per Sanger-Sequenzierung bei 24 Patient\*innen durchgeführt, welche sich in vorherigen Untersuchungen von bekannten SHFM Loci und Genen unauffällig gezeigt hatten.

In einem Fall, I14, welcher von unilateraler SHFM betroffen ist, konnte eine Variante in *UBA2* (NM\_005499.3:c.149A>G;p.Asp50Gly) nachgewiesen werden. Der Sohn und die Tochter weisen ebenfalls Ektrodaktylien auf (PP4), standen aber leider nicht für eine molekular-genetische Diagnostik zur Verfügung.

Asp50, welches durch die oben beschriebene Variante durch Gly ersetzt wird, ist Teil einer 15-stelligen Aminosäuresequenz (Ile47 bis Phe61), welche von allen *UBA2* homologen der Nephrozoen getragen wird. Olsen *et al.* zeigten, dass Mutationen von Teilen dieses Elements (Asn56Ala, Leu57Ala, Arg59Ala) in einem Funktionsverlust von *UBA2* resultieren. Weiterhin stellten sie fest, dass die in dem Fall von I14 veränderte Aminosäure Asp50 Wasserstoffbrückenbindungen mit Asn177 und Thr178 eingeht. Diese sind für die korrekte Faltung von *UBA2* unerlässlich und wirken sich somit direkt auf die Funktion aus (PS3; Olsen et al., 2010). Zusätzlich ist die Variante nicht in Vergleichsdatenbanken auffindbar (PM2) und wird von MutationTaster als pathogen klassifiziert (PP3). Aufgrund dieser Informationen ordneten wir die Variante unter Berücksichtigung der ACMG-Richtlinien als höchstwahrscheinlich pathogen ein (1PS+1PM+2PP).

### 3.4.2 *HMGB1*-Mutation als Ursache von BPTAS

Im Genom des Feten I15 konnte eine *de novo* frameshift Variante in *HMGB1* (NM\_002128:c.551\_554delAGAA;p.Lys184Argfs\*44) identifiziert werden. Das Kind der phänotypisch unauffälligen Eltern leidet unter anderem an einer mirror-image Poly-syndaktylie. Die von uns aufgefundene Variante führt dazu, dass die gesamte saure C-terminale Domäne des Proteins, welche 30 Aminosäuren umfasst, durch 41 andere Aminosäuren ersetzt wird. Diese Domäne wird normalerweise durch Aspartat/Glutamat (Asp/Glu) Repeats gebildet und ist in *HMGB1* und seinen orthologen Genen besonders stark konserviert. Sie spielt eine wichtige Rolle bei der Stabilisierung der Sekundärstruktur von *HMGB1* und ermöglicht eine korrekte Proteinfunktion (Knapp et al., 2004). Vor allem ist hierbei das Beugen von spezifischen DNA-Abschnitte zu nennen, was erst durch eine korrekte Sekundärstruktur ermöglicht wird (Anggayasti et al., 2020; Belgrano et al., 2013).

Die Variante, welche in I15 identifiziert werden konnte, findet sich in keiner Vergleichsdatenbank. In gnomAD sind zwar Varianten der C-terminalen Domäne vorhanden, allerdings verändern diese bis auf eine Ausnahme nie die eigentliche Asp/Glu Repeat Struktur dieser Domäne. Weiterhin konnten Itou *et al.* 2011 zeigen, dass *HMGB1* in Maus- und

Zebrafischmodellen mit WNT, BMP und SHH während der Extremitätenentwicklung interagiert. Knock-Out Versuche führten zu Extremitätenfehlbildungen (Itou et al., 2011). *HMGB1* besitzt einen pLI-Score von 1. Eine andere Variante in einem für mirror-image Fehlbildungen bekannten krankheitsauslösenden Gen oder einem anderen Kandidaten-Locus konnte nicht identifiziert werden. Mensah *et al.* lieferten kürzlich weitere Ergebnisse, die den Locus von *HMGB1* und den mit final-exon frameshift Mutationen (wie im Fall von I15) assoziieren Phänotyp und Pathomechanismus genauer beleuchteten (Mensah et al., 2023). Hinsichtlich dieser Erkenntnisse definieren wir die Variante von I15 als pathogen für den Phänotyp BPTAS.

#### 3.4.3 *SEMA3D* als Kandidatengen für Short-Stature-Limb-Abnormality-Syndrome

Bezüglich des Patienten I16 zeigte sich eine *de novo* missense Variante in *SEMA3D* (NM\_152754:c.1918G>A;p.Asp640Asn). Der Patient zeigt short stature, absent distal phalanges of the 5th fingers and toes und dysplastic middle phalanges of the toes. Bisher waren keine Fälle von Extremitätenfehlbildungen in der Familie aufgetreten.

Die Variante ist in der Datenbank des 1000 Genomes Projects nicht auffindbar und tritt lediglich ein Mal in ExAC auf. Weiterhin wird sie von MutationTaster als pathogen klassifiziert. Die Aminosäure Asp640, welche durch die Variante des Patienten durch Asn ausgetauscht wird, befindet sich in der Immunglobulin-ähnlichen Domäne von *SEMA3D* und weist unter Wirbeltieren eine starke Konservierung auf. *SEMA3D* ist bekannt dafür, Zelldifferenzierungen in der Neuralleiste zu regulieren. Weiterhin nimmt es eine wichtige Rolle in der Organogenese (Sanchez-Castro et al., 2015) und der Entwicklung von Extremitäten ein (Govindan et al., 2016). Aufgrund dieser Erkenntnisse wurde *SEMA3D* als potenziell krankheitsauslösendes Gen bei Kleinwuchs mit Extremitätenfehlbildungen bewertet.

#### 3.4.4 Varianten in *ALDH1A2* als potenzieller Auslöser für häutige Syndaktylien

Eine der Patientinnen, I17, trug eine paternal vererbte frameshift Variante des Gens *aldehyde dehydrogenase 1 family member A2 (ALDH1A2)*, welches für die Retinalaldehyddehydrogenase 2 kodiert. Sowohl die Index-Patientin, als auch deren Vater weisen eine häutige Syndaktylie des dritten bis vierten Fingers sowie des zweiten bis dritten Zehs auf.

Die von uns identifizierte Variante *ALDH1A2* (NM\_003888:c.35\_delT;p.Val12Glyfs\*31) ist weder in der Datenbank des 1000 Genomes Projects noch in ExAC aufgeführt. Weiterhin wird sie von MutationTaster als pathogen bewertet. Physiologisch bildet *ALDH1A2* einen direkten Regulationspartner von *HOXA13* und spielt somit eine wichtige Rolle während der Entwicklung der Finger und Zehen von Wirbeltieren. Insbesondere die Steuerung des programmierten Zelltods von interdigitalen Zellen während der embryonalen Entwicklung von Fingern bzw. Zehen stellt eine Aufgabe von *ALDH1A2* dar (Shou et al., 2013). Zhao *et al.* konnten außerdem zeigen, dass gerettete *ALDH1A2* knock-out Mäuse einen reduzierten interdigitalen Zelltod aufweisen und dadurch eine unvollständige Separation der Finger und Zehen stattfindet (Zhao et al., 2010).

Unter Berücksichtigung dieser Informationen, klassifizierten wir die paternal vererbte Variante als potenziell krankheitsauslösend.

### 3.5 Varianten unklarer Signifikanz

Neben den bereits beschriebenen Varianten konnten dreizehn weitere VUS in zehn Genen identifiziert werden. Alle diese Varianten wurden von MutationTaster als pathogen eingestuft und qualifizierten sich in den wöchentlich stattfindenden Fallbesprechungen für eine ausführlichere Analyse. Die Gene, in denen die Varianten lagen, zeigten im Mausmodell eine Expression in den Extremitäten oder waren in der Fachliteratur im Zusammenhang mit der Entwicklung von Extremitäten erwähnt worden. Bei manchen ließ die Fachliteratur außerdem einen Zusammenhang mit Regulationswegen der Extremitätenentwicklung vermuten. Unter der Anwendung der Association for Clinical Genomic Science (ACGS) Best Practice Guidelines for Variant Classification in Rare Disease 2020 (<https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf>), konnten sechs kalte, drei lauwarme, zwei warme und zwei heiße Varianten identifiziert werden (siehe Tabelle 1).

Tabelle 1: VUS-Klassifikation

Individuum	Gen	Genbank Transkript ID	DNA-Veränderung	Aminosäure-Veränderung	VUS - ACMG Kriterien	ACGS VUS Temperatur
I23	GABPB1	NM_005254	c.589G>T	V197L	PM2 + BP4	Kalt
I24	BPTF	NM_182641	c.7360G>A	A2454T	PM2 + BP1	Kalt
I24	BPTF	NM_004459	c.*1914A>T	-	PM2 + BP4	Kalt
I25	CHD9	NM_025134	c.5483T>C	M1828T	PM2 + BP4	Kalt
I25	CHD9	NM_025134	c.*1144C>T	-	PM2 + BP4	Kalt
I26	PKIG	NM_007066	c.142_144delGAG	p.E48del	PS2 + PM2 + PM4 + PP3	Heiß
I27	ASAP2	NM_003887	c.2450G>A	R817Q	PM2 + BP4	Kalt
I28	AMBRA1	NM_017749	c.305C>A	T102N	PM2 + PP4	Lauwarm
I29	RBL2	NM_005611	c.2294A>C	N765T	PM2 + PM6 + BP4	Warm
I29	FLNB	NM_001164317	c.7587_7588delAG	E2530Dfs*4	PVS1 + PM2 + PP3	Heiß
I30	BOC	NM_033254	c.1664C>T	A555V	PM2 + PM6 + BP1	Warm
I31	BAZ2B	NM_013450	c.*230C>A	-	PM2	Lauwarm
	BAZ2B	NM_013450	c.2778G>A	M926I	PM2 + BP1	Lauwarm

Darstellung der VUS entsprechend ACMG- und ACGS-Kriterien; *eigene Darstellung*

Zwei der von uns aufgefundenen VUS wurden als heiß klassifiziert. Eine der Varianten trat im Patienten I26, welcher unter Amelien der Extremitäten sowie Syndaktylien der Finger leidet, auf. Diese *de novo* (PS2) aufgetretene *in-frame* Deletion (NM\_007066:c.142\_144delGAG;p.E48del), ist in gnomAD nicht auffindbar (PM2) und verkürzt die Proteinstruktur um eine Aminosäure (PM4). Außerdem wird sie vom Genomic Evolutionary Rate Profiling (GERP; Cooper et al., 2005) als pathogen eingeordnet (PP3). Allerdings zeigten sich in der Recherche zum betroffenen Gen *PKIG* keine potenziellen Zusammenhänge zwischen der Proteinfunktion und dem Phänotyp des Patienten. Somit wurde die Variante als heiße VUS klassifiziert.

Eine weitere VUS konnte im Genom der Patientin I29 identifiziert werden. Diese Kleinwuchs und Skelettdysplasien aufweisende Patientin, trägt eine frameshift Variante in

*FLNB* (NM\_001164317:c.7587\_7588delAG;E2530Dfs\*4). Diese Variante führt zu einem nonsense mediated decay (NMD) des translatierten Proteins, also zu einem Funktionsverlust von *FLNB* (PVS1). Mutationen in *FLNB* sind mit dem Larsen-Syndrom (MIM 150250) und dem Spondylocarpotarsal-Synostosis-Syndrom (SCT; MIM 272460) assoziiert, welche sich teilweise mit dem Phänotyp von I29 decken. Zusätzlich ist die Variante in gnomAD unauffindbar (PM2) und wurde von GERP als pathogen klassifiziert (PP3). I29 besitzt konsanguine, phänotypisch gesunde Eltern und drei Geschwister, die einen ähnlichen Phänotyp aufweisen. Dies lässt ein autosomal rezessives Vererbungsmuster vermuten, welches bei SCT vorkommt. Die identifizierte Variante war allerdings heterozygot, jedoch könnte beispielsweise eine Variante auf dem anderen Allel von *FLNB* liegen, welche in der vorliegenden Analyse nicht aufgefunden werden konnte. Genauere Validierungen konnten nicht durchgeführt werden, da keinerlei genetische Informationen der Eltern oder Geschwister vorliegen.

Darüber hinaus weist *FLNB* einen pLI-Score von 0 auf. Dies bedeutet, dass das Gen von ExAC als LoF-resistent klassifiziert wird. Außerdem finden sich in gnomAD 38 LoF Varianten. Aufgrund dieser Informationslage klassifizierten wir die Variante in *FLNB* nicht als pathogen, sondern als heiße VUS.

Weiterhin konnten mit Hilfe der VCFs, welche ohne Berücksichtigung der Trio-Information erstellt wurden, und eines hauseigenen Filters 95 potenzielle VUS identifiziert werden. 38 hiervon waren potenzielle LoF, welche in zwei unabhängigen Patient\*innen, die einen ähnlichen Phänotyp zeigten, im gleichen Gen auftraten. 40 weitere traten in zwei unabhängigen Patient\*innen ohne ähnlichen Phänotyp auf. 13 der 95 wahrscheinlichen LoF Varianten lagen in Genen, welche mit Extremitätenfehlbildungen assoziiert sind. In vier Fällen konnte bei dem\*der betroffenen Patient\*in bereits eine andere Variante als pathogen identifiziert werden.

Alle VUS-Kombinationen wurden einer weiterführenden Recherche unterzogen, wobei lediglich bei einer Variante (I13: *UBA2*; s.o.) eine Pathogenität entdeckt werden konnte.

### 3.6 Analyse des nicht-kodierenden Bereichs

Parallel zur Analyse des kodierenden Bereiches wurden die WGS-Daten genutzt, um eventuelle Varianten im nicht-kodierenden Bereich aufzufinden. Letzterer wurde in der

Entwicklung der genetischen Forschung, relativ zum kodierenden Bereich lange vernachlässigt und erfährt erst in jüngster Zeit, seit dem Bekanntwerden seines enormen Interaktions- und Regulationspotentials, eine genauere Betrachtung (Pennacchio et al., 2006). In **Methodik** wurde bereits die eigenständige Erstellung eines sog. limb regulomes beschrieben, welches die Möglichkeit bieten soll, eine große Anzahl von Varianten im nicht-kodierenden Bereich auszuwerten (siehe Abbildung 5). Die massive Anzahl von Varianten, die bei der WGS-Analyse auftritt, wird mit Hilfe von funktionellen epigenetischen und evolutionären Daten von Maus und Mensch in Kombination mit *in vivo* validierten Enhancer-Sequenzen gefiltert. Dieses Regulom umfasst 5.591.007 potenziell bedeutsame Loci, die 7.294.220 Basenpaare abdecken, was ungefähr 0,24% des menschlichen Genoms (hg19) entspricht.

Insgesamt konnten 326.142.377 SNVs in nicht-kodierenden Bereichen des Genoms identifiziert werden. Von diesen waren 19.362 nicht in entsprechenden Vergleichsdatenbanken vorhanden und wurden deshalb als selten bewertet. Davon wiederum lagen 143 im definierten limb regulome.

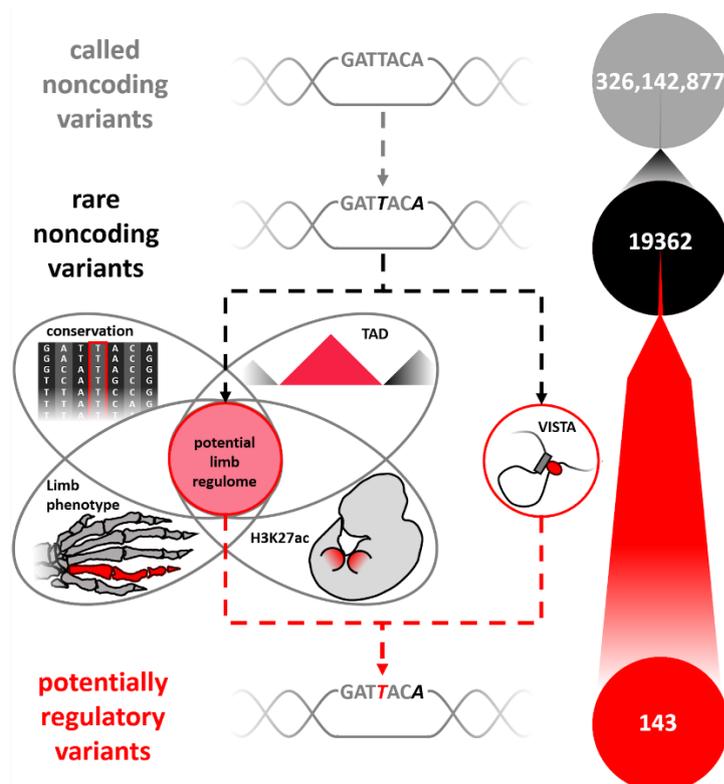


Abbildung 5: Visualisierung der Analyse des nicht-kodierenden Bereichs. Diese Abbildung wurde bereits in dem publizierten Artikel *Elsner et al., 2021* verwendet.

### 3.6.1 *De novo* Varianten in nicht-kodierenden Bereichen

In den genomischen Daten von 69 Patient\*innen konnten sieben seltene, im von uns vordefinierten regulatorischen Bereich befindliche, *de novo* Varianten identifiziert werden. Zwei dieser Varianten wurden von der weiteren Analyse ausgeschlossen, da sich während der vorherigen Analyse der Genome der entsprechenden Patient\*innen bereits höchstwahrscheinlich pathogene SNVs bzw. SVs gezeigt hatten.

Im Genom des Patienten I18, welcher unter bilateralen Syndaktylien der Finger zwei bis vier leidet, wurde eine *de novo* Variante (chr1:41,948,304AAG>A) im zweiten Intron von *EDN2* identifiziert. An dieser Position kann eine verstärkte Acetylierung von H3K27 in Extremitätenknospen nachgewiesen werden. Die zwei Basenpaare große Deletion entfernt einen von sechs AG-Repeats, dessen Länge in Wirbeltieren nicht konserviert ist. Weiterhin kodiert *EDN2* das Protein Endothelin 2, welches zur Klasse der Vasokonstriktoren gehört. Hinsichtlich Extremitätenfehlbildungen konnte bisher kein Zusammenhang gezeigt werden.

Eine weitere Patientin (I19), welche Amelien der oberen Extremität aufweist, zeigte drei Varianten (chr5:157,285,900CACGTGGG>C, chr5:157,285,909CTCGG>C, chr5:157,285,915CACAACCTG>C), welche derselben einzelnen InDel in *CLINT1* manuell zugeordnet werden konnten. Diese Variante befindet sich im ersten Intron des Gens, 15 Basenpaare abwärts der ersten Exon-Intron Grenze und besitzt eine verstärkte Acetylierung von H3K27 in Extremitätenknospen. Allerdings hat *CLINT1* lediglich einen pLI-Score von 0,54 und weist keinerlei Evidenz bezüglich eines Zusammenhangs mit der Entwicklung von Extremitäten auf.

Neben diesen Varianten konnten weitere im Genom der Patientin I20 identifiziert werden. Diese Patientin, welche unter hypoplastischen Fehlbildungen der Zehen leidet, trug die *de novo* Variante chr9:2,161,639G>T, welche 46 Basenpaare aufwärts der Grenze zwischen dem 27. Intron und dem 28. Exon des Gens *SMARCA2* liegt. Pathogene Mutationen von *SMARCA2* lösen das Nicolaidis-Baraitser-Syndrom (MIM 601358) aus, welches sich neben leichten skelettalen Fehlbildungen vor allem durch mentale Retardierung und typische Gesichtszüge auszeichnet (Sousa et al., 2009). Dies konnte bei I20 nicht nachgewiesen werden.

### 3.6.2 Nicht-kodierende Varianten in VISTA Elementen

Insgesamt konnten drei seltene Varianten in VISTA Elementen identifiziert werden. Die erste Variante war in Element 181 (chr15:37,242,490C>G (rs186583386)) lokalisiert. Deren Träger, Patient I32, hat eine fehlende Ulna sowie eine Oligodaktylie der Hand. Element 181 zeigt allerdings weder eine verstärkte Acetylierung von H3K27 in menschlichen Extremitätenknospen, noch gibt es embryologische Daten, die eine Aktivität in den sich entwickelnden Extremitäten vermuten lassen.

Eine weitere Variante konnte im Element 628 der Patientin I29 nachgewiesen werden, welche unter einer vererbten Form der skelettalen Dysplasie leidet. Außerdem zeigte I31 eine vererbte Variante (chr9:133,355,065C>T (rs142193709)) im Element 1328. Diese Patientin besitzt aplastische Radii und Daumen, das entsprechende Elternteil war phänotypisch unauffällig. Sowohl Element 628 als auch 1328 wurden vom VISTA Konsortium als neuronale Enhancer gekennzeichnet, was keinen Zusammenhang zur Entwicklung von Extremitäten vermuten lässt.

### 3.6.3 Analyse von räumlich benachbarten Varianten

In einer weiteren Analyse konnten 3.425 Varianten identifiziert werden, welche maximal 300 Basenpaare von einer anderen nicht-kodierenden Variante eines\*r anderen Patient\*in entfernt waren. Diese Varianten wurden von uns als benachbart eingeschätzt, sodass ein ähnlicher Einfluss auf eine naheliegende Region möglich wäre.

16 dieser Varianten lagen im von uns definierten regulatorischen Bereich, wovon fünf ausgeschlossen wurden, da sich die benachbarte Partner-Variante außerhalb des Bereichs befand.

In zwei Fällen konnte eine Variante nachgewiesen werden, die den festgelegten Kriterien genügte, allerdings nicht exakt dieselbe Position betraf: I21 und I22 trugen die überlappenden Deletionen, chr22:24,552,064GGGGCCGGGACTGGGGCCGGGACT>G und chr22:24,552,086ACTGGGGCCGGG>A. Diese Varianten liegen im 29. Intron von *CABIN1*, welches eine erhöhte Acetylierung von H3K27 Elementen in menschlichen Extremitätenknospen aufweist. Zusätzlich haben beide Patienten eine Syndaktylie der Finger. Allerdings wurde die Variante in beiden Fällen von phänotypisch unauffälligen Eltern vererbt und befindet sich in einem lediglich teilweise konservierten Abschnitt von *CABIN1*.

Acht weitere Varianten waren Dopplungen. Das bedeutet, dass die gleiche Variante des regulatorischen Bereichs in zwei unterschiedlichen Patient\*innen aufgefunden werden konnte. Keines dieser Paare wies einen gleichen oder ähnlichen Phänotyp auf.

Eine Suche nach kodierenden Varianten in bekannten, mit Extremitätenfehlbildungen assoziierten Genen *in trans* mit seltenen und konservierten, nicht-kodierenden Varianten in derselben TAD blieb ohne Ergebnisse.

Zusammenfassend lässt sich sagen, dass es uns nicht gelang, SNVs im nicht-kodierenden Bereich zu identifizieren, welche sich als potenziell pathogen im Zusammenhang mit Extremitätenfehlbildungen einordnen lassen würden.

### 3.7 Genotyp-Phänotyp-Korrelationsanalyse von *GLI3*

#### 3.7.1 Analyse der hauseigenen Fälle

Zur genaueren Analyse der Genotyp-Phänotyp-Korrelation am Locus von *GLI3*, erfolgte die Phänotypisierung und Sequenzierung von insgesamt 94 IPs, welche dem IMG zur Verfügung standen. Die so gewonnenen Daten wurden mit Hilfe von klinischen und genotypischen Datenbanken sowie mittels bioinformatischer Pathogenitäts-Vorhersage-tools ausgewertet.

Insgesamt konnten 15 verschiedene krankheitsauslösende Varianten in 16 Familien (vgl. tab. 1, fig. 1 (Sczakiel et al., 2021)) mit insgesamt 39 Individuen identifiziert werden (siehe Tabelle 2), welche sich über die gesamte kodierende Region von *GLI3* verteilen (siehe Abbildung 6). Hierbei handelte es sich in 14 Fällen um Varianten, welche laut ACMG-Klassifikation als pathogen einzustufen sind. In einem weiteren Fall wurde die Variante als wahrscheinlich pathogen klassifiziert.

Tabelle 2: GLI3 Varianten

Individuum	DNA-Veränderung	Aminosäure-Veränderung	ACMG-Klassifikation	ClinVar-ID
GLI3-I1	c.2374C>T	Arg792*	Pathogenic	VCV000013828
GLI3-I2	c.4172delG	Gly1391Alafs*28	Pathogenic	VCV001120232
GLI3-I3	c.2059delG	Glu687Lysfs*6	Pathogenic	VCV001120244
GLI3-I4	c.1999C>T	Arg667*	Pathogenic	VCV001120233
GLI3-I5	c.1880_1881delAT	His627Argfs*48	Pathogenic	VCV001120234
GLI3-I6	c.1793dupA	Asn598Lysfs*7	Pathogenic	VCV001120235
GLI3-I7	c.1028+1G>A	/	Pathogenic	VCV000528802
GLI3-I8	c.366C>A	Tyr122*	Pathogenic	VCV001120236
GLI3-I9	c.1033_1048del	Ala345Thrfs*15	Pathogenic	VCV001120238
GLI3-I10	c.2103+2T>A	/	Pathogenic	VCV001120239
GLI3-I11	c.2594C>G	Ser865*	Pathogenic	VCV001120240
GLI3-I12	c.868C>T	Arg290*	Pathogenic	VCV000013826
GLI3-I13	c.1133_1134insC	Pro379Serfs*33	Pathogenic	VCV001120241
GLI3-I14	c.1880A>C	His627Pro	Likely pathogenic	VCV001120242
GLI3-I15	c.2374C>T	Arg792*	Pathogenic	VCV000013828
GLI3-I16	c.3667_3670delin-sATCAA	Tyr1223lIefs*24	Pathogenic	VCV001120243

Klassifikationen und Varianten der 16 IPs im Gen *GLI3* (Transkript: NM\_000168.6). *modifiziert nach Sczakiel et al., 2021*

Im nächsten Schritt wurden die 16 IPs, bei denen eine pathogene Varianten in *GLI3* identifiziert werden konnte, hinsichtlich ihrer Phänotypen einem der drei bekannten, durch Mutationen in *GLI3* ausgelösten Syndrome (GCPS, IPD, PHS) zugeordnet. Acht Individuen zeigten ausschließlich phänotypische Merkmale einer IPD (GLI3-I1, -I3, -I5, -I8, -I9, -I10, -I14, -I15), wobei bei fünf weiteren zusätzlich phänotypische Merkmale des GCPS (GLI3-I2, -I4, -I6, -I13, -I16) vorhanden waren. Keinem der IPs konnte der Phänotyp PHS zugeordnet werden, allerdings war bei drei Individuen keine zusätzliche phänotypische

Klassifikation und somit auch keine Zuordnung zu einem Syndrom möglich. Der entsprechende Phänotyp (IPD oder GCPS) war bis auf GLI3-I6, innerhalb der Familien konstant. Im Falle von GLI3-I6, zeigte der IP die typische Ausprägung von GCPS: Er präsentierte sich mit Polysyndaktylien der Hände und Füße sowie Makrozephalie, Hypertelorismus und einer breiten Stirn. Die Mutter hingegen wies lediglich eine isolierte Polysyndaktylie auf. Insgesamt konnte keine Korrelation zwischen der Art einer Variante und einem entsprechenden Phänotyp festgestellt werden.

Eine Subgruppenanalyse der Polydaktylien der Hände zeigte, dass eine postaxiale Polydaktylie Typ B (PAP-B; MIM 174200) in sieben, eine präaxiale Polydaktylie Typ IV (PPD-IV; MIM 174700) in sechs und beide Unterformen in acht der 39 Patient\*innen aufgefunden werden konnten. Sieben weitere zeigten eine postaxiale Polydaktylie Typ A (PAP-A; MIM 174200), bei elf war keine weitere Klassifikation möglich. Eine Analyse der Polysyndaktylietypen der Füße zeigte eine PPD-IV in 22 der 39 Fälle, PAP-A in einem Fall und beide Phänotypen in einem weiteren. Zwei Fälle zeigten eine isolierte zentrale Polydaktylie, in 13 Fällen war keine genauere Klassifikation möglich.

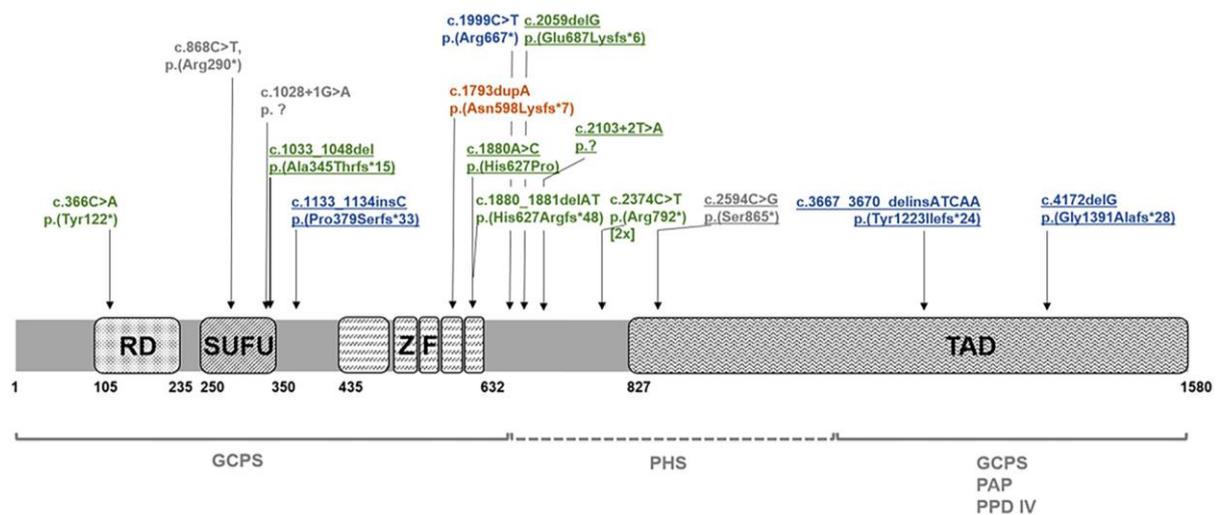


Abbildung 6: Lokalisation von 15 Varianten in *GLI3* in insgesamt 16 Familien. Unterstrichen sind solche, die erstmals beschrieben wurden. RD: Repressor-Domäne. SUFU: Suppressor-von-fusioniertem-Protein-Domäne. ZF: Zink-Finger-Domäne. Farbliche Kodierung nach, durch entsprechende Variante, ausgelöstem Phänotyp: Grün: IPD, Blau: GCPS, Orange: IPD und GCPS, Grau: keine genaue Phänotypisierung möglich. Klammern unter dem Gen markieren Domänen, welche bisher als korrelierend zum entsprechenden Phänotyp gelten. *modifiziert nach Sczakiel et al., 2021*

### 3.7.2 Analyse bereits publizierter Fälle

Nach der Analyse der hauseigenen Daten wurde eine retrospektive Analyse von bereits publizierten Fällen vorgenommen, in denen eine Mutation in *GLI3* als krankheitsauslösend klassifiziert worden war (siehe Abbildung 7). Hierbei wurden insgesamt 309 Fälle gesichtet. 65 hiervon zeigten den Phänotyp IPD, wobei in 32 Fällen (49,2%) die Variante in der N-terminalen Domäne von *GLI3* lag. In 20 (30,8%) weiteren war sie in der zentralen Domäne und in 12 Fällen (18,5%) im C-terminalen Bereich zu finden, wobei es sich bei keiner der beschriebenen Varianten der C-terminalen Domäne um eine missense Variante handelte. Ein einzelnes Individuum (1,5%) war von einer großen Strukturvariante betroffen, welche eine IPD auslöste.

In 157 von 309 Fällen trat der Phänotyp GCPS auf. Ähnlich wie bereits beim Phänotyp IPD, zeigten sich die Varianten über das gesamte Gen *GLI3* verteilt: 79 (50,3%) in der N-terminalen, 24 (15,3%) in der zentralen und 37 (23,6%) in der C-terminalen Domäne. In weiteren 17 Fällen traten Strukturvarianten auf, die mehrere Bereiche von *GLI3* betrafen. 74 der 309 Varianten lösten den Phänotyp PHS aus. 68 dieser 74 (91,9%) Varianten lagen in der zentralen Domäne von *GLI3*.

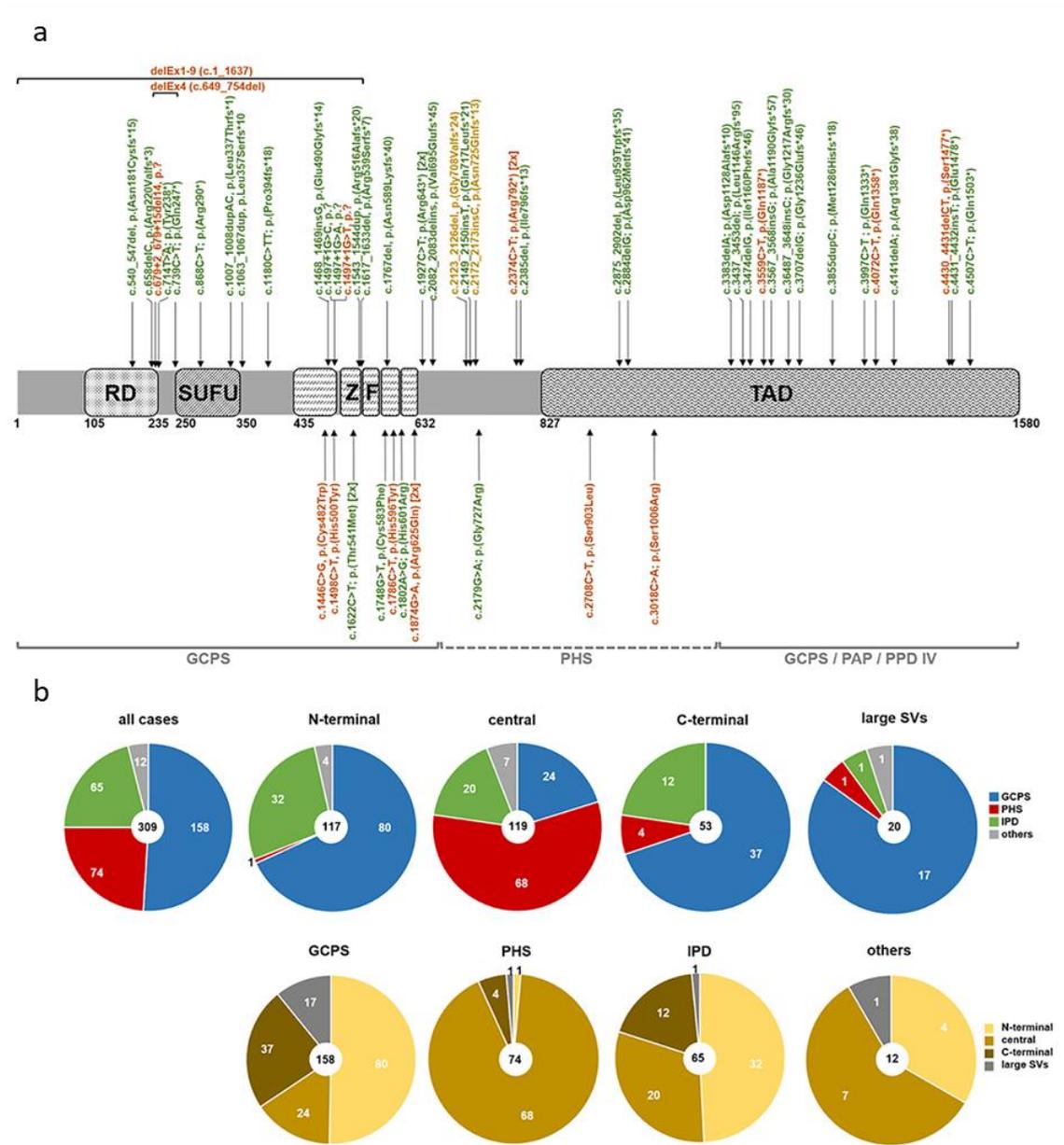


Abbildung 7: Übersicht der bereits publizierten *GLI3* Varianten und entsprechender Phänotypen. a: Lokalisation IPD auslösender *GLI3* Varianten. Oberhalb: truncating Varianten, Unterhalb: missense Varianten. Farbliche Kodierung nach, durch entsprechende Variante, ausgelöstem Phänotyp: Grün: IPD, Orange: IPD und GCPS, Gelb: IPD und PHS. Klammern unter dem Gen markieren Domänen, welche bisher als korrelierend zum entsprechenden Phänotyp gelten. b: Darstellung der absoluten Zahlen der Lokalisationen von *GLI3* Varianten im Zusammenhang mit dem entsprechenden Phänotyp. Obere Reihe: Betroffene *GLI3*-Region mit entsprechendem Phänotyp. Untere Reihe: Phänotyp mit entsprechender Lokalisation der pathogenen Varianten in *GLI3*. *modifiziert nach Sczakiel et al., 2021*

### 3.8 Validierung der Pathogenität von *HMGB1*-Mutationen

#### 3.8.1 *HMGB1*-Varianten bei Patient\*innen mit BPTAS

Bei insgesamt fünf *HMGB1*-IPs konnten zwei verschiedene Varianten gefunden werden. Im Falle von *HMGB1*-I1 konnte eine heterozygote *de novo* frameshift Mutation identifiziert werden (NM\_002128.7; c.556\_559delGAAG: p.Glu186Argfs\*42)(siehe Abbildung 8b).

Bei *HMGB1*-I2, -I3 und -I4 konnte die gleiche heterozygote frameshift Variante (NM\_002128.7; c.551\_554delAGAA: p.Lys184Argfs\*44) identifiziert werden. Die Varianten von *HMGB1*-I3 und -I4 zeigten sich nach der Analyse der Eltern als *de novo* (siehe Abbildung 8c). Die Eltern von *HMGB1*-I2 standen nicht zur Testung zur Verfügung.

Eine genauere Analyse der cDNA von *HMGB1*-I3 zeigte das Vorhandensein von Transkripten des Wildtyps sowie der *HMGB1*-Mutation. Die Expression des Gens *HMGB1* glich der Kontrolle (siehe Abbildung 8e).

Zusätzlich zu *HMGB1*-I2, -I3 und -I4 zeigte das zuvor beschriebene I15/*HMGB1*-I5 die gleiche Variante in *HMGB1*. Diese wurde bereits ausführlich beschrieben und als krankheitsauslösend für den Phänotyp BPTAS bewertet (siehe Abbildung 8d, siehe 3.4.2). Alle hier beschriebenen Varianten fanden sich im letzten Exon von *HMGB1*.

Zuletzt konnte bei *HMGB1*-I6 eine heterozygote Mikrodeletion im aCGH nachgewiesen werden (arr[hg19]13q12.3(31,035,934\_31,117,112)x1). Diese Daten wurden durch eine qPCR ergänzt. Hier zeigte sich, dass nicht wie angenommen nur die Exons eins bis vier von der Deletion betroffen waren. Vielmehr handelte es sich um eine Deletion des kompletten Gens *HMGB1*, welche sich *de novo* präsentierte (vgl. extended data fig. 2i+j; (Mensah et al., 2023)).

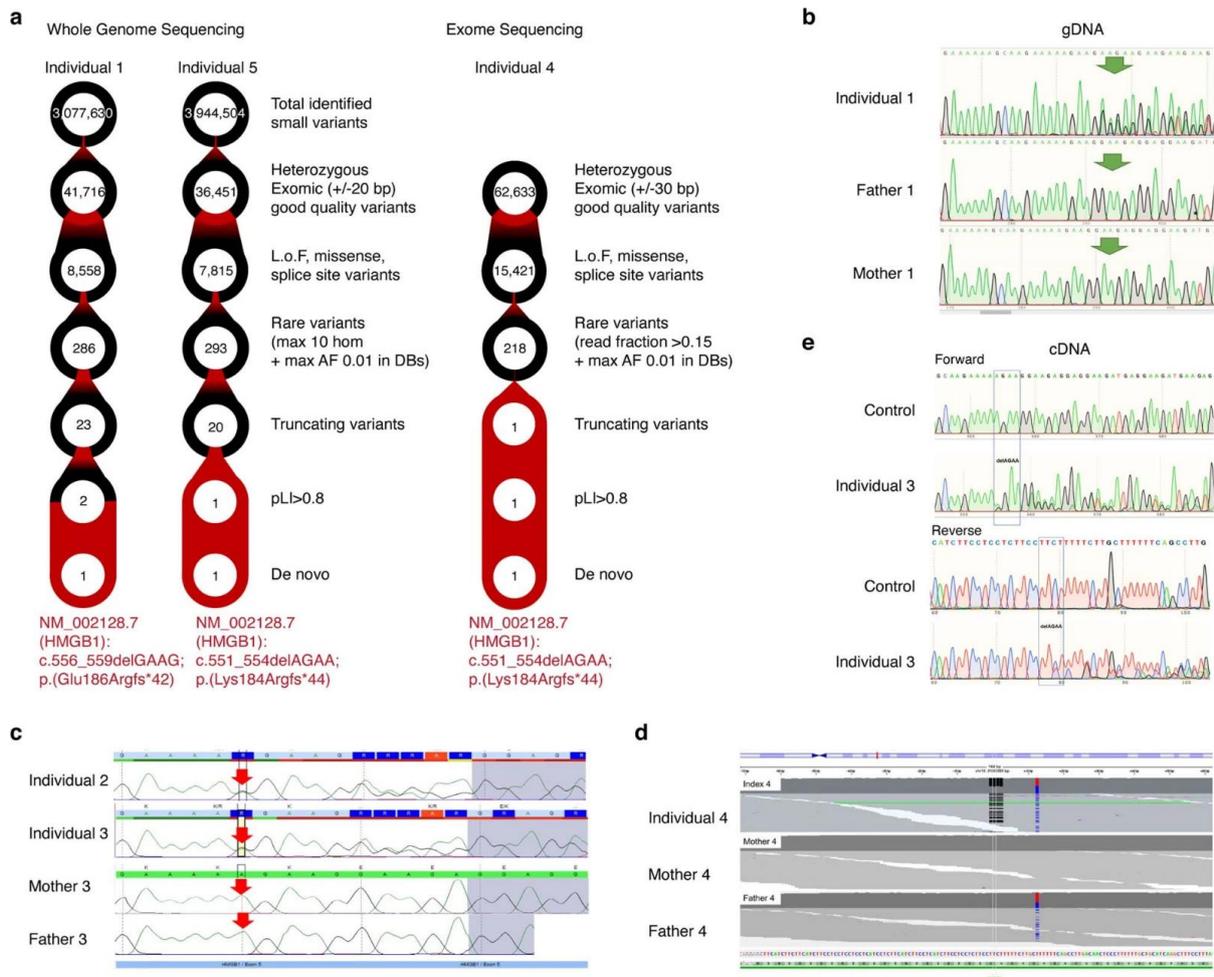


Abbildung 8: Workflow und Ergebnisse der Genotyp-Analyse der HMGB1-Individuen 1-5. a: Schema des Workflows mit Variantenanzahl der WGS (HMGB1-I1 und I5) sowie der Exom-Sequenzierung (HMGB1-I4). b: gDNA Sanger-Sequenzierung von *HMGB1* in HMGB1-I1. c: gDNA Sanger-Sequenzierung von *HMGB1* in HMGB1-I2 und I3. *De novo* in HMGB1-I3. d: Exom-Sequenzierung von HMGB1-I4. Variante gleich HMGB1-I2 und -I3 sowie *de novo*. e: Sequenzierungsdaten von *HMGB1*-cDNA bei HMGB1-I3. Sowohl Wildtyp- als auch Varianten-Transkript von *HMGB1* nachgewiesen. *modifiziert nach Mensah et al., 2023*

### 3.8.2 Face2Gene Analyse von BPTAS Individuen

Aufgrund der Ähnlichkeit der Gesichtsmorphologie von Patient\*innen mit BPTAS erfolgte eine Analyse mittels F2GC. Obwohl die Software noch nicht zur Erkennung von BPTAS trainiert wurde, zeigte sie ähnliche Ergebnisse für die einzelnen Individuen. So wurde in

drei von neun Fällen die Diagnose Blepharophimosis-ptosis-epicanthus inversus Syndrom (BPES; MIM 110100) von der Software auf Rang eins vorgeschlagen. In sechs von neun Fällen erreichten BPES oder das 22q11.2-Deletions Syndrom die ersten zehn Ränge. Darüber hinaus zeigte sich BPES in acht von neun Fällen unter den Top-30 der von F2GC vorgeschlagenen Diagnosen.

Normalerweise wird BPES nur selten von F2GC als Diagnose in Betracht gezogen (Pantel et al., 2020). Zusätzlich können die beiden charakteristischen Merkmale von BPES (Blepharophimose und Telekanthus/Hypertelorismus) auch auf der Gesamt-Maske von BPTAS erkannt werden (vgl. extended data fig. 1s; (Mensah et al., 2023)). Dies lässt vermuten, dass Patienten mit BPTAS charakteristische faziale Merkmale besitzen.

### 3.8.3 Phasentrennung abhängig von *HMGB1*

Im Anschluss an die beschriebenen Ergebnisse erfolgte eine detaillierte Analyse der Eigenschaften des Proteins HMGB1. Dieses DNA-bindende Protein besitzt zwei high mobility group (HMG) Boxen für DNA-Bindung (Lambert et al., 2018) sowie einen sauren Schwanz am Ende der C-terminalen Domäne. Diese saure Domäne ist hoch konserviert und liegt nach aktuellen Erkenntnissen in einer sogenannten intrinsically disordered region (IDR) des Proteins (siehe Abbildung 9). IDRs finden sich in verschiedenen anderen Proteinen, wie zum Beispiel Transkriptionsfaktoren oder der RNA-Polymerase II, in welchen sie für eine Phasentrennung durch die Beeinflussung der Interaktion der Proteine untereinander sorgen. Aufgrund der beobachteten Veränderung in der IDR von HMGB1 durch die beschriebenen Varianten, konnte eine Veränderung der phasentrennenden Eigenschaften des Proteins vermutet werden. Diese Vermutung konnte von Mensah *et al.* in diversen Experimenten bestätigt werden (vgl. fig. 2; (Mensah et al., 2023)).

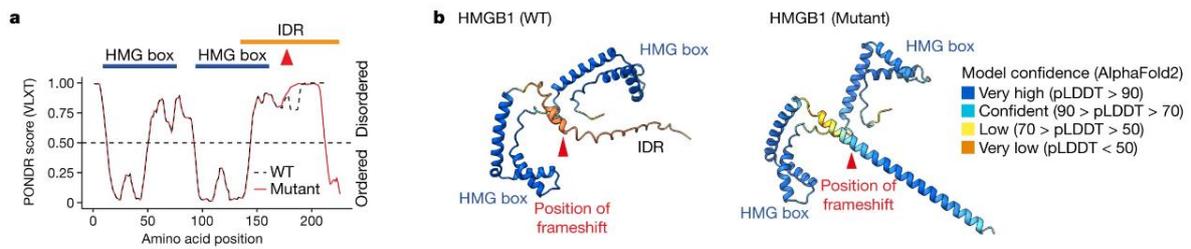


Abbildung 9: Analyse des Proteins HMGB1. a: Darstellung der inneren Unordnung von HMGB1. Der rote Pfeil markiert die Lokalisation der BPTAS final-exon frameshift Variante. Blaue Bereiche markieren die HMG-Box, der orange Bereich die IDR. WT: Wildtyp. b: Strukturvorhersage von HMGB1 als Wildtyp und Variante mittels AlphaFold2 (Jumper et al., 2021; Varadi et al., 2021). Die Farben Blau bis Orange zeigen die Sicherheit der Vorhersage an verschiedenen Lokalisationen (pLDDT-Wert). *modifiziert nach Mensah et al., 2023*

## 4. Diskussion

### 4.1 Identifizierte Varianten

Durch strukturiertes Filtern und eine gezielte Analyse der Genom-Daten der Kohorte aus 69 Familien, war es möglich fünf SNVs in bekannten Krankheitsgenen und drei in Kandidatengenen zu identifizieren. Des Weiteren konnte in zwei Fällen eine Strukturvariante als Ursache der Extremitätenfehlbildung der jeweiligen Patient\*innen aufgezeigt werden. Außerdem gelang der Nachweis einer pathogenen Repeat-Expansion in *HOXD13* in einem weiteren Fall.

Im Genom der Patientin I10 befindet sich eine komplexe Strukturvariante. Es handelt sich hierbei um eine circa 100kb umfassende Inversion, welche von einer 10kb großen Deletion flankiert wird. In der aCGH, welche die Patientin bereits 2013 erhalten hatte, war lediglich die Deletion aufgefallen, die allerdings nicht als pathogen klassifiziert wurde. Die wesentlich größere Inversion, welche sich direkt an die Deletion anschließt, konnte in der aCGH nicht nachgewiesen werden, da sich die Kopienanzahl bei einer Inversion nicht verändert. Deshalb ergab die Diagnostik zuerst keinen Befund, der den Phänotyp der Patientin erklären konnte. Eine Analyse der Region via Sanger-Sequenzierung wäre vor WGS nicht möglich gewesen, da die basengenaue Lokalisation der Inversion für das entsprechende Primer-Design hätte bekannt sein müssen. Auch eine Exom-Sequenzierung hätte keine Ergebnisse liefern können, weil sich die Bruchpunkte außerhalb des exonischen Bereichs befinden. Schlussendlich wäre auch eine Chromosomenanalyse nicht aufschlussreich gewesen. Bei diesem Verfahren liegt die maximale Auflösung bei 5 Mega-Basen, wodurch die deutlich kleinere Inversion von I10 übersehen worden wäre. Im Rahmen der vorliegenden Arbeit erhielt die Patientin ein WGS, wodurch beide Strukturvarianten identifiziert und als pathogen klassifiziert werden konnten. Herkömmliche molekular-genetische Herangehensweisen hätten diese Variante aus bereits genannten Gründen nicht auffinden können. Nur das WGS war in der Lage entscheidende Erkenntnisse zu liefern, welche zur Diagnose der Variante führten.

Des Weiteren lässt diese Variante vermuten, dass der SHFM3 Locus auch durch komplexe SVs beeinflusst werden kann. Bisher waren nur Duplikationen als pathogen beschrieben worden. Es darf allerdings angenommen werden, dass im Fall I10 ein verändertes Zusammenspiel von Enhancer-Elementen dieser und einer benachbarten TAD

zum entsprechenden Phänotyp SHFM3 führt. Um einen genauen Eindruck der exakten Regulationsmechanismen an diesem Locus zu erhalten, bedarf es jedoch weiterer Untersuchungen.

Im Falle des Patienten I11 konnte eine komplexe Translokation als pathogen identifiziert werden. In der aCGH war bereits 2015 eine *de novo* Duplikation des Locus 7q36.3 gezeigt worden, diese erklärte den Phänotyp des Patienten jedoch nicht vollständig. Durch die Durchführung des WGS konnte neben der Duplikation eine komplexe Translokation zwischen Chromosom 7q36.3 und Chromosom 9p24.1 nachgewiesen werden, welche die Struktur der Loci fundamental verändert. Wir definierten diese Variante als pathogen, konnten also auch in diesem Fall eine aus dem Array bestehende Information erweitern und konkretisieren, welche schlussendlich zu einer molekular-genetischen Diagnose führte.

In einem weiteren Fall, I9, konnte eine *HOXD13* Repeat-Expansion bestimmt werden. Diese war durch vorherige molekular-genetische Untersuchungen nicht aufgefallen, konnte durch WGS allerdings identifiziert und durch herkömmliche *HOXD13* Mikrosatellitenanalyse validiert werden. Hier zeigt sich ein weiterer Vorteil des WGS, nämlich eine nahezu unbeschränkte Sicht auf das gesamte Genom eines Menschen. Natürlich ist es von Vorteil, wenn man anhand des Phänotyps den Bereich der möglichen pathogenen Loci eingrenzen kann, sodass nur noch diese analysiert werden müssen. Allerdings ist dies vor allem im Bereich molekular-genetischer Forschung bei großen Kohorten nicht immer möglich. Eine unvoreingenommene Analyse des gesamten Genoms weitet also den Blick, um einerseits Varianten in unbekanntem bzw. unbeachteten Loci zu identifizieren und andererseits das Wissen über bestehende Krankheitsloci zu erweitern.

All diese unterschiedlichen, zum Teil komplexen, Arten von Mutationen konnten durch ein einziges Testverfahren, das Whole-Genome-Sequencing, identifiziert werden. Dadurch wurde in 12 von 69 Fällen eine Variante als pathogen klassifiziert (17,4%), was sich mit Ergebnissen anderer Studien deckt. So beschreiben Turro *et al.* 2020, dass in einer Kohorte von 7.065 Patienten, 1.138 (16,11%) durch WGS molekular-genetisch diagnostiziert werden konnten. Die Größe der Gesamtkohorte betrug 13.037, allerdings wurde nur bei 9.802 eine seltene Erkrankung erkannt und nur 7.065 davon phänotypisiert.

Hierbei ist allerdings anzumerken, dass die Kohorte vor Einschluss keiner molekular-genetischen Diagnostik unterzogen wurde. Dadurch wurden auch Patient\*innen eingeschlossen, welche durch eine herkömmliche molekular-genetische Untersuchung, wie zum Beispiel eine aCGH, diagnostiziert worden wären (Turro et al., 2020).

Dies stellt einen Gegensatz zum Vorgehen dieser Arbeit dar, da hier lediglich Patient\*innen eingeschlossen wurden, welche sich in herkömmlicher molekular-genetischer Diagnostik ohne pathologischen Befund gezeigt hatten. Daraus kann man schließen, dass der wahre Anteil von Patient\*innen, die durch WGS eine molekular-genetische Diagnose erhalten können, höher liegt als von Turro *et al.* gezeigt.

#### 4.1.1 *UBA2* als neues Krankheitsgen

In zwei Patienten der Kohorte sowie in einem weiteren Fall des IMGs, konnten pathogene Varianten in *UBA2* identifiziert werden. Alle Patienten weisen den Phänotyp SHFM auf und besitzen laut den Kriterien des ACMG pathogene (I12; I13), bzw. höchstwahrscheinlich pathogene (I14) Varianten. Außerdem konnte gezeigt werden, dass der Phänotyp, welcher durch eine monoallelische Veränderung von *UBA2* auftritt, variabel ist und Mutationen eine unterschiedliche Penetranz besitzen. Dieses Ergebnis deckt sich mit aktuellen Erkenntnissen (Schnur et al., 2021). So beschreiben auch Yamoto *et al.* klinisch unauffällige Träger von hypo- bzw. amorphen *UBA2* Mutationen (Yamoto et al., 2019). Dies ist auch bei der Mutter von I13 der Fall. Weiterhin zeigt sich die entsprechende SHFM in den Patienten weder bilateral identisch noch an allen Extremitäten. Der genaue Grund für diese Variabilität sowie die unvollständige Penetranz sollte in weiteren Studien untersucht werden.

#### 4.1.2 Varianten unklarer Signifikanz

Neben den bereits beschriebenen Varianten gelang es 174 VUS zu identifizieren, wovon 13 als besonders vielversprechende Kandidaten gehandelt wurden. Die restlichen Varianten waren zwar selten und wurden von den genutzten Vorhersage-Tools als potenziell pathogen identifiziert, wiesen allerdings immer mindestens ein Kriterium auf, welches sie von der weiteren Analyse ausschloss. Dies war dann der Fall, wenn für den\*die betroffene\*n Patient\*in bereits eine andere, pathogene Variante identifiziert werden konnte oder zu wenig Informationen über das betroffene Gen vorlagen.

Durch weitere Recherche konnten die 13 vielversprechendsten Varianten entsprechend der Klassifikation und Terminologie des ACGS als sechs kalte, drei lauwarme, zwei warme und zwei heie Varianten klassifiziert werden. Die heien Varianten wiesen hierbei zwar Kriterien auf, die eine Pathologie derselben vermuten lieen, allerdings konnte in keinem der Flle ein direkter Zusammenhang zwischen mutmalich pathogener VUS und Phnotyp beziehungsweise Vererbungsmuster hergestellt werden.

Alle gefilterten VUS sowie die 95 VUS, die durch die Einzel-Analyse der VCFs identifiziert werden konnten, wurden auf der Website Genematcher.org eingetragen und mit der wissenschaftlichen Gemeinschaft geteilt. Bisher konnte in zwei Fllen ein vielversprechender Kontakt aufgebaut werden. Der Beweis der Pathogenitt einer Variante war allerdings noch nicht mglich.

## **4.2 Analyse des nicht-kodierenden Bereichs**

Der in dieser Arbeit konzipierte Ansatz zur Analyse des nicht-kodierenden Bereiches des menschlichen Genoms bestand darin, mit Hilfe von bereits vorhandenen Erkenntnissen (Position von Enhancern in der VISTA Datenbank, H3K27 Acetylierung-Peaks in menschlichen Extremittenknospen, cis-regulatorische Domnen) sowie hauseigenen Daten potenziell relevante Bereiche zu definieren. Das so entstandene limb regulome umfasst circa 0,24% des gesamten Genoms, von denen angenommen werden darf, dass sich dort mit einer hheren Wahrscheinlichkeit Varianten befinden, welche im Zusammenhang mit Extremittenfehlbildungen stehen. Leider konnten nur unsichere Erkenntnisse ber urschliche Varianten von Extremittenfehlbildungen im nicht-kodierenden Bereich des Genoms gewonnen werden.

Die Ursachen hierfür sind vielfltig. Zum einen wurde ein definiertes Filter genutzt, welches das Sichtfeld fr Varianten einschrnkte. Des Weiteren wurde ein willkrlich gewhlter Abstand von 300 Basenpaaren fr die Definition der Varianten verwendet, die einen gleichen oder hnlichen regulatorischen Effekt aufweisen knnten. Dadurch wurde vorausgesetzt, dass zwei Varianten nah beieinanderliegen mussten, um den gleichen Enhancer und somit das gleiche Gen zu beeinflussen. Die Expression eines Gens kann aber auch von mehreren Enhancern reguliert werden, welche grere Abstnde zueinander

aufweisen (Pennacchio et al., 2013). Zusätzlich liegen weit komplexere Regulationsmechanismen im Genom vor, welche durch den hier gewählten Ansatz nur zum Teil abgedeckt werden können (Spielmann et al., 2018). Weiterhin wurden lediglich Mendel'sche Erbgänge in der Analyse berücksichtigt.

In weiteren Studien sollte versucht werden, das gesamte Genom inklusive maximal vieler bekannter Regulationsmechanismen zu betrachten. Somit könnten nicht-kodierende Varianten identifiziert werden, die über komplexe Mechanismen die Expression von einem oder mehreren Genen beeinflussen. Dadurch bestünde zum Beispiel die Möglichkeit der Diagnostik von oligo- beziehungsweise polygenen Erkrankungen.

### 4.3 Qualität der Analyse

Durch VarFish war es möglich SNVs und SVs genomweit zu betrachten. Sowohl durch die Voranalyse und Datenspeicherung in SODAR als auch durch das benutzerfreundliche Interface war eine strukturierte Analyse der 334.719.069 Varianten möglich. Allerdings befindet sich die Software zum Zeitpunkt der Erstellung dieser Arbeit im Beta-Stadium, wodurch noch einige Fehler in der Arbeitsweise bestehen könnten.

Bei den Strukturvarianten wurde eine weitere Schwäche der Software sichtbar: Die von VarFish genutzten Algorithmen, welche aus den genomischen Rohdaten die SVs identifizieren, präsentierten sich vor allem bei Inversionen als teilweise unpräzise. Von den 633 SVs zeigten sich nach manueller Analyse 417 falsch-positive Calls, wovon 194 als Inversionen gecallt wurden. Falsch-positiv bedeutet hierbei, dass sie entweder fälschlicherweise als *de novo* identifiziert wurden oder an der Stelle, an der die Algorithmen eine Variante vermuteten, keine zufriedenstellende Datenlage (zum Beispiel zu wenige reads) im BAM-File vorhanden war. Dieses Problem beim Auffinden von Varianten lässt zusätzlich vermuten, dass ein Teil von SVs durch die Algorithmen nicht gefunden und somit auch nicht von VarFish angezeigt wurde.

Durch die hohe Rate an falsch-positiven Filterergebnissen, welche unter anderem durch die oben beschriebenen Schwachstellen der Software auftraten, mussten die Filter bezüglich der minimalen Größe der Strukturvarianten enger eingestellt werden. Dadurch konnten nur SVs größer als 1,5kb betrachtet werden, wodurch eine manuelle Auswertung jeder Strukturvariante, welche den Filterkriterien genügte, ermöglicht werden konnte.

Somit wird allerdings ein großer Vorteil der Genomanalyse gegenüber der aCGH abgeschwächt. Während eine aCGH eine Genauigkeit von circa 3kb aufweist, ist es mittels Genomanalyse theoretisch möglich, CNVs basengenau zu identifizieren. Dieser Vorteil musste allerdings durch das nachfolgende Filter begrenzt werden, um eine manuelle Auswertung der SVs gewährleisten zu können.

Wie in Abbildung 1 zu erkennen, befanden sich im Datensatz neun Ausreißer, in deren Fällen eine abnormal hohe Anzahl an seltenen SNVs von VarFish identifiziert wurden. Dies kann mehrere Ursachen haben. Zum einen handelt es sich bei den genutzten Datenbanken von ExAC, gnomAD und dem 1000 Genomes Project um solche, die größtenteils europäische, süd-asiatische und mittel-amerikanische Exom- bzw. Genomdaten als Referenz enthalten (Karczewski et al., 2020; Lek et al., 2016). Dadurch besteht die Möglichkeit, dass bei einem Patienten mit zum Beispiel indischer Herkunft, mehr Varianten als selten beurteilt werden, da sie nicht häufig in den Vergleichspopulationen vorkommen. Zusätzlich lässt sich nicht ausschließen, dass in diesen neun Fällen eine schlechte Qualität der Daten bestand oder dass die SNV-callende Software fehlerhaft war.

Bei den SVs konnten insgesamt fünf Ausreißer mit einer erhöhten Anzahl an Calls festgestellt werden. Insgesamt mussten in neun Fällen strengere Filterkriterien angewandt werden, da hier initial zu viele Calls gefiltert wurden, um sie manuell auszuwerten. Darüber hinaus war in einem Fall sogar eine manuelle Auswertung der Daten durch die Vielzahl an Calls unmöglich. Die soeben für SNVs beschriebenen Fehlerquellen, lassen sich hierbei zum Teil auf die SVs übertragen. Darüber hinaus spielt natürlich die zuvor angesprochene Ungenauigkeit der Calling-Algorithmen eine große Rolle. Im Falle des ausgeschlossenen Patienten lässt sich aufgrund der außergewöhnlich hohen Anzahl von SV-Calls eine Chromotrypsis als weitere Ursache nicht ausschließen. Hierbei kommt es innerhalb von kurzer Zeit zu intra- und interchromosomalen Neuordnungen, welche sich unter anderem in Duplikationen, Deletion, Inversionen und Break-Ends (also potenziellen Translokationen) äußern (Jones & Jallepalli, 2012). Diese Varianten würden von VarFish als SVs identifiziert werden und könnten somit zu der hohen Anzahl an Calls bei I4 führen. Eine Testung auf eine Chromotrypsis sollte im weiteren Verlauf erfolgen.

Die Filtereinstellungen, welche in VarFish verwendet wurden, waren so gewählt, dass eine geeignete Balance zwischen der Anzahl der gefilterten Calls und deren Qualität be-

stand. Zu Beginn dieser Arbeit, wurden verschiedene Einstellungen angewandt und deren Ergebnisse durch den Autor der vorliegenden Arbeit ausgewertet und angepasst. So konnte ein Filter entwickelt werden, welches hochqualitative Calls seltener Varianten lieferte, ohne die Kapazitäten der manuellen Auswertung zu erschöpfen. Könnte man diese Prozesse weiter oder vollständig automatisieren und die Qualität der Rohdaten sowie die Algorithmen zum Auffinden von Varianten verbessern, wird eine noch effektivere und ausführlichere Auswertung von Genomdaten möglich sein. Damit könnte in Zukunft eine noch größere Prozentzahl von Fällen gelöst werden.

#### **4.4 Limitationen von WGS bei der Untersuchung von 69 IPs mit Extremitätenfehlbildungen**

WGS besitzt neben seinem enormen Potential diverse Schwächen. Bei der Sequenzierung werden überwiegend short-reads verwendet, wodurch zwar SNVs und kleine InDels exakt bestimmt werden können, größere SVs jedoch häufig nicht. Darüber hinaus besitzt das Genom viele repetitive Elemente, also Abschnitte in denen sich eine bestimmte Basenabfolge häufig wiederholt. Bei short-read basierten Sequenzieretechnologien kann es während des Mappings zu einer Fehlzuordnung oder auch zu Lücken kommen, welche die Analyse aufgrund inkompletter Datensätze erschweren (van Dijk et al., 2018). Außerdem ist die Länge bzw. Längenveränderung sehr langer repetitiver Abschnitte nicht sicher zu bestimmen. Um diese Schwachstelle auszugleichen, könnten in Zukunft long-read Sequenzierungsmethoden verwendet werden. Vielversprechende Ansätze sind hierbei Nanopore-Sequenzierung oder die single-molecule real-time Sequenzierung, welche zusätzlich das Potential besitzen, epigenetische Veränderungen der DNA zu detektieren (van Dijk et al., 2018).

Neben der Sequenzierungsmethode weist auch die nachfolgende Analyse der Daten Verbesserungsmöglichkeiten auf. Die in diesem Projekt verwendete Software zur Analyse der Genom-Daten deckt nur einen Teil der möglichen pathogenen Veränderungen des Genoms ab. So wurden zum Beispiel lediglich monogene Erbgänge betrachtet, wodurch oligo- bzw. polygene ausgeschlossen worden. Des Weiteren konnten durch zusätzliche Analysen der Rohdaten, neben dem Filtern in VarFish, zwei Varianten identifiziert werden, bei denen eine reduzierte Penetranz vermutet werden kann. Die eigentliche Analyse war hierfür nicht ausgelegt, sodass initial die Variante von I13 in *UBA2* nicht identifiziert

werden konnte. Dies war erst durch die zusätzliche, vergleichende Analyse der VCF-Daten mehrerer Familien möglich.

Außerdem können epigenetische Vorgänge, welche immer häufiger als Ursachen für Fehlbildungen des Menschen vermutet werden (Garland et al., 2020), durch ein WGS unzureichend betrachtet werden.

All dies legt nahe, dass durch die Implementierung zusätzlicher Methoden zur Identifikation von Varianten und deren Regulationsmechanismen weitere Fälle molekular-genetisch diagnostiziert werden können.

#### **4.5 Genotyp-Phänotyp-Korrelation von *GLI3***

In der durchgeführten Untersuchung konnten 15 Varianten in 16 Familien identifiziert werden. Acht Individuen präsentierten eine IPD, wobei fünf weitere zusätzliche Merkmale des GCPS zeigten und keines dem Phänotyp PHS zugeordnet werden konnte.

Bisherige Studien legen nahe, dass im Falle von *GLI3* eine Genotyp-Phänotyp-Korrelation besteht. Demnach sollen durch Varianten in *GLI3* insgesamt neun Syndrome bzw. Erkrankungen entstehen, in denen das Merkmal Polydaktylie vorkommt (Al-Qattan et al., 2017). Es wurden dabei sowohl Mutationsarten (frameshift/nonsense, missense, splicing usw.), als auch Lokalisation der Variante im Gen einzelnen Phänotypen zugeordnet. Varianten, welche vor allem im vorderen Drittel des Gens liegen, sollen in GCPS resultieren, während Varianten des mittleren Drittels PHS auslösen (Johnston et al., 2005).

Ein solche eindeutige Korrelation konnte im Fall der vorgelegten Studie nicht nachgewiesen werden. Die aufgefundenen Varianten, die GCPS auslösen, finden sich zwar in der C-terminalen, allerdings auch in der N-terminalen sowie in der zentralen Domäne. Darüber hinaus konnten fünf Varianten identifiziert werden, die in der zentralen Region liegen. Neben GCPS (1/5) zeigten die IPs den Phänotyp IPD (3/5). Im letzten Fall konnte keine genauere Analyse des Phänotyps stattfinden. Keiner dieser IPs zeigte das für Varianten in der zentralen Domäne typische PHS, wobei keine Bildgebung des Kopfes erfolgte, sodass asymptotische Hamartome nicht ausgeschlossen werden können.

Zusammenfassend steht die Verteilung der GCPS auslösenden Varianten, die in dieser Arbeit im gesamten Bereich von *GLI3* nachgewiesen werden konnten, im Einklang mit anderen bisher publizierten Ergebnissen (Kalff-Suske et al., 1999). Eine eindeutige *GLI3*-

GCPS-Genotyp-Phänotyp-Korrelation, wie etwa von Johnston *et al.* beschrieben, konnte nicht bestätigt werden (Johnston *et al.*, 2005).

Im Fall des Individuums GLI3-I6, konnte ein veränderter Phänotyp in unterschiedlichen Generationen einer Familie festgestellt werden. Der IP zeigte die typische Ausprägung von GCPS, während die Mutter lediglich eine isolierte Polysyndaktylie aufwies. Dies legt die Vermutung nahe, dass GCPS und IPD unterschiedliche Schweregrade der gleichen Grunderkrankung sind, was sich mit anderen Hypothesen der Genotyp-Phänotyp-Korrelation von *GLI3* deckt (Johnston *et al.*, 2010). Darüber hinaus lässt die Lokalisation der fünf Varianten in der zentralen Domäne die Vermutung zu, dass sich IPD hier als milde Form von PHS präsentiert (Johnston *et al.*, 2010).

Nach der Analyse der hauseigenen Fälle, erfolgte die retrospektive Auswertung von bereits publizierten pathogenen *GLI3*-Varianten und deren Beurteilung bezüglich einer möglichen Genotyp-Phänotyp-Korrelation. Varianten, die eine IPD oder GCPS auslösten, konnten in allen drei Bereichen von *GLI3* identifiziert werden. Dies deckt sich mit den Ergebnissen anderer Studien größerer Kohorten (Kalff-Suske *et al.*, 1999). Eine genauere Analyse der IPD-auslösenden Varianten zeigte, dass nonsense Varianten über die gesamte kodierende Region von *GLI3* (c.540–c.4507) zu finden waren, während missense Varianten vermehrt in der zentralen Region (c.1446–c.3018) auftraten. Bei letzterem war vor allem die Zink-Finger-Domäne betroffen. Insgesamt konnte gezeigt werden, dass IPD und GCPS von einer Vielzahl an Varianten unterschiedlicher Lokalisation ausgelöst werden, was mit Ergebnissen anderer Publikationen übereinstimmt (Jamsheer *et al.*, 2012; Kalff-Suske *et al.*, 1999).

Lediglich PHS wies in der retrospektiven Analyse eine Locus-abhängige Genotyp-Phänotyp-Korrelation auf, da 91,9% der beschriebenen Varianten im zentralen Drittel des Proteins zu finden waren. Diese Korrelation zeigt sich in vergleichbaren Studien und deckt sich mit der Beschreibung eines sogenannten PHS-Bereichs von *GLI3* (Démurger *et al.*, 2015; Johnston *et al.*, 2005). Diese Beobachtungen zeigen erneut, dass *GLI3*-Varianten unterschiedlicher Art und unterschiedlicher Lokalisation mit einem breiten phänotypischen Spektrum assoziiert sind. Vor allem GCPS und PHS differenzieren sich phänotypisch voneinander und können im Falle von PHS einem spezifischen Bereich zugeordnet werden. IPD präsentiert sich im Zwischenbereich als milde Manifestation von GCPS oder seltener PHS. Darüber hinaus konnten auch asymptotische Träger\*innen

von *GLI3*-Varianten identifiziert werden, was die phänotypische Bandbreite zusätzlich unterstreicht (Démurger et al., 2015).

#### 4.6 Kleine intragenische Deletion von *GLI3*

Im weiteren Verlauf konnte in einem zusätzlichen Fall mit *GLI3*-assoziierter Polydaktylie (*GLI3*-I17), welcher dem IMG im Rahmen der Routinediagnostik zugesandt wurde, eine 130 Basenpaare große Deletion identifiziert werden. Der Patient hatte zuvor eine genetische Standarddiagnostik sowie eine Ganzgenomsequenzierung und entsprechende Auswertung erhalten. Beide Methoden zeigten keine pathologisch auffällige Variante in *GLI3*. Bei einer genaueren manuellen Auswertung konnte jedoch eine 130bp große Deletion (NM\_000168.6; c.420\_473+76del) im Übergang vom dritten Exon zum dritten Intron gefunden werden. Diese Variante wurde per Sanger-Sequenzierung bestätigt.

Die herkömmlichen Methoden versagten in diesem Falle aus zweierlei Gründen. Die übliche Sanger-Sequenzierung des Gens *GLI3* erfasste die Mutation nicht, da die Primer (*GLI3*-fwd: TTGCTTTGTGAATCGGAATG, *GLI3*-rev:TGGTAAAAGCCAGCATCTCG) in der Deletion lokalisiert waren. Deshalb wurde lediglich der Wildtyp amplifiziert.

Das WGS wiederum identifizierte die Variante nicht, da eine minimale SV-Größe von 3kb gewählt wurde. Dies geschah um häufig auftretende Artefakte, die oftmals kleiner als 3kb sind, herauszufiltern. Da die Variante in *GLI3* lediglich 130bp groß war, wurde sie mit diesen Filtereinstellungen nicht erkannt. In einer nachfolgenden Untersuchung mit an die Deletion angepassten Primern (*GLI3*-fwd:ACAGGCATTCCTCAGGAAGC, *GLI3*-rev:GCTTTGTGAATCGGAATGCTGT), konnte die Variante *de novo* nachgewiesen und anschließend als pathogen klassifiziert werden (siehe Abbildung 10).

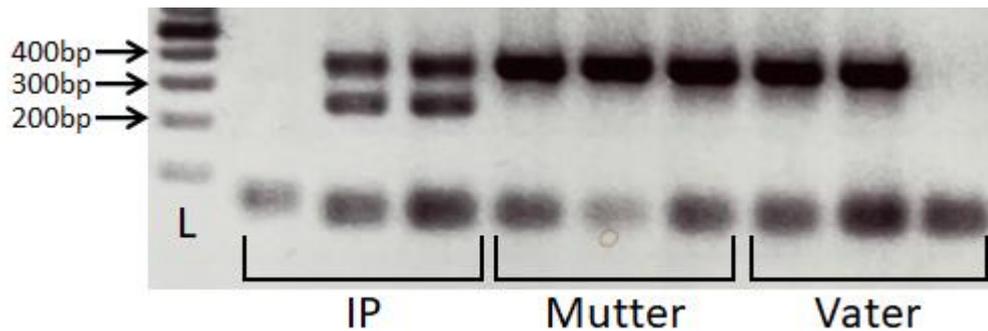


Abbildung 10: Ergebnis der Sanger-Sequenzierung von *GLI3* Exon 2-4 des Trios von IP GLI3-I17. Es zeigen sich je drei Banden der jeweiligen Probanden. Im Falle der IP erkennt man das Vorliegen zwei unterschiedlich langer Amplifikate. Der Wildtyp des Allels (obere Bande), welcher auch bei Vater und Mutter homozygot vorliegt, besitzt eine Länge von 372bp. Darunter findet man bei der IP das von der Deletion betroffene Allel (untere Bande) mit einer Länge von 242bp. Ansatz Nummer eins des IP und drei des Vaters blieben ohne Amplifikat. L: Leiter. *eigene Darstellung*

Anhand von GLI3-I17 lassen sich Stärken und Schwächen des WGS verdeutlichen. Einerseits können Varianten identifiziert werden, die durch eine genetische Standarddiagnostik, wie eine Sanger-Sequenzierung mit einem festen Primer-Design, zunächst übersehen worden waren. Dies geschieht durch den theoretisch uneingeschränkten Blick der WGS sowohl auf die Art als auch auf die Größe und Lokalisation der Varianten. In diesem Spezifikum liegt jedoch gleichzeitig auch eine Schwäche, da die Anzahl der Varianten, die pro Genom generiert werden, durch eine Person im klinischen Setting nicht sinnvoll auswertbar sind. Das bedeutet, dass Filter benötigt werden, um die Anzahl der Varianten zu verringern und die manuelle Auswertung zu ermöglichen. Zwar kann man so bei korrekter Einstellung die Wahrscheinlichkeit, eine pathogene Variante zu finden, erhöhen, allerdings zeigt das oben genannte Beispiel, dass durchaus auch pathogene Varianten herausgefiltert und somit übersehen werden können.

Darüber hinaus stellt die im Falle von Varianten in *GLI3* beschriebene variable Expressivität und Penetranz des Phänotyps eine weitere Herausforderung im Zusammenhang mit WGS dar. Zwar bestehen bei *GLI3* gewisse Genotyp-Phänotyp-Korrelationen, diese sind allerdings nur im Falle bestimmter Arten von Varianten und Lokalisationen eindeutig. Manche Faktoren, wie epigenetische Effekte, welche die Genotyp-Phänotyp-Korrelation beeinflussen, werden vom WGS nicht erfasst und können somit nicht in die Auswertung eingeschlossen werden (Kingdom & Wright, 2022). Um einen gesamtheitlichen Überblick

zu vervollständigen, bedarf es weiterer Forschung und Optimierung von Auswertungspipelines.

#### 4.7 Final-exon frameshift Varianten in *HMGB1* lösen BPTAS aus

Bei der Analyse des Locus von *HMGB1* konnten insgesamt dreizehn Individuen mit BPTAS untersucht werden. Sechs hiervon wurden durch das IMG sequenziert und analysiert, die übrigen wurden aus publizierter Literatur in die Betrachtung mit einbezogen. Alle Individuen zeigten einen charakteristischen Phänotyp mit skelettalen, kraniofazialen, neurologischen und urogenitalen Komponenten (vgl. extended data fig. 1; (Mensah et al., 2023)).

Bei der Analyse zeigten vor allem die Fehlbildungen der Extremitäten eine auffällige Ähnlichkeit. So präsentierten alle elf Individuen, bei denen eine entsprechend detaillierte Phänotypisierung stattfinden konnte, tibiale A- oder Hypoplasien sowie präaxiale Polysyndaktylien und Kontrakturen großer Gelenke. Darüber hinaus zeigte sich bei allen ein unterschiedliches Ausmaß der Fehlbildungen zwischen unterer und oberer Extremität, wobei die untere Extremität schwerwiegender betroffen war.

Der digital-akrale Phänotyp bedarf hierbei besonderer Aufmerksamkeit. Neben einer Brachydaktylie zeigten sich hypoplastische beziehungsweise lateral dislozierte, rudimentäre Nägel, welche, zusammen mit einer Vergrößerung der distalen Phalangen, als Störung der dorso-ventralen Strukturierung der Hand gewertet wurden. Radiologische Aufnahmen der Individuen *HMGB1*-I1, -I2, -I4 und -I5 zeigten eine ähnliche Form der Brachydaktylie. In Kombination mit den radiologischen Aufnahmen der Individuen, welche aus bestehender Literatur in die Analyse eingeschlossen wurden, konnten die Fehlbildungen der Brachydaktylie A1 zugeordnet werden. Diese werden durch Mutationen in *IHH* ausgelöst (Gao & He, 2004), wobei jedoch keine Verbindung zwischen *IHH* und *HMGB1* in STRING (v.11.0; Szklarczyk et al., 2019) nachgewiesen werden konnte.

Neben den beschriebenen charakteristischen Fehlbildungen konnte durch die Auswertung der kraniofazialen Eigenschaften und das Erstellen eines durchschnittlichen fazialen Phänotyps von BPTAS mittels F2GR eine Ähnlichkeit zwischen den eingeschlossenen Individuen mit BPTAS festgestellt werden. Im Vordergrund standen hierbei fehlgebildete



Mutationsfrequenzen von *HMGB1* lassen statistisch (pLI 0,82; %HI-Score 4,78; LOEUF-Score 0,47; sHet-Score 0,136) vermuten, dass ein Funktionsverlust des Proteins nicht toleriert wird (Karczewski et al., 2020; Lek et al., 2016; Weghorn et al., 2019). Für das gesamte Gen *HMGB1* sind insgesamt 1191 Varianten in gnomAD gelistet, wovon 55 den sauren Schwanz betreffen. Allerdings verändern diese kaum die Asp-Glu Struktur, der Schwanz wird durch die beschriebenen Varianten maximal um fünf Aminosäuren verkürzt. Zusammen mit der hohen Konservierung der sauren Endregion in Wirbeltieren lässt dies auf die Bedeutsamkeit der Integrität derselben schließen. Darüber hinaus klassifiziert MutationTaster die Variante als pathogen. All das legt nahe, dass Varianten in *HMGB1*, die den sauren Schwanz maßgeblich modulieren und bei mehreren Patient\*innen mit einem charakteristischen, seltenen Phänotyp identifiziert wurden, als pathogen klassifiziert werden müssen.

#### 4.8 Genotyp-Phänotyp-Korrelationen von *HMGB1*

Neben den zwei beschriebenen final-exon frameshift Varianten, konnte eine Mikrodeletion bei *HMGB1*-I6 identifiziert werden. Der Phänotyp des Patienten imponierte mit einer geistigen Entwicklungsverzögerung sowie einer Microcephaly, einer Klinodaktylie des fünften Fingers und Syndaktylien der zweiten und dritten Zehe. Dies deckt sich mit der Diagnose eines 13q12.3 Deletions-Syndroms, dessen kritische Region vor kurzem auf den Bereich chr13:31,021,687–31,042,716, welcher *HMGB1* enthält, reduziert werden konnte. Die Variante trat bei *HMGB1*-I6 *de novo* auf. All dies unterstützt die Annahme von Uguen *et al.*, die *HMGB1* als entscheidendes Element des 13q12.3 Deletions-Syndroms definieren (Uguen et al., 2021).

Neben den beschriebenen Individuen mit final-exon frameshift Varianten von *HMGB1*, zeigt auch das Individuum, welches die Mikrodeletion trägt, einen eigenen, differenzierbaren Phänotyp. Dies legt die Vermutung nahe, dass verschiedene Varianten von *HMGB1* (LoF Varianten vs. final-exon frameshifts) verschiedene Phänotypen auslösen. Der Grund hierfür lässt sich in den unterschiedlichen Pathomechanismen annehmen. Während frameshift Varianten des letzten Exons den sauren Schwanz von *HMGB1* und damit die phasentrennenden Eigenschaften des Proteins ändern, resultieren Mikrodele-

tionen und LoF Varianten in einem nonsense mediated decay. Dadurch werden fehlerhafte oder unvollständige Proteine abgebaut, sodass keine Proteinfunktion des mutierten Allels vorhanden ist. Untersuchungen von Mensah *et al.* legen nahe, dass es im Falle von frameshift Mutationen des letzten Exons nicht zu einem NMD kommt (Mensah *et al.*, 2023).

Daraus lässt sich hypothetisieren, dass Varianten, welche die HMG-Boxen des Proteins beeinflussen, einen systemischen Einfluss auf das Individuum in Form von geistiger Entwicklungsverzögerung haben. Final-exon frameshift Varianten, welche nur den sauren Schwanz des Proteins verändern, scheinen hauptsächlich ektope bzw. quantitative Auswirkungen zu besitzen, sodass es zum Phänotyp BPTAS kommt (Mensah *et al.*, 2023). All dies legt eine Genotyp-Phänotyp-Korrelation nahe, deren genaue Aspekte zukünftige Forschung spezifizieren sollte.

#### **4.9 Veränderte Proteinfunktion von HMGB1**

Die Ergebnisse von Mensah *et al.* zeigen, dass die C-terminale saure Domäne (intrinsisch unorganisierter Asp/Glu-Repeat) von HMGB1 eine wichtige Rolle bei der Lokalisation des Proteins im Zellkern spielt. Wildtyp-HMGB1 verteilt sich im gesamten Zellkern, während sich mutiertes HMGB1, dessen C-Terminale Domäne durch einen größtenteils basischen Abschnitt ersetzt wurde, in der äußeren Phase des Nukleolus wiederfindet. Die Mutation verändert also nachweislich die Phasentrennungseigenschaften von HMGB1. Derartige frameshift-induzierte Austausch intrinsisch ungeordneter Proteinregionen scheinen ein neuer Pathomechanismus zu sein, der eine Reihe bereits bekannter und möglicherweise noch unbekannter pathogener Mutationen funktionell erklärt (Mensah *et al.*, 2023; X. Wang *et al.*, 2023).

#### **4.10 WGS ermöglicht Auffinden neuer Varianten**

Im Falle der hier dargestellten Analyse konnte durch das Auffinden einer potenziell pathogenen Variante im WGS eine Reihe von Erkenntnissen gewonnen werden. Die Variante sowie der entsprechende Pathomechanismus, waren bisher weder in der Literatur beschrieben noch durch herkömmliche genetische Diagnostik identifiziert worden. Der

unvoreingenommene Blick des WGS ermöglichte das Auffinden der Variante von HMGB1-I5 und legte somit den Grundstein für die weitere Analyse. Daraus resultierte die Neubeschreibung einer Gen-Syndrom-Assoziation (HMGB1 - BPTAS) sowie die Untersuchung weiterer Gene beziehungsweise Proteine, welche eine ähnliche Struktur aufwiesen. Darüber hinaus konnten weitere Erkenntnisse bezüglich der IDRs von Proteinen und der damit verbundenen Kondensationsphysiologie innerhalb des Nucleolus gewonnen werden. Durch WGS konnte also die Basis geschaffen werden, von der aus sich neue Ergebnisse in der genetischen Forschung gewinnen ließen. Man kann vermuten, dass mit der Weiterentwicklung der Methodik des WGS und Verbesserung von Analysesoftware und -prozessen viele solcher Ausgangspunkte geschaffen werden können, von denen aus neues Wissen gewonnen werden kann.

## 5. Schlussfolgerungen

In der vorliegenden Arbeit konnte gezeigt werden, dass durch die Analyse der Genomdaten von 69 Patient\*innen, die bereits herkömmliche molekular-genetische Untersuchungen durchlaufen hatten, in 12 Fällen (17,4%) die Identifikation einer pathogenen Variante gelang. Hierbei handelt es sich um neun SNVs, eine komplexe SV, eine Repeat-Verlängerung und eine komplexe Translokation. Darüber hinaus konnten die durch das WGS gewonnenen Erkenntnisse als Grundlage verwendet werden, von der aus zwei weitere Forschungsprojekte konzipiert und durchgeführt wurden. Dies ermöglichte die genauere Bewertung von zwei Genloci sowie die Beschreibung eines unbekanntes Pathomechanismus (frameshift-induced-IDR-swapping).

Der vorgestellte Ansatz des WGS besitzt die dargestellten Limitationen, sowohl von Seiten der Software als auch in Bezug auf die manuelle Analyse. Allerdings wird die Analyse der Genomdaten, durch Weiterentwicklung von Calling-Algorithmen und Sequenziermethoden sowie durch neue Ansätze für die Analyse von nicht-kodierenden Bereichen oder komplexen Regulationsmechanismen in Zukunft weiter gesteigert werden.

WGS sollte in der alltäglichen klinischen Diagnostik von genetischen Krankheiten etabliert werden, da es ermöglicht, Varianten unterschiedlichster Art aufzufinden, auch solche, die durch herkömmliche molekular-genetische Diagnostik nicht identifiziert werden können.

## Literaturverzeichnis

- Abe, K. T., Rizzo, I. M. P. O., Coelho, A. L. V., Sakai, N., Jr, Carvalho, D. R., & Speck-Martins, C. E. (2018). 19q13.11 microdeletion: Clinical features overlapping ectrodactyly ectodermal dysplasia-clefting syndrome phenotype. *Clinical Case Reports*, *6*(7), 1300–1307.
- 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. *Nature Methods*, *7*(4), 248–249.
- Aerden, M., Bauters, M., Van Den Bogaert, K., Vermeesch, J. R., Holvoet, M., Plasschaert, F., & Devriendt, K. (2020). Genotype-phenotype correlations of UBA2 mutations in patients with ectrodactyly. *European Journal of Medical Genetics*, *63*(11), 104009.
- Al-Qattan, M. M., Shamseldin, H. E., Salih, M. A., & Alkuraya, F. S. (2017). GLI3-related polydactyly: a review. *Clinical Genetics*, *92*(5), 457–466.
- Anggayasti, W. L., Ogino, K., Yamamoto, E., Helmerhorst, E., Yasuoka, K., & Mancera, R. L. (2020). The acidic tail of HMGB1 regulates its secondary structure and conformational flexibility: A circular dichroism and molecular dynamics simulation study. In *Computational and Structural Biotechnology Journal* (Vol. 18, pp. 1160–1172). <https://doi.org/10.1016/j.csbj.2020.05.012>
- Antaki, D., Brandler, W. M., & Sebat, J. (2018). SV2: accurate structural variation genotyping and de novo mutation detection from whole genomes. *Bioinformatics*, *34*(10), 1774–1777.
- Basel, D., DePaepe, A., Kilpatrick, M. W., & Tsipouras, P. (2003). Split hand foot malformation is associated with a reduced level of Dactylin gene expression. *Clinical Genetics*, *64*(4), 350–354.
- Belgrano, F. S., de Abreu da Silva, I. C., Bastos de Oliveira, F. M., Fantappiè, M. R., & Mohana-Borges, R. (2013). Role of the acidic tail of high mobility group protein B1 (HMGB1) in protein stability and DNA bending. *PloS One*, *8*(11), e79572.
- Berisha, S. Z., Shetty, S., Prior, T. W., & Mitchell, A. L. (2020). Cytogenetic and molecular diagnostic testing associated with prenatal and postnatal birth defects. *Birth Defects Research*, *112*(4), 293–306.
- Brison, N., Debeer, P., & Tylzanowski, P. (2014). Joining the fingers: a HOXD13 Story. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, *243*(1), 37–48.
- Bult, C. J., Blake, J. A., Smith, C. L., Kadin, J. A., Richardson, J. E., & Mouse Genome Database Group. (2019). Mouse Genome Database (MGD) 2019. *Nucleic Acids Research*, *47*(D1), D801–D806.
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C., & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*(7745), 496–502.
- Cheung, S. W., & Bi, W. (2018). Novel applications of array comparative genomic hybridization in molecular diagnostics. *Expert Review of Molecular Diagnostics*, *18*(6), 531–542.
- Chokdeemboon, C., Mahatumarat, C., Rojvachiranonda, N., Tongkobetch, S., Suphapeetiporn, K., & Shotelersuk, V. (2013). FGFR1 and FGFR2 mutations in Pfeiffer syndrome. *The Journal of Craniofacial Surgery*, *24*(1), 150–152.
- Chun, K., Teebi, A. S., Jung, J. H., Kennedy, S., Laframboise, R., Meschino, W. S., Nakabayashi, K., Scherer, S. W., Ray, P. N., & Teshima, I. (2002). Genetic analysis of patients with the Saethre-Chotzen phenotype. *American Journal of Medical Genetics*, *110*(2), 136–143.

- Cooper, G. M., & Brown, C. D. (2008). Qualifying the relationship between sequence conservation and molecular function. *Genome Research*, *18*(2), 201–205.
- Cooper, G. M., Stone, E. A., Asimenos, G., NISC Comparative Sequencing Program, Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, *15*(7), 901–913.
- Cotney, J., Leng, J., Yin, J., Reilly, S. K., DeMare, L. E., Emera, D., Ayoub, A. E., Rakic, P., & Noonan, J. P. (2013). The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell*, *154*(1), 185–196.
- Démurger, F., Ichkou, A., Mougou-Zerelli, S., Le Merrer, M., Goudefroye, G., Delezoide, A.-L., Quélin, C., Manouvrier, S., Baujat, G., Fradin, M., Pasquier, L., Megarbané, A., Faivre, L., Baumann, C., Nampoothiri, S., Roume, J., Isidor, B., Lacombe, D., Delrue, M.-A., Mercier, S., Philip, N., Schaefer, E., Holder, M., Krause, A., Laffargue, F., Sinico, M., Amram, D., André, G., Liquier, A., Rossi, M., Amiel, J., Giuliano, F., Boute, O., Dieux-Coeslier, A., Jacquemont, M.L., Afenjar, A., Van Maldergem, L., Lackmy-Port-Lis, M., Vincent-Delorme, C., Chauvet, M. L., Cormier-Daire, V., Devisme, L., Geneviève, D., Munnich, A., Viot, G., Raoul, O., Romana, S., Gonzales, M., Encha-Razavi, F., Odent, S., Vekemans, M., Attie-Bitach, T. (2015). New insights into genotype–phenotype correlation for GLI3 mutations. *European Journal of Human Genetics: EJHG*, *23*(1), 92.
- Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M., & Bérout, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, *37*(9), e67.
- Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., Baker, C.N., Bower, L., Brown, J.M., Caddle, L.B., Chiani, F., Clary, D., Cleak, J., Daly, M.J., Denegre, J.M., Doe, B., Dolan, M.E., Edie Helmut Fuchs, S.M., Gailus-Durner, V., Galli, A., Gambadoro, A., Gallegos, J., Guo, S., Horner, N.R., Hsu, C.W., Johnson, S.J., Kalaga, S., Keith, L.C., Lanoue, L., Lawson, T.N., Lek, M., Mark, M., Marschall, S., Mason, J., McElwee, M.L., Nutter, S.N.L.M.J., Peterson, K.A., Ramirez-Solis, R., Rowland, D.J., Ryder, E., Samocha, K.E., Seavitt, J.R., Selloum, M., Szoke-Kovacs, Z., Tamura, M., Trainor, A.G., Tudose, I., Wakana, S., Warren, J., Wendling, O., West, D.B., Wong, L., Yoshiki, A.; International Mouse Phenotyping Consortium, Wurst, W., MacArthur, D.G., Tocchini-Valentini, G.P., Gao, X., Flicek, P., Bradley, A., Skarnes, W.C., Justice, M.J., Parkinson, H.E., Moore, M., Wells, S., Braun, R.E., Svenson, K.L., de Angelis, M.H., Herault, Y., Mohun, T., Mallon, A.M., Henkelman, R.M., Brown, S.D.M., Adams, D.J., Lloyd, K.C.K., McKerlie, C., Beaudet, A.L. & Murray, M.B.S.A. (2017). Corrigendum: High-throughput discovery of novel developmental phenotypes. *Nature*, *551*(7680), 398.
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380.
- Elliott, A. M., & Evans, J. A. (2006). Genotype–phenotype correlations in mapped split hand foot malformation (SHFM) patients. *American Journal of Medical Genetics. Part A*, *140A*(13), 1419–1427.
- Elsner, J., Mensah, M. A., Holtgrewe, M., Hertzberg, J., Bigoni, S., Busche, A., Coutelier, M., de Silva, D. C., Elçioğlu, N., Filges, I., Gerkes, E., Girisha, K. M., Graul-Neumann, L., Jamsheer, A., Krawitz, P., Kurth, I., Markus, S., Megarbane, A., Reis, A., Reuter, M.S., Svoboda, D., Teller, C., Tuysuz, B., Türkmen, S., Wilson, M., Woitschach, R., Vater, I., Caliebe, A., Hülsemann, W., Horn, D., Mundlos, S. & Spielmann, M. (2021). Genome sequencing in families with congenital limb malformations. *Human Genetics*, *140*(8), 1229–1239.

- Flöttmann, R., Kragesteen, B. K., Geuer, S., Socha, M., Allou, L., Sowińska-Seidler, A., Bosquillon de Jarcy, L., Wagner, J., Jamsheer, A., Oehl-Jaschkowitz, B., Wittler, L., de Silva, D., Kurth, I., Maya, I., Santos-Simarro, F., Hülsemann, W., Klopocki, E., Mountford, R., Fryer, A., Borck, G., Horn, D., Lapunzina, P., Wilson, M., Mascrez, B., Duboule, D., Mundlos, S. & Spielmann, M. (2018). Noncoding copy-number variations are associated with congenital limb malformation. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 20(6), 599–607.
- Garland, M. A., Sun, B., Zhang, S., Reynolds, K., Ji, Y., & Zhou, C. J. (2020). Role of epigenetics and miRNAs in orofacial clefts. *Birth Defects Research*.  
<https://doi.org/10.1002/bdr2.1802>
- Gao, B., & He, L. (2004). Answering a century old riddle: brachydactyly type A1. *Cell Research*, 14(3), 179–187.
- Govindan, J., Tun, K. M., & Iovine, M. K. (2016). Cx43-Dependent Skeletal Phenotypes Are Mediated by Interactions between the Hapln1a-ECM and Sema3d during Fin Regeneration. *PloS One*, 11(2), e0148202.
- Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*, 33(suppl\_1), D514–D517.
- Hehir-Kwa, J. Y., Pfundt, R., & Veltman, J. A. (2015). Exome sequencing and whole genome sequencing for the detection of copy number variation. In *Expert Review of Molecular Diagnostics* (Vol. 15, Issue 8, pp. 1023–1032). <https://doi.org/10.1586/14737159.2015.1053467>
- Holder-Espinasse, M., Jamsheer, A., Escande, F., Andrieux, J., Petit, F., Sowinska-Seidler, A., Socha, M., Jakubiuk-Tomaszuk, A., Gerard, M., Mathieu-Dramard, M., Cormier-Daire, V., Verloes, A., Toutain, A., Plessis, G., Jonveaux, P., Baumann, C., David, A., Farra, C., Colin, E., Jacquemont, S., Rossi, A., Mansour, S., Ghali, N., Moncla, A., Lahiri, N., Hurst, J., Pollina, E., Patch, C., Ahn, J.W., Valat, A.S., Mezel, A., Bourgeot, P., Zhang, D. & Manouvrier-Hanu, S. (2019). Duplication of 10q24 locus: broadening the clinical and radiological spectrum. *European Journal of Human Genetics: EJHG*, 27(4), 525–534.
- Holtgrewe, M., Stolpe, O., Nieminen, M., Mundlos, S., Knaus, A., Kornak, U., Seelow, D., Segebrecht, L., Spielmann, M., Fischer-Zirnsak, B., Boschann, F., Scholl, U., Ehmke, N., & Beule, D. (2020). VarFish: comprehensive DNA variant analysis for diagnostics and research. *Nucleic Acids Research*, 48(W1), W162–W169.
- Hombach, D., Schuelke, M., Knierim, E., Ehmke, N., Schwarz, J. M., Fischer-Zirnsak, B., & Seelow, D. (2019). MutationDistiller: user-driven identification of pathogenic DNA variants. In *Nucleic Acids Research* (Vol. 47, Issue W1, pp. W114–W120).  
<https://doi.org/10.1093/nar/gkz330>
- International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. In *Nature* (Vol. 409, Issue 6822, pp. 860–921).  
<https://doi.org/10.1038/35057062>
- International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931–945.
- Itou, J., Taniguchi, N., Oishi, I., Kawakami, H., Lotz, M., & Kawakami, Y. (2011). HMGB factors are required for posterior digit development through integrating signaling pathway activities. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 240(5), 1151–1162.
- Jäger, M., Wang, K., Bauer, S., Smedley, D., Krawitz, P., & Robinson, P. N. (2014). Jannovar: a java library for exome annotation. *Human Mutation*, 35(5), 548–555.

- Jamsheer, A., Sowińska, A., Trzeciak, T., Jamsheer-Bratkowska, M., Geppert, A., & Latos-Bieleńska, A. (2012). Expanded mutational spectrum of the GLI3 gene substantiates genotype–phenotype correlations. *Journal of Applied Genetics*, *53*(4), 415.
- Johnston, J. J., Olivos-Glander, I., Killoran, C., Elson, E., Turner, J. T., Peters, K. F., Abbott, M. H., Aughton, D. J., Aylsworth, A. S., Bamshad, M. J., Booth, C., Curry, C. J., David, A., Dinulos, M. B., Flannery, D. B., Fox, M. A., Graham, J. M., Grange, D. K., Guttmacher, A. E., Hannibal, M.C., Henn, W., Hennekam, R.C, Holmes, L.B., Hoyme, H.E., Leppig, K.A., Lin, A.E., Macleod, P., Manchester, D.K., Marcelis, C., Mazzanti, L., McCann, E., McDonald, M.T., Mendelsohn, N.J., Moeschler, J.B., Moghaddam, B., Neri, G., Newbury-Ecob, R., Pagon, R.A., Phillips, J.A., Sadler, L.S., Stoler, J.M., Tilstra, D., Walsh Vockley, C.M., Zackai, E.H., Zadeh, T.M., Brueton, L., Black, G.C., Biesecker, L.G. (2005). Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: robust phenotype prediction from the type and position of GLI3 mutations. *American Journal of Human Genetics*, *76*(4), 609–622.
- Johnston, J. J., Sapp, J. C., Turner, J. T., Amor, D., Aftimos, S., Aleck, K. A., Bocian, M., Bodurtha, J. N., Cox, G. F., Curry, C. J., Day, R., Donnai, D., Field, M., Fujiwara, I., Gabbett, M., Gal, M., Graham, J. M., Hedera, P., Hennekam, R. C. M., Hersh, J. H., Hopkin, R. J., Kayserili, H., Kidd, A. M., Kimonis, V., Lin, A. E., Lynch, S. A., Maisenbacher, M., Mansour, S., McGaughran, J., Mehta, L., Murphy, H., Raygada, M., Robin, N. H., Rope, A. F., Rosenbaum, K. N., Schaefer, G. B., Shealy, A., Smith, W., Soller, M., Sommer, A., Stalker, H. J., Steiner, B., Stephan, M. J., Tilstra, D., Tomkins, S., Trapane, P., Tsai, A. C., Van Allen, M. I., Vasudevan, P. C., Zabel, B., Zunich, J., Black, G. C., Biesecker, L. G. (2010). Molecular analysis expands the spectrum of phenotypes associated with GLI3 mutations. *Human Mutation*, *31*(10), 1142–1154.
- Jones, M. J. K., & Jallepalli, P. V. (2012). Chromothripsis: chromosomes in crisis. *Developmental Cell*, *23*(5), 908–917.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.
- Kalff-Suske, M., Wild, A., Topp, J., Wessling, M., Jacobsen, E. M., Bornholdt, D., Engel, H., Heuer, H., Aalfs, C. M., Ausems, M. G., Barone, R., Herzog, A., Heutink, P., Homfray, T., Gillessen-Kaesbach, G., König, R., Kunze, J., Meinecke, P., Müller, D., Rizzo, R., Strenge, S., Superti-Furga, A., Grzeschik, K. H. (1999). Point mutations throughout the GLI3 gene cause Greig cephalopolysyndactyly syndrome. *Human Molecular Genetics*, *8*(9). <https://doi.org/10.1093/hmg/8.9.1769>
- Kalia, S. S., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., Evans, J. P., Herman, G. E., Hufnagel, S. B., Klein, T. E., Korf, B. R., McKelvey, K. D., Ormond, K. E., Richards, C. S., Vlangos, C. N., Watson, M., Martin, C. L., & Miller, D. T. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *19*(2). <https://doi.org/10.1038/gim.2016.190>

- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G. & Talkowski, M.E.; Genome Aggregation Database Consortium, Neale, B.M., Daly, M.J., MacArthur, D.G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443.
- Karousis, E. D., & Mühlemann, O. (2019). Nonsense-Mediated mRNA Decay Begins Where Translation Ends. *Cold Spring Harbor Perspectives in Biology*, *11*(2). <https://doi.org/10.1101/cshperspect.a032862>
- 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. In *Genome Research* (Vol. 12, Issue 6, pp. 996–1006). <https://doi.org/10.1101/gr.229102>.
- Kingdom, R., & Wright, C. F. (2022). Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts. *Frontiers in Genetics*, *13*, 920390.
- Knapp, S., Müller, S., Digilio, G., Bonaldi, T., Bianchi, M. E., & Musco, G. (2004). The Long Acidic Tail of High Mobility Group Box 1 (HMGB1) Protein Forms an Extended and Flexible Structure That Interacts with Specific Residues within and between the HMG Boxes†. In *Biochemistry* (Vol. 43, Issue 38, pp. 11992–11997). <https://doi.org/10.1021/bi049364k>
- Köhler, S., Vasilevsky, N. A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., Baynam, G., Bello, S. M., Boerkoel, C. F., Boycott, K. M., Brudno, M., Buske, O. J., Chinnery, P. F., Cipriani, V., Connell, L. E., Dawkins, H. J. S., DeMare, L. E., Devereau, A. D., de Vries, B. B. A., Firth, H.V., Freson, K., Greene, D., Hamosh, A., Helbig, I., Hum, C., Jähn, J.A., James, R., Krause, R., Laudederkind, S.J.F., Lochmüller, H., Lyon, G.J., Ogishima, S., Olry, A., Ouwehand, W.H., Pontikos, N., Rath, A., Schaefer, F., Scott, R.H., Segal, M., Sergouniotis, P.I., Sever, R., Smith, C.L., Straub, V., Thompson, R., Turner, C., Turro, E., Veltman, M.W., Vulliamy, T., Yu, J., von Ziegenweidt, J., Zankl, A., Züchner, S., Zemojtel, T., Jacobsen, J.O., Groza, T., Smedley, D., Mungall, C.J., Haendel, M. & Robinson, P.N. (2017). The Human Phenotype Ontology in 2017. *Nucleic Acids Research*, *45*(D1), D865–D876.
- Kroeldrup, L., Kjaergaard, S., Kirchhoff, M., Kock, K., Brasch-Andersen, C., Kibaek, M., & Ousager, L. B. (2012). Duplication of 7q36.3 encompassing the Sonic Hedgehog (SHH) gene is associated with congenital muscular hypertrophy. *European Journal of Medical Genetics*, *55*(10), 557–560.
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. In *Cell* (Vol. 172, Issue 4, pp. 650–665). <https://doi.org/10.1016/j.cell.2018.01.029>
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J. B., Kattman, B. L., Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), D1062–D1067.

- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J. & MacArthur, D.G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285–291.
- Letunic, I., Khedkar, S., & Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Research*, 49(D1), D458–D460.
- Lohan, S., Spielmann, M., Doelken, S. C., Flöttmann, R., Muhammad, F., Baig, S. M., Wajid, M., Hülsemann, W., Habenicht, R., Kjaer, K. W., Patil, S. J., Girisha, K. M., Abarca-Barriga, H. H., Mundlos, S., & Klopocki, E. (2014). Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clinical Genetics*, 86(4), 318–325.
- MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L., & Scherer, S. W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research*, 42(Database issue), D986–D992.
- Malik, S., Percin, F. E., Bornholdt, D., Albrecht, B., Percesepe, A., Koch, M. C., Landi, A., Fritz, B., Khan, R., Mumtaz, S., Akarsu, N. A., & Grzeschik, K.-H. (2014). Mutations affecting the BHLHA9 DNA-binding domain cause MSSD, mesoaxial synostotic syndactyly with phalangeal reduction, Malik-Percin type. *American Journal of Human Genetics*, 95(6), 649–659.
- McGinn, S., & Gut, I. G. (2013). DNA sequencing - spanning the generations. *New Biotechnology*, 30(4), 366–372.
- Mensah, M. A., Niskanen, H., Magalhaes, A. P., Basu, S., Kircher, M., Sczakiel, H. L., Reiter, A. M. V., Elsner, J., Meinecke, P., Biskup, S., Chung, B. H. Y., Dombrowsky, G., Eckmann-Scholz, C., Hitz, M. P., Hoischen, A., Holterhus, P.-M., Hülsemann, W., Kahrizi, K., Kalscheuer, V. M., Kann, A., Krumbiegel, M., Kurth, I., Leubner, J., Longardt A. C., Moritz, J. D., Najmabadi, H., Skipalova, K., Snijders Blok, L., Tzschach, A., Wiedersberg, E., Zenker, M., Garcia-Cabau, C., Buschow, R., Salvatella, X., Kraushar, M. L., Mundlos, S., Caliebe, A., Spielmann, M., Horn, D., Hnisz, D. (2023). Aberrant phase separation and nucleolar dysfunction in rare genetic diseases. *Nature*, 614(7948), 564–571.
- de Mollerat, X. J., Gurrieri, F., Morgan, C. T., Sangiorgi, E., Everman, D. B., Gaspari, P., Amiel, J., Bamshad, M. J., Lyle, R., Blouin, J. L., Allanson, J. E., Le Marec, B., Wilson, M., Braverman, N. E., Radhakrishna, U., Delozier-Blanchet, C., Abbott, A., Elghouzzi, V., Antonarakis, S., Stevenson, R. E., Munnich, A., Neri, G. & Schwartz, C.E. (2003). A genomic rearrangement resulting in a tandem duplication is associated with split hand-split foot malformation 3 (SHFM3) at 10q24. In *Human Molecular Genetics* (Vol. 12, Issue 16, pp. 1959–1971). <https://doi.org/10.1093/hmg/ddg212>
- Mundlos, S., & Horn, D. (2014). *Limb Malformations: An Atlas of Genetic Disorders of Limb Development*. Springer.

- Newman, S. A., Glimm, T., & Bhat, R. (2018). The vertebrate limb: An evolving complex of self-organizing systems. *Progress in Biophysics and Molecular Biology*.  
<https://doi.org/10.1016/j.pbiomolbio.2018.01.002>
- Ng, P. C., & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Research*, 11(5). <https://doi.org/10.1101/gr.176601>
- Niehus, S., Jónsson, H., Schönberger, J., Björnsson, E., Beyter, D., Eggertsson, H. P., Sulem, P., Stefánsson, K., Halldórsson, B. V., & Kehr, B. (2021). PopDel identifies medium-size deletions simultaneously in tens of thousands of genomes. *Nature Communications*, 12(1), 730.
- Nieminen, M., Stolpe, O., Schumann, F., Holtgrewe, M., & Beule, D. (2020). SODAR Core: a Django-based framework for scientific data management and analysis web apps. In *Journal of Open Source Software* (Vol. 5, Issue 55, p. 1584). <https://doi.org/10.21105/joss.01584>
- Nucleotide sequence of bacteriophage  $\lambda$  DNA. (1982). *Journal of Molecular Biology*, 162(4), 729–773.
- Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., & Petersen, G.B. (1982). Nucleotide sequence of bacteriophage  $\lambda$  DNA. In *Journal of Molecular Biology*, 162(4), 729–773.
- Olsen, S. K., Capili, A. D., Lu, X., Tan, D. S., & Lima, C. D. (2010). Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature*, 463(7283), 906–912.
- Pantel, J. T., Hajjir, N., Danyel, M., Elsner, J., Abad-Perez, A. T., Hansen, P., Mundlos, S., Spielmann, M., Horn, D., Ott, C.-E., & Mensah, M. A. (2020). Efficiency of Computer-Aided Facial Phenotyping (DeepGestalt) in Individuals With and Without a Genetic Syndrome: Diagnostic Accuracy Study. In *Journal of Medical Internet Research* (Vol. 22, Issue 10, p. e19263). <https://doi.org/10.2196/19263>
- Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., & Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, 444(7118), 499–502.
- Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. *Nature Reviews. Genetics*, 14(4), 288–295.
- Petersen, B.-S., Fredrich, B., Hoepfner, M. P., Ellinghaus, D., & Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genetics*, 18(1), 14.
- Phadke, S. R., Kar, A., Bhowmik, A. D., & Dalal, A. (2016). Complex Camptosynpolydactyly and Mesoaxial synostotic syndactyly with phalangeal reduction are allelic disorders. *American Journal of Medical Genetics. Part A*, 170(6), 1622–1625.
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, 20(1), 110–121.
- Prokop, J. W., May, T., Strong, K., Bilinovich, S. M., Bupp, C., Rajasekaran, S., Worthey, E. A., & Lazar, J. (2018). Genome sequencing in the clinic: the past, present, and future of genomic medicine. *Physiological Genomics*, 50(8), 563–579.
- Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28(18), i333–i339.
- Rentsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886–D894.

- 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., & ACMG Laboratory Quality Assurance Committee. (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. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 17(5), 405–424.
- Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., & Mesirov, J. P. (2017). Variant Review with the Integrative Genomics Viewer. In *Cancer Research* (Vol. 77, Issue 21, pp. e31–e34). <https://doi.org/10.1158/0008-5472.can-17-0337>
- Robinson, P. N., Kohler, S., Oellrich, A., Wang, K., Mungall, C. J., Lewis, S. E., Washington, N., Bauer, S., Seelow, D., Krawitz, P., Gilissen, C., Haendel, M., Smedley, D., & Sanger Mouse Genetics Project. (2014). Improved exome prioritization of disease genes through cross-species phenotype comparison. In *Genome Research* (Vol. 24, Issue 2, pp. 340–348). <https://doi.org/10.1101/gr.160325.113>
- Rodriguez-Mateos, P., Azevedo, N. F., Almeida, C., & Pamme, N. (2020). FISH and chips: a review of microfluidic platforms for FISH analysis. *Medical Microbiology and Immunology*. <https://doi.org/10.1007/s00430-019-00654-1>
- Russo, C. D., Di Giacomo, G., Cignini, P., Padula, F., Mangiafico, L., Mesoraca, A., D'Emidio, L., McCluskey, M. R., Paganelli, A., & Giorlandino, C. (2014). Comparative study of aCGH and Next Generation Sequencing (NGS) for chromosomal microdeletion and microduplication screening. *Journal of Prenatal Medicine*, 8(3-4), 57–69.
- Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., Sirota-Madi, A., Olender, T., Golan, Y., Stelzer, G., Harel, A., & Lancet, D. (2010). GeneCards Version 3: the human gene integrator. In *Database* (Vol. 2010, Issue 0, pp. baq020–baq020). <https://doi.org/10.1093/database/baq020>
- Sanchez-Castro, M., Pichon, O., Briand, A., Poulain, D., Gournay, V., David, A., & Le Caignec, C. (2015). Disruption of the SEMA3D Gene in a Patient with Congenital Heart Defects. In *Human Mutation* (Vol. 36, Issue 1, pp. 30–33). <https://doi.org/10.1002/humu.22702>
- Schnur, R. E., Yousaf, S., Liu, J., Chung, W. K., Rhodes, L., Marble, M., Zambrano, R. M., Sobreira, N., Jayakar, P., Pierpont, M. E., Schultz, M. J., Pichurin, P. N., Olson, R. J., Graham, G. E., Osmond, M., Contreras-García, G. A., Campo-Neira, K. A., Peñaloza-Mantilla, C. A., Flage, M., Kuppa, S., Navarro, K., Sacoto, M.J.G., Wentzensen, I.M., Scarano, M.I., Juusola, J., Prada, C.E. & Hufnagel, R.B. (2021). UBA2 variants underlie a recognizable syndrome with variable aplasia cutis congenita and ectrodactyly. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 23(9), 1624–1635.
- Schwarze, K., Buchanan, J., Fermont, J. M., Dreau, H., Tilley, M. W., Taylor, J. M., Antoniou, P., Knight, S. J. L., Camps, C., Pentony, M. M., Kvikstad, E. M., Harris, S., Popitsch, N., Pagnamenta, A. T., Schuh, A., Taylor, J. C., & Wordsworth, S. (2019). The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 22(1), 85–94.
- Schwarz, J. M., Hombach, D., Köhler, S., Cooper, D. N., Schuelke, M., & Seelow, D. (2019). RegulationSpotter: annotation and interpretation of extratranscriptomic DNA variants. *Nucleic Acids Research*, 47(W1), W106–W113.
- Schwarz, J. M., Rödelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. In *Nature Methods* (Vol. 7, Issue 8, pp. 575–576). <https://doi.org/10.1038/nmeth0810-575>

- Sczakiel, H. L., Hülsemann, W., Holtgrewe, M., Abad-Perez, A. T., Elsner, J., Schwartzmann, S., Horn, D., Spielmann, M., Mundlos, S., & Mensah, M. A. (2021). GLI3 variants causing isolated polysyndactyly are not restricted to the protein's C-terminal third. *Clinical Genetics*, *100*(6), 758–765.
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature*, *550*(7676), 345–353
- Shi, Y., Zhang, B., Kong, F., & Li, X. (2018). Prenatal limb defects: Epidemiologic characteristics and an epidemiologic analysis of risk factors. *Medicine*, *97*(29), e11471.
- Shou, S., Carlson, H. L., Perez, W. D., & Stadler, H. S. (2013). HOXA13 regulates Aldh1a2 expression in the autopod to facilitate interdigital programmed cell death. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, *242*(6), 687–698.
- Sousa, S. B., Abdul-Rahman, O. A., Bottani, A., Cormier-Daire, V., Fryer, A., Gillessen-Kaesbach, G., Horn, D., Josifova, D., Kuechler, A., Lees, M., MacDermot, K., Magee, A., Morice-Picard, F., Rosser, E., Sarkar, A., Shannon, N., Stolte-Dijkstra, I., Verloes, A., Wakeling, E., Wilson, L. & Hennekam, R.C. (2009). Nicolaides-Baraitser syndrome: Delineation of the phenotype. *American Journal of Medical Genetics. Part A*, *149A*(8), 1628–1640.
- Spielmann, M., Lupiáñez, D. G., & Mundlos, S. (2018). Structural variation in the 3D genome. In *Nature Reviews Genetics* (Vol. 19, Issue 7, pp. 453–467). <https://doi.org/10.1038/s41576-018-0007-0>
- Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., Abeyasinghe, S., Krawczak, M., & Cooper, D. N. (2003). Human Gene Mutation Database (HGMD): 2003 update. *Human Mutation*, *21*(6). <https://doi.org/10.1002/humu.10212>
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1). <https://doi.org/10.1093/nar/gky1131>
- Tanaka, M. (2013). Molecular and evolutionary basis of limb field specification and limb initiation. *Development, Growth & Differentiation*, *55*(1), 149–163.
- Turro, E., Aistle, W. J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H. L., Sanchis-Juan, A., Frontini, M., Thys, C., Stephens, J., Mapeta, R., Burren, O. S., Downes, K., Haimel, M., Tuna, S., Deevi, S. V. V., Aitman, T. J., Bennett, D. L., Calleja, P., Carss, K., Caulfield, M.J., Chinnery, P.F., Dixon, P.H., Gale, D.P., James, R., Koziell, A., Laffan, M.A., Levine, A.P., Maher, E.R., Markus, H.S., Morales, J., Morrell, N.W., Mumford, A.D., Ormondroyd, E., Rankin, S., Rendon, A., Richardson, S., Roberts, I., Roy, N.B.A., Saleem, M.A., Smith, K.G.C., Stark, H., Tan, R.Y.Y., Themistocleous, A.C., Thrasher, A.J., Watkins, H., Webster, A.R., Wilkins, M.R., Williamson, C., Whitworth, J., Humphray, S., Bentley, D.R.; NIHR BioResource for the 100,000 Genomes Project, Kingston, N., Walker, N., Bradley, J.R., Ashford, S., Penkett, C.J., Freson, K., Stirrups, K.E., Raymond, F.L. & Ouwehand, W.H. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, *583*(7814), 96–102.

- Uguen, K., Krysiak, K., Audebert-Bellanger, S., Redon, S., Benech, C., Viora-Dupont, E., Tran, M.-T. F., Rondeau, S., Elsharkawi, I., Granadillo, J. L., Neidich, J., Soares, C. A., Tkachenko, N., Amudhavalli S, M., Engleman, K., Boland, A., Deleuze, J. F., Bezieau, S., Odent, S., Toutain, A., Bonneau, D., Gilbert-Dussardier, B., Faivre, L., Rio, M., Le Marchal, C., Ferec, C., Repnikova, E., Cao, Y. (2021). Heterozygous HMGB1 loss-of-function variants are associated with developmental delay and microcephaly. *Clinical Genetics*, 100(4). <https://doi.org/10.1111/cge.14015>
- Ulfig, N. (2009). *Kurzlehrbuch Embryologie: 1. Auflage*, Georg Thieme Verlag.
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. In *Trends in Genetics* (Vol. 30, Issue 9, pp. 418–426). <https://doi.org/10.1016/j.tig.2014.07.001>
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics: TIG*, 34(9), 666–681.
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., Velankar, S. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1), D439–D444.
- Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., & Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), 854–858.
- Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Research*, 35(Database issue), D88–D92.
- Wang, Q., Pierce-Hoffman, E., Cummings, B. B., Alföldi, J., Francioli, L. C., Gauthier, L. D., Hill, A. J., O'Donnell-Luria, A. H., Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Karczewski, K. J., & MacArthur, D. G. (2021). Author Correction: Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. *Nature Communications*, 12(1), 827.
- Wang, X., Bigman, L. S., Greenblatt, H. M., Yu, B., Levy, Y., & Iwahara, J. (2023). Negatively charged, intrinsically disordered regions can accelerate target search by DNA-binding proteins. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkad045>
- Weghorn, D., Balick, D. J., Cassa, C., Kosmicki, J. A., Daly, M. J., Beier, D. R., & Sunyaev, S. R. (2019). Applicability of the Mutation-Selection Balance Model to Population Genetics of Heterozygous Protein-Truncating Variants in Humans. *Molecular Biology and Evolution*, 36(8). <https://doi.org/10.1093/molbev/msz092>
- Yamoto, K., Saitou, H., Nishimura, G., Kosaki, R., Takayama, S., Haga, N., Tonoki, H., Okumura, A., Horii, E., Okamoto, N., Suzumura, H., Ikegawa, S., Kato, F., Fujisawa, Y., Nagata, E., Takada, S., Fukami, M., & Ogata, T. (2019). Comprehensive clinical and molecular studies in split-hand/foot malformation: identification of two plausible candidate genes (LRP6 and UBA2). *European Journal of Human Genetics: EJHG*, 27(12), 1845–1857.
- Zeller, R., López-Ríos, J., & Zuniga, A. (2009). Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nature Reviews. Genetics*, 10(12), 845–858.
- Zhao, X., Brade, T., Cunningham, T. J., & Duester, G. (2010). Retinoic acid controls expression of tissue remodeling genes Hmgn1 and Fgf18 at the digit-interdigit junction. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 239(2), 665–671.

- Zhu, M., Need, A. C., Han, Y., Ge, D., Maia, J. M., Zhu, Q., Heinzen, E. L., Cirulli, E. T., Pelak, K., He, M., Ruzzo, E. K., Gumbs, C., Singh, A., Feng, S., Shianna, K. V., & Goldstein, D. B. (2012). Using ERDS to infer copy-number variants in high-coverage genomes. *American Journal of Human Genetics*, *91*(3), 408–421.
- Zuniga, A., Zeller, R., & Probst, S. (2012). The molecular basis of human congenital limb malformations. *Wiley Interdisciplinary Reviews. Developmental Biology*, *1*(6), 803–822.

## Eidesstattliche Versicherung

„Ich, Jonas Elsner, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Untersuchung des diagnostischen Potenzials der Ganzgenomsequenzierung am Beispiel von 69 Fällen mit angeborenen Extremitätenfehlbildungen; Investigation into the diagnostic potential of Whole-Genome-Sequencing by the example of 69 cases with congenital limb malformations“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; [www.icmje.org](http://www.icmje.org)) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

## Anteilserklärung an den erfolgten Publikationen

Jonas Elsner hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: Jonas Elsner, Martin A. Mensah, Manuel Holtgrewe, Jakob Hertzberg, Stefania Bigoni, Andreas Busche, Deepthi C. de Silva, Nursel Elçioglu, Isabel Filges, Erica Gerkes, Katta Girisha, Luitgard Graul-Neumann, Aleksander Jamsheer, Peter Krawitz, Ingo Kurth, Susanne Markus, Andre Megarbane, André Reis, Miriam Reuter, Daniel Svoboda, Christopher Teller, Beyhan Tuysuz, Seval Türkmen, Meredith Wilson, Rixa Woitschach, Inga Vater, Almuth Caliebe, Wiebke Hülsemann, Denise Horn, Stefan Mundlos, & Malte Spielmann, Genome sequencing in families with congenital limb malformations, Human Genetics, 2021

Beitrag im Einzelnen:

Jonas Elsner konzipierte gemeinsam mit Martin Mensah die Studie. Er wählte geeignete Trios bzw. Singletons aus und trug bereits bestehende geno- und phänotypische Informationen zusammen. Dabei bediente er sich digitaler Datenbanken sowie klassischen Papierakten. Im Anschluss daran phänotypisierte Jonas Elsner die einzelnen Patient\*innen und klassifizierte sie mittels der Human Phenotype Ontology. Diese Informationen wurden in einem Excel-Sheet gespeichert. Nach dem Zusammentragen der Daten, konzipierte er ein Filter und nachfolgenden Workflow für die bevorstehende Analyse, welcher durch ihn ausführlich getestet und angepasst wurde. Daraufhin erfolgte die Analyse der einzelnen Trios bzw. Singletons. Hierfür filterte Jonas Elsner die einzelnen Whole-Genome-Daten in VarFish, prozessierte sie mittels dem in der Arbeit beschriebenen Workflow, und trug die Ergebnisse in Excel-Dateien zusammen. Weiterhin erstellte er VCF-Dateien für jede\*n Patient\*in mittels VarFish, welche mittels eines von ihm geschriebenen Python-Skript analysiert wurden. Auch diese Ergebnisse wurden von ihm in Excel-Dateien organisiert.

Jonas Elsner erstellte wöchentliche Präsentation der Patient\*innen und ihrer vielversprechenden Varianten, welche in Konferenzen mit Martin Mensah und Malte Spielmann besprochen wurden. Varianten, im protein-kodierenden, wie nicht protein-kodierenden Bereich, bei denen eine Assoziation zum bestehenden Phänotyp des\*der Patient\*in vermutet wurde, wurden von Jonas Elsner genauer recherchiert und aufbereitet. Dabei bediente er sich diversen Quellen, welche in der Arbeit näher ausgeführt werden.

Jonas Elsner schrieb den Erstentwurf des Manuskripts. Die Abbildungen figure 1, 2 und 3 wurden vom ihm erstellt. Die supplementary figure 1 bis 9 sowie alle supplementary tables wurden von Jonas Elsner konzipiert und erstellt. Die Überarbeitung und Fertigstellung erfolgten durch Jonas Elsner, in Zusammenarbeit mit Martin Mensah. Die Formatierung vor der Veröffentlichung erfolgte durch Jonas Elsner.

Nach der Publikation teilte Jonas Elsner die gesammelten Informationen der VUS auf der Website <https://genematcher.org/> mit der wissenschaftlichen Gemeinschaft. Der E-Mail-Verkehr sowie potenzielle Kollaborationen und damit in Zusammenhang stehende klinische Tabellen wurden von ihm koordiniert und bearbeitet.

---

Publikation 2: Henrike Lisa Sczakiel, Wiebke Hülsemann, Manuel Holtgrewe, Angela Teresa Abad-Perez, Jonas Elsner, Sarina Schwartzmann, Denise Horn, Malte Spielmann, Stefan Mundlos & Martin Atta Mensah, GLI3 variants causing isolated polysyndactyly are not restricted to the protein's C-terminal third, *Clinical Genetics*, 2021

Beitrag im Einzelnen:

Jonas Elsner unterstützte die Phänotypisierung sowie die Auswertung der WGS-Daten der hausintern analysierten Patient\*innen. Im Falle des Individuums GLI3-17, wurde die WGS-Auswertung und die nachfolgende Analyse von Jonas Elsner durchgeführt. Jonas Elsner nahm aktiv an der Überarbeitung und dem Review Prozess des Manuskripts teil.

Publikation 3: Martin A Mensah, Henri Niskanen, Alexandre P Magalhaes, Shaon Basu, Martin Kircher, Henrike L Sczakiel, Alisa M V Reiter, Jonas Elsner, Peter Meinecke, Saskia Biskup, Brian H Y Chung, Gregor Dombrowsky, Christel Eckmann-Scholz, Marc Phillip Hitz, Alexander Hoischen, Paul-Martin Holterhus, Wiebke Hülsemann, Kimia Kahrizi, Vera M Kalscheuer, Anita Kan, Mandy Krumbiegel, Ingo Kurth, Jonas Leubner, Ann Carolin Longardt, Jörg D Moritz, Hossein Najmabadi, Karolina Skopalova, Lot Snijders Blok, Andreas Tzschach, Eberhard Wiedersberg, Martin Zenker, Carla Garcia-Cabau, René Buschow, Xavier Salvatella, Matthew L Kraushar, Stefan Mundlos, Almuth Caliebe, Malte Spielmann, Denise Horn & Denes Hnisz, Aberrant phase separation and nucleolar dysfunction in rare genetic diseases, *Nature* 2023

Beitrag im Einzelnen:

Jonas Elsner phänotypisierte die eingeschlossenen Patient\*innen mit BPTAS. Das daraufhin stattfindende WGS wurde von Jonas Elsner analysiert und im Einklang mit den anderen Sequenzierungsergebnissen ausgewertet. Im Falle von HMGB1-I5 erfolgte die gesamte Klassifikation des Phänotyps sowie die Auswertung des WGS, Präsentation und Nachbereitung des Falls durch Jonas Elsner. Er unterstützte im Weiteren den Upload der Daten und das Erstellen einer Maske bei Face2Gene. Zuletzt wirkte er bei der Erstellung und Überarbeitung des Manuskripts mit und nahm aktiv am Review Prozess teil.

---

Unterschrift, Datum und Stempel des/der erstbetreuenden Hochschullehrers/in

---

Unterschrift des Doktoranden/der Doktorandin

## Auszug aus der Journal Summary List Publikation 1

Journal Data Filtered By: **Selected JCR Year: 2019** Selected Editions: SCIE,SSCI  
Selected Categories: **"GENETICS and HEREDITY"** Selected Category

Scheme: WoS

**Gesamtanzahl: 177 Journale**

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE REVIEWS GENETICS	37,390	33.133	0.070320
2	NATURE GENETICS	97,380	27.603	0.207900
3	TRENDS IN ECOLOGY & EVOLUTION	38,175	14.764	0.032330
4	TRENDS IN GENETICS	12,530	11.333	0.022030
5	Annual Review of Genetics	7,800	11.146	0.009710
6	GENOME RESEARCH	41,755	11.093	0.076940
7	MOLECULAR BIOLOGY AND EVOLUTION	50,486	11.062	0.084810
8	GENOME BIOLOGY	44,221	10.806	0.133970
9	Genome Medicine	6,463	10.675	0.026510
10	AMERICAN JOURNAL OF HUMAN GENETICS	36,785	10.502	0.060140
11	GENES & DEVELOPMENT	55,192	9.527	0.063070
12	MOLECULAR THERAPY	17,977	8.986	0.030980
13	GENETICS IN MEDICINE	13,045	8.904	0.040880
14	ONCOGENE	66,303	7.971	0.068320
15	Annual Review of Genomics and Human Genetics	2,679	7.243	0.005190
16	AMERICAN JOURNAL OF MEDICAL GENETICS PART C- SEMINARS IN MEDICAL GENETICS	2,288	7.101	0.005340
17	GENOMICS PROTEOMICS & BIOINFORMATICS	2,224	7.051	0.005980
18	GENOMICS	9,595	6.205	0.005790

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
19	Molecular Autism	2,510	5.869	0.007450
20	MUTATION RESEARCH-REVIEWS IN MUTATION RESEARCH	3,621	5.803	0.003120
21	HUMAN GENETICS	8,803	5.743	0.013370
22	npj Genomic Medicine	574	5.631	0.003320
23	CURRENT OPINION IN GENETICS & DEVELOPMENT	7,707	5.512	0.016170
24	Horticulture Research	1,360	5.404	0.003360
25	CRISPR Journal	218	5.343	0.001170
26	PLoS Genetics	45,571	5.174	0.114250
27	HUMAN MOLECULAR GENETICS	39,652	5.100	0.064170
28	Journal of Genetics and Genomics	2,271	5.065	0.004310
29	Clinical Epigenetics	3,787	5.028	0.010640
30	JOURNAL OF MEDICAL GENETICS	12,049	4.943	0.015100
31	Genes & Diseases	1,081	4.803	0.003310
32	Microbial Genomics	1,139	4.632	0.005810
33	CANCER GENE THERAPY	2,914	4.534	0.002860
33	Circulation-Cardiovascular Genetics	3,090	4.534	0.008600
35	THEORETICAL AND APPLIED GENETICS	21,335	4.439	0.014120
36	JOURNAL OF MOLECULAR MEDICINE-JMM	7,543	4.427	0.009210
37	Biotechnology & Genetic Engineering Reviews	590	4.375	0.000380
38	HUMAN GENE THERAPY	5,631	4.273	0.008140
39	Genes and Nutrition	1,585	4.258	0.002000

# Druckexemplar der Publikation

Human Genetics (2021) 140:1229–1239  
<https://doi.org/10.1007/s00439-021-02295-y>

ORIGINAL INVESTIGATION



## Genome sequencing in families with congenital limb malformations

Jonas Elsner<sup>1</sup> · Martin A. Mensah<sup>1,2</sup> · Manuel Holtgrewe<sup>3</sup> · Jakob Hertzberg<sup>4</sup> · Stefania Bigoni<sup>5</sup> · Andreas Busche<sup>6</sup> · Marie Coutelier<sup>1,7</sup> · Deepthi C. de Silva<sup>8</sup> · Nursel Elçioglu<sup>9,10</sup> · Isabel Filges<sup>11</sup> · Erica Gerkes<sup>12</sup> · Katta M. Girisha<sup>13</sup> · Luitgard Graul-Neumann<sup>1</sup> · Aleksander Jamsheer<sup>14</sup> · Peter Krawitz<sup>15</sup> · Ingo Kurth<sup>16</sup> · Susanne Markus<sup>17</sup> · Andre Megarbane<sup>18</sup> · André Reis<sup>19</sup> · Miriam S. Reuter<sup>19</sup> · Daniel Svoboda<sup>20</sup> · Christopher Teller<sup>21</sup> · Beyhan Tuysuz<sup>22</sup> · Seval Türkmen<sup>1,23</sup> · Meredith Wilson<sup>24</sup> · Rixa Woitschach<sup>25</sup> · Inga Vater<sup>26</sup> · Almuth Caliebe<sup>26</sup> · Wiebke Hülsemann<sup>27</sup> · Denise Horn<sup>1</sup> · Stefan Mundlos<sup>1,4</sup> · Malte Spielmann<sup>4,26,28</sup>

Received: 18 March 2021 / Accepted: 10 May 2021 / Published online: 22 June 2021  
 © The Author(s) 2021

### Abstract

The extensive clinical and genetic heterogeneity of congenital limb malformation calls for comprehensive genome-wide analysis of genetic variation. Genome sequencing (GS) has the potential to identify all genetic variants. Here we aim to determine the diagnostic potential of GS as a comprehensive one-test-for-all strategy in a cohort of undiagnosed patients with congenital limb malformations. We collected 69 cases (64 trios, 1 duo, 5 singletons) with congenital limb malformations with no molecular diagnosis after standard clinical genetic testing and performed genome sequencing. We also developed a framework to identify potential noncoding pathogenic variants. We identified likely pathogenic/disease-associated variants in 12 cases (17.4%) including four in known disease genes, and one repeat expansion in *HOXD13*. In three unrelated cases with ectrodactyly, we identified likely pathogenic variants in *UBA2*, establishing it as a novel disease gene. In addition, we found two complex structural variants (3%). We also identified likely causative variants in three novel high confidence candidate genes. We were not able to identify any noncoding variants. GS is a powerful strategy to identify all types of genomic variants associated with congenital limb malformation, including repeat expansions and complex structural variants missed by standard diagnostic approaches. In this cohort, no causative noncoding SNVs could be identified.

### Introduction

The repertoire of diagnostic tests in human genetics is as diverse as the types of genetic alterations they were developed to detect (Berisha et al. 2020). Through the development of Next Generation Sequencing technologies (NGS) sequencing has become several orders of magnitude faster and cheaper. This has led to an enormous increase in the efficiency of genetic testing (Levy and Myers 2016). NGS quickly found its way from research applications to the clinic: today, panel and exome sequencing are elements

of the routine diagnostics in genetic medicine (Deciphering Developmental Disorders Study 2017). Despite these significant advances, classical genetic testing methods such as chromosomal microarray analysis (CMA) and Sanger sequencing remain part of the standard diagnostic arsenal. This is because NGS-based gene panels often do not detect structural variants such as inversions and translocations, or fail to determine repeat lengths (Berisha et al. 2020). The goal of detecting all types of genetic variation in a single test can theoretically be achieved by short-read based genome sequencing (GS) (Xue et al. 2015). While there are some very encouraging proof of concept studies for the use of GS in individuals with intellectual disability (Lindstrand et al. 2019), GS is not yet part of the clinical routine and there is a lack of systematic studies on the benefits of such tests for individuals with congenital malformations.

A major limitation of panel and exome sequencing approaches is that they usually do not cover 98% of the genome which is noncoding, and are, hence, unable to detect deep intronic splice variants or intergenic regulatory

Jonas Elsner and Martin A. Mensah shared first authorship.

Stefan Mundlos and Malte Spielmann shared senior authorship.

✉ Stefan Mundlos  
[stefan.mundlos@charite.de](mailto:stefan.mundlos@charite.de)

✉ Malte Spielmann  
[malte.spielmann@uksh.de](mailto:malte.spielmann@uksh.de)

Extended author information available on the last page of the article

variants. Therefore, over 40% of individuals with genetic diseases receive no molecular diagnosis after standard testing (Gilissen et al. 2014). This is likely because the noncoding sequence has largely been ignored despite most nucleotides and single nucleotide variants being noncoding. The two main challenges that currently hamper the medical interpretation of noncoding variants are the poor understanding of the “regulatory code” of the noncoding genome and the large number of noncoding variants in each individual that renders classical functional work-up strategies impossible.

In this study, we aimed to determine the diagnostic potential of GS as a comprehensive one-test-for-all strategy in a cohort of 69 unsolved patients with congenital limb malformations. We also attempted to develop a framework to prioritize the large number of noncoding variants identified in the GS studies by combining mouse genetic and human functional epigenetic data with *in vivo*-validated enhancer sequences.

## Materials and methods

### Study design

Patients affected with malformations of two limbs, or two individuals from a family, each affected with a malformation of at least one limb were recruited (Supplementary Fig. 1). Exclusion criteria included a molecularly established genetic diagnosis, a suspected diagnosis of amniotic band syndrome, or an isolated fifth finger clinodactyly. A convenience set of samples was collected from the patients of the Department of Hand Surgery of the Katholisches Kinderkrankenhaus Wilhelmstift Hamburg and the Institute of Medical Genetics and Human Genetics of the Charité (IMG)—Universitätsmedizin Berlin. This sample-set was compiled with cases that were sent to the IMG by external physicians for diagnostic purposes. The sample-set was fixed before conducting GS.

### Included patients

We included 69 patients in this study (Supplementary Table 1). We sequenced the index case and both parents in 64 cases, the index and one parent in one case, and only the index in four cases (parental DNA was not available for testing). In one case, we additionally sequenced a sibling. In five cases, one parent showed a limb malformation comparable to the index. In one case featuring ectrodactyly and apparently unaffected parents and grandparents, a maternal grand-uncle was affected, who was also sequenced. In 60 cases no family member other than the index was reported to show a limb malformation.

### Phenotyping and conventional genetic testing

Limb malformations were phenotyped based on photographs and radiographs by a panel of medical professionals including expert clinical geneticists. Phenotypes were described as per Human Phenotype Ontology (HPO) terminology.

Based on a patient’s phenotype, genes were selected for sequencing by medical geneticists. Sample preparation and Sanger-sequencing were performed using standard procedures. High resolution (1 M oligo) CMA was performed as described previously (Flöttmann et al. 2018).

### Genome sequencing and variant calling

Paired-end PCR-free GS was performed by MacroGen Inc. (South-Korea) using a HiSeq X Ten platform. DNA preparation, sequencing, and sequence data processing were performed according to MacroGen’s standard protocol (coverage 30x and read length 150 bp).

The FASTQ files were transferred to the Core Unit Bioinformatics of the Berlin Institute of Health (CUBI) for variant calling. Files were further processed and securely stored in the System for Omics Data Analysis and Retrieval (SODAR) (Nieminen et al. 2020). GATK HC was used to call simple nucleotide variants, while structural variants were called using Delly2, PopDel, and ERDS/SV2. Afterwards, variants were processed and annotated by the VarFish platform (Holtgrewe et al. 2020). Variants were mapped according to the hg19 reference genome.

### Variant filtration

#### Coding SNVs and SVs

Each index case was filtered as a singleton, regardless of the availability of family data. If parental samples were available, a trio-based filtration approach was additionally performed. Male-index trio-cases were also filtered for hemizygous X-chromosomal variants.

Simple nucleotide variant filtration was performed on the VarFish platform (Version v0.17.2) (Holtgrewe et al. 2020). We filtered GS data for non-synonymous exonic and splice variants using default settings for read depth, allelic balance, and read quality. Allele counts were set as described in Supplementary Table 2. SNVs that passed filtration were exported as variant calling files (VCF). For evaluation of variant pathogenicity, VCFs were uploaded to MutationDistiller and Exomiser. The first ten results were exported for semi-automated, in-depth analysis (see Supplementary Fig. 2 for details). We also tested for truncating or probable

LoF (CADD > 20) variants affecting the same gene with a pLI > 0.9 in at least two independent patients.

Structural variant filtration was performed as described in Supplementary Table 3. The minimal size for structural variant filtration was 1500 bp. Whenever we obtained more than 30 structural variants after initial filtering, we increased the number of minimally covered informative reads from 2 to 5. Each SV passing filtration was judged manually with the information provided by the IGV-Browser and UCSC (see Supplementary Fig. 3 for details).

All findings were evaluated at weekly clinical meetings.

#### Analysis of noncoding variants: limb regulome

We defined a limb-specific potential regulome to filter and interpret non-coding variants. For this purpose, we created a list of 1719 genes involved in embryonic limb development based on data from the Mouse Genome Informatics (MGI) database and entries in OMIM. We defined the human limb regulome as the following: 1. all conserved (phyloP > 1.3) variants, 2. those located within the same topologically associating domain (TAD) (as determined in human fibroblasts (Dixon et al. 2012)) as a limb gene, 3. those that were marked by an H3K27 acetylation peak in human limb buds (Cotney et al. 2013). We also included the validated enhancer elements of the VISTA Enhancer Browser. Coding and noncoding SNVs were filtered for rare variants (frequency < 0.1%) and were considered to be potentially affecting the same regulatory element if they were either less than 300 bp apart or positioned in the same established enhancer.

Whenever we identified rare, potentially pathogenic heterozygous coding variants of genes associated with a recessive limb phenotype in individuals featuring at least a partial overlap with that phenotype, we also screened for *in trans* conserved non-coding variants with a MAF < 3% affecting the same TAD.

## Results

We collected a cohort of 69 individuals affected with congenital limb malformations. All individuals had previously gone through our clinical genetics routine pipeline including clinical examination, candidate gene testing, and CMA. We then performed GS as a comprehensive one-test-for-all strategy.

In total, we identified 333,163,643 single nucleotide variants (SNVs) among the 69 sequenced index patients, of which 7,020,766 were either coding or flanking coding elements by 10 bp or less. 326,142,877 were noncoding SNVs, of which 19,362 were rare (gnomAD AF < 0.01). 21,369 of the coding SNV calls were classified by Janovar to be of at least moderate relevance (missense and

truncating), and 1429 to be of high relevance (truncating only). VarFish filtering returned 5761 SNVs. Filtering by Exomizer and MutationTaster identified 433 potentially pathogenic coding calls among these, of which 174 were high-quality calls suitable for further evaluation.

30,062 SNVs resulted in the potential loss-of-function of probably haploinsufficient genes. 49 of these affected the same gene in two unrelated index individuals (Supplementary Fig. 2).

We also analyzed the structural variants in 68 of the 69 index patients. 55 cases were filtered as trios with unaffected parents and moderate filter settings. Five were filtered as trios with another affected relative and moderate filter settings. Stricter filter settings were chosen for 9 trios because moderate filter settings produced an unmanageable amount of SV calls. 3 cases were analyzed as singletons. Individuals I1, I2, I3 did not yield any results, I4 was excluded from the SV analysis because too many SVs were called even with stricter filter settings due to poor data quality.

Of the 1,555,426 SVs, 633 SVs passed the filtering by VarFish, of which 222 were inversions, 288 deletions, 76 duplications, and 47 breakpoints of potential translocations.

417 of these SVs were excluded because they were of poor calling quality or because they were inherited from an unaffected parent. We then manually inspected the remaining 216 SVs. Segregation analysis in the parents was performed by qPCR after comparing candidate CNVs with known limb genes according to the Human Phenotype Ontology, cross-species phenotype comparison, mouse models, gene expression data (Cao et al. 2019), limb enhancer elements (Visel et al. 2009), and the local topological associating domain (TAD) architecture of the locus (Dixon et al. 2012; Cao et al. 2019). As a result, we identified 30 promising variants (Supplementary Fig. 3).

#### Variants in four known, limb malformation associated genes

We identified pathogenic variants in established disease genes in four individuals (I5, I6, I7, I8, Supplementary Fig. 4), which we confirmed by Sanger sequencing. These included a missense variant in *FGFR1*, already described in the literature (Muenke et al. 2014), and three previously undescribed variants in the genes *FGFR2*, *GLI3*, and *BHLHA9*. In all four cases, we classified the variants as (likely) pathogenic according to the criteria of the American College of Medical Genetics and Genomics (ACMG), based on the type of variant and the phenotype of the patient. All variants were inherited (note that the mother of I8, was not radiographically phenotyped, which is necessary to

diagnose mild *FGFR2*-associated phenotypes (Flöttmann et al. 2015)).

### Repeat expansions of *HOXD13*

In individual I9, featuring brachy-poly-syndactyly, we detected a repeat expansion by eight alanines on the *HOXD13* allele inherited from his affected mother, already described as pathogenic in the literature (Brison et al. 2014), and a polymorphic repeat expansion by only one alanine on the paternal *HOXD13* allele (Supplementary Fig. 5). These findings were confirmed by conventional *HOXD13* micro-satellite analysis.

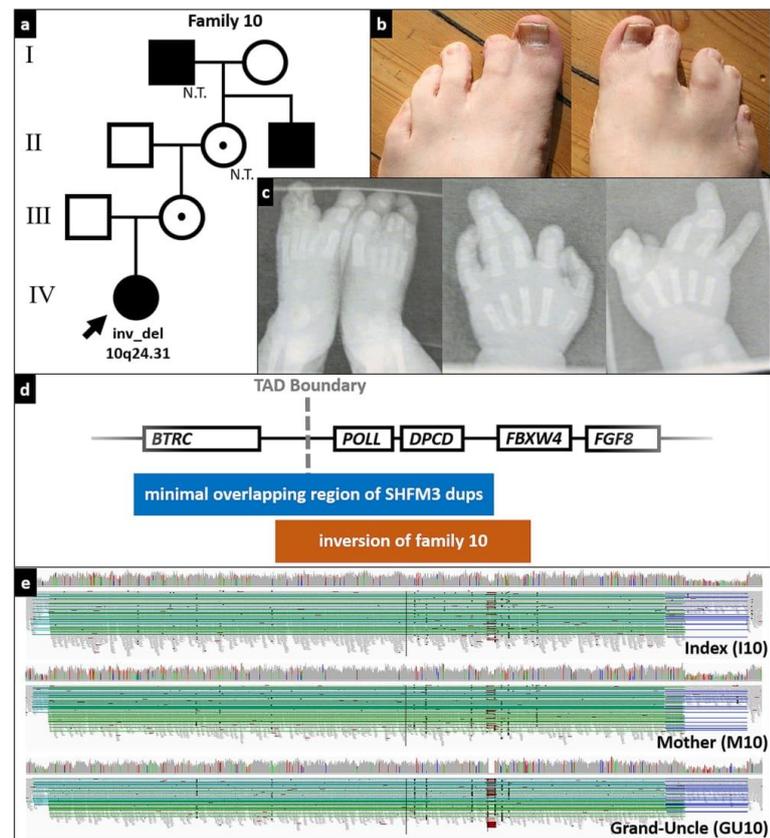
### Structural variants at known disease loci

In individual I10 with bilateral upper and lower limb ectrodactyly, we identified an inversion of 105 kb (chr10: 103,321,526–103,426,609) flanked by two

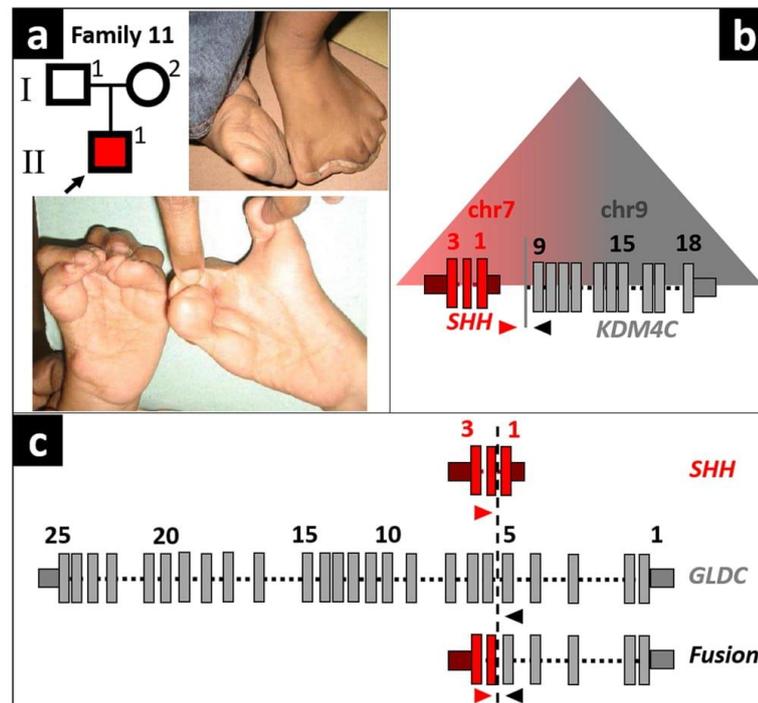
deletions (chr10:103,319,219–103,321,525 and chr10:103,426,610–103,436,718) at the split-hand foot malformation locus 3 (SHFM3) on chr10q24 inherited from his unaffected mother (Fig. 1, Supplementary Fig. 6). His affected great-uncle also carried the inversion. The variant overlaps with the most common duplications associated with ectrodactyly (de Mollerat et al. 2003; Klopocki et al. 2012). The minimal overlapping region of pathogenic SHFM3 duplications includes *BTRC*, *POLL*, and *DPCD* (Holder-Espinasse et al. 2019). The inversion described here is copy number neutral, suggesting that positional effects rather than gene dosage might be responsible for the phenotype. It includes a topologically associating domain boundary (Holder-Espinasse et al. 2019) and is likely to change the enhancer landscape at the SHFM3 locus leading to *FGF8* misregulation causing ectrodactyly.

In individual I11, featuring bilateral mirror-image polydactyly of the hands and feet (Fig. 2a), CMA had detected a 300 kb amplification on chr7q36.1. We initially classified

**Fig. 1** Inversion-deletion at SHFM3 locus: **a** pedigree, N.T. not tested. **b** feet of grand-uncle (II-3). **c** hands and feet of the index patient (IV-1). **d** genomic architecture of SHFM3. **e** GS data of the family, note the presence of an inversion (chr10: 103,321,526–103,426,609) flanked by deletions (chr10:103,319,219–103,321,525 and chr10:103,426,610–103,436,718) on either site



**Fig. 2** **a** Pedigree and phenotype of individual I11. **b** Potential neo-TAD at the fusion site. **c** Breakpoint and fusion sites between regions from chr7 and chr9



the variant as a variant of unknown significance, because the individual's phenotype did not match that of an individual with muscular hypertrophy, reported to have a similarly sized and positioned duplication (Kroeldrup et al. 2012).

The amplification was identified again using GS. However, sequencing revealed that it was part of a complex structural variant containing two overlapping duplications (dup1 and dup2) at chr7q36.1 (Supplementary Fig. 7). The smaller dup2 shares the central breakpoint with dup1. The distal breakpoint of dup2 is positioned within dup1 in intron 1 of *SHH*, at chr7:155,603,964. GS data showed that the duplicons were not positioned *in tandem*, but are both fused to sequences originating from chr9p24.1. The distal breakpoint of dup2 was fused to intron 5 of *GLDC* and the distal breakpoint of dup1 to intron 8 of *KDM4C* (Fig. 2c; Supplementary Fig. 7). Analysis of the parents by Sanger sequencing showed that the structural variant occurred *de novo* in individual I11.

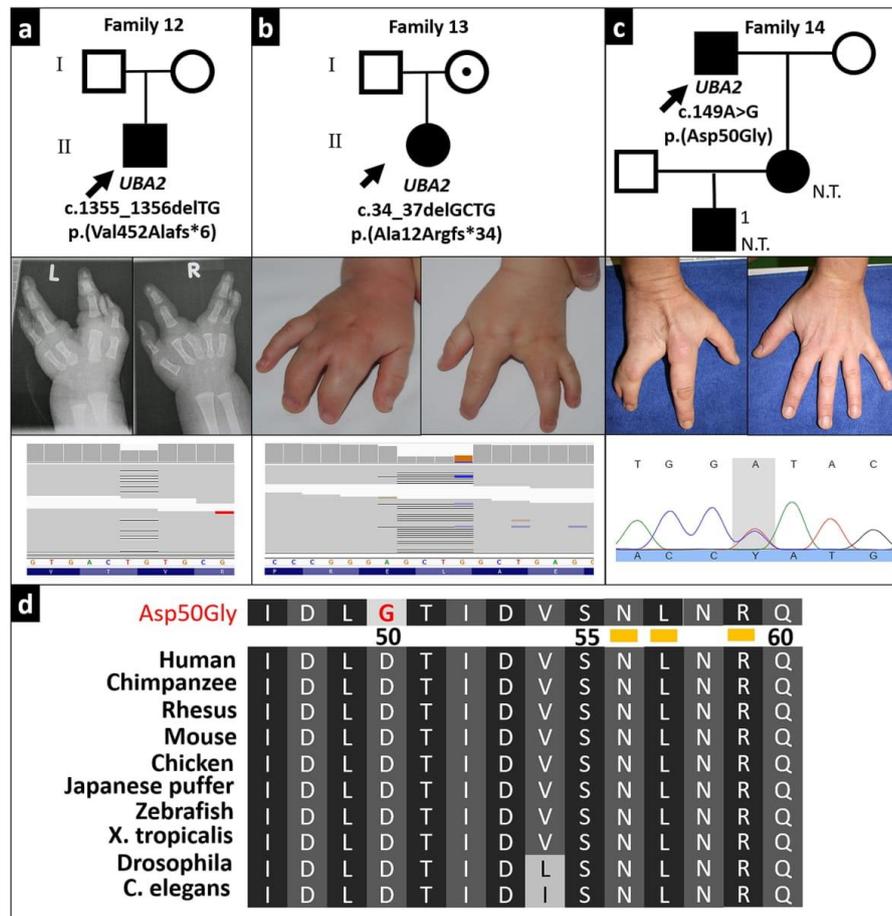
The mirror-image polydactyly of individual I11 shows striking phenotypic overlap with Laurin–Sandrow syndrome, which is caused by duplications of the *SHH* regulator ZRS, positioned in intron 5 of *LMBR1* on chr7p36.3, resulting in ectopic expression of *SHH* in the embryonic limb (Lohan

et al. 2014). Both duplications do not include the ZRS and duplications of *SHH* itself have not been described to cause Laurin–Sandrow syndrome. However, a duplicated fragment containing *SHH* that is inserted into another domain, as observed in the *de novo* SV of I11, makes an ectopic expression of *SHH* in the embryonic limb very likely. We assume that the formation of an *SHH-KDM4C* neo-TAD, resulting in the misregulation of *SHH* by *KDM4C*-enhancers in the limb mesenchyme (note the known expression of *KDM4C* in embryonic vertebrate limb buds) is the most likely explanation for such an ectopic *SHH*-expression (Fig. 2b).

Therefore, we re-classified the complex SV involving *SHH* in individual I11 with bilateral mirror-image polydactyly as causative.

#### Establishing *UBA2* as a novel disease gene

We also identified variants in new candidate genes. Two unrelated individuals with isolated split hand malformation featured different heterozygous frameshift variants in the *ubiquitin-like modifier-activating enzyme 2 (UBA2)* (Fig. 3a, b). Individual I12 harboured the *de novo* variant NM\_005499.3(*UBA2*):c.1355\_1356delTG;p.(Val452A1afs\*6).



**Fig. 3** *UBA2* variants and ectrodactyly. **a–c** Patients with likely pathogenic *UBA2* variants *upper panels*: pedigrees, N.T.: not tested; *middle panels*: characteristic limb malformations, *lower panels*: sequencing data. **d** conservation of Asp50 mutated in individual I14, numbers

indicate amino acid residues, *yellow bars* highlight positions tested by Olsen et al. to cause loss of function when substituted by alanine (Olsen et al. 2010)

Individual I13 inherited the variant NM\_005499.3(*UBA2*):c.34\_37delGCTG;p.(Ala12Argfs\*34) from his apparently unaffected mother (no radiographs of her hands were available).

We classified the pathogenicity of these variants according to the ACMG guidelines. Both are null variants of *UBA2* which has a pLI-score of 1 (PVS1). c.1355\_1356delTG occurred de novo in an individual with a negative family history (PS2). The variants are absent from the 1000 Genomes Project and the Exome Aggregation Consortium databases (PM2) and were predicted to be

pathogenic by MutationTaster (PP3). *UBA2* variants have recently been described in individuals with ectrodactyly (Chowdhury et al. 2014; Abe et al. 2018; Yamoto et al. 2019; Aerden et al. 2020). Hence, we regarded *UBA2* as a disease-associated gene and these variants as pathogenic (1PVS(+1PS)+1PM+1PP).

These findings prompted subsequent Sanger sequencing of *UBA2* in 24 unrelated families with ectrodactyly, who have been tested negative for variants in the established SHFM loci/genes. In one individual (I14) with unilateral split-hand malformation (Fig. 3c), we identified the missense

variant NM\_005499.3(*UBA2*):c.149A > G;p (Asp50Gly). The daughter of I14 and her son also feature ectrodactyly (PP4), but were unavailable for testing. Asp50 is part of a consecutive 15 amino acid sequence (Ile47 to Phe61) shared amongst all nephrozoan *UBA2* orthologues (Fig. 3d). Olsen et al. showed that variants of residues (Asn56Ala, Leu57Ala, Arg59Ala) of this element result in loss of *UBA2* function, and found that the very residue mutated in individual I14, Asp50, forms hydrogen bonds with Asn177 and Thr178 essential for proper *UBA2* folding and thus its function (PS3) (Olsen et al. 2010). The variant was also absent in the databases (PM2) and predicted to be pathogenic by MutationTaster (PP3). Hence, we classified these variants as likely pathogenic according to the ACMG's guidelines (1PS + 1PM + 2PP).

### Novel candidate genes

Our analysis also revealed several novel, high-confidence candidate genes associated with limb defects. In individual I15 featuring severe mirror image foot polydactyly, we found a de novo frameshift variant in the gene encoding the high mobility group box 1 protein (*HMGB1*) (Supplementary Fig. 8). NM\_002128.7(*HMGB1*):c.551\_554delAGAA;p. (Lys184Argfs\*44) leads to the replacement of the protein's entire C-terminal 30-residue acidic tail by 41 other unrelated residues. The tail is normally formed by an Asp/Glu-repeat element, which is highly conserved among *HMGB1* orthologues. This repeat element stabilizes *HMGB1*'s secondary structure and is crucial for its DNA-bending capacity (Belgrano et al. 2013; Anggayasti et al. 2020). The variant is not only absent from the databases but also no variant listed in gnomAD contains an amino acid residue except Glu or Asp in the acidic tail domain. *HMGB1* has a pLI score of 1. In mouse and zebrafish studies, *HMGB1* has been shown to regulate digit number during embryonic limb development by interacting with WNT, BMP, and SHH (Itou et al. 2011). We, therefore, consider *HMGB1* to be a novel candidate gene for mirror image foot polydactyly.

Individual I16, who featured short stature, absent distal phalanges of the 5th fingers and toes, and dysplastic middle phalanges of the toes carried a de novo missense variant in the gene encoding semaphorin 3D (*SEMA3D*) (Supplementary Fig. 9). NM\_152754.3(*SEMA3D*):c.1918G > A;p.(Asp640Asn) is absent from the 1000 Genomes Project database and is listed only 4 times in the Genome Aggregation Database (gnomAD). The Asp640 residue in the immunoglobulin-like domain of *SEMA3D* is highly conserved amongst vertebrates. The variant is predicted to be pathogenic by MutationTaster. *SEMA3D* regulates neural crest cell differentiation and is involved in the organogenesis of the heart (Sanchez-Castro et al. 2015), parathyroid gland (Singh et al. 2019), and, notably, limbs (Govindan

et al. 2016). We, therefore, consider it a candidate gene for short stature with limb abnormalities.

In individual I17 we identified a paternally inherited frameshift variant in the *aldehyde dehydrogenase 1 family member A2* gene (*ALDH1A2*) encoding retinaldehyde dehydrogenase 2 (Supplementary Fig. 10). Both, the patient and her father feature isolated cutaneous syndactyly of the fingers III–IV and the toes II–III. The variant NM\_003888.4(*ALDH1A2*):c.35delT;p.(Val12Glyfs\*31) is absent from the databases and is predicted to be disease-causing by MutationTaster. *ALDH1A2* is a direct target of *HOXA13* and plays a key role in vertebrate digit development by regulating, in particular, interdigital programmed cell death (Shou et al. 2013). Rescued *ALDH1A2* knockout mice show reduced interdigital cell death and thus impaired digit separation during limb development resulting in syndactyly (Zhao et al. 2010). It is, therefore, likely that *ALDH1A2*:c.35delT caused the phenotype of syndactyly in individual I17 and her father.

### Identification of noncoding variants

So far, the interpretation of disease-related variation has been focused on protein-coding DNA and the identification of variants that directly result in the disruption of specific gene functions.

Here, we aimed to develop a framework to prioritize a large number of noncoding variants from GS studies, by combining mouse genetic and human functional epigenetic data with in vivo-validated enhancer sequences. We then defined a limb-specific regulome that we used to filter all noncoding variants (Materials and Methods). Our potentially disease-relevant limb-specific regulome consists of 5,591,007 sites covering in total 7,294,220 bp, i.e. 0.24% of the human genome (hg19).

Overall, we identified 19,362 rare noncoding SNVs in the 69 index patients, of which 143 were located within the limb regulome (Fig. 4). First, we focused on the de novo variants and identified 6 calls located in potential regulatory elements. Two variants were excluded because they were called in cases with (likely) pathogenic coding or structural variants.

Individual I18 presenting with bilateral syndactyly of fingers II–V featured the de novo call chr1:41948304AAG > A in intron 2 of *EDN2*. The position shows increased acetylation of H3K27 in human limb buds. The 2 bp deletion also removes one element of a 6-AG-repeat whose length is not conserved in vertebrates. Furthermore, *EDN2* encodes endothelin 2, a potent vasoconstrictor with no evident link to limb development.

Individual I19 showing upper limb amelia featured three calls (chr5:157285900CACGTGGG > C, chr5:157285909CTCGG > C, chr5:157285915CACAAC

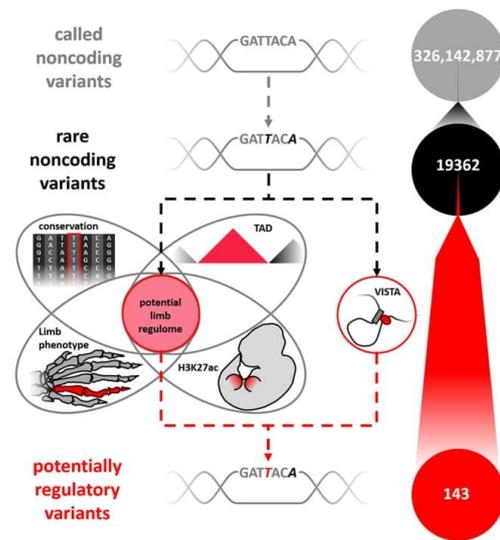


Fig. 4 Pipeline of noncoding data analysis

TG > C) referring to the same indel in intron 1 (15 bp downstream of the first exon–intron boundary) of *CLINT1*. However, *CLINT1* shows only a moderate pLI score (0.54) and there is no evidence other than increased H3K27ac marks of its promoter region in human limb buds linking it to limb development.

We were not able to identify any rare variants in validated VISTA enhancers that showed enhancer activity in the limb bud.

Next, we focused on noncoding variants that were located close to one another in more than one case. In total, 3425 rare noncoding variants in the unsolved cases were positioned 300 bp or less apart from a variant in another unsolved case. 16 of these calls were located within the limb regulome, but in five of these variants, the other variant was positioned outside of the limb regulome.

In two cases both variants were positioned within the limb regulome and within 300 bp: individual I21 and individual I22, both showing finger syndactyly, harbored the overlapping deletions chr22:24552064GGGGGCCGG GACTGGGGCCGGGACT > G and chr22:24552086ACTGGGGCCGGGG > A, respectively. The deletions are positioned in intron 29 of *CABINI*, in an evolutionarily partially conserved element, that shows increased H3K27ac marks in human embryonic limb buds. However, both deletions were inherited from unaffected parents.

Eight of the close variants were double hits (i.e. we detected rare calls not in just one but two index patients at four positions of the potential limb regulome). However, none of these four pairs of index patients showed overlapping phenotypes.

We identified no coding variant of a known limb disease gene in trans with a conserved, rare noncoding variant of the same TAD.

In summary, despite extensive efforts, we were not able to identify any noncoding SNVs that showed convincing evidence to be causal in congenital limb malformations.

## Discussion

In this study, we set out to determine the potential of GS as a comprehensive diagnostic tool to determine all kinds of genetic variants associated with congenital limb malformation.

In our cohort of patients with congenital limb malformations, GS was able to detect both previously described and novel causative genetic variants in already established limb malformation associated genes. In addition, it enabled the identification of three candidate genes and the independent verification of the novel disease gene *UBA2* for causing ectrodactyly (Yamoto et al. 2019). Our approach was able to detect SNVs and structural variants. Finally, GS proved to be a powerful strategy to identify genomic variants previously missed by most other approaches, including repeat expansions and complex structural variants. In total, we identified variants that we consider to be likely pathogenic/disease-associated in 12 of 69 cases (17.4%). This diagnostic yield is comparable to the recent landmark study conducted by the British National Health Service that used GS in cohorts with other congenital disease entities (Turro et al. 2020). A clear advantage of GS compared to most other technologies is the ability to detect copy number neutral variants and to gain position information on CNVs. In our cohort of only 69, we were able to detect two complex variants, an inversion at the *FGF8* locus and a translocated triplication including the *SHH* gene. Both were missed by standard technologies. Further research is necessary to clarify their exact pathomechanisms. The variants identified in the genes *HMGBl*, *SEMA3D* and *ALDH1A2* are all likely to cause loss of function. The genes were previously associated with vertebrate limb development in animal studies and the variants either arose de novo or segregate with the respective phenotype. However, we could not identify unrelated individuals featuring comparable variants in these candidate genes and similar phenotypes. Future research is necessary to identify such to establish the described candidates as disease genes. Our findings once again highlight the role of GS as an attractive

one-test-for-all strategy for clinically very heterogeneous cohorts such as congenital malformation syndromes or intellectual disability (Gilissen et al. 2014; Turro et al. 2020). The total cost of the various conventional tests currently used in clinical routine far exceeds that of trio GS.

One of the main challenges of GS data is the medical interpretation of changes in the noncoding DNA. While most clinical GS studies tend to ignore noncoding SNVs (Gilissen et al. 2014) there are recent anecdotal reports of noncoding variants as the cause of Mendelian disorders (Lettice et al. 2003; Jeong et al. 2008; Albers et al. 2012; Bhatia et al. 2013; Weedon et al. 2014; Bae et al. 2014), although there is no established systematic approach, yet. Therefore, we set out to develop a framework to prioritize such noncoding variants associated with congenital limb malformation. We used a combinatorial approach of mouse and human epigenetic data, *in vivo* validated enhancer sequences, knock-out mice, and the recent knowledge about 3D genome folding, and the *cis*-regulatory architecture of the genome to define a limb regulome. This limb regulome consists of 0.24% of the genome and includes all known *in vivo*-validated limb enhancer elements. Contrary to our expectation, we could only identify candidate loci, but no definitely pathogenic noncoding variants. These findings are in stark contrast to our recent study where we demonstrated that CNVs affecting noncoding regulatory elements are a major cause of congenital limb malformations (Flöttmann et al. 2018).

While our results suggest that GS is sensitive to classical sequence variants, it is noteworthy that the method cannot detect epigenetic variants. Epimutations (e.g. imprinting defects) are known to cause inheritable human disease. However, to our knowledge, no epimutation has been linked to congenital limb malformation yet.

Tools for the analysis of GS data are continuously being developed further and the precision of algorithms to call structural variants can certainly be improved. We expect the diagnostic rate to increase steadily with the accuracy of the instruments invoked to analyze GS data.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00439-021-02295-y>.

**Acknowledgements** We would like to thank all families for their collaboration and contribution to this project. J.E. is supported by a scholarship of the Studienstiftung des deutschen Volkes e.V. M.A.M. is a participant in the BIH Charité Digital Clinician Scientist Program founded by the late Prof. Duska Dragun and funded by the Charité-Universitätsmedizin Berlin and the Berlin Institute of Health.

**Author contributions** Jonas Elsner, Martin A. Mensah, Stefan Mundlos and Malte Spielmann designed the study. Jonas Elsner, Martin A. Mensah, Manuel Holtgrewe, Jakob Hertzberg and Marie Coutelier performed the variant filtration and processed data. Jonas Elsner, Martin A. Mensah and Malte Spielmann investigated and together with Stefan Mundlos and Denise Horn critically evaluated variants. Martin A. Mensah, Stefania Bigoni, Andreas Busche, Deepthi C. de Silva,

Nursel Elçioğlu, Isabel Filges, Erica Gerkes, Katta Girisha, Luitgard Graul-Neumann, Jane Hurst, Aleksander Jamsheer, Peter Krawitz, Ingo Kurth, Susanne Markus, Andre Megarbane, André Reis, Miriam Reuter, Daniel Svoboda, Christopher Teller, Beyhan Tuysuz, Seval Türkmén, Meredith Wilson, Rixa Woitschach, Inga Vater, Almuth Caliebe, Wiebke Hülsemann, Denise Horn, Stefan Mundlos and Malte Spielmann examined the patients and provided samples. Jonas Elsner and Martin A. Mensah visualised the workflow and results. Jonas Elsner, Martin A. Mensah, Stefan Mundlos and Malte Spielmann wrote and reviewed the manuscript with the approval of all authors. Stefan Mundlos and Malte Spielmann supervised the project.

**Funding** Open Access funding enabled and organized by Projekt DEAL. A.J. was supported by the grant from the Polish National Science Centre UMO-2016/22/E/NZ5/00270. M.S. is supported by grants from the Deutsche Forschungsgemeinschaft (DFG) (SP1532/3-1, SP1532/4-1 and SP1532/5-1), the Max Planck Foundation and the Deutsches Zentrum für Luft- und Raumfahrt (DLR 01GM1925).

**Code and data availability** Code and data (pseudonymized, grouped where possible, and minimized) are available upon reasonable request.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethics approval** This study was approved by the ethics committees of the Charité-Universitätsmedizin Berlin (EA2/087/15) and the Medical Chamber (Ärztekammer) Hamburg and conducted in accordance with the declaration of Helsinki. The Charité's ethic commission served as the central IRB.

**Consent to participate** Participants or their legal guardians provided written informed consent for participation in this study. Consents of all participants were archived. All published data has been pseudonymized, grouped where possible, and minimized.

**Consent for publication** Participants or their legal guardians provided written informed consent for pseudonymized publication of their data. Consents of all participants were archived. All published data have been pseudonymized, grouped where possible, and minimized.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Abe KT, Rizzo IMPO, Coelho ALV et al (2018) 19q13.11 microdeletion: clinical features overlapping ectrodactyly ectodermal

- dysplasia-clefting syndrome phenotype. *Clin Case Rep* 6:1300–1307. <https://doi.org/10.1002/ccr3.1600>
- Aerden M, Bauters M, Van Den Bogaert K et al (2020) Genotype-phenotype correlations of UBA2 mutations in patients with ectrodactyly. *Eur J Med Genet* 63:104009. <https://doi.org/10.1016/j.ejmg.2020.104009>
- Albers CA, Paul DS, Schulze H et al (2012) Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet* 44(435–9):S1–2. <https://doi.org/10.1038/ng.1083>
- Anggayasti WL, Ogino K, Yamamoto E et al (2020) The acidic tail of HMGB1 regulates its secondary structure and conformational flexibility: a circular dichroism and molecular dynamics simulation study. *Comput Struct Biotechnol J* 18:1160–1172
- Bae B-I, Tietjen I, Atabay KD et al (2014) Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science* 343:764–768. <https://doi.org/10.1126/science.1244392>
- Belgrano FS, de Abreu da Silva IC, Bastos de Oliveira FM et al (2013) Role of the acidic tail of high mobility group protein B1 (HMGB1) in protein stability and DNA bending. *PLoS ONE* 8:e79572. <https://doi.org/10.1371/journal.pone.0079572>
- Berisha SZ, Shetty S, Prior TW, Mitchell AL (2020) Cytogenetic and molecular diagnostic testing associated with prenatal and postnatal birth defects. *Birth Defects Res* 112:293–306. <https://doi.org/10.1002/bdr2.1648>
- Bhatia S, Bengani H, Fish M et al (2013) Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am J Hum Genet* 93:1126–1134. <https://doi.org/10.1016/j.ajhg.2013.10.028>
- Brisson N, Debeer P, Tylzanowski P (2014) Joining the fingers: a HOXD13 story. *Dev Dyn* 243:37–48. <https://doi.org/10.1002/dvdy.24037>
- Cao J, Spielmann M, Qiu X et al (2019) The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566:496–502. <https://doi.org/10.1038/s41586-019-0969-x>
- Chowdhury S, Bandholz AM, Parkash S et al (2014) Phenotypic and molecular characterization of 19q12q13.1 deletions: a report of five patients. *Am J Med Genet A* 164A:62–69. <https://doi.org/10.1002/ajmg.a.36201>
- Cotney J, Leng J, Yin J et al (2013) The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* 154:185–196. <https://doi.org/10.1016/j.cell.2013.05.056>
- de Mollerat XJ, Gurrieri F, Morgan CT et al (2003) A genomic rearrangement resulting in a tandem duplication is associated with split hand-split foot malformation 3 (SHFM3) at 10q24. *Hum Mol Genet* 12:1959–1971
- Deciphering Developmental Disorders Study (2017) Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542:433–438. <https://doi.org/10.1038/nature21062>
- Dixon JR, Selvaraj S, Yue F et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380. <https://doi.org/10.1038/nature11082>
- Flöttmann R, Knaus A, Zemojtel T et al (2015) FGFR2 mutation in a patient without typical features of Pfeiffer syndrome—the emerging role of combined NGS and phenotype based strategies. *Eur J Hum Genet* 58:376–380. <https://doi.org/10.1016/j.ejmg.2015.05.007>
- Flöttmann R, Kragesteen BK, Geuer S et al (2018) Noncoding copy-number variations are associated with congenital limb malformation. *Genet Med* 20:599–607. <https://doi.org/10.1038/gim.2017.154>
- Gilissen C, Hehir-Kwa JY, Thung DT et al (2014) Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511:344–347. <https://doi.org/10.1038/nature13394>
- Govindan J, Tun KM, Iovine MK (2016) Cx43-dependent skeletal phenotypes are mediated by interactions between the Hapln1a-ECM and Sema3d during Fin regeneration. *PLoS ONE* 11:e0148202. <https://doi.org/10.1371/journal.pone.0148202>
- Holder-Espinasse M, Jamsheer A, Escande F et al (2019) Duplication of 10q24 locus: broadening the clinical and radiological spectrum. *Eur J Hum Genet* 27:525–534. <https://doi.org/10.1038/s41431-018-0326-9>
- Holtgrewe M, Stolpe O, Nieminen M et al (2020) VarFish: comprehensive DNA variant analysis for diagnostics and research. *Nucleic Acids Res* 48:W162–W169. <https://doi.org/10.1093/nar/gkaa241>
- Itou J, Taniguchi N, Oishi I et al (2011) HMGB factors are required for posterior digit development through integrating signaling pathway activities. *Dev Dyn* 240:1151–1162
- Jeong Y, Leskow FC, El-Jaick K et al (2008) Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat Genet* 40:1348–1353. <https://doi.org/10.1038/ng.230>
- Klopocki E, Lohan S, Doelken SC et al (2012) Duplications of BHLHA9 are associated with ectrodactyly and tibia hemimelia inherited in non-Mendelian fashion. *J Med Genet* 49:119–125. <https://doi.org/10.1136/jmedgenet-2011-100409>
- Kroeldrup L, Kjaergaard S, Kirchhoff M et al (2012) Duplication of 7q36.3 encompassing the Sonic Hedgehog (SHH) gene is associated with congenital muscular hypertrophy. *Eur J Med Genet* 55:557–560. <https://doi.org/10.1016/j.ejmg.2012.04.009>
- Lettice LA, Heaney SJH, Purdie LA et al (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12:1725–1735. <https://doi.org/10.1093/hmg/ddg180>
- Levy SE, Myers RM (2016) Advancements in next-generation sequencing. *Annu Rev Genomics Hum Genet* 17:95–115. <https://doi.org/10.1146/annurev-genom-083115-022413>
- Lindstrand A, Eisfeldt J, Pettersson M et al (2019) From cytogenetics to cytogenomics: whole-genome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability. *Genome Med* 11:68. <https://doi.org/10.1186/s13073-019-0675-1>
- Lohan S, Spielmann M, Doelken SC et al (2014) Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clin Genet* 86:318–325. <https://doi.org/10.1111/cge.12352>
- Muenke M, Schell U, Hehr A et al (2014) A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. *Nat Genet* 8:269–274. <https://doi.org/10.1038/ng1194-269>
- Nieminen M, Stolpe O, Schumann F et al (2020) SODAR core: a Django-based framework for scientific data management and analysis web apps. *J Open Source Softw* JOSS 55:1584. <https://doi.org/10.21105/joss.01584>
- Olsen SK, Capili AD, Lu X et al (2010) Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature* 463:906–912. <https://doi.org/10.1038/nature08765>
- Sanchez-Castro M, Pichon O, Briand A et al (2015) Disruption of the SEMA3D gene in a patient with congenital heart defects. *Hum Mutat* 36:30–33. <https://doi.org/10.1002/humu.22702>
- Shou S, Carlson HL, Perez WD, Stadler HS (2013) HOXA13 regulates Aldh1a2 expression in the autopod to facilitate interdigital programmed cell death. *Dev Dyn* 242:687–698. <https://doi.org/10.1002/dvdy.23966>
- Singh A, Mia MM, Cibi DM et al (2019) Deficiency in the secreted protein Semaphorin3d causes abnormal parathyroid development in mice. *J Biol Chem* 294:8336–8347. <https://doi.org/10.1074/jbc.RA118.007063>
- Turro E, Astle WJ, Megy K et al (2020) Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* 583:96–102. <https://doi.org/10.1038/s41586-020-2434-2>

- Visel A, Blow MJ, Li Z et al (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457:854–858. <https://doi.org/10.1038/nature07730>
- Weedon MN, Cebola I, Patch A-M et al (2014) Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* 46:61–64. <https://doi.org/10.1038/ng.2826>
- Xue Y, Ankala A, Wilcox WR, Hegde MR (2015) Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* 17:444–451. <https://doi.org/10.1038/gim.2014.122>
- Yamamoto K, Saitsu H, Nishimura G et al (2019) Comprehensive clinical and molecular studies in split-hand/foot malformation: identification of two plausible candidate genes (LRP6 and UBA2). *Eur J Hum Genet* 27:1845–1857. <https://doi.org/10.1038/s41431-019-0473-7>
- Zhao X, Brade T, Cunningham TJ, Duester G (2010) Retinoic acid controls expression of tissue remodeling genes Hmgn1 and Fgf18 at the digit-interdigit junction. *Dev Dyn* 239:665–671

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Jonas Elsner<sup>1</sup> · Martin A. Mensah<sup>1,2</sup> · Manuel Holtgrewe<sup>3</sup> · Jakob Hertzberg<sup>4</sup> · Stefania Bigoni<sup>5</sup> · Andreas Busche<sup>6</sup> · Marie Coutelier<sup>1,7</sup> · Deepthi C. de Silva<sup>8</sup> · Nursel Elçioğlu<sup>9,10</sup> · Isabel Filges<sup>11</sup> · Erica Gerkes<sup>12</sup> · Katta M. Girisha<sup>13</sup> · Luitgard Graul-Neumann<sup>1</sup> · Aleksander Jamsheer<sup>14</sup> · Peter Krawitz<sup>15</sup> · Ingo Kurth<sup>16</sup> · Susanne Markus<sup>17</sup> · Andre Megarbane<sup>18</sup> · André Reis<sup>19</sup> · Miriam S. Reuter<sup>19</sup> · Daniel Svoboda<sup>20</sup> · Christopher Teller<sup>21</sup> · Beyhan Tuysuz<sup>22</sup> · Seval Türkmen<sup>1,23</sup> · Meredith Wilson<sup>24</sup> · Rixa Woitschach<sup>25</sup> · Inga Vater<sup>26</sup> · Almuth Caliebe<sup>26</sup> · Wiebke Hülsemann<sup>27</sup> · Denise Horn<sup>1</sup> · Stefan Mundlos<sup>1,4</sup> · Malte Spielmann<sup>4,26,28</sup>

<sup>1</sup> Institute of Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

<sup>2</sup> Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>3</sup> Core Unit Bioinformatics, Berlin Institute of Health (BIH), Berlin, Germany

<sup>4</sup> Max Planck Institute for Molecular Genetics, RG Development and Disease, Berlin, Germany

<sup>5</sup> Medical Genetics Unit, Department of Mother and Child, Ferrara Sant'Anna University Hospital, Ferrara, Italy

<sup>6</sup> Institut Für Humangenetik, Universitätsklinikum Münster, Münster, Germany

<sup>7</sup> Department of Human Genetics, Faculty of Medicine, Jewish General Hospital, McGill University, Montreal, QC, Canada

<sup>8</sup> Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka

<sup>9</sup> Department of Pediatric Genetics, Marmara University Medical School, Istanbul, Turkey

<sup>10</sup> Eastern Mediterranean University Medical School, Cyprus, Mersin 10, Turkey

<sup>11</sup> Institut für Medizinische Genetik und Pathologie, Universitätsspital Basel, Basel, Switzerland

<sup>12</sup> Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>13</sup> Department of Medical Genetics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India

<sup>14</sup> Department of Medical Genetics, Poznan University of Medical Sciences, Poznan, Poland

<sup>15</sup> Institute for Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany

<sup>16</sup> Institute of Human Genetics, Medical Faculty, RWTH Aachen University Hospital, Aachen, Germany

<sup>17</sup> Fachärztin Für Humangenetik, Bischof-von-Henle-Straße 2a, Regensburg, Germany

<sup>18</sup> Department of Human Genetics, Gilbert and Rose-Marie Chagoury School of Medicine, Lebanese American University, Byblos, Lebanon

<sup>19</sup> Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

<sup>20</sup> Kinderhandchirurgie, Medizinische Fakultät Mannheim der Universität Heidelberg, Heidelberg, Germany

<sup>21</sup> Synlab MVZ Bad Nauheim, Mondorfstr. 1761231, Bad Nauheim, Germany

<sup>22</sup> Department of Pediatric Genetics, Cerrahpasa Medical Faculty, Istanbul University-Cerrahpasa, Istanbul, Turkey

<sup>23</sup> National Center of Genetics (NCG), Laboratoire National de Santé 1, Rue Louis Rech, 3555 Dudelange, Luxembourg

<sup>24</sup> Genetic Medicine, Children's Hospital at Westmead, Paediatrics and Child Health, Sydney, Australia

<sup>25</sup> Institute of Human Genetics, University Medical Center Hamburg, Eppendorf, Germany

<sup>26</sup> Institute of Human Genetics, University of Kiel, Kiel, Germany

<sup>27</sup> Katholisches Kinderkrankenhaus Wilhelmstift, Hamburg, Germany

<sup>28</sup> Institute of Human Genetics, University of Lübeck, Lübeck, Germany

## Auszug aus der Journal Summary List Publikation 2

Journal Data Filtered By: **Selected JCR Year: 2020** Selected Editions: SCIE,SSCI  
 Selected Categories: **"GENETICS and HEREDITY"** Selected Category  
 Scheme: WoS

**Gesamtanzahl: 175 Journale**

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE REVIEWS GENETICS	42,803	53.242	0.058930
2	NATURE GENETICS	113,066	38.330	0.186810
3	TRENDS IN ECOLOGY & EVOLUTION	44,127	17.712	0.028180
4	Annual Review of Genetics	9,053	16.830	0.009270
5	MOLECULAR BIOLOGY AND EVOLUTION	61,557	16.240	0.082270
6	GENOME BIOLOGY	54,758	13.583	0.098990
7	TRENDS IN GENETICS	14,465	11.639	0.018800
8	MOLECULAR THERAPY	24,333	11.454	0.030250
9	GENES & DEVELOPMENT	61,885	11.361	0.048660
10	Genome Medicine	8,115	11.117	0.023450
11	AMERICAN JOURNAL OF HUMAN GENETICS	43,477	11.025	0.054200
12	ONCOGENE	77,576	9.867	0.059180
13	GENOME RESEARCH	47,141	9.043	0.064690
14	Annual Review of Genomics and Human Genetics	3,238	8.929	0.004800
15	GENETICS IN MEDICINE	17,957	8.822	0.041550
16	npj Genomic Medicine	971	8.617	0.003910
17	GENOMICS PROTEOMICS & BIOINFORMATICS	3,241	7.691	0.006360
18	Molecular Autism	3,579	7.509	0.007400
19	Genes & Diseases	1,850	7.103	0.003170
20	Horticulture Research	2,607	6.793	0.004610

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
21	Clinical Epigenetics	5,526	6.551	0.011550
22	JOURNAL OF MEDICAL GENETICS	15,071	6.318	0.013500
23	HUMAN MOLECULAR GENETICS	47,192	6.150	0.047520
24	CRISPR Journal	453	6.071	0.002240
25	Circulation-Genomic and Precision Medicine	858	6.054	0.003910
26	CANCER GENE THERAPY	3,768	5.987	0.002720
27	PLoS Genetics	52,461	5.917	0.085820
28	GENOMICS	11,771	5.736	0.006800
29	THEORETICAL AND APPLIED GENETICS	25,124	5.699	0.012580
30	HUMAN GENE THERAPY	7,074	5.695	0.006860
31	MUTATION RESEARCH-REVIEWS IN MUTATION RESEARCH	4,185	5.657	0.003170
32	CURRENT OPINION IN GENETICS & DEVELOPMENT	8,648	5.578	0.012100
33	Genes and Nutrition	2,090	5.523	0.001740
34	Environmental Microbiome	56	5.286	0.000100
35	GENE THERAPY	8,112	5.250	0.004130
36	CHROMOSOME RESEARCH	2,752	5.239	0.002250
37	Microbial Genomics	1,939	5.237	0.006440
38	Biology of Sex Differences	2,153	5.027	0.004070
39	GENES CHROMOSOMES & CANCER	6,108	5.006	0.004810
40	JOURNAL OF INHERITED METABOLIC DISEASE	8,207	4.982	0.006970
41	Epigenetics & Chromatin	2,461	4.954	0.007240

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
42	DNA REPAIR	7,382	4.913	0.010500
43	Forensic Science International-Genetics	6,013	4.882	0.006510
44	HUMAN MUTATION	16,381	4.878	0.019060
45	MOLECULAR GENETICS AND METABOLISM	9,612	4.797	0.008610
46	Epigenomics	3,848	4.778	0.006970
47	Human Genomics	2,039	4.639	0.003030
48	Frontiers in Genetics	18,705	4.599	0.034820
48	JOURNAL OF MOLECULAR MEDICINE-JMM	9,445	4.599	0.008480
50	JOURNAL OF GENE MEDICINE	2,294	4.565	0.001380
51	GENETICS	46,408	4.562	0.034850
52	Epigenetics	6,347	4.528	0.006530
53	DNA RESEARCH	3,726	4.458	0.003800
54	CLINICAL GENETICS	9,278	4.438	0.012790
55	CURRENT GENE THERAPY	1,576	4.391	0.001330
56	CHROMOSOMA	3,633	4.316	0.003660
57	GENETICS SELECTION EVOLUTION	4,473	4.297	0.004770
58	MOLECULAR PHYLOGENETICS AND EVOLUTION	22,497	4.286	0.022160
59	Journal of Genetics and Genomics	2,754	4.275	0.003760
60	EUROPEAN JOURNAL OF HUMAN GENETICS	12,623	4.246	0.017880
61	Briefings in Functional Genomics	2,186	4.241	0.002850
62	MITOCHONDRION	4,982	4.160	0.004690
63	HUMAN GENETICS	10,191	4.132	0.009650

# Druckexemplar der Publikation



Received: 21 July 2021 | Revised: 30 August 2021 | Accepted: 1 September 2021  
DOI: 10.1111/cge.14059

## SHORT REPORT



# *GLI3* variants causing isolated polysyndactyly are not restricted to the protein's C-terminal third

Henrike Lisa Sczakiel<sup>1,2</sup> | Wiebke Hülsemann<sup>3</sup> | Manuel Holtgrewe<sup>4</sup> |  
Angela Teresa Abad-Perez<sup>1</sup> | Jonas Elsner<sup>1</sup> | Sarina Schwartzmann<sup>1</sup> |  
Denise Horn<sup>1</sup> | Malte Spielmann<sup>2,5,6</sup> | Stefan Mundlos<sup>1,2</sup> | Martin Atta Mensah<sup>1,7</sup>

<sup>1</sup>Institute of Medical Genetics and Human Genetics, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

<sup>2</sup>RG Development & Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany

<sup>3</sup>Katholisches Kinderkrankenhaus Wilhelmstift, Handchirurgie, Hamburg, Germany

<sup>4</sup>Core Facility Bioinformatics, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>5</sup>Institute of Human Genetics, University of Lübeck, Lübeck, Germany

<sup>6</sup>Institute of Human Genetics, University of Kiel, Kiel, Germany

<sup>7</sup>BH Biomedical Innovation Academy, Digital Clinician Scientist Program, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

### Correspondence

Martin Atta Mensah, Institute of Medical Genetics and Human Genetics, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Augustenburger Platz 1, Berlin 13353, Germany.  
Email: martin-atta.mensah@charite.de

### Funding information

Berlin Institute of Health; Charité – Universitätsmedizin Berlin

### Abstract

Loss of function variants of *GLI3* are associated with a variety of forms of polysyndactyly: Pallister-Hall syndrome (PHS), Greig-Cephalopolysyndactyly syndrome (GCPS), and isolated polysyndactyly (IPD). Variants affecting the N-terminal and C-terminal thirds of the *GLI3* protein have been associated with GCPS, those within the central third with PHS. Cases of IPD have been attributed to variants affecting the C-terminal third of the *GLI3* protein. In this study, we further investigate these genotype–phenotype correlations. Sequencing of *GLI3* was performed in patients with clinical findings suggestive of a *GLI3*-associated syndrome. Additionally, we searched the literature for reported cases of either manifestation with mutations in the *GLI3* gene. Here, we report 48 novel cases from 16 families with polysyndactyly in whom we found causative variants in *GLI3* and a review on 314 previously reported *GLI3* variants. No differences in location of variants causing either GCPS or IPD were found. Review of published data confirmed the association of PHS and variants affecting the *GLI3* protein's central third. We conclude that the observed manifestations of *GLI3* variants as GCPS or IPD display different phenotypic severities of the same disorder and propose a binary division of *GLI3*-associated disorders in either PHS or GCPS/polysyndactyly.

### KEYWORDS

GCPS, genotype–phenotype correlations, *GLI3*, PHS, polydactyly, syndactyly

## 1 | INTRODUCTION

The Gli-Kruppel family member 3 (*GLI3*) gene encodes a zinc finger transcription factor that plays an important role in the sonic hedgehog signaling pathway and thus in various developmental processes

including limb development (MIM: \*165240).<sup>1,2</sup> Genetic variants in *GLI3* are well known causes of the allelic disorders Pallister-Hall syndrome (PHS; MIM: #146510), Greig-Cephalopolysyndactyly syndrome (GCPS; MIM: #175700), postaxial polydactyly type A and B (PAP-A and -B; MIM: #174200), and preaxial polydactyly type IV (PPD-IV;

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Clinical Genetics* published by John Wiley & Sons Ltd.

TABLE 1 Sixteen novel families with *GLI3*-associated disorders: Indices and their affected family members

Family ID	Mutation	Inherited variant?	Polydactyly hands	Syndactyly hands	Polydactyly feet	Syndactyly feet	Macrocephaly	Hypertelorism	Broad forehead	Broad nasal bridge	Hamartoma	Headache	Seizures	Developmental delay	Others
1	1-1 c.2374C>T, p.(Arg792)	+	PAP-A both	-	PPD-IV both	I/II both	-	-	-	-	-	-	-	-	-
1-2			PAP-A both	-	PPD-IV both	-	-	-	-	-	-	-	-	-	-
1-3			PAP-A both	-	PPD-IV both	-	-	-	-	-	-	-	-	-	-
1-4			PAP-A both	-	PPD-IV both	-	-	-	-	-	-	-	-	-	-
2	2-1 c.4172delG, p.(Gly1391Alafs*28)	+	PAP-A both	-	PAP-A and PPD-IV R, PAP-A L	-	+	+	-	-	-	-	-	-	Speech delay
2-2			-	-	+	-	+	+	+	-	-	-	-	-	Speech delay
2-3			+	-	+	-	+	+	-	-	-	-	-	-	-
3	3-1 c.2059delG, p.(Glu687Lysfs*6)	+	PAP-A both, duplicated IV R	R	Duplicated IV L, duplicated III R	both	-	-	-	-	-	-	-	-	-
3-2		n.t.	+	-	One foot	-	-	-	-	-	-	-	-	-	-
3-2		+	PAP-B both	-	+	-	-	-	-	-	-	-	-	-	-
3-4		n.t.	+	-	+	-	-	-	-	-	-	-	-	-	-
3-5		n.t.	+	-	-	-	-	-	-	-	-	-	-	-	-
3-6		n.t.	+	-	One foot	-	-	-	-	-	-	-	-	-	-
3-7		n.t.	+	-	-	-	-	-	-	-	-	-	-	-	-
4	4-1 c.1999C>T, p.(Arg667)	+	PAP-B both	II-V both	PPD-IV both	I-IV R, I-III L	-	+	-	+	-	-	-	-	-
4-2		n.t.	+	-	+	+	-	-	-	-	-	-	-	-	-
4-3		+	PAP-B both	-	-	I-II both	-	-	+	-	-	-	-	-	-
5	5-1 c.1880_1881delAT, p.(His627Argfs*48)	De novo	PAP-B both	-	PPD-IV both	I-III both	-	-	-	-	-	-	-	-	-
6	6-1 c.1793dupA, p.(Asn598Lysfs*7)	+	PPD-IV both, PAP-B both	-	PPD-IV both	II-IV both	+	+	-	-	-	-	-	-	-
6-2		+	PPD-IV both, PAP-B both	-	PPD-IV both	II-IV both	-	-	-	-	-	-	-	-	-
7	7-1 c.1028 + 1G>A, p.?	+	-	-	PPD-IV both	I-III both	-	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	CAH (simple-virilizing)
7-2		n.t.	PAP-B both	-	PPD-IV both	I-II both	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
7-3		n.t.	-	IV-V both	PPD-IV both	I-II, I-III R, IV-V both	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
7-4		n.t.	-	IV-V both	-	I-III both	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
7-5		n.t.	-	IV-V both	-	I-II both	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
8	8-1 c.366C>A, p.(Tyr122*)	+	PPD-IV both, PAP-B both	III-IV both	PPD-IV both	I-IV both	-	-	-	-	-	-	-	-	-

(Continues)

TABLE 1 (Continued)

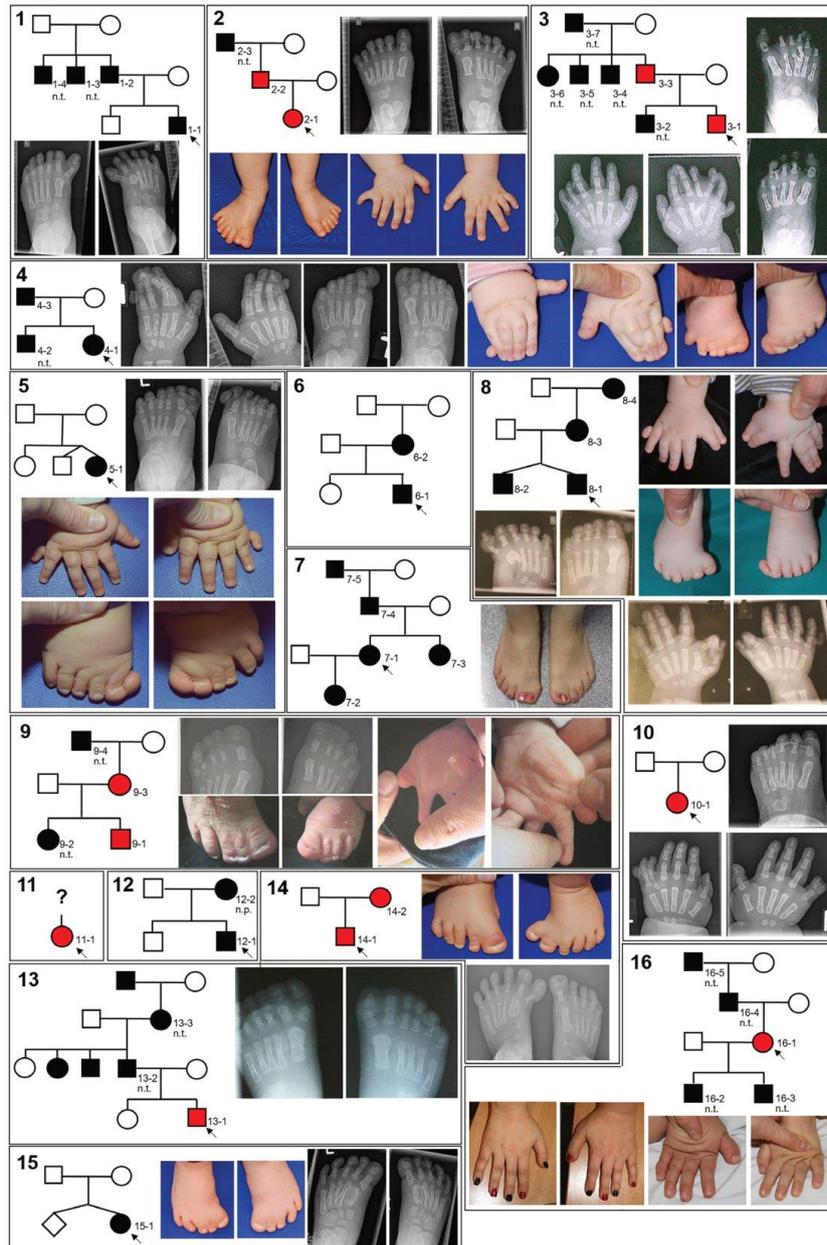
Family ID	Mutation	Inherited variant?	Polydactyly hands	Syndactyly hands	Polydactyly feet	Syndactyly feet	Macrocephaly	Hypertelorism	Broad forehead	Broad nasal bridge	Hamartoma	Headache	Seizures	Developmental delay	Others
8-2		n.t.	PPD-IV both	II-IV L, III-V R	PPD-IV both	II-III both	-	-	-	-	-	-	-	-	-
8-3		+	Synpolydactyly both hands and feet	-	Synpolydactyly both hands and feet	-	-	-	-	-	-	-	-	-	-
8-4		n.t.	Synpolydactyly both hands and feet	-	Synpolydactyly both hands and feet	-	-	-	-	-	-	-	-	-	-
9	9-1 c.1033_1048del; p.(Asn345Trpfs 15)	+	-	II-III R	-	II-III both	-	-	-	-	-	-	-	-	-
9-2		n.t.	-	+	-	+	-	-	-	-	-	-	-	-	-
9-3		+	PAP L	II-IV both	PPD-IV both	-	-	-	-	-	-	-	-	-	-
9-4		n.t.	Hexadactyly both	-	-	-	-	-	-	-	-	-	-	-	-
10	10-1 c.2103 + 2 T>A, p.?	De novo	PAP-A both	-	PAP-A L	-	-	-	-	-	-	-	-	-	-
11	11-1 c.2595C>G, p.(Ser865)	?	PAP both	Synostosis II-IV both	-	-	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
12	12-1 c.848C>T, p.(Arg290)	+	PPD-IV both, PAP-B both	III-IV L	Central polydactyly both	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
12-2		+	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
13	13-1 c.1133_1134insC, p.(Pro379Serfs 33)	?	PAP both	II-IV R	PPD-IV both	II-IV both	-	-	-	+	-	-	-	-	Elongated head
13-2		n.t.	Broad thumbs both, PAP-B both	III-IV R	PPD-IV both	II-IV both	-	-	-	+	-	-	-	-	Elongated head
13-3		n.t.	Broad thumbs, polydactyly	+	PPD-IV both	II-IV both	-	-	-	+	-	-	-	-	Elongated head
14	14-1 c.1880A>C, p.(His627Pro)	+	PAP-B both	-	PPD-IV both	III-R	-	-	-	-	-	-	-	-	Malposition of the feet
14-2			PAP-B both	-	-	-	-	-	-	-	-	-	-	-	-
15	15-1 c.2374C>T, p.(Arg792)	de novo	PPD-IV both, PAP-B both	III-V both	PPD-IV both	III-III both	-	-	-	-	-	-	-	-	-
16	16-1 c.3667_3670delinsATCAA, p.(Trp1228Ilefs 24)	?	+	-	+	-	+	-	-	-	-	-	-	-	-
16-2		n.t.	PPD-IV both, PAP-B both	+	+	+	+	-	-	-	-	-	-	-	+
16-3		n.t.	-	+	+	+	+	-	-	-	-	-	-	-	+
16-4		+	+	-	+	-	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.
16-5		n.t.	PPD-IV both	-	-	-	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.

Note: GLI3 variants refer to the transcript NM\_000148.6.

Abbreviations: I-V, respective digits; L, left; R, right; n.p., not phenotyped; n.t., not tested; PAP-A/B, postaxial polydactyly type A/B; PPD-IV, preaxial polydactyly type IV.

MIM: #174700).<sup>3,4</sup> They all feature poly- and syndactylies of varying severity. While PAP-A, PAP-B, and PPD-IV are defined by non-syndromic polysyndactyly, PHS and GCPS depict more complex

syndromic phenotypes. GCPS is characterized by polysyndactyly, macrocephaly, and facial dysmorphisms (especially hypertelorism, broad nasal bridge, high forehead, and frontal bossing). Mild mental



**FIGURE 1** Family pedigrees and images of index patients. Pedigrees: red: affected individuals carrying novel variants; solid black: affected individuals; n.t.: not tested; n.p.: not phenotyped; arrow: index patient [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

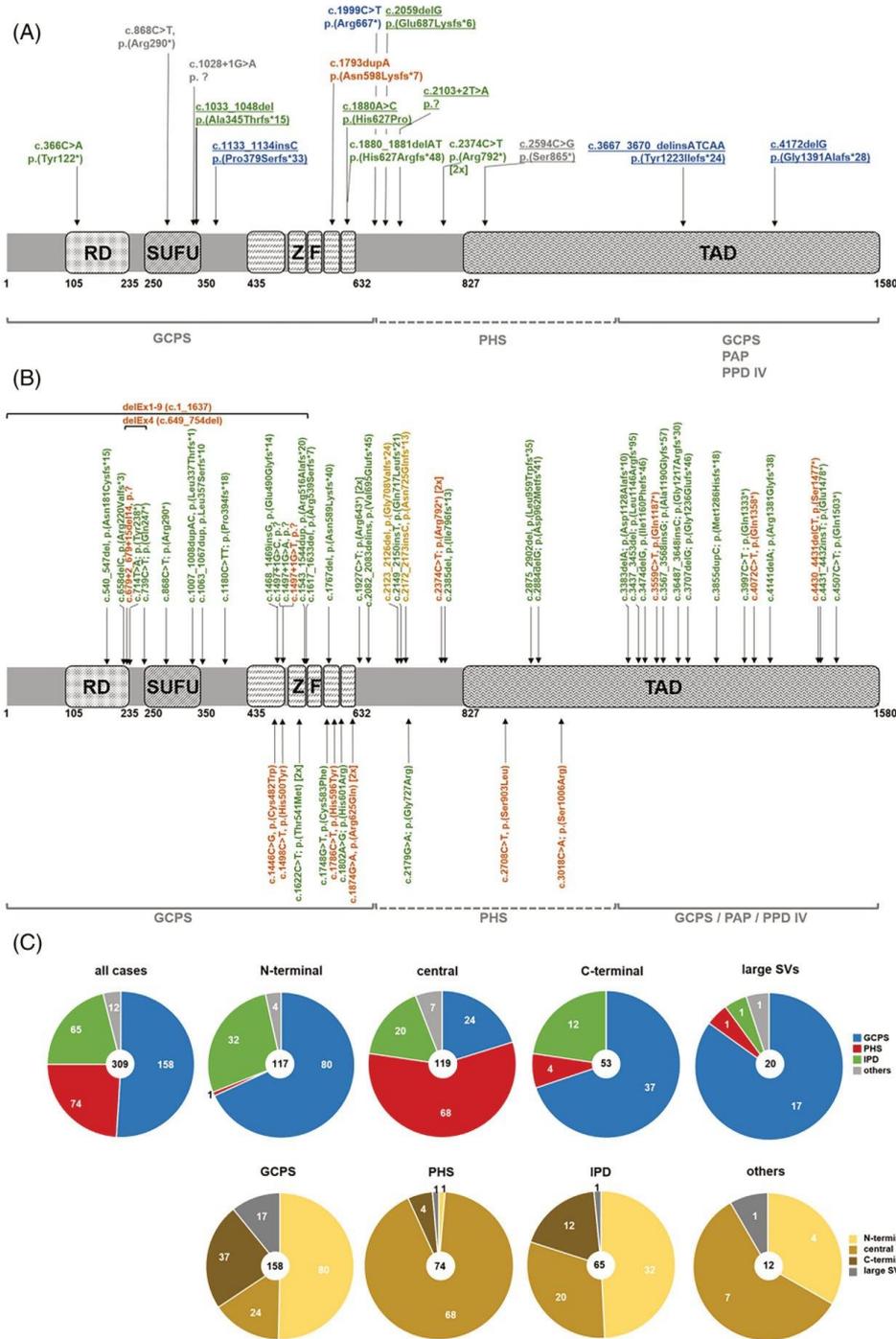


FIGURE 2 Legend on next page.

retardation, trigonocephaly and craniosynostosis as well as further rare anomalies may occur. By contrast, abnormalities characteristic for PHS are a (mostly postaxial) polysyndactyly associated with hypothalamic hamartoma, hypopituitarism, bifid epiglottis, imperforate anus, headaches, seizures, and developmental delay. Mild, incidental forms to lethal courses have been described.

The *GLI3* protein harbors two N-terminal transcriptional repressor domains (RD and SUFU), five DNA binding zinc finger domains (ZF), and a C-terminal transcriptional activator domain (TAD).<sup>1,5</sup> A certain genotype–phenotype correlation has been proposed: GCPS is predominantly attributed to variants affecting the N-terminal (amino acid position 1–660) and C-terminal third (amino acid position 1160–1580) of the *GLI3* protein, whereas PHS is attributed to variants affecting the central third (amino acid position 661–1159) of *GLI3*.<sup>6</sup> Isolated polysyndactyly (IPD), however, is supposed to result from C-terminal variants of the *GLI3* protein.<sup>7,8</sup>

## 2 | PATIENTS AND METHODS

Patients with either syndromic or nonsyndromic polysyndactyly were included in our study and phenotyped by clinical geneticists and/or primary physicians. Patients with clinical findings suggestive of a *GLI3*-associated polysyndactyly syndrome were selected for targeted Sanger sequencing of the *GLI3* gene performed on genomic DNA isolated from whole blood samples. Whenever possible trio sequencing of the affected index and both parents was applied. Raw Sanger sequencing data were analyzed using SeqPilot (JSI medical systems, USA) and variants were evaluated using ClinVar (NCBI, USA), HGMD (Qiagen Digital Insights, Denmark), and gnomAD<sup>9</sup> databases. For missense variants, an additional pathogenicity prediction was conducted using the bioinformatic prediction tools MutationTaster,<sup>10</sup> Polyphen2,<sup>11</sup> and SIFT.<sup>12</sup>

## 3 | RESULTS

We tested 94 individuals with polysyndactyly and detected 15 different causing, mostly amorphic, *GLI3* variants in 16 families (Table 1 and Table S1; Figures 1 and 2A). Eight were novel variants and seven had been reported previously. The variants spanned almost the entire coding region of *GLI3* (c.366–c.4172).

Upon identification of a causing variant in *GLI3* further and more detailed phenotypic information was collected. Eight families showed no further abnormalities beyond IPD (families 1, 3, 5, 8, 9, 10, 14, 15) whereas five families showed additional features in line with GCPS

(families 2, 4, 6, 13, 16). Three families did not take part in further phenotypic characterization (families 7, 11, 12). We did not observe a distinct correlation between type of mutation (nonsense vs. missense) and phenotype. Notably, specific phenotypes and severity varied within families. However, in the majority of families diagnosis of either IPD or GCPS was uniform within families. Only in family 6 both GCPS and IPD occurred (Table 1; Figure 2A). Subgroup analysis of polydactyly subtypes revealed a co-occurrence of PAP-B and PPD-IV in hands. In feet PPD-IV was the leading manifestation of polydactyly, in one case also accompanied by PAP-A (Figure 1 and Figure S1).

Conducting a retrospective analysis of studies on *GLI3*-associated disorders, we analyzed 309 published cases of *GLI3* variants (Table S2). Of those, 65 cases showed IPD. In 32 out of these 65 cases (49.2%) of IPD, variants affected the N-terminal third of the *GLI3* protein, while 20 cases (30.8%) harbored variants affecting the central and 12 cases (18.5%) the C-terminal third of the *GLI3* protein. One case (1.5%) was caused by a large structural variant. Interestingly, none of the identified missense variants (12 out of 65 cases) affected the C-terminal third of the *GLI3* protein (Figure 2B,C).

While GCPS cases were also associated with variants spread across the entire *GLI3* protein (79/157 (50.3%) in the N-terminal third, 24/157 (15.3%) in the central third, 37/157 (23.6%) in the C-terminal third and 17/157 (10.8%) large SVs), PHS cases showed a distinct genotype–phenotype correlation with the majority of cases (68/74 (91.9%)) harboring variants affecting the central third of the *GLI3* protein (Figure 2C).

## 4 | DISCUSSION AND CONCLUSION

In 16 families with identified *GLI3* variants, we observed GCPS as well as IPD. Notably, none of the individuals in our study presented with PHS. This can most likely be attributed to a prior selection bias since classic PHS presents with a severe phenotype.

Previous studies suggested a genotype–phenotype correlation of *GLI3*-associated disorders with GCPS being associated with variants affecting the protein's N-terminal or C-terminal third and IPD only occurring when variants affected the C-terminal third.<sup>6–8</sup> Yet, in this study, we could not observe such a distinct genotype–phenotype correlation. Strikingly, *GLI3* variants associated with IPD even occurred exclusively in the N-terminal and central third of the *GLI3* protein. Interestingly, we observed only one variant affecting the protein's C-terminal third which was associated with the presence of GCPS. Three further variants causing GCPS affect the N-terminal third of the *GLI3* protein, whereas only one variant associated with GCPS affects the protein's central third, so that locations of variants causing GCPS are

**FIGURE 2** (A) *GLI3* variants identified in 16 families and their location within the *GLI3* protein. Green: variants causing IPD. Blue: GCPS. Orange: IPD as well as GCPS. Gray: Patients could not be fully phenotyped. Underscored: novel variants. Brackets below indicate regions previously described to be associated with the respective *GLI3* disorders (B) and (C) Review of published *GLI3* variants and associated phenotypes (B) Location of *GLI3* variants causing IPD. Top: truncating variants, bottom: missense variants. Green: variants causing IPD. Orange: IPD as well as GCPS. Yellow: IPD as well as PHS. (C) Descriptive statistics of location of *GLI3* variants and the associated phenotypes. Top: affected regions of the *GLI3* protein and the phenotypes associated with each affected region. Bottom: vice versa, different *GLI3*-associated phenotypes and regions of the *GLI3* protein the respective causing variant is found in [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

overall compatible with previous reports (Figure 2A). Additionally, one individual showed GCPS whereas another individual from the same family presented with IPD. This phenomenon could also be observed in previously published cases (Figure 2B). Thus, we follow previous observations<sup>13–15</sup> and hypothesize that GCPS and IPD might be different extents of severity of the same disorder rather than distinct entities of disease. Also, we identified several variants affecting the central third of the GLI3 protein that were associated with IPD but not with PHS. These cases possibly represent mild forms of PHS similar to those reported previously.<sup>13</sup> Since brain imaging results of patients with IPD were not available, we cannot exclude the presence of such asymptomatic hamartomas.

A recent study focusing on the limb phenotypes associated with *GLI3* variants found a correlation of variants affecting the N-terminal half of *GLI3* with an anterior polydactyly phenotype (leading manifestation of PPD-IV in feet) and variants affecting the C-terminal TAD region with a posterior polydactyly phenotype (leading manifestation of PAP in hands).<sup>16</sup> The majority of the variants identified in our study located to the N-terminal half of *GLI3*, whereas only three variants affected the C-terminal TAD. With our data, we can neither confirm nor exclude these observations by Baas et al. (Figure S2).

To further evaluate a genotype–phenotype correlation regarding GCPS versus IPD, we performed a retrospective analysis of published cases with *GLI3* variants. In line with observations from our study, we found no clear genotype–phenotype correlation for IPD in these cases – with 49.2% of *GLI3* variants associated with IPD affecting the protein's N-terminal, 30.8% the central and 18.5% the C-terminal third. The same holds true for GCPS cases, whereas we could confirm the genotype–phenotype correlation for PHS with 91.9% of causing variants affecting the central third of the *GLI3* protein. Notably, nonsense variants causing IPD affected almost the entire coding region of *GLI3* (c.540–c.4507) whereas missense variants clustered in the central part of the *GLI3* protein (c.1446–c.3018) and especially the zinc-finger domains (Figure 2B). We could also confirm the previously stated observation that both GCPS and IPD are caused by varying types of variants (missense, nonsense, splice site, larger deletions).<sup>6,15</sup>

These results from our broad retrospective analysis of 314 published cases with confirmed *GLI3* variants are in line with previous studies of larger cohorts: Kalf-Suske et al. also showed that nonsense variants in *GLI3* leading to GCPS are distributed over the entire protein while Johnston et al. as well as Démurger et al. found that variants in the central third of *GLI3* cause PHS.<sup>6,17,18</sup> Also, previous studies reported variants affecting all thirds of the *GLI3* protein associated with a spectrum from IPD to GCPS.<sup>14,15</sup> Neither in our cohort nor in previous studies patients with *GLI3* variants and an isolated non-limb phenotype only (e.g., macrocephaly, hypertelorism, broad forehead) could be identified. Interestingly however, Démurger et al. reported an even asymptomatic carrier of a familial *GLI3* variant (c.427G>T, p.(Glu143\*)) highlighting the clinical variability of the phenotypes.<sup>17</sup>

Thus, IPD appears not to be restricted to cases with variants affecting the C-terminal third of the *GLI3* protein but may manifest independently of variant location. PHS, however, is strongly

correlated to variants affecting the protein's central third. Taken together, *GLI3* variants are associated with a phenotypically broad spectrum of only two distinct entities: GCPS on the one hand and PHS on the other, with IPD occurring as a mild manifestation of GCPS or in rarer cases of PHS.

#### ACKNOWLEDGEMENTS

We thank all families for their collaboration and contribution and Valerie Johnston and Gabriele Hildebrand for technical support. M.A.M. is a participant in the BIH Charité Digital Clinician Scientist Program founded by the late Prof. Duska Dragun and funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/cge.14059>.

#### DATA AVAILABILITY STATEMENT

Original sequencing data are available upon reasonable request.

#### ETHICS STATEMENT

Tested individuals provided written informed consent for participating in this study as approved by the ethical review board of the Charité – Universitätsmedizin Berlin.

#### ORCID

Martin Atta Mensah  <https://orcid.org/0000-0001-8080-8779>

#### REFERENCES

- Ruppert JM, Vogelstein B, Arheden K, Kinzler KW. *GLI3* encodes a 190-kilodalton protein with multiple regions of *GLI* similarity. *Mol Cell Biol*. 1990;10(10):5408–5415.
- Wang B, Fallon JF, Beachy PA. Hedgehog-regulated processing of *GLI3* produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell*. 2000;100(4):423–434.
- Kang S, Graham JM Jr, Olney AH, Biesecker LG. *GLI3* frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat Genet*. 1997;15(3):266–268.
- Radhakrishna U, Blouin JL, Mehenni H, et al. Mapping one form of autosomal dominant postaxial polydactyly type A to chromosome 7p15–q11.23 by linkage analysis. *Am J Hum Genet*. 1997;60(3):597–604.
- Matissek SJ, Elsawa SF. *GLI3*: a mediator of genetic diseases, development and cancer. *Cell Commun Signal*. 2020;18(1):54.
- Johnston JJ, Olivos-Glander I, Killoran C, et al. Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: robust phenotype prediction from the type and position of *GLI3* mutations. *Am J Hum Genet*. 2005;76(4):609–622.
- Biesecker LG. Strike three for *GLI3*. *Nat Genet*. 1997;17(3):259–260.
- Al-Qattan MM, Shamseldin HE, Salih MA, Alkuraya FS. *GLI3*-related polydactyly: a review. *Clin Genet*. 2017;92(5):457–466.
- Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020; 581(7809):434–443.

10. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014; 11(4):361-362.
11. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4): 248-249.
12. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016;11(1):1-9.
13. Johnston JJ, Sapp JC, Turner JT, et al. Molecular analysis expands the spectrum of phenotypes associated with GLI3 mutations. *Hum Mutat*. 2010;31(10):1142-1154.
14. Debeer P, Peeters H, Driess S, et al. Variable phenotype in Greig cephalopolysyndactyly syndrome: clinical and radiological findings in 4 independent families and 3 sporadic cases with identified GLI3 mutations. *Am J Med Genet A*. 2003;120A(1):49-58.
15. Jamsheer A, Sowińska A, Trzeciak T, Jamsheer-Bratkowska M, Geppert A, Latos-Bieleńska A. Expanded mutational spectrum of the GLI3 gene substantiates genotype-phenotype correlations. *J Appl Genet*. 2012;53(4):415-422.
16. Baas M, Burger EB, van den Ouweland AM, et al. Variant type and position predict two distinct limb phenotypes in patients with GLI3-mediated polydactyly syndromes. *J Med Genet*. 2021;58(6): 362-368.
17. Démurger F, Ichkou A, Mougou-Zerelli S, et al. New insights into genotype-phenotype correlation for GLI3 mutations. *Eur J Hum Genet*. 2015;23(1):92-102.
18. Kalf-Suske M, Wild A, Topp J, et al. Point mutations throughout the GLI3 gene cause Greig cephalopolysyndactyly syndrome. *Hum Mol Genet*. 1999;8(9):1769-1777.

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Sczakiel HL, Hülsemann W, Holtgrewe M, et al. *GLI3* variants causing isolated polysyndactyly are not restricted to the protein's C-terminal third. *Clinical Genetics*. 2021;100(6):758-765. doi: 10.1111/cge.14059

## Auszug aus der Journal Summary List Publikation 3

Journal Data Filtered By: **Selected JCR Year: 2021** Selected Editions: SCIE,SSCI  
 Selected Categories: **"MULTIDISCIPLINARY SCIENCES"** Selected Category  
 Scheme: WoS  
**Gesamtanzahl: 73 Journale**

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
1	NATURE	1,008,544	69.504	1.11428
2	SCIENCE	883,834	63.714	0.89813
3	Nature Human Behaviour	11,204	24.252	0.04187
4	National Science Review	10,508	23.178	0.01957
5	Science Bulletin	13,517	20.577	0.02141
6	Nature Communications	604,735	17.694	1.29690
7	Science Advances	104,068	14.957	0.28119
8	Journal of Advanced Research	8,207	12.822	0.00826
9	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	860,450	12.779	0.74016
10	Research	3,467	11.036	0.00574
11	Research Synthesis Methods	5,479	9.308	0.01057
12	Scientific Data	17,754	8.501	0.04575
13	GigaScience	8,120	7.658	0.01707
14	ANNALS OF THE NEW YORK ACADEMY OF SCIENCES	53,642	6.499	0.01954
15	iScience	13,293	6.107	0.02797
16	Frontiers in Bioengineering and Biotechnology	16,204	6.064	0.02051
17	Machine Learning-Science and Technology	791	6.013	0.00139
18	GLOBAL CHALLENGES	1,876	5.135	0.00443
19	Scientific Reports	696,320	4.996	1.17671
20	NPJ Microgravity	867	4.970	0.00142

## Article

# Aberrant phase separation and nucleolar dysfunction in rare genetic diseases

<https://doi.org/10.1038/s41586-022-05682-1>

Received: 9 November 2021

Accepted: 22 December 2022

Published online: 8 February 2023

Open access

 Check for updates

Martin A. Mensah<sup>1,2,3,32</sup>, Henri Niskanen<sup>4,32</sup>, Alexandre P. Magalhaes<sup>4</sup>, Shaon Basu<sup>4</sup>, Martin Kircher<sup>5,6</sup>, Henrike L. Sczakiel<sup>1,2,3</sup>, Alisa M. V. Reiter<sup>1</sup>, Jonas Elsner<sup>1</sup>, Peter Meinecke<sup>7</sup>, Saskia Biskup<sup>8</sup>, Brian H. Y. Chung<sup>9</sup>, Gregor Dombrowsky<sup>10,11</sup>, Christel Eckmann-Scholz<sup>12</sup>, Marc Phillip Hitz<sup>10,11</sup>, Alexander Hoischen<sup>13,14</sup>, Paul-Martin Holterhus<sup>15</sup>, Wiebke Hülsemann<sup>16</sup>, Kimia Kahrizi<sup>17</sup>, Vera M. Kalscheuer<sup>3</sup>, Anita Kan<sup>18</sup>, Mandy Krumbiegel<sup>19</sup>, Ingo Kurth<sup>20</sup>, Jonas Leubner<sup>21</sup>, Ann Carolin Longardt<sup>22</sup>, Jörg D. Moritz<sup>23</sup>, Hossein Najmabadi<sup>17</sup>, Karolina Skipalova<sup>1</sup>, Lot Snijders Blok<sup>14</sup>, Andreas Tzschach<sup>24</sup>, Eberhard Wiedersberg<sup>25</sup>, Martin Zenker<sup>26</sup>, Carla Garcia-Cabau<sup>27</sup>, René Buschow<sup>28</sup>, Xavier Salvatella<sup>27,29</sup>, Matthew L. Kraushar<sup>4</sup>, Stefan Mundlos<sup>1,2,3,30</sup>, Almuth Caliebe<sup>9</sup>, Malte Spielmann<sup>3,6,31,33,33</sup>, Denise Horn<sup>1,33,33</sup> & Denes Hnisz<sup>4,33,33</sup>

Thousands of genetic variants in protein-coding genes have been linked to disease. However, the functional impact of most variants is unknown as they occur within intrinsically disordered protein regions that have poorly defined functions<sup>1–3</sup>. Intrinsically disordered regions can mediate phase separation and the formation of biomolecular condensates, such as the nucleolus<sup>4,5</sup>. This suggests that mutations in disordered proteins may alter condensate properties and function<sup>6–8</sup>. Here we show that a subset of disease-associated variants in disordered regions alter phase separation, cause mispartitioning into the nucleolus and disrupt nucleolar function. We discover de novo frameshift variants in *HMGB1* that cause brachyphalangy, polydactyly and tibial aplasia syndrome, a rare complex malformation syndrome. The frameshifts replace the intrinsically disordered acidic tail of *HMGB1* with an arginine-rich basic tail. The mutant tail alters *HMGB1* phase separation, enhances its partitioning into the nucleolus and causes nucleolar dysfunction. We built a catalogue of more than 200,000 variants in disordered carboxy-terminal tails and identified more than 600 frameshifts that create arginine-rich basic tails in transcription factors and other proteins. For 12 out of the 13 disease-associated variants tested, the mutation enhanced partitioning into the nucleolus, and several variants altered rRNA biogenesis. These data identify the cause of a rare complex syndrome and suggest that a large number of genetic variants may dysregulate nucleoli and other biomolecular condensates in humans.

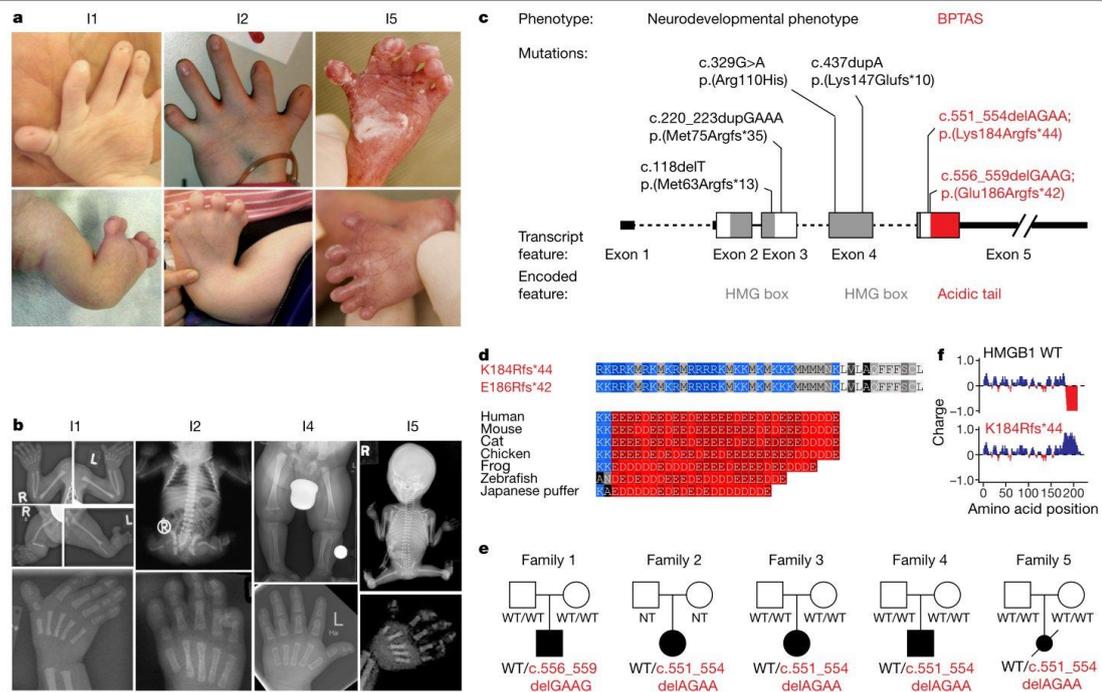
Monogenic and common diseases are frequently associated with mutations in transcriptional regulatory proteins, including DNA-binding transcription factors. However, the functional impact of the majority of such mutations is unknown, and many complex diseases still lack a clear underlying genetic component<sup>1,9–12</sup>. We initially set out to identify the molecular basis of brachyphalangy, polydactyly and tibial aplasia/hypoplasia syndrome (BPTAS; Online Mendelian Inheritance in Man database identifier: 609945), an extremely rare complex malformation syndrome with an as yet unknown molecular aetiology<sup>13–19</sup>. During the study, five individuals (II–I5) were diagnosed with BPTAS. All five exhibited a distinct skeletal phenotype, including short and malformed lower limbs characterized by tibia aplasia or hypoplasia, preaxial polysyndactyly and contractures of large joints (Fig. 1a,b). In all five individuals, anomalies of the upper limbs were less severe compared with those of the lower limbs, and included brachydactyly or brachyphalangy of fingers with an irregular finger length (Fig. 1a,b).

Short radius and ulna and contractures or pterygia of the elbow joints were present in four out of five individuals. All individuals with BPTAS diagnosed during our study or described in previous reports also presented with distinct craniofacial, neurological and genitourinary features. Phenotypic findings are summarized in Supplementary Table 1. Detailed clinical and family histories are provided in Supplementary Note and Extended Data Fig. 1.

## De novo *HMGB1* frameshifts in BPTAS

We performed genome sequencing of II and detected a potentially pathogenic variant: the heterozygous frameshift NM\_002128.7(*HMGB1*): c.556\_559delGAAG;p.(Glu186Argfs\*42) in the final exon of *HMGB1* (Extended Data Fig. 2a). *HMGB1* encodes a highly conserved, low-specificity DNA binding factor associated with cell signalling<sup>20</sup>, cell motility<sup>21,22</sup>, base excision repair<sup>23,24</sup> and chromatin looping<sup>25</sup>.

A list of affiliations appears at the end of the paper.



**Fig. 1 | De novo frameshifts in *HMGB1* cause BPTAS. a**, Photographs of individuals diagnosed with BPTAS. Top row, hands of I1, I2 and I5. Note brachydactyly, irregular finger length and hypoplasia of the nails. Bottom row, lower extremities of I1, I2 and I5, presenting with malformed legs, joint contractures, preaxial polysyndactyly and hypoplasia of the nails. **b**, Radiographs of I1, I2, I4 and I5. Top far left, limb radiograms (at newborn age) of I1 showing brachydactyly and brachyphalangy, tibial aplasia, hypoplastic fibulae and preaxial polysyndactyly. Top middle left, babygram of I2. Note tibial aplasia, hypoplastic and absent fibulae, hypoplastic pelvic bones and hypoplastic right femur. Top middle right, lower extremities of I4 (at 6 months) showing asymmetric shortness of tibiae and fibulae. Top far right, fetogram of I5 showing tibial aplasia, hypoplastic and absent fibulae, hypoplastic pelvic bones and contractures of joints. Bottom row, hand radiographs of I1, I2 (both at newborn age), I4 (at 6 months)

Sanger sequencing of I1 and his parents confirmed the presence of the frameshift variant and revealed de novo occurrence (Fig. 1c–e and Extended Data Fig. 2b). Sanger sequencing of *HMGB1* in I2 and I3, and trio exome sequencing of I4 identified a similar de novo heterozygous frameshift: NM\_002128.7 (*HMGB1*):c.551\_554delAGAA;p.(Lys184Argfs\*44), a variant also detected in a previously described female fetus<sup>26</sup> (I5) (Fig. 1c–e and Extended Data Fig. 2c,d). Sequencing of cDNA from a lymphoblastoid cell line derived from peripheral blood cells from I3 confirmed the presence of both wild-type and mutant *HMGB1* transcripts (Extended Data Fig. 2e). The two frameshift mutations result in almost identical, positively charged sequences (Fig. 1d,f).

#### Altered *HMGB1* phase separation in vitro

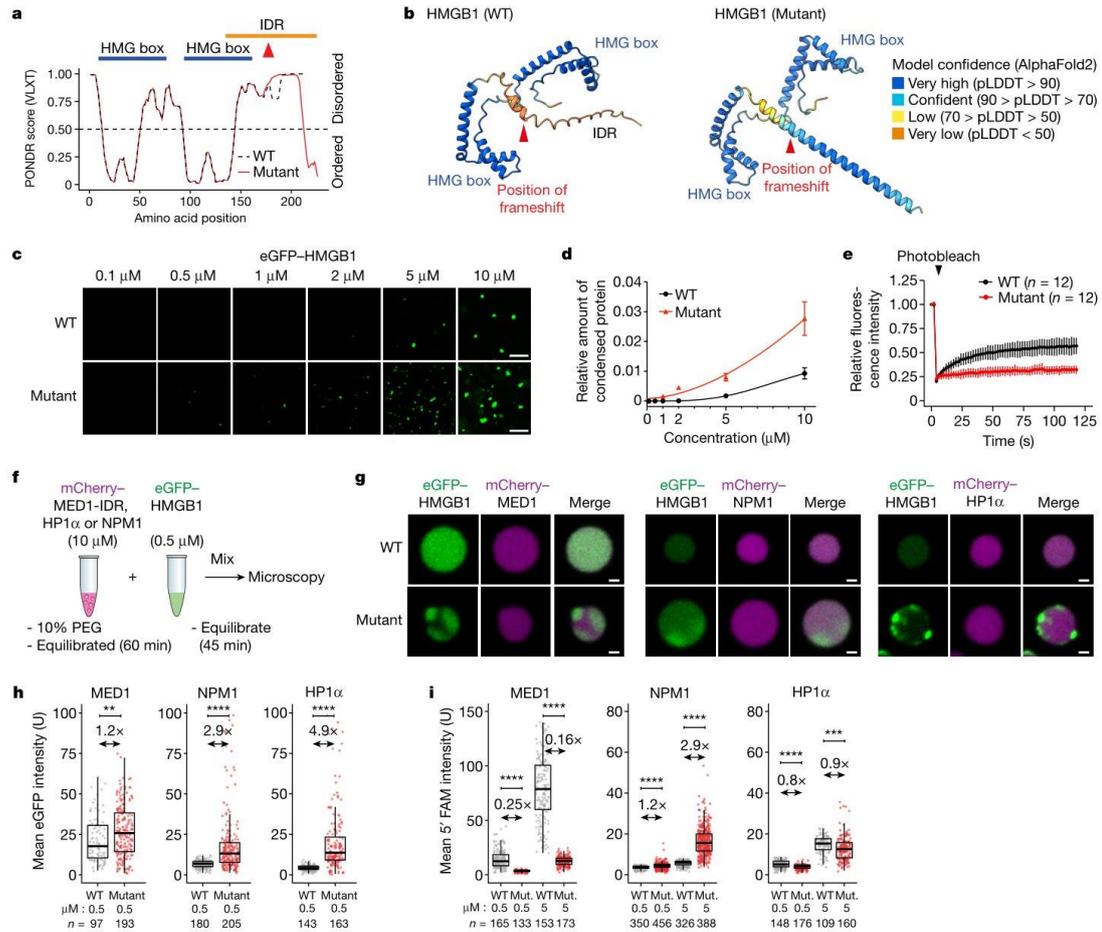
To investigate the potential pathogenic role of the frameshift variant in *HMGB1*, we first explored structural and sequence features of the wild-type and mutant proteins. *HMGB1* is a low-specificity DNA-binding protein that contains two HMG boxes that are responsible for DNA binding<sup>27</sup> and a C-terminal acidic tail (Fig. 2a,b). The acidic tail is

and of I5 (at 21 weeks of gestation). Note the short middle phalanges and short proximal phalanges of the thumbs. **c**, Pathogenic frameshift variants in the acidic tail of *HMGB1* in the individuals with BPTAS reported in this article are highlighted in red. Previously reported variants associated with developmental delay are in black. Note the genotype–phenotype correlation: C-terminal frameshifts result in BPTAS, whereas other variants lead to a neurodevelopmental phenotype. **d**, Amino acid sequence of the C terminus of *HMGB1* in individuals with BPTAS and in selected vertebrates. Acidic residues glutamate and aspartate are shaded in red, basic residues arginine and lysine are shaded in blue. Note the replacement of the conserved acidic tail in individuals with BPTAS. **e**, Family pedigrees. Individuals with BPTAS are highlighted with black boxes, and the genotypes are below the boxes. **f**, Charge plots of WT and mutant *HMGB1*. I, individual; L, left; NT, not tested; R, right; WT, wild type.

predicted to be intrinsically disordered and resides within an approximately 60-amino-acid long conserved intrinsically disordered region (IDR), as revealed by AlphaFold2 and PONDR analyses (Fig. 2a,b and Extended Data Fig. 3a). Both algorithms predicted a slight propensity of the C-terminal portion of the IDR to assume a helical conformation in the frameshift mutant *HMGB1* (Fig. 2a,b and Extended Data Fig. 3b). This prediction was confirmed by circular dichroism experiments on synthetic peptides that corresponded to the C-terminal 80–90 amino acid region (Extended Data Fig. 3c–e). IDRs of numerous proteins, including transcription factors, co-activators (for example, Mediator) and RNA polymerase II (RNAPII), contribute to phase separation by mediating multivalent low-affinity interactions<sup>6,28–31</sup>. Therefore, we hypothesized that the potentially BPTAS-causing frameshift may alter the phase-separation capacity of *HMGB1*.

To test the phase-separation capacity of *HMGB1*, we purified recombinant *HMGB1* proteins tagged with enhanced green fluorescent protein (eGFP) and examined their behaviour in vitro. Wild-type full-length *HMGB1* formed droplets in the presence of a crowding agent (10% polyethylene glycol (PEG)), and the number and size of droplets scaled with

## Article



**Fig. 2 | ABPTAS-causing frameshift alters HMGB1 phase separation in vitro.**

**a**, Graph plotting the intrinsic disorder of HMGB1. Red arrowhead shows the position of the BPTAS frameshift. The position of the IDR is highlighted with an orange bar and the position of HMG boxes with blue bars. **b**, Structures of WT and mutant HMGB1 predicted with AlphaFold2. Colours ranging from blue to orange depict the per-residue measure of local confidence (pLDDT) for the model. **c**, Representative images from droplet formation assays of eGFP-HMGB1 variants at the indicated concentrations. The experiment was repeated three times, with similar results obtained. **d**, Quantification of the relative amount of condensed protein at the indicated concentrations. Data displayed as the mean ± s.d. **e**, Relative fluorescence intensity of the bleached area from

eGFP-HMGB1 condensates before and after photobleaching. Data displayed as the mean ± s.d. **f**, Scheme of co-droplet assays. **g**, Representative images of eGFP-HMGB1 proteins mixed with preassembled mCherry-labelled MED1-IDR, HP1α or NPM1 droplets. **h, i**, Quantification of eGFP (**h**) and 5' FAM (**i**) fluorescence intensity in mCherry-labelled MED1-IDR, HP1α and NPM1 droplets mixed with full-length mEGFP-HMGB1 proteins (**h**) or 5' FAM-HMGB1-IDR peptides (**i**). Fold change values between the mean intensities of WT and mutants (Mut.) are indicated above the plot. Median is shown as a line within the boxplot, which spans from the 25th to 75th percentiles. Whiskers depict a 1.5× interquartile range. *P* values are from two-tailed Welch's *t*-test. \*\**P* < 1 × 10<sup>-2</sup>, \*\*\**P* < 1 × 10<sup>-3</sup>, \*\*\*\**P* < 1 × 10<sup>-4</sup>. Scale bars, 5 μm (**e**) and 10 μm (**g**).

the concentration of the protein (Fig. 2c, d and Extended Data Fig. 4a–c). The droplets were spherical, settled on the surface and occasionally underwent fusion (Supplementary Video 1), which are hallmarks of phase separation<sup>32</sup>. By contrast, the frameshift mutant HMGB1 formed amorphous condensates that appeared at a lower saturation concentration (Fig. 2c, d) and, after photobleaching, recovered fluorescence slower than wild-type HMGB1 droplets (Fig. 2e). Similar results were observed using synthetic peptides that corresponded to the C-terminal 80–90 amino acid region of wild-type and frameshift mutant HMGB1 (Extended Data Fig. 4d–g). These results indicate that the frameshift in HMGB1 enhances condensate formation and alters condensate properties in vitro.

Mammalian nuclei contain numerous biomolecular condensates, for example, the nucleolus, heterochromatin, co-activator and RNAPII condensates<sup>4,5</sup>. IDRs play important roles in the partitioning of proteins into nuclear condensates<sup>4,5</sup>. We therefore tested whether the frameshift mutation alters the partitioning of HMGB1 into nuclear condensates. Using purified marker proteins, we assembled the following model condensates: recombinant mCherry-tagged MED1 IDR droplets as an in vitro model for Mediator co-activator condensates<sup>29,31,33</sup>; mCherry-tagged HP1α droplets as an in vitro model for heterochromatin<sup>34</sup>; and mCherry-tagged NPM1 droplets as an in vitro model for the granular component of the nucleolus<sup>35</sup>. Wild-type and



## Article

mutant HMGB1 localized to discrete nuclear inclusions (Fig. 3a, Extended Data Fig. 5a,b and Supplementary Video 2). Ectopic expression of the mutant HMGB1 IDR also led to the formation of nuclear inclusions in live U2OS cells (Fig. 3a and Extended Data Fig. 5a–c), which indicated that the replaced IDR of the mutant HMGB1 is responsible for its altered subnuclear localization. Nuclear inclusions were observed in several other human cell types expressing mutant HMGB1 (Extended Data Fig. 5d).

Mutant HMGB1 nuclear inclusions frequently contained cavities and resembled nucleoli. Nucleoli are phase-separated multiphasic condensates that contain an outer granular component enriched in NPM1 and an inner dense fibrillar component enriched in FIB1 (ref. 35) (Fig. 3b). To gain initial insights into the nature of the mutant HMGB1 nuclear inclusions, we expressed FIB1 tagged with red fluorescent protein (RFP–FIB1) and eGFP–HMGB1 in live U2OS cells. The cavities in the mutant HMGB1 inclusions tended to encapsulate FIB1 (Fig. 3c). Fluorescence recovery after photobleaching (FRAP) experiments revealed that the HMGB1 shell displayed arrested dynamics around the FIB1 cores (Fig. 3d and Extended Data Fig. 5e). These results suggest that the mutant HMGB1 inclusions may be abnormal, arrested nucleoli.

To further probe the identity of mutant HMGB1 nuclear inclusions, we performed immunofluorescence against various nuclear proteins known to form condensates. The immunofluorescence analyses revealed that mutant HMGB1 inclusions were distinct from RNAPII and MED1 puncta, nuclear speckles and heterochromatin (Extended Data Fig. 5f–h). However, they overlapped with NPM1 and FIB1 (Extended Data Fig. 5h). The NPM1 signal within the HMGB1 inclusions inversely correlated with the HMGB1 signal (Pearson's  $r = -0.70$ ) (Extended Data Fig. 5i). Moreover, the amount of diffuse NPM1 outside nucleoli correlated with the amount of HMGB1 in the inclusions (Pearson's  $r = 0.50$ ) (Extended Data Fig. 5j). These results indicate that the mutant HMGB1 inclusions replace the NPM1-enriched granular component of nucleoli.

Targeted mutagenesis experiments revealed that arginine residues in the mutant HMGB1 tail drive nucleolar mispartitioning, and a hydrophobic patch drives nucleolar arrest. Various mutant HMGB1 sequences were expressed in live U2OS cells. A HMGB1 protein lacking the entire IDR (Del IDR) or the sequence after the frameshift position (Del FS) was not enriched in the nucleolus (Fig. 3e,f). Deletion of arginine residues (R del), substitution of arginine residues with alanine residues (R>A), substitution of arginine and lysine residues with alanine residues (R&K>A), and substitution of arginine residues with lysine residues (R>K) within the sequence created by the frameshift led to failure of the mutant protein to partition into the nucleolus (Fig. 3e,f). Furthermore, deletion of the short hydrophobic patch at the C terminus of the frameshifted sequence (Patchless) did not alter nucleolar mispartitioning (Fig. 3e,f and Extended Data Fig. 5k), but it did rescue the arrested dynamics of the mutant HMGB1 nucleoli assessed by FRAP (Fig. 3e,g). These results demonstrate that nucleolar mispartitioning of the frameshift mutant HMGB1 depends on arginine residues within the sequence created by the frameshift and that the hydrophobic patch contributes to nucleolar arrest.

### Mutant HMGB1 and nucleolar dysfunction

To test whether the nucleolar mispartitioning of HMGB1 affects nucleolar function, we investigated ribosomal RNA (rRNA) production using quantitative PCR with reverse transcription (RT–qPCR)<sup>36</sup>. The level of 28S rRNA in U2OS cells expressing the frameshift mutant HMGB1 was significantly reduced by about 1.5-fold ( $P < 0.05$ , Student's  $t$ -test) (Extended Data Fig. 6a). Ribosomal dysfunction was subsequently probed using an assay of nascent translation that measures puromycin incorporation<sup>37</sup>. U2OS cells expressing mutant eGFP–HMGB1 consistently displayed lower levels of puromycin intensity than non-transfected (that is, GFP<sup>-</sup>) cells and cells transfected with wild-type eGFP–HMGB1 (Fig. 3h,i and Extended Data Fig. 6b,c). Furthermore,

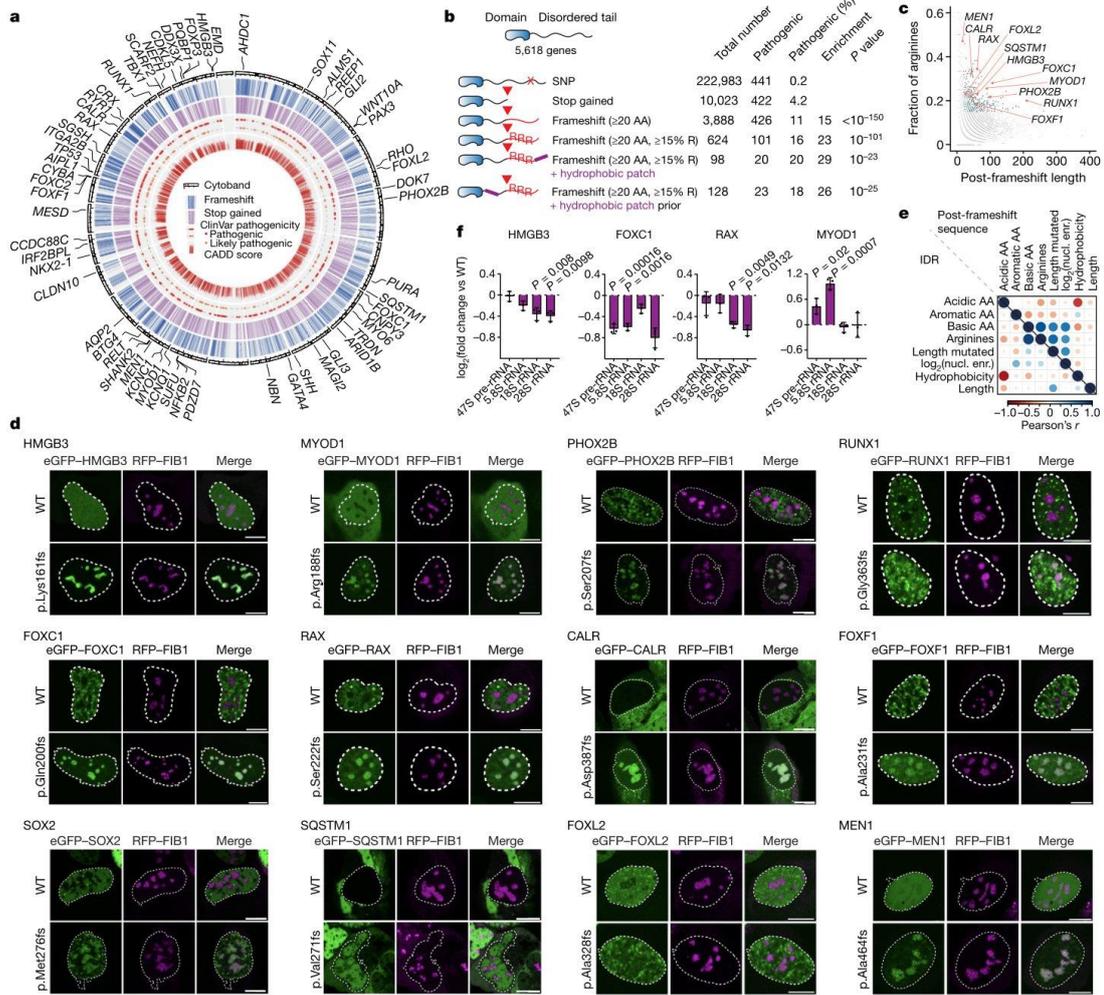
U2OS cells expressing the mutant HMGB1 exhibited substantially reduced viability after several days of culture compared with cells expressing wild-type HMGB1 ( $P < 5 \times 10^{-4}$ , one-way analysis of variance (ANOVA)) (Fig. 3j and Extended Data Fig. 6d). The reduced viability was associated with nucleolar arrest, as transfection of cells with the Patchless mutant did not compromise viability (Fig. 3j and Extended Data Fig. 6d). The findings of nucleolar mispartitioning, nucleolar arrest and viability were corroborated using cell lines expressing stably integrated eGFP–HMGB1 transgenes from a PiggyBac transposon (Extended Data Fig. 6e–j). These results indicate that the presence of the HMGB1 frameshift mutant in cells disrupts nucleolar function and is cytotoxic.

### ACMG classification of HMGB1 variants

The clinical and genetic information of the five individuals with BPTAS and the functional data were used to classify the *HMGB1* frameshift variants as pathogenic. This classification was made in accordance with the criteria of the American College of Medical Genetics and Genomics (ACMG)<sup>38</sup>. Both frameshifts observed in individuals with BPTAS result in the replacement of the highly conserved acidic tail of the protein (ACMG criterion PM1), and were classified as pathogenic by MutationTaster (ACMG criterion PP3). Notably, of the 43 nonsynonymous variants in the HMGB1 tail (1,123 alleles) listed in the gnomAD database (v.2.1.1)<sup>39</sup>, only 4 variants (5 alleles) introduce amino acids other than aspartate and glutamate (ACMG criterion PM2) (Extended Data Fig. 2f,g). All previously described pathogenic *HMGB1* variants are associated with neurodevelopmental phenotypes without severe skeletal anomalies, which are therefore distinct from BPTAS (Fig. 1c and Supplementary Table 2), including a chromosomal microdeletion encompassing the *HMGB1* locus in an individual (I6) diagnosed in our study (Extended Data Fig. 2h–j, Supplementary Note and Supplementary Table 2). The functional data presented in this study suggest a deleterious effect (ACMG criterion PS3). In summary, the identification of almost the same (ACMG criterion PS4) *HMGB1* frameshift variants in five (shown to be de novo in four; ACMG criteria PS2 and PM6) unrelated individuals with the same ultrarare diagnosis of BPTAS (ACMG criterion PP4) argues for the classification of these variants as pathogenic (ACMG evidence level: 3S+2M+2P; Supplementary Note).

### Catalogue of variants in C-terminal IDRs

We then sought to investigate whether replacement of a disordered C-terminal tail with an arginine-rich basic tail and the consequent nucleolar mispartitioning and dysfunction could occur in other diseases. To this end, we generated a catalogue of genetic variants in intrinsically disordered tails of cellular proteins. First, we annotated 9,303 isoforms of 5,618 genes that have a C-terminal IDR consisting of at least 20 amino acids (Supplementary Table 3). We then identified genetic variants that occur in the disordered tails of the 5,618 genes annotated in the 1000 Genomes Project, ClinVar, COSMIC and dbSNP databases. These analyses revealed 249,464 genetic variants in C-terminal IDRs, including 10,023 truncating variants and 3,888 frameshifts that replace the C-terminal sequence with  $\geq 20$  amino acids (Fig. 4a, Extended Data Fig. 7a–c and Supplementary Tables 4 and 5). Of the 3,888 frameshifts, 426 were annotated as pathogenic in ClinVar, 763 were common variants curated in the 1000 Genomes Project and 189 in the dbSNP databases (Fig. 4b, Extended Data Fig. 7b,c and Supplementary Table 5). The frameshifts were associated with higher-than-average pathogenicity (Fig. 4a), and frameshifts were enriched for pathogenic variants (Fig. 4b and Extended Data Fig. 7c). Genes encoding transcription factors were highly enriched among those that contained C-terminal IDR mutations (Extended Data Fig. 7d). Among the 3,888 frameshift variants, 624 were predicted to result in a sequence consisting of at least 15% arginine residues, of which 101 were classified as pathogenic in ClinVar (Fig. 4b,c



**Fig. 4 | A catalogue of variants in C-terminal IDRs reveals frameshifts associated with nucleolar mispartitioning and dysfunction. a**, Circos plot of the IDR variant catalogue. The circles indicate the location of genes that contain a truncation (stop gained) or frameshift variant in the dbSNP, 1000 Genomes Project, COSMIC and ClinVar databases. The highlighted genes contain a pathogenic frameshift that creates a sequence of  $\geq 20$  amino acids comprising  $\geq 15\%$  arginine residues. **b**, Summary statistics and features of variant types in C-terminal IDRs. *P* values are from hypergeometric tests. **c**, Identification of frameshifts creating a sequence of  $\geq 20$  amino acids that consist of  $\geq 15\%$  arginine residues. Plotted is the fraction of arginine residues against the length of the sequence created by the frameshift. The genes containing the variants selected for further validation are highlighted orange. Pathogenic gene variants are in blue. **d**, Representative images of U2OS cells co-expressing RFP-FIB1 and the indicated eGFP-tagged proteins. Nuclear area revealed by Hoechst staining is shown as dashed white lines. Mutations in the following genes are associated with the indicated conditions: microphthalmia (*HMGB3* and *RAX*); myopathy (*MYOD1*); congenital central hypoventilation (*PHOX2B*); myelodysplasia (*RUNX1*); Axenfeld-Rieger syndrome type 3 (*FOXC1*);

myofibrosis (*CALR*); alveolar capillary dysplasia (*FOXF1*); anophthalmia/microphthalmia-oesophageal atresia syndrome (*SOX2*); Paget disease of bone 2, early-onset frontotemporal dementia and amyotrophic lateral sclerosis (*SQSTM1*); blepharophimosis, ptosis and epicanthus inversus (*FOXL2*); and hereditary cancer predisposing syndrome (*MEN1*). **e**, Nucleolar mispartitioning strongly correlates with the fraction of arginine residues in the frameshift sequence. Plotted are Pearson's correlation coefficients of the extent of nucleolar mispartitioning of mutant proteins with protein features of their IDRs (left triangle) and features of the sequences created by the frameshifts (right triangle). The colour corresponds to the value of Pearson's correlation coefficients, and the size of the circles is proportional to the *P* value of the Pearson's *r*. **f**, RT-qPCR analysis of rRNA species in U2OS cells expressing the indicated WT and mutant proteins. rRNA levels are normalized against an RNAPII transcript (*GAPDH*), and fold changes are calculated against the rRNA/*GAPDH* level measured in the cells expressing WT protein. Data are shown as mean  $\pm$  s.d., *P* values are from two-tailed Welch's *t*-test. AA, amino acid; SNP, single nucleotide polymorphism; nucl. enr., nucleolar enrichment.

## Article

and Supplementary Fig. 2a–g). Overall, 29 out of 66 genes containing arginine-rich frameshift variants had a probability of loss-of-function intolerance (pLI) score of <0.05, which is consistent with a potential gain-of-function effect of the variants (Extended Data Fig. 7e). The variants were associated with various pathogenic conditions, including neurodevelopmental diseases and cancer predisposition (Extended Data Fig. 7f–h). Moreover, 98 of the frameshifts also created a sequence resembling the short hydrophobic patch encoded by the HMGB1 frameshift (Fig. 4b), and 128 of the frameshifts occurred in genes that contained at least one hydrophobic patch in their IDR (Fig. 4b). Overall, the catalogue revealed >200,000 variants in C-terminal IDRs, including 624 frameshifts that replace a C-terminal tail with an arginine-rich basic tail, of which 101 frameshifts were classified as pathogenic.

Genes containing pathogenic frameshift variants that create an arginine-rich basic tail were expressed in U2OS cells. As such frameshifts are highly enriched in genes that encode transcription factors, we selected nine transcription factors (HMGB3, FOXC1, FOXF1, MYOD1, RAX, RUNX1, SOX2, PHOX2B and FOXL2) and four additional proteins (MEN1, SQSTM1, CALR and DVL1) for functional testing (Fig. 4d, Extended Data Figs. 8 and 9a–d and Supplementary Fig. 3a–d). The frameshift mutants of 12 out of the 13 proteins formed nuclear inclusions that overlapped the FIB1–RFP-labelled dense fibrillar component of the nucleolus in live cells (Fig. 4d and Extended Data Fig. 9e–i). The extent of mispartitioning into the nucleolus strongly correlated with the length of the IDR sequence replaced by the frameshift and the fraction of arginine residues in the sequence created by the frameshifts (Fig. 4e and Supplementary Fig. 4a,b). For six variant proteins, cavities enriched in FIB1–RFP were apparent (Fig. 4d and Extended Data Fig. 10a). FRAP experiments showed that condensate properties for 7 out of the 13 variants were affected (Extended Data Figs. 9g and 10b). Six of the mutant proteins that showed significant nucleolar enrichment were further analysed. For four out of six, changes in the level of rRNA species in cells expressing the frameshift mutants were detectable (Fig. 4f and Extended Data Fig. 10c,d). These results indicate that disease-associated frameshifts that generate an arginine-rich basic tail in C-terminal IDRs can cause nucleolar mispartitioning and dysfunction.

## Discussion

We propose that disease-associated and common variants in disordered regions may alter phase separation and partitioning of proteins into biomolecular condensates. In particular, the results presented here indicate that frameshift variants that substantially increase the arginine content of various proteins lead to mispartitioning into the nucleolus and disruption of nucleolar function. Our data identified the replacement of the disordered tail with an arginine-rich basic tail in HMGB1 as the pathomechanism underlying BPTAS, a rare complex malformation syndrome<sup>13</sup>. The HMGB1 variant appears to encode a sequence that combines high arginine content, reminiscent of the phase-separation grammar of native nucleolar proteins<sup>40</sup>, and a hydrophobic patch that predominantly contributes to nucleolar arrest and dysfunction (Fig. 3e–j). The frameshift therefore interferes with the ‘molecular grammar’ of phase separation encoded in HMGB1, and the resulting mutant protein disrupts condensate features and function of the nucleolus where it accumulates. The extent to which the minimal propensity of the HMGB1 mutant sequence to form a helix contributes to these effects remains to be tested.

We provided evidence that arginine-rich frameshifts occur in hundreds of proteins, which implies that there is a common mechanism for hundreds of disease-associated and common genetic variants with previously unknown functions. The organismal effects of such frameshifts are probably influenced by tissue-specific expression and haploinsufficiency (or haploinsufficiency) of the genes in which they occur. For example, BPTAS is associated with a frameshift in HMGB1 that is broadly expressed and haploinsufficient (pLI score of 0.83),

which is consistent with phenotypic features presenting in multiple organ systems and partially overlapping with those seen when the locus is deleted (Supplementary Note). Of note, mispartitioning into the nucleolus and nucleolar dysfunction have been reported for poly-(proline:arginine)-dipeptides produced by repeat-expanded variants of *C9orf72* linked to amyotrophic lateral sclerosis<sup>36,41,42</sup>. Aberrant phase separation and nucleolar dysfunction may therefore occur in a wide range of genetic conditions as a shared underlying molecular pathomechanism.

Finally, the IDR variant catalogue provides a resource for exploring further models of how disease-associated variants may alter biomolecular condensates. For example, the >10,000 variants that truncate a C-terminal IDR may inhibit biogenesis of condensates, and several such variants have been associated with condensate dissolution in cultured cells<sup>43</sup>. Disease-associated alanine repeat expansions in a few transcription factors have been shown to alter the composition of their condensates<sup>9</sup>, and our catalogue contains >200 frameshift sequences consisting of at least 25% alanine residues. In summary, we propose that disruption of phase separation may frequently occur in genetic diseases. Further investigation of the underlying molecular basis may lead to future strategies that alter phase separation with therapeutic intent.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-05682-1>.

- Clausnitzer, M. et al. A brief history of human disease genetics. *Nature* **577**, 179–189 (2020).
- Tsang, B., Pritisnanac, I., Scherer, S. W., Moses, A. M. & Forman-Kay, J. D. Phase separation as a missing mechanism for interpretation of disease mutations. *Cell* **183**, 1742–1756 (2020).
- Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
- Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
- Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
- Basu, S. et al. Unblending of transcriptional condensates in human repeat expansion disease. *Cell* **181**, 1062–1079.e30 (2020).
- Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
- Patel, A. et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).
- Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2015).
- Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
- Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Kandath, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
- Baraitser, M. et al. A syndrome of brachyphalangy, polydactyly and absent tibiae. *Clin. Dysmorphol.* **6**, 111–121 (1997).
- Faravelli, F., Di Rocco, M., Stella, G., Selicorni, A. & Camera, G. Brachyphalangy, feet polydactyly, absent/hypoplastic tibiae: a further case and review of main diagnostic findings. *Clin. Dysmorphol.* **10**, 101–103 (2001).
- Pierson, D. M. et al. Total anomalous pulmonary venous connection and a constellation of craniofacial, skeletal, and urogenital anomalies in a newborn and similar features in his 36-year-old father. *Clin. Dysmorphol.* **10**, 95–99 (2001).
- Olney, R. S. et al. Limb/pelvis hypoplasia/aplasia with skull defect (Schinzel phocomelia): distinctive features and prenatal detection. *Am. J. Med. Genet.* **103**, 295–301 (2001).
- Wechsler, S. B., Lehoczy, J. A., Hall, J. G. & Innis, J. W. Tibial aplasia, lower extremity mirror image polydactyly, brachyphalangy, craniofacial dysmorphism and genital hypoplasia: further delineation and mutational analysis. *Clin. Dysmorphol.* **13**, 63–69 (2004).
- Bernardi, P. et al. Additional features in a new case of a girl presenting brachyphalangy, polydactyly and tibial aplasia/hypoplasia. *Am. J. Med. Genet. A* **149A**, 1532–1538 (2009).
- Shafeghati, Y. et al. Brachyphalangy, polydactyly and tibial aplasia/hypoplasia syndrome (OMIM 609945): case report and review of the literature. *Eur. J. Pediatr.* **169**, 1535–1539 (2010).
- Itou, J. et al. HMGB factors are required for posterior digit development through integrating signaling pathway activities. *Dev. Dyn.* **240**, 1151–1162 (2011).
- Yanai, H. et al. HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature* **462**, 99–103 (2009).

22. Scaffidi, P., Misteli, T. & Bianchi, M. E. Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature* **418**, 191–195 (2002).
23. Bianchi, M. E., Beltrame, M. & Paonessa, G. Specific recognition of cruciform DNA by nuclear protein HMGB1. *Science* **243**, 1056–1059 (1989).
24. Prasad, R. et al. HMGB1 is a cofactor in mammalian base excision repair. *Mol. Cell* **27**, 829–841 (2007).
25. Sofiadis, K. et al. HMGB1 coordinates SASP-related chromatin folding and RNA homeostasis on the path to senescence. *Mol. Syst. Biol.* **17**, e9760 (2021).
26. Elsner, J. et al. Genome sequencing in families with congenital limb malformations. *Hum. Genet.* **140**, 1229–1239 (2021).
27. Lambert, S. A. et al. The human transcription factors. *Cell* **175**, 598–599 (2018).
28. Boehning, M. et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
29. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).
30. Cho, W. K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
31. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
32. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and challenges in studying liquid–liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).
33. Asimi, V. et al. Hijacking of transcriptional condensates by endogenous retroviruses. *Nat. Genet.* **54**, 1238–1247 (2022).
34. Larson, A. G. et al. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
35. Feric, M. et al. Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016).
36. Kwon, I. et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139–1145 (2014).
37. Aviner, R. The science of puromycin: from studies of ribosome function to applications in biotechnology. *Comput. Struct. Biotechnol. J.* **18**, 1074–1083 (2020).
38. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
39. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
40. Mitrea, D. M. et al. Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *eLife* **5**, e13571 (2016).
41. White, M. R. et al. C9orf72 poly(PR) dipeptide repeats disturb biomolecular phase separation and disrupt nucleolar function. *Mol. Cell* **74**, 713–728.e6 (2019).
42. Lee, K. H. et al. C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. *Cell* **167**, 774–788.e17 (2016).
43. Banani, S. F. et al. Genetic variation associated with condensate dysregulation in disease. *Dev. Cell* **57**, 1776–1788.e8 (2022).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your

intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>1</sup>Institute of Medical Genetics and Human Genetics, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany. <sup>2</sup>BIH Biomedical Innovation Academy, Berlin Institute of Health at Charité–Universitätsmedizin Berlin, Berlin, Germany. <sup>3</sup>RG Development and Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany. <sup>4</sup>Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany. <sup>5</sup>Exploratory Diagnostic Sciences, Berlin Institute of Health at Charité–Universitätsmedizin Berlin, Berlin, Germany. <sup>6</sup>Institute of Human Genetics, University Hospitals Schleswig-Holstein, University of Lübeck and Kiel University, Lübeck, Kiel, Germany. <sup>7</sup>Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. <sup>8</sup>Center for Genomics and Transcriptomics (CeGaT), Tübingen, Germany. <sup>9</sup>Department of Pediatrics and Adolescent Medicine, School of Clinical Medicine, LKS Faculty of Medicine, The University of Hong Kong, Pok Fu Lam, Hong Kong. <sup>10</sup>Department of Congenital Heart Disease and Pediatric Cardiology, University Hospital Schleswig-Holstein, Kiel, Germany. <sup>11</sup>Department of Medical Genetics, Carl von Ossietzky University, Oldenburg, Germany. <sup>12</sup>Department of Obstetrics and Gynecology, University Hospital Schleswig-Holstein, Kiel, Germany. <sup>13</sup>Department of Internal Medicine, Radboud Institute for Molecular Life Sciences, Radboud Expertise Center for Immunodeficiency and Autoinflammation and Radboud Center for Infectious Disease (RCI), Radboud University Medical Center, Nijmegen, The Netherlands. <sup>14</sup>Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. <sup>15</sup>Department of Pediatrics, Pediatric Endocrinology and Diabetes, University Hospital Schleswig-Holstein, Schleswig-Holstein, Germany. <sup>16</sup>Handchirurgie, Katholisches Kinderkrankenhaus Wilhelmstift, Hamburg, Germany. <sup>17</sup>Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. <sup>18</sup>Department of Obstetrics and Gynaecology, Queen Mary Hospital, Pok Fu Lam, Hong Kong. <sup>19</sup>Institute of Human Genetics, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. <sup>20</sup>Institute for Human Genetics and Genomic Medicine, Medical Faculty, RWTH Aachen University Hospital, Aachen, Germany. <sup>21</sup>Department of Pediatric Neurology, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany. <sup>22</sup>Department of Pediatrics, University Hospital Center Schleswig-Holstein, Kiel, Germany. <sup>23</sup>Department of Radiology and Neuroradiology, Pediatric Radiology, University Hospital Schleswig-Holstein, Kiel, Germany. <sup>24</sup>Institute of Human Genetics, Medical Center, University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. <sup>25</sup>Zentrum für Kinder- und Jugendmedizin, Helios Kliniken Schwerin, Schwerin, Germany. <sup>26</sup>Institute of Human Genetics, University Hospital, Otto-von-Guericke University, Magdeburg, Germany. <sup>27</sup>Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain. <sup>28</sup>Microscopy Core Facility, Max Planck Institute for Molecular Genetics, Berlin, Germany. <sup>29</sup>ICREA, Passeig Lluís Companys 23, Barcelona, Spain. <sup>30</sup>BCRT-Berlin Institute of Health Center for Regenerative Therapies, Berlin, Germany. <sup>31</sup>DZHK (German Centre for Cardiovascular Research), partner site Hamburg, Lübeck, Kiel, Lübeck, Germany. <sup>32</sup>These authors contributed equally: Martin A. Mensah, Henri Niskanen. <sup>33</sup>These authors jointly supervised this work: Malte Spielmann, Denise Horn, Denes Hnisz. <sup>34</sup>e-mail: Malte.Spielmann@uksh.de; Denise.Horn@charite.de; hnisz@molgen.mpg.de

## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

## Komplette Publikationsliste

- Elsner, Jonas, Martin A. Mensah, Manuel Holtgrewe, Jakob Hertzberg, Stefania Bigoni, Andreas Busche, Marie Coutelier, et al. 2021. "Genome Sequencing in Families with Congenital Limb Malformations." *Human Genetics* 140 (8): 1229–39.
- Mensah, Martin A., Henri Niskanen, Alexandre P. Magalhaes, Shaon Basu, Martin Kircher, Henrike L. Sczakiel, Alisa M. V. Reiter, et al. 2023. "Aberrant Phase Separation and Nucleolar Dysfunction in Rare Genetic Diseases." *Nature* 614 (7948): 564–71.
- Pantel, Jean Tori, Nurulhuda Hajjir, Magdalena Danyel, Jonas Elsner, Angela Teresa Abad-Perez, Peter Hansen, Stefan Mundlos, et al. 2020. "Efficiency of Computer-Aided Facial Phenotyping (DeepGestalt) in Individuals With and Without a Genetic Syndrome: Diagnostic Accuracy Study." *Journal of Medical Internet Research* 22 (10): e19263.
- Sczakiel, Henrike Lisa, Wiebke Hülsemann, Manuel Holtgrewe, Angela Teresa Abad-Perez, Jonas Elsner, Sarina Schwartzmann, Denise Horn, Malte Spielmann, Stefan Mundlos, and Martin Atta Mensah. 2021. "GLI3 Variants Causing Isolated Polysyndactyly Are Not Restricted to the Protein's C-Terminal Third." *Clinical Genetics* 100 (6): 758–65.

## Danksagung

Ich danke Dr. Martin A. Mensah für die herausragende Betreuung während des Projektes. Weiterhin danke ich ihm für die Vermittlung an Prof. Malte Spielmann, welcher einen wichtigen Teil der Supervision übernahm. Darüber hinaus möchte ich beiden für das Überlassen des Themas und die Ermöglichung der Arbeit danken.

Prof. Stefan Mundlos danke ich für die Möglichkeit der Bearbeitung dieses Projekts sowie für wertvolle Hinweise auf dem Weg zur Publikation.

Allen an der Publikation Beteiligten danke ich für ihre gehaltvollen Beiträge und ihren Rat, ohne den diese Arbeit nicht möglich gewesen wäre.

Darüber hinaus danke ich allen Mitarbeiter\*innen des Instituts für Medizinische Genetik und Humangenetik der Charité, welche mir während der Erstellung der Arbeit für Rückfragen zur Verfügung standen.

Darüber hinaus möchte ich meinen Eltern, meiner Schwester sowie meiner Freundin Rebecca danken, welche mir während der Erstellung der Arbeit zur Seite standen.

Nicht zuletzt danke ich der Studienstiftung des deutschen Volkes e. V., welche mich finanziell unterstützt hat und mir somit die fokussierte Erstellung dieser Arbeit ermöglichte.