

**DESIGN UND ETABLIERUNG
VON
NEXT GENERATION SEQUENCING-METHODEN
ZUR DIAGNOSTIK
VERSCHIEDENER ERBKRAKHEITEN**



Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von

Julia Elisa Kuhn, geb. Bach

Geboren in Düsseldorf

Würzburg, Juni 2015

Eingereicht am:

Mitglieder der Promotionskommission:

Vorsitzender:

Gutachterin: Dr. habil. Simone Rost

Gutachter: Prof. Dr. Jörg Schultz

Tag des Promotionskolloquiums:

Doktorurkunde ausgehändigt am:

Die vorliegende Arbeit wurde von Februar 2012 bis Juni 2015 am Institut für Humangenetik der Universität Würzburg unter Betreuung von Frau Dr. habil. Simone Rost angefertigt.

Eidesstattliche Versicherung gemäß § 4 Abs. 3 Satz 3, 5 ,8 der Promotions-Ordnung

Hiermit erkläre ich, dass ich diese Doktorarbeit in allen Teilen selbständig und ohne Hilfe einer kommerziellen Promotionsberatung sowie ausschließlich unter Verwendung der angegebenen Quellen und Hilfsmittel verfasst habe.

Weiterhin versichere ich an Eides statt, dass ich die Gelegenheit zum Promotionsvorhaben nicht kommerziell vermittelt bekommen habe und insbesondere nicht eine Person oder Organisation eingeschaltet habe, die gegen Entgelt Betreuer bzw. Betreuerinnen für die Anfertigung von Dissertationen sucht.

Ich erkläre hiermit, dass die Regeln der Universität Würzburg über gute wissenschaftliche Praxis eingehalten wurden.

Die Dissertation wurde bisher weder in gleicher noch in ähnlicher Form einer anderen Hochschule oder in einem anderen Prüfungsverfahren mit dem Ziel, einen akademischen Grad zu erwerben, vorgelegt.

Am 01. Februar 2012 wurde mir von der Bayerischen Julius-Maximilians-Universität Würzburg der akademische Grad „Diplombiologin“ verliehen. Weitere akademische Grade habe ich weder erworben noch zu erwerben versucht.

Würzburg, Juni 2015

Elisa Kuhn

INHALTSVERZEICHNIS

1	ZUSAMMENFASSUNG	1
2	SUMMARY	3
3	EINLEITUNG.....	5
3.1	Next Generation Sequencing in der humangenetischen Diagnostik	5
3.2	NGS-Methoden	8
3.2.1	Probenanreicherung	8
3.2.2	Sequenzierung.....	9
3.3	Krankheiten und Gene.....	11
3.3.1	Hörstörungen und das <i>COL4A6</i> -Gen	11
3.3.2	Maligne Hyperthermie, <i>RYR1</i> und andere Kandidatengene	13
3.3.3	Gliedergürteldystrophien und die Gene <i>ANO5</i> , <i>DYSF</i> und <i>CAPN3</i>	14
3.3.4	Myofibrilläre Myopathien und ihre Kandidatengene.....	16
3.3.5	Hämophilie A und das <i>F8</i> -Gen	18
3.4	Zielsetzung und Inhalt der Arbeit	21
	Projekt 1: Schnelles Screening eines einzelnen Kandidaten-Gens	21
	Projekt 2: Ein-Schritt-Analyse eines großen Gens	21
	Projekt 3: Screening mehrerer Kandidatengene für heterogene Krankheitsbilder.....	22
	Projekt 4: Exomanalyse	22
	Projekt 5: Analyse der kompletten genomischen Sequenz eines Gens.....	23
4	METHODISCHE DETAILS	24
4.1	Polymerase-Kettenreaktion (PCR).....	24
4.2	Marker- bzw. Haplotypanalyse	24
4.3	DNA-Sequenzierung nach Sanger	24
4.4	Probenanreicherung für Next Generation Sequencing.....	25
4.4.1	Access Array-System (Fluidigm)	25
4.4.2	SureSelect-System (Agilent).....	26
4.4.3	TruSeq-System (Illumina)	26
4.5	Next Generation Sequencing.....	26
4.5.1	GS Junior (Roche).....	26
4.5.2	MiSeq (Illumina).....	26
4.5.3	HiSeq2000 (Illumina)	26

4.6	Molekulargenetische Diagnostikmethoden der HA-Routinediagnostik	26
4.7	<i>In vitro</i> -Spleißanalyse über Minigen-Konstrukte.....	26
5	ZUSAMMENFASSENDE DARSTELLUNG UND DISKUSSION DER PUBLIZIERTEN ERGEBNISSE.....	27
5.1	Schnelles Screening eines einzelnen Kandidaten-Gens	27
5.2	Ein-Schritt-Analyse eines großen Gens	30
5.3	Screening mehrerer Kandidatengene für heterogene Krankheitsbilder	34
5.3.1	LGMD-Panel	34
5.3.2	MFM-Panel.....	35
5.4	Exomanalyse	40
5.5	Analyse der kompletten genomischen Sequenz eines Gens	43
5.5.1	Mutationsstatistik und Diagnostik der Hämophilie A.....	43
5.5.2	Identifizierung tiefer intronischer Mutationen im <i>F8</i> -Gen mittels NGS	45
5.5.3	Minigen-Konstrukte zur Analyse tiefer intronischer Varianten im <i>F8</i> -Gen.....	51
5.6	Fazit und Ausblick	55
6	LITERATURVERZEICHNIS	57
7	ABKÜRZUNGSVERZEICHNIS	69
8	VERWENDETE DATENBANKEN UND ONLINE-TOOLS.....	72
9	LISTE DER VORGELEGTEN PUBLIKATIONEN	73
10	SONDERDRUCKE DER VORGELEGTEN PUBLIKATIONEN	75
	Anlage 1	75
	Anlage 2	93
	Anlage 3	103
	Anlage 4	119
	Anlage 5	130
	Anlage 6	135
	Anlage 7	151

1 ZUSAMMENFASSUNG

Innerhalb des letzten Jahrzehnts entstanden zahlreiche neue Anreicherungs- und Sequenzier-technologien der zweiten (und dritten) Generation, die in rasantem Tempo weiterentwickelt und schon jetzt in vielen Bereichen als neuer Goldstandard für molekulargenetische Forschung und Diagnostik angesehen werden. Als Hochdurchsatz-Verfahren ermöglichen diese Next Generation Sequencing-Methoden (NGS) in immer kürzerer Zeit die parallele Analyse zahlreicher Proben und immer größerer Zielregionen bis hin zum ganzen Genom und führten in der Humangenetik dadurch zu Forschungsansätzen in neuen Dimensionen.

In dieser Doktorarbeit, die im molekulargenetischen Diagnostik-Labor der Humangenetik Würzburg durchgeführt wurde, wurden in fünf Projekten NGS-Ansätze unterschiedlicher Stufen bzw. Größenordnungen für verschiedene erblich bedingte Erkrankungen konzipiert und etabliert und in Forschungsprojekten sowie der Routinediagnostik eingesetzt. Dabei wurden verschiedene Methoden zur Anreicherung der Zielsequenzen und zur NGS-Sequenzierung erprobt und auf ihre Effizienz beurteilt. Die Ergebnisse des NGS und darauf basierender Nachweis-Experimente wurden in sieben Veröffentlichungen dokumentiert, auf denen diese Dissertation aufbaut.

In den drei ersten Projekten wurden das Access Array-System (Fluidigm) zur Anreicherung der Zielsequenzen und der GS Junior (Roche) zur Erzeugung der Sequenzen verwendet.

In Projekt 1 wurde *COL4A6* als neues Kandidatengen für nicht-syndromale Hörstörungen identifiziert. Um mögliche weitere Mutationsträger zu detektieren, wurde erfolgreich ein kleiner NGS-Ansatz für das zügige Screening dieses Gens bei knapp 100 weiteren Patienten etabliert. Diese und weitere Ergebnisse bestätigten die Kausalität der *COL4A6*-Mutation eines Index-Patienten mit schwerer, X-chromosomal-rezessiver Hörstörung.

Ein geeigneter NGS-Ansatz für die Analyse des großen *RYR1*-Gens wurde in Projekt 2 gesucht. Der erste Ansatz mit Access Array-System und GS Junior führte zwar bei 39 von 87 Patienten mit Maligner Hyperthermie und/oder Central Core Disease zu dem Auffinden einer (potentiell) pathogenen Variante, allerdings mit hohen Ausfallquoten. Mit der zweiten Methode (Anreicherung: SureSelect-System custom design, Agilent; Sequenzierung: HiSeq, Illumina) wurden neben *RYR1* noch 63 weitere Gene analysiert, was zu deutlich besseren Ergebnissen und vier Mutationsfunden führte.

Projekt 3 beinhaltete die Etablierung zwei kleiner Panels für Muskelkrankheiten. Ein Panel für drei Gene für Gliedergürteldystrophien wurde sogar erfolgreich in die akkreditierte Routinediagnostik übernommen. Mit dem zweiten Panel für acht Kandidatengene myofibrillärer Myopathien (MFM) wurde u.a. eine neue Mutation im *BAG3*-Gen identifiziert.

Das Exom eines MFM-Patienten wurde in Projekt 4 nach Anreicherung mit dem SureSelect-System (Agilent) auf dem HiSeq (Illumina) sequenziert. Nach Auswertung und Beurteilung der identifizierten Varianten wurde ein neuer Erbgang für Myotilinopathien entdeckt. Verschiedene Nachweisexperimente bestätigten die Kausalität der Mutation im Myotilin-Gen.

In Projekt 5 wurde die komplette genomische Sequenz des *F8*-Gens nach tiefen intronischen Mutationen bei Hämophilie-Patienten abgesucht (Anreicherung SureSelect custom design, Agilent; Sequenzierung MiSeq, Illumina). Bei jedem der analysierten Patienten konnte mindestens eine verdächtige Variante identifiziert werden, die zu verändertem Spleißverhalten führen könnte. Drei Mutationen waren schon durch Publikationen bekannt, bei einer weiteren konnten *in vitro*-Spleißanalysen die Kausalität bestätigen.

Die Ergebnisse dieser Arbeit zeigen, dass die zur Verfügung stehenden Methoden zur Anreicherung von Zielsequenzen aus dem menschlichen Genom und zu deren Sequenzierung je nach Komplexität der Fragestellung, d.h. der Anzahl und Größe der Gene sowie der Anzahl der zu untersuchenden Proben, sinnvoll und effizient kombiniert werden können. Im Verlauf der Arbeit haben sich die NGS-Techniken rasant weiterentwickelt. So sind PCR-basierte Ansätze zur Anreicherung der Zielsequenzen für die meisten Anwendungen von hybridisierungs-basierten Methoden verdrängt worden. Von den ursprünglich drei konkurrierenden Verfahren zur Hochdurchsatzsequenzierung hat sich die Methode des „sequencing-by-synthesis“ (Illumina) weitgehend durchgesetzt. Diese Entwicklung spiegelt sich auch in den während dieser Arbeit erhobenen Daten wider.

2 SUMMARY

Several enrichment and sequencing technologies of the second (and third) generation have been developed in the past decade, were rapidly refined and are already considered as new state of the art method in several fields of molecular genetic research and diagnostics. Considered as high-throughput technologies, these next-generation sequencing methods (NGS) allow the parallel analysis of several samples and regions of interests up to whole genomes in decreasing time and thus permitted research projects with novel dimensions in human genetics.

This doctoral thesis was performed at the molecular genetic laboratory at the Department of Human Genetics in Würzburg. In five projects, NGS approaches of variable scale and for different hereditary diseases were designed, established and applied in research and routine diagnostics. Different methods for target enrichment and NGS analysis were tested and evaluated concerning their efficiency. The results of NGS and subsequent verification experiments were documented in seven publications forming the basis of this dissertation.

In project 1 - 3, the Access Array system (Fluidigm) was used for target enrichment and the GS Junior (Roche) for sequence generation.

COL4A6 has been identified as novel candidate gene for non-syndromic hereditary hearing loss in project 1. A small NGS approach was established to screen this gene in approx. 100 patients with hearing loss in order to search for additional carriers of *COL4A6* mutations. The results of this and further experiments confirmed the causality of the *COL4A6* mutation found in the index patient with severe X-linked hearing loss.

Project 2 aimed at finding a convenient NGS method for the analysis of the large *RYR1* gene. A first approach with the Access Array system and the GS Junior lead to the identification of a (potential) pathogenic mutation in 39 out of 87 patients with malignant hyperthermia and / or central core disease, but with high failure rates. *RYR1* and 63 further genes were then analyzed in a second approach (target enrichment with SureSelect custom design, Agilent; sequence analysis on a HiSeq, Illumina) providing considerably improved results and the identification of four mutations in five patients.

Two small panels for muscular diseases were established in project 3. A panel for three genes associated with limb-girdle muscular dystrophies were even successfully applied in accredited routine diagnostics. A novel mutation in the *BAG3* gene could be identified using the second panel established for eight candidate genes of myofibrillar myopathies (MFMs).

The exome of a patient with MFM was analyzed in project 4 after target enrichment with the SureSelect system (Agilent) and sequence analysis on a HiSeq (Illumina). A novel inheritance pattern of myotilinopathy was identified after analysis and evaluation of the de-

tected variants. Several experiments confirmed the causality of the mutation in the myotilin gene.

In project 5, the whole genomic sequence of the *F8* gene was analyzed for deep intronic mutations in haemophilic patients (target enrichment with SureSelect custom design, Agilent; sequence analysis on a MiSeq, Illumina). In each of the patients at least one conspicuous variant was identified probably leading to alternative splicing. Three mutations were known by publications and for another one causality could be proven by an *in vitro* splicing assay.

The results of this doctoral thesis show that the available methods for target enrichment and sequence analysis of specific targets of the human genome can be combined in a reasonable and efficient way considering the number and size of the targeted genes and probes. During the course of this doctoral thesis, NGS technologies have been further developed in a rapid way. For most applications, PCR-based technologies for target enrichment have been displaced by hybridization-based methods. Of the originally three competing techniques of high-throughput sequencing the “sequencing-by-synthesis” method (Illumina) has become the widely accepted standard. This development is reflected in the data generated in this doctoral thesis.

3 EINLEITUNG

3.1 Next Generation Sequencing in der humangenetischen Diagnostik

In der molekulargenetischen Forschung und Diagnostik stellt die Sequenzierung von DNA eine bedeutende Methode dar. Während klassische Polymerase-Kettenreaktion (PCR) und Sanger-Sequenzierung hierbei lange als Goldstandard galten, wurden innerhalb des letzten Jahrzehnts zahlreiche Probenanreicherungs- und Sequenzieretechnologien der zweiten Generation entwickelt, die unter dem Begriff Second oder Next Generation Sequencing (NGS) zusammengefasst werden. Die NGS-Technologien haben als Hochdurchsatz-Verfahren den klaren Vorteil, dass große Zielregionen und zahlreiche Proben in einem Schritt analysiert werden können. So konnten und können immer mehr genetische Variationen entdeckt, neue krankheitsassoziierte Gene identifiziert und dadurch die genetische Ursache unspezifischer Phänotypen und neuer Krankheitsbilder geklärt sowie allgemein Erkenntnisse über das Genom zahlreicher Organismen inklusive des Menschen gewonnen werden [1].

Beim NGS gibt es je nach Vorhaben verschiedene Ansätze und Größenordnungen. NGS kann im kleinen Maßstab in Form kleiner Panels für einzelne Kandidatengene eines genetisch heterogenen Krankheitsbildes oder in Form größerer Panels mit einer großen Auswahl klinisch relevanter Gene angewandt werden. Die Sequenzierung des ganzen Exoms (Whole Exome Sequencing, WES) bietet sich an, um neue Kandidatengene zu identifizieren. Dabei wird im Gegensatz zum Whole Genome Sequencing (WGS) nur der sehr kleine kodierende Teil (ca. 1 %) des menschlichen Genoms analysiert [2], was dementsprechend mit einem geringeren Zeit- und Kostenaufwand verbunden ist und weniger schwer klassifizierbare Varianten v.a. in nicht-kodierenden Bereichen des Genoms liefert.

Vor allem die großen NGS-Ansätze ergeben eine beträchtliche Datenmenge, die es bioinformatisch zu bewältigen gilt. Nicht jede der gefundenen Varianten ist pathogen, das menschliche Genom trägt etwa 10 Millionen Polymorphismen in sich (Single Nucleotide Polymorphism, SNP), davon mindestens 20 000 im kodierenden Bereich, und das Genom zweier nicht-verwandter Menschen unterscheidet sich in Millionen von Basen [3, 4]. Die Herausforderung besteht darin, aus der Fülle an Varianten die pathogene(n) herauszufiltern, was sich v.a. in nicht-kodierenden Bereichen des Genoms als schwierig gestaltet, da ein potentieller pathogener Effekt hier oft schwer einzuschätzen ist. Doch auch tiefe intronische Mutationen wurden als kausal für verschiedene Krankheitsbilder beschrieben [5-8]. Für die Recherche stehen verschiedene Datenbanken zur Verfügung, in denen bekannte Mutationen oder Polymorphismen sowie krankheits- bzw. gen-spezifische Informationen gespeichert werden wie z.B. in der Human Gene Mutation Database (HGMD [9]), der Single Nucleotide

Polymorphism Database (dbSNP [10]) oder der Datenbank Online Mendelian Inheritance in Man (OMIM), in der alle mendelschen Krankheiten und die assoziierten Gene aufgeführt und mit einer sog. MIM-Nummer versehen sind [11]. Auch diverse *in silico*-Vorhersageprogramme, die anhand verschiedener funktioneller, klinischer oder anderer Kriterien den potentiellen Effekt einer Variante berechnen, können bei der Bewertung einer unbekanntem Variante helfen und werden oft als Hilfsmittel zur ersten Einschätzung herangezogen [12-16].

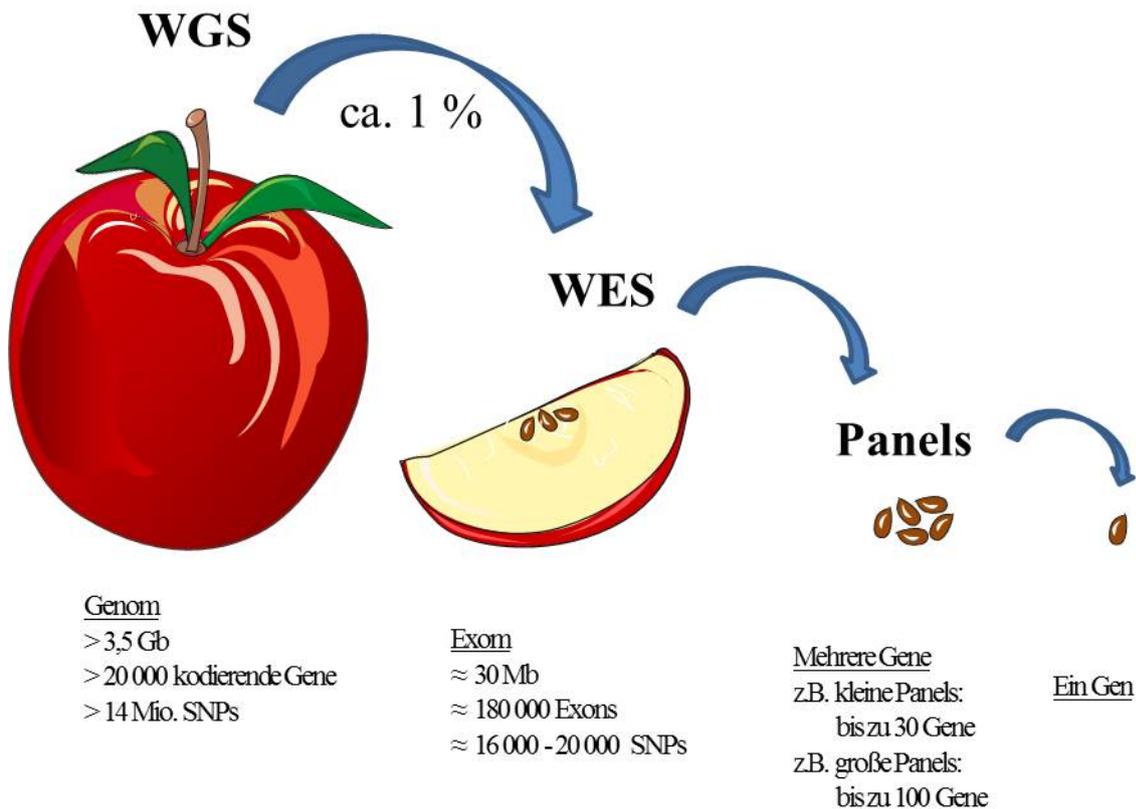


Abb. 1 Vergleich WGS, WES und Panels für mehrere oder einzelne Gene. Quellen: Ensembl (<http://www.ensembl.org/index.html>, Human genome assembly: GRCh38.p2) und [2]. (Abbildung erstellt mit Servier Medical Art, Servier, Suresnes, Frankreich.)

Das menschliche Genom gilt seit der Veröffentlichung durch das Konsortium des Humangenomprojektes im Jahr 2004 als entschlüsselt [17]. Seitdem wurden andere Großprojekte mit dem Ziel, das menschliche Genom weiter zu enträtseln, gestartet wie z.B. das 1 000-Genome-Projekt, das International HapMap-Projekt oder das Personal Genome-Projekt, an welchem verschiedene Länder beteiligt sind [18, 19]. Die Veröffentlichung von menschlichen Genomen wird dabei kontrovers diskutiert. Einerseits bietet WGS den Komfort in einem Schritt eine riesige Menge an Informationen zu erhalten. Außerdem können neue Kenntnisse über den Zusammenhang zwischen natürlicher genetischer Variation und Krankheiten gewonnen werden. Darüber hinaus spielen Genom-Studien z.B. eine wichtige Rolle in der personalisierten und präventiven Medizin sowie der Pharmakogenetik [20-22]. Jedoch bedeuten die Handhabung, Sicherung und Speicherung der enormen Datenmengen

eine neue bioinformatische Herausforderung [23]. Gegnerische Bedenken und Argumente betreffen auch die Gefahr der genetischen Diskriminierung und den Datenschutz. Durch die Veröffentlichung eines Genoms werden nicht nur Informationen über die sequenzierte Person offengelegt, sondern es können auch Rückschlüsse auf den Genotyp und Informationen z.B. über Vaterschaftsverhältnisse, die Prädisposition für Krankheiten oder die Suszeptibilität für Medikamente von Verwandten gezogen werden [24, 25]. Ein Problem v.a. bei WES und WGS, aber auch bei größeren Panels sind auch Zufallsbefunde bzw. der richtige Umgang damit unter Beachtung gesetzlicher Regelungen und Einverständniserklärungen des Patienten [26, 27].

3.2 NGS-Methoden

Innerhalb des vergangenen Jahrzehnts wurden verschiedene Probenvorbereitungs-Technologien und NGS-Plattformen entwickelt, die zu immer schnelleren und präziseren Ergebnissen, größeren Laufweiten, einer höheren Ausbeute und geringeren Kosten im Vergleich zur traditionellen Sanger-Sequenzierung führten [1].

3.2.1 Probenanreicherung

Den NGS-Methoden gemeinsam ist die Notwendigkeit, vor der Sequenzierung eine DNA-Fragment-„Bibliothek“ herzustellen, die aus gen- oder panel-spezifisch angereicherten und mit plattform-spezifischen Adaptoren versehenen DNA-Fragmenten besteht. Neben PCR-basierten Ansätzen werden v.a. hybridisierungs-basierte Methoden zur Probenanreicherung angeboten; über die wichtigsten soll dieser Abschnitt einen groben Überblick schaffen.

Eine klassische PCR mit einem einzelnen Primer-Paar ist zwar für Sanger-Sequenzierungen einzelner Exons oder Gene eine ideale Anreicherungsmethode, sie kann jedoch die Anforderungen des NGS an einen hohen Durchsatz und massive Parallelität nicht ausreichend erfüllen [28]. Für kleinere Projekte mag eine Multiplex-PCR mit mehreren Primer-Paaren ein guter Ansatz sein, das Finden einer passenden Kombination und optimaler PCR-Bedingungen ist allerdings häufig schwierig und es kommt leicht zu unspezifischen Produkten oder Primer / Primer-Interaktionen [29]. Diese Problematik wird bei einigen kommerziellen PCR-basierten Anreicherungsmethoden für NGS umgangen. Die Firma Qiagen (Hilden, Deutschland) bietet z.B. mit den GeneRead DNAseq Panels ein System an, das auf einer Multiplex-PCR in mehreren Reaktionsansätzen basiert (<https://www.qiagen.com/>). Eine weitere Methode, das Access Array-System, das in einigen Projekten dieser Doktorarbeit verwendet wurde, wurde von der Firma Fluidigm (South San Francisco, USA) entwickelt. Es basiert auf der parallelen PCR von bis zu 48 x 48 verschiedenen Amplikons auf einem Chip, in den ein System winziger Flüssigkeitsleitungen mit separaten Reaktionskammern eingelassen ist. Dabei können bis zu 48 DNA-Proben mit bis zu 48 Primerpaaren kombiniert werden. Nach der PCR kann je DNA-Probe ein Pool der bis zu 48 PCR-Produkte entnommen werden, die dann wiederum für die Sequenzierung zusammengeführt werden können, da sie mit molekularen Barcodes versehen sind [30], was aufgrund der meist parallelen Sequenzierung mehrerer Proben beim NGS generell die Regel ist.

Zwei große Anbieter hybridisierungs-basierter Anreicherungssysteme sind Agilent Technologies (Santa Clara, CA, USA) und Illumina (San Diego, CA, USA). Bei beiden müssen die (genomischen) DNA-Proben zunächst fragmentiert werden, ehe sie angereichert und mit Adaptoren versehen werden. Die Firma Agilent bietet z.B. das SureSelect-System an. Hier wird nach der mechanischen oder enzymatischen (SureSelect XT bzw. QXT) Fragmentierung genomischer DNA über RNA-Sonden die gewünschte Zielregion durch Hybridisierung herausgefischt und vervielfältigt (<http://www.genomics.agilent.com/>; [31]). Je nach Vorha-

ben können so kleine maßgeschneiderte oder größere vorgefertigte Panels für krankheitsspezifische oder alle klinisch relevanten Gene bis hin zum ganzen Exom angereichert und auf einer der gängigen NGS-Plattformen sequenziert werden. Ähnlich funktioniert auch die Anreicherung mit dem TruSight-System von Illumina, bei dem jedoch im Zuge der sog. Tagmentierung in einem einzigen Schritt die Fragmentierung und das Anhängen der Adapter stattfinden. Auch hier gibt es die Möglichkeit, maßgeschneiderte oder vorgefertigte kleinere und größere Panels zu nutzen (<http://www.illumina.com/>).

3.2.2 Sequenzierung

Das erste kommerzielle Sequenziergerät für NGS wurde 2005 von 454 Life Sciences (Roche Applied Science, Mannheim, Deutschland) vorgestellt [32]. Die 454-Geräte (GS FLX, GS FLX+, GS Junior, GS Junior+) nutzen das Prinzip der Pyrosequenzierung und sind daher anfällig für falsch-positive Ergebnisse innerhalb längerer Homopolymere, ermöglichen jedoch große Leselängen, was einen Vorteil bei *de novo*-Sequenzierungen darstellt. Vor der Sequenzierung muss die vorbereitete DNA-Bibliothek über eine Emulsions-PCR (emPCR) mit klonaler Amplifikation an kleinen Beads in einer Emulsion angereichert werden. Die Beads mit den daran immobilisierten DNA-Fragmenten werden in die mikroskopisch kleinen Vertiefungen einer PicoTiter-Platte verteilt und dort sequenziert [32-34]. Ähnlich funktionieren auch der Ion PGM und Ion Proton (Personal Genome Machine) von Ion Torrent (Life Technologies, Darmstadt, Deutschland), auch hier ist eine emPCR zur Anreicherung nötig. Die Sequenzierung basiert jedoch auf einer Halbleiter-Technologie, bei der die Änderung des pH-Wertes gemessen wird, die bei der Abspaltung von Protonen beim Einbau eines Nukleotides während der Sequenzierreaktion entsteht [35].

Auch bei den SOLiD-Sequenziergeräten (5500xl und 5500; Sequencing by Oligo Ligation and Detection) von Applied Biosystems (Life Technologies) werden die Proben zunächst klonal an Beads amplifiziert, die dann für die Sequenzierung auf die Glasoberfläche des Sequenzierträgers aufgebracht werden. Die Sequenzierung funktioniert über die Ligation fluoreszenzmarkierter Oligonukleotide. Jede Position wird dabei zweimal sequenziert, was präzisere Ergebnisse liefert, wobei ein Nachteil allerdings kurze Laufweiten sind [34, 36].

Neben den schon genannten NGS-Plattformen wird schließlich noch Illumina zu den Marktführern der letzten Jahre gezählt und momentan aufgrund des breiten Angebots und der vergleichsweise geringen Sequenzierkosten sogar als *der* Marktführer betrachtet. Die Firma Solexa brachte mit dem Genome Analyzer eines der frühen NGS-Geräte auf den Markt und wurde kurze Zeit später von Illumina aufgekauft. Die aktuellen Geräte von Illumina sind verschiedene Varianten des HiSeq, MiSeq und NextSeq. Die Technologie beruht auf der Anreicherung (Clusterbildung) der DNA-Bibliothek durch Brückenamplifikation auf dem Sequenzierträger (FlowCell) und einer Sequenzierung der DNA-Cluster durch zyklische, reversible Termination [1, 34, 37, 38].

Weitere Entwicklungen des NGS sind Technologien, die auf der Sequenzierung einzelner Moleküle basieren (Single Molecule Sequencing) und teilweise schon als Sequenziertechno-

logien der dritten Generation (Third generation sequencing) bezeichnet werden, wenn außerdem noch in Echtzeit (real-time) sequenziert wird. Da keine Anreicherung der DNA-Fragmente an Beads oder dem Sequenzierträger notwendig ist, entfällt ein Schritt, bei dem leicht Amplifikationsfehler in der Sequenz entstehen können. Pacific Biosciences (Menlo Park, CA, USA) brachte mit dem PacBio RS und der SMRT-Technologie (Single Molecule, Real-Time) das erste Gerät der dritten Generation auf den Markt. Hierbei wird die Polymerase zusammen mit einem zirkularisierten DNA-Fragment in winzigen Kammern vereinzelt und durch schmale Öffnungen der Einbau des jeweiligen fluoreszenzmarkierten Nukleotids detektiert [34, 39, 40].

Eine weitere Technologie ist die Detektion der Sequenz über Nanoporen, durch die das zu analysierende DNA-Molekül wandert. Bei Oxford Nanopore Technologies (Oxford, UK) z.B. wird die Sequenz durch eine Spannungsänderung detektiert, die beim Abspalten der einzelnen Nukleotide durch eine Exonuklease an der Nanopore entsteht. Die Firma hat das MinION-System entwickelt, ein kleines tragbares Gerät, kaum größer als ein USB-Stick, welches direkt an den Computer angeschlossen werden kann. Momentan wird das System noch als Beta-Version getestet [1].

Im Rahmen der vorliegenden Doktorarbeit wurden verschiedene NGS-Projekte mit unterschiedlichen Größenmaßstäben getestet, die größtenteils im molekulargenetischen Diagnostiklabor des Instituts für Humangenetik der Universität Würzburg (im Folgenden Humangenetik Würzburg) durchgeführt werden konnten. Hier stehen aktuell zwei NGS-Plattformen zur Verfügung, der GS Junior von Roche sowie ein MiSeq und ein Next-Seq von Illumina.

3.3 Krankheiten und Gene

Der folgende Abschnitt soll einen groben Überblick über die Krankheiten und Krankheitsbilder und die jeweils damit assoziierten Gene bieten, um die es in den einzelnen in dieser Doktorarbeit bearbeiteten Projekten ging. Insgesamt handelte es sich primär um eine methodisch orientierte Arbeit und sekundär um den medizinischen Hintergrund oder die Aufklärung molekulargenetischer Mechanismen. Daher liegt der Fokus hier nun weniger auf einer detaillierten klinischen Beschreibung der Pathologie, Mechanismen, Symptome oder Therapiemöglichkeiten als vielmehr auf den Auswahlkriterien der Krankheiten für die zu etablierende Diagnostik mittels Next Generation Sequencing.

3.3.1 Hörstörungen und das *COL4A6*-Gen

3.3.1.1 Erblich bedingte Hörstörungen

Mit einer Inzidenz von 1 - 2 von 1 000 Kindern werden kongenitale bzw. in frühester Kindheit auftretende schwere Hörstörungen zu den häufigsten sensorischen Erkrankungen gezählt [41, 42]. Etwa 50 - 60 % der Fälle haben einen genetischen Hintergrund, aber auch verschiedene Umwelteinflüsse wie Infektionen, z.B. mit dem Cytomegalievirus (CMV), oder Traumata können ursächlich für einen kindlichen Hörverlust sein [42, 43]. In industrialisierten Ländern mit einer immer besseren medizinischen Versorgung und Vorsorge nimmt der Anteil umwelt- und v.a. infektionsbedingten Hörverlusts jedoch kontinuierlich ab [44]. Erblich bedingte Hörstörungen sind genetisch stark heterogen und lassen sich durch eine Einteilung u.a. in syndromale und nicht-syndromale, prä- und postlingual manifestierende sowie autosomal-rezessiv (AR), autosomal-dominant (AD) oder X-chromosomal-rezessiv (XR) vererbte Formen genauer charakterisieren. Darüber hinaus sind auch einige mitochondrial vererbte Hörstörungen bekannt [45].

Die über 400 bekannten Syndrome mit einer schweren Hörstörung als eines der Leitsymptome machen etwa 30 % der angeborenen Hörstörungen aus [42]. Darunter zählen z.B. das Pendred- (PDS, MIM #274600) und das Usher-Syndrom (z.B. USH1, MIM #276900; beide AR), das Branchio-otorenale (z.B. BOR1, MIM #113650) und das Waardenburg-Syndrom (z.B. WS1, MIM # 193500; beide AD) sowie das Alport-Syndrom (hauptsächlich XR; MIM #301050) zu den häufigeren Formen [41].

Der weitaus größere Teil schwerer Hörstörungen ist jedoch nicht-syndromal [42] und auf eine Mutation in einem von 60 bekannten Genen für autosomal-rezessive, 32 Genen für autosomal-dominante und 4 Genen für X-chromosomal-rezessive Formen zurückzuführen, wobei jeweils noch mehr Loci beschrieben sind (<http://hereditaryhearingloss.org>, Stand 25.05.2015). Nahezu 50 % der autosomal-rezessiven Fälle können auf Mutationen allein im *GJB2*-Gen (MIM *121011) zurückgeführt werden, welches das Protein Connexin 26 kodiert [46]. Drei besonders häufige Mutationen – drei kleine Deletionen – sind hier bekannt [45],

denen in bestimmten Populationen ein Foundereffekt zugesprochen wird: c.35delG mit dem höchsten Auftreten in Südeuropa bzw. der kaukasischen Bevölkerung [47, 48], c.167delT bei den aschkenasischen Juden [49] und c.235delC in der japanischen Bevölkerung [50].

Die Diagnostik bei Hörstörungen erfordert ein interdisziplinäres Zusammenarbeiten verschiedener Spezialisten [42]. Vor allem bei syndromalem Hörverlust kann eine Früherkennung und korrekte Einstufung des Syndroms hilfreich sein, um ggf. weitere mit dem Syndrom assoziierte Symptome rechtzeitig erkennen und den richtigen Therapieansatz wählen zu können [46]. In industrialisierten Ländern wurde das routinemäßige Neugeborenen-Screening für Hörstörungen zunehmend standardisiert und verfeinert, sodass das Alter, in dem ein Defekt erkannt und spezifisch therapiert und die Patienten besonders gefördert werden können, immer weiter gesunken ist [51, 52]. Um die genetische Ursache dieses stark heterogenen Erkrankungsbildes finden zu können, bietet sich Next Generation Sequencing an, da so schnell eine ganze Reihe krankheitsrelevanter Gene in Form von Panels oder gleich das ganze Exom nach der kausalen Mutation abgesucht werden können [46, 53]. Liegt z.B. ein klarer mendelscher X-chromosomal-rezessiver Erbgang vor, bietet sich nach Ausschluss der wenigen bekannten X-chromosomalen „Hörstörungs-Gene“ als nächster Schritt ein Screening des ganzen X-chromosomalen Exoms an, um auf mögliche neue Kandidatengene zu stoßen. Dies war der Fall in einem der in dieser Doktorarbeit bearbeiteten Projekte über eine Familie mit einer nicht-syndromalen, offenbar XR vererbten Hörstörung (Projekt 1).

3.3.1.2 Das *COL4A6*-Gen

In der untersuchten Familie fiel eine Variante im *COL4A6*-Gen (MIM *303631) auf. Dieses Gen wurde bis dato nur in Verbindung mit Deletionen, die *COL4A6* und das benachbarte *COL4A5* (MIM *303630) umfassen, mit Alport-Syndrom mit kosegregierender Leiomyomatose (DL-ATS, MIM #308940) assoziiert [54]. Auch Mutationen in *COL4A5* allein wurden mit dem (X-chromosomal gekoppelten) Alport-Syndrom (ATS, MIM #301050) assoziiert, bei dem neben der Hörstörung v.a. auch Defekte der Nieren und der Retina auftreten [55]. Jedoch waren keine krankheitsverursachenden Mutationen in *COL4A6* allein bekannt.

Das Gen liegt auf Xq22.3, Kopf an Kopf direkt neben *COL4A5* und kodiert die alpha-6 Isoform der Typ IV-Kollagene [56]. Beide Gene haben eine gemeinsame Promotor-Region, die Promotoren werden jedoch auch gewebespezifisch verwendet. *COL4A6* umfasst 46 Exons, darunter 2 alternative erste Exons: 1A (Transkript Variante A, ENST00000372216 / NM_001847.3) und 1B (Transkript Variante B, ENST00000334504 / NM_033641) für die gewebespezifische Verwendung [57].

Nun sollte ein geeigneter Ansatz gefunden und etabliert werden, um das *COL4A6*-Gen effizient und praktikabel bei einer größeren Anzahl weiterer Hörstörungspatienten mittels Next Generation Sequencing analysieren zu können (s. Projekt 1), da dieses Vorhaben angesichts der Anzahl der zu analysierenden Exons und Patienten mittels klassischer Sanger-Sequenzierung einen hohen Aufwand bedeuten würde.

3.3.2 Maligne Hyperthermie, *RYR1* und andere Kandidatengene

3.3.2.1 Maligne Hyperthermie

Die Maligne Hyperthermie (MH) ist eine pharmakogenetisch bedingte Veranlagung, bei der unter der Einwirkung bestimmter gasförmiger (Inhalations-)Anästhetika oder depolarisierender Muskelrelaxantien lebensbedrohliche Situationen auftreten können [58]. Während Betroffene im Alltag keinerlei Symptome zeigen, gerät der Stoffwechsel der Skelettmuskulatur bei einer MH-Krise völlig aus dem Gleichgewicht, da unkontrolliert Kalziumionen (Ca^{2+}) aus dem Sarkoplasmatischen Reticulum (SR) in die Muskelzelle ausgeschüttet werden [59]. Kalzium wird für den sog. „molekularen Ruderschlag“, die Konformationsänderung der Myofibrillen (Myosin und Aktin), benötigt, die zur Kontraktion des Muskels führt. Der Überschuss an Ca^{2+} -Ionen und die Behinderung ihres Rücktransports in das SR führen dann im Wesentlichen zu den typischen hypermetabolischen MH-Symptomen wie Versteifung der Skelettmuskulatur und Rhabdomyolyse, Erhöhung der Körpertemperatur, erhöhte Herzschlagfrequenz und Übersäuerung des Organismus, die in ihrer Reihenfolge und Stärke variieren können [60, 61]. Ein schnelles Erkennen und Handeln bei einer auftretenden MH-Krise ist unbedingt notwendig und lebensrettend, jedoch sind Substanzen der Notfalltherapie wie Dantrolen in schlecht entwickelten Ländern oft nicht verfügbar, weshalb die Sterblichkeitsrate bei MH-Krisen dort deutlich höher ist [62]. Auch unabhängig von einer Narkose kann eine MH-Krise bei prädisponierten Personen induziert werden; als Auslöser wurden u.a. emotionaler, körperlicher oder Hitze-Stress sowie Infektionen beschrieben [63-65].

Die Inzidenz für eine MH-Krise liegt bei 1 : 5 000 bis 1 : 50 000 - 100 000 Anästhesien, die Prävalenz bei 1 : 3 000 Personen [59]. Die tatsächliche Häufigkeit der genetischen Prädisposition kann jedoch nur geschätzt werden, da trotz Veranlagung eine MH-Krise nicht bei jeder Exposition mit Triggersubstanzen auftritt; bei manchen prädisponierten Personen können mehrere Narkosen ereignislos verlaufen [59, 60]. Auch in Zusammenhang mit bestimmten kongenitalen Myopathien, z.B. der Central Core-Krankheit (CCD, MIM #117000) oder der Multiminicore-Myopathie (MIM #255320), wurde ein erhöhtes Risiko für das Auftreten einer MH-Krise beschrieben [66, 67].

Aufgrund des im Alltag unauffälligen Phänotyps gestaltet sich die Diagnostik der MH als schwierig, da Betroffene oft erst beim Auftreten der ersten MH-Krise erkannt werden können. Daher sollte bei bekannter familiärer Belastung vor einer anstehenden Narkose auf jeden Fall auf eine Veranlagung zur MH getestet werden, da ggf. alternative Anästhetika eingesetzt werden können, die bei Betroffenen keine MH-Krise auslösen, und lebensbedrohliche Situationen einfach vermieden werden können [59, 62]. Das sehr sensitive und spezifische Standard-Diagnostikverfahren für die Veranlagung zur MH ist der *in vitro*-Kontrakturtest (IVCT), bei dem biopsiertes Muskelgewebe den triggernden Substanzen Halothan und Koffein ausgesetzt und auf eine übermäßige Kontraktion getestet wird [68, 69]. Mittels molekulargenetischer Diagnostik, v.a. DNA-Sequenzierung, kann bei positivem IVCT und / oder positiver Familienanamnese die zugrunde liegende Mutation gesucht werden [70].

3.3.2 *RYR1* und andere Kandidatengene

Die Vererbung der Veranlagung zur MH (Malignant Hyperthermia Susceptibility, MHS) ist heterogen und meist autosomal-dominant. Es sind derzeit 6 MHS-Loci bekannt, die in Zusammenhang stehen mit dem Ca^{2+} -Stoffwechsel der Skelettmuskulatur bzw. den beiden wichtigsten Ionenkanälen für dessen Regulation – dem Dihydropyridin-Rezeptor (DHPR), der in den transversalen Tubuli der Muskelzellmembran lokalisiert ist und bei einem ankommenden Aktionspotenzial den Ryanodin-Rezeptor Subtyp 1 (RYR1) in der Membran des SR aktiviert, sodass durch diesen Ca^{2+} freigesetzt wird [71, 72]. Die meisten Mutationen wurden im MHS1-Locus (MIM #145600) bzw. dem *RYR1*-Gen (MIM *180901) gefunden, welches 106 Exons umfasst (Transkript Variante 1, ENST00000359596 / NM_000540.2) [73, 74]. Auch die CCD wird mit dem *RYR1*-Gen assoziiert. Für MH und CCD sind dabei unterschiedliche Hotspots beschrieben worden, in denen jeweils gehäuft Mutationen lokalisiert werden konnten [66, 67].

MHS5 (MIM #601887) wird mit dem Gen *CACNA1S* (auch *CACNL1A3*; MIM *114208) assoziiert, welches eine der Untereinheiten des DHPR kodiert [75]. Für die Loci MHS2 - 4 und 6 wurde bisher noch kein Gen endgültig identifiziert. Als Kandidaten wurden v.a. die Gene für die anderen Untereinheiten des DHPR diskutiert sowie das Gen *SCN4A*, welches eine Untereinheit des adulten Natriumkanals der Skelettmuskulatur kodiert [76].

In der molekulargenetischen MH- bzw. CCD-Diagnostik werden die Hotspots oder das komplette *RYR1*-Gen standardmäßig mittels Sanger-Sequenzierung analysiert. Im Diagnostiklabor der Humangenetik Würzburg werden als Hotspots Exon 17, 39 - 41, 45 und 46 für MH und Exon 95 sowie 100 - 102 für CCD sequenziert. Die Komplettssequenzierung der 106 Exons des *RYR1*-Gens gestaltet sich als recht aufwändig, einerseits wegen der Größe des Gens und andererseits wegen einiger meist aufgrund eines hohen GC-Gehaltes in methodischer Hinsicht kniffliger Exons. Zudem können auf diese Weise nur MHS1-Fälle aufgeklärt werden. Um einen einfacheren Weg für die Analyse des *RYR1*-Gens zu finden und weitere Kandidatengene zu identifizieren, bietet sich auch hier NGS als praktische Hochdurchsatz-Methode an. Das Etablieren einer effizienten molekulargenetischen MH-Diagnostik mittels NGS war das Ziel in Projekt 2 dieser Doktorarbeit.

3.3.3 Gliedergürteldystrophien und die Gene *ANO5*, *DYSF* und *CAPN3*

3.3.3.1 Gliedergürteldystrophien

Unter dem Begriff der Gliedergürteldystrophien (limb-girdle muscular dystrophy, LGMD) wird eine Reihe genetisch heterogener Muskelerkrankungen zusammengefasst, die charakteristischerweise die proximalen Muskelgruppen im Schulter- und Beckenbereich (Glieder-gürtel) betreffen. Typischerweise sind LGMDs nicht kongenital und nicht-syndromal und haben einen progressiven Charakter. Momentan sind acht autosomal-dominant vererbte LGMDs bekannt, die unter dem Typ LGMD1 zusammengefasst werden (Subtypen

LGMD1A - H). Dem Typ 2 werden die autosomal-rezessiven LGMDs mit aktuell 23 Formen zugeordnet (Subtypen LGMD2A - W).

Die Pathomechanismen der LGMDs sind häufig bedingt durch strukturelle Defekte im Dystrophin-Dystroglycan-Komplex der Muskelfaser, andere durch Membran-, Sarkomer-, Kernlamina- oder enzymatische Defekte [77-79]. Das klinische Erscheinungsbild ist sehr heterogen und variiert z.B. bezüglich des Erkrankungsalters, der Progressionsgeschwindigkeit oder der Erhöhung des CK-Wertes (Serum-Kreatinkinase). Bei manchen Formen sind auch die Herz- und Atemmuskulatur betroffen. LGMDs können von anderen Muskelkrankheiten wie der Fazioscapulohumeralen Dystrophie (FSHD) oder den X-gebundenen Dystrophinopathien vom Typ Duchenne oder Becker (DMD/BMD) z.B. über die klinischen Symptome und histologische oder immunologische Methoden abgegrenzt werden.

Die exakte Diagnose des LGMD-Subtyps lässt sich allerdings meist erst durch molekulargenetische Methoden (v.a. DNA-Sequenzierung) oder Western Blots stellen. Angesichts der großen Anzahl an Kandidatengen für LGMDs bietet sich hier v.a. das Screening mehrerer Gene mittels NGS an [78-80].

Zu den häufigsten rezessiven LGMDs werden die Subtypen LGMD2A und 2B weltweit und 2L in Nordeuropa gezählt, denen Mutationen in den Genen *CAPN3*, *DYSF* bzw. *ANO5* (MIM *114240, *603009 bzw. *608662) zugrunde liegen [81-85]. Bei einer LGMD2A (MIM #253600) sind aufgrund des häufig besonders betroffenen Schultergürtels klinisch vor allem die hervortretenden Schulterblätter („Scapulae alatae“) prominent. Eine Beteiligung der Atem- und Herzmuskulatur wird selten beobachtet. Das Erkrankungsalter ist häufig in der Jugend [78, 80].

Kennzeichnend für eine LGMD2B (MIM #253601) ist die Erkrankung im jungen Erwachsenenalter und eine unterschiedliche Schwere der Phänotypen. Die Krankheit schreitet meist nur langsam voran. Eine Beteiligung der Herzmuskulatur ist möglich. Betroffene sind charakteristischerweise unfähig auf Zehenspitzen zu stehen. Die gemessenen CK-Werte sind meist deutlich stärker erhöht als bei einer LGMD2A [80, 85].

Auch die LGMD2L (MIM #61130) geht einher mit stark erhöhten CK-Werten und ist langsam progredient. Das Erkrankungsalter zeigt eine breite Spanne von der Kindheit bis zum späten Erwachsenenalter und Frauen scheinen seltener betroffen zu sein als Männer. Eine Beteiligung der Herz- oder Atemmuskulatur wurde nicht beobachtet. [79, 81].

Aufgrund der Häufigkeit der drei LGMD-Subtypen 2A, 2B und 2L in Zentraleuropa und des ähnlichen klinischen Phänotyps wurden die assoziierten Gene *CAPN3*, *DYSF* und *ANO5* im Rahmen dieser Doktorarbeit für die Etablierung eines kleinen NGS-Panels für autosomal-rezessive Gliedergürteldystrophien mit hohem CK-Wert ausgewählt (siehe 3.4, Projekt 3).

3.3.3.2 *CAPN3*, *DYSF* und *ANO5*

Das *CAPN3*-Gen liegt auf dem langen Arm des Chromosoms 15 (15q15.1) und kodiert Calpain 3, eine muskelspezifische Protease, die vermutlich eine Bindung mit Titin, einem wichtigen Strukturprotein der Muskelfaser, eingeht. *CAPN3* besteht aus 24 Exons (Transkript

Variante 1, ENST00000397163 / NM_000070.2), die insgesamt 821 Aminosäuren kodieren [84]. Eine in (Südost-) Europa häufige Mutation ist die kleine Deletion c.550delA in Exon 5 von *CAPN3* [86-88].

Auf Chromosom 2p13.3 liegt das *DYSF*-Gen (Transkript Variante 8, ENST00000258104 / NM_003494.3) welches aus 55 Exons besteht, die die 2080 Aminosäuren des Dysferlin-Proteins kodieren, welches membran-assoziiert ist und eine Rolle bei Reparaturmechanismen in der Muskelzellmembran spielt [85, 89].

Aus 22 Exons besteht das *ANO5*-Gen (Transkript Variante 1, ENST00000324559 / NM_213599.2), welches auf 11p14.3 lokalisiert ist. Es kodiert die 913 Aminosäuren des Anoctamin 5-Proteins, ein calcium-aktivierter Chlorid-Kanal im Muskel [81]. Eine häufige Mutation ist eine Duplikation in Exon 5 (c.191dupA), die zu einem Frameshift mit vorzeitigem Stopp-Codon führt, der den Abbau der mRNA durch Nonsense-mediated mRNA Decay (NMD) auslöst. Der Mutation wird ein Foundereffekt zugesprochen und sie tritt meist homozygot, gelegentlich jedoch auch heterozygot zusammen mit einer weiteren Mutation auf (compound heterozygot) [81, 90].

3.3.4 Myofibrilläre Myopathien und ihre Kandidatengene

3.3.4.1 Myofibrilläre Myopathien

Die myofibrillären Myopathien (MFM) sind eine Gruppe klinisch und genetisch heterogener neuromuskulärer Erkrankungen mit progredientem Verlauf. Initiales Symptom ist häufig eine distal beginnende Muskelschwäche. Neben der Skelettmuskulatur ist oft auch der Herzmuskel betroffen, auch eine Beteiligung der Atemmuskulatur sowie periphere Neuropathien werden berichtet. Das Erkrankungsalter liegt meist im frühen bis späten Erwachsenenalter, aber auch Kinder mit MFM wurden beschrieben [91-93].

Pathologisch charakteristisch für MFM sind Proteinaggregate in den Muskelfasern, die von Abbauprodukten defekter Myofibrillen und Z-Scheiben sowie der ektopischen Expression bestimmter Proteine im Muskel herrühren. Außerdem treten typische Muster von sowohl degenerierenden und nekrotischen als auch regenerierenden und fibrösen Fasern bzw. Regionen auf.

Aufgrund dieser pathologischen Befunde wird die MFM häufig über immunhistologische und elektronenmikroskopische Analysen von Muskelbiopsien diagnostiziert. Dabei zeigen sich für einzelne Subtypen der MFM zum Teil individuell charakteristische proteinbiochemische und strukturelle Muster [94-97]. Weitere diagnostische Methoden sind laser-basierte Mikrodissektion und massenspektrometrie-basierte Proteomanalysen von betroffenem Muskelgewebe. Die genaue Mutation kann durch DNA-Sequenzierung der bekannten MFM-Kandidatengene detektiert werden, wofür sich NGS als Hochdurchsatz-Methode anbietet [93]. Allerdings lassen sich dadurch nicht alle MFM-Fälle klären, sodass es noch weitere Kandidatengene geben muss [98].

3.3.4.2 Kandidatengene für myofibrilläre Myopathien

Die Vererbung der MFM ist meist autosomal-dominant, selten auch X-gebunden oder autosomal-rezessiv. Tab. 1 zeigt einen Überblick über die Kandidatengene für MFMs. Die Genprodukte sind zumeist Bestandteile des Sarkomers oder stehen in enger Verbindung damit. Zudem werden mit einigen der Gene auch andere Muskelerkrankungen, v.a. LGMD-Subtypen, assoziiert [93, 94].

Tab. 1 Übersicht über die MFM-Kandidatengene. Gegeben sind Genname, Lokalisation, verwendete(s) Transkript(e) und MIM-Nummer des Phänotyps sowie das kodierte Protein und Referenzen. *Transkript in der Skelettmuskulatur, **Transkript in der Herzmuskulatur, RBM = reducing body myopathy, HMERF = hereditary myopathy with early respiratory failure

Gen Genort	Transkript MIM	Protein	Referenz
DES 2q35	ENST00000373960 / NM_001927.3 MFM1, MIM #601419	Desmin	[93, 99, 100]
CRYAB 11q23.1	ENST00000526180 / NM_001885.2 MFM2, MIM #608810	alpha-B Crystallin	[101]
MYOT 5q31	ENST00000239926 / NM_006790.2 MFM3, MIM #609200	Myotilin	[102]
ZASP (LDB3) 10q23.2	ENST00000372066 / NM_001080116.1* ENST00000361373 / NM_007078.2** MFM4, MIM #609452	LIM domain binding 3	[103]
FLNC 7q32.1	ENST00000325888 / NM_001458.3 MFM5, MIM #609524	Filamin C	[104]
BAG3 10q26.11	ENST00000369085 / NM_004281.3 MFM6, MIM #612954	BCL2-associated atha- nogene 3	[105]
FHL1 Xq26.3	ENST00000394155 / NM_001159702.2 RBM, MIM #300717 / #300718	Four and a half LIM domains 1	[106]
TTN 2q31.2	ENST00000589042 / NM_001267550.1 HMERF, MIM #603689	Titin	[107]
DNAJB6 7q36.3	ENST00000429029 / NM_005494.2 LGMD1D, MIM #603511	DnaJ (Hsp40) Homolog, subfamily B, member 6	[108]

Die Desminopathie ist die häufigste Form der MFM. Das Protein Desmin ist ein Intermediärfilament der Muskelzelle und hat somit strukturelle Bedeutung [93, 99, 100].

Alpha-B Crystallin gehört zu der Familie der Heat shock-Proteine und fungiert als Chaperon [101]. Beschriebene Fälle von α B-Crystallinopathien sind klinisch heterogen mit unterschiedlich schweren Symptomen und unterschiedlich frühem Erkrankungsalter [93, 94].

Myotilin spielt als Komponente der Z-Scheibe eine wichtige Rolle für die Stabilität und Integrität des Sarkomers. Myotilinopathien zeichnen sich meist durch einen spät auftretenden Phänotyp aus. Exon 2 stellt einen Mutations-Hotspot im *MYOT*-Gen dar, selten wurde jedoch auch eine Mutation in Exon 9 des Gens beschrieben [93, 102, 109].

Vom *ZASP*-Gen (auch *LDB3*) werden in der Skelett- und Herzmuskulatur unterschiedliche Transkripte verwendet, die sich jedoch zum Teil überschneiden. Schon vor der Assoziation von *ZASP* mit MFM wurden Mutationen in dem Gen als ursächlich für Kardiomyopathie

beschrieben, ehe es aufgrund seiner strukturellen Funktion im Muskel als Interaktionspartner von α -Actinin als Kandidatengen für MFM untersucht wurde [103].

Filamin C dient als Bindungspartner u.a. von Aktin der Verankerung des Zytoskeletts der Muskelfaser in der Membran [110]. Exon 48 des *FLNC*-Gens wird als Mutations-Hotspot für MFM beschrieben [111].

Eine *BAG3*opathie ist eine eher seltene Form der MFM, die zu einem früh auftretenden, schweren Phänotyp führt. In den meisten Fällen wurde die Mutation c.626C>T, p.Pro209Leu in Exon 3 gefunden [93, 105]. Das von *BAG3* kodierte Enzym hat als Ko-Chaperon eine regulatorische Funktion in der Muskelzelle [112].

Mutationen im *FHL1*-Gen führen zu der einzigen X-gebundenen MFM. Das Gen wurde zuerst mit einer Reihe anderer Myopathien, v.a. der Reducing-Body-Myopathie (RBM) assoziiert, bis schließlich auch bei Patienten mit typischen Zeichen einer MFM Mutationen im *FHL1*-Gen identifiziert werden konnten [106, 113].

Das *TTN*-Gen besteht aus 363 Exons und kodiert das große Titin-Protein, das im Sarkomer aufgrund seiner Elastizität eine wichtige Rolle für die strukturelle Integrität hat. In Exon 343 wurde bei Patienten mit einer MFM mit Beteiligung der Atemmuskulatur als früh auftretendes Symptom (HMERF) eine Mutation identifiziert [107, 114].

Bei Myopathie-Patienten mit zytoplasmatischen Proteinaggregaten wurden Mutationen im *DNAJB6*-Gen detektiert, welches ein Chaperon kodiert [108].

Auch das *PLEC1*-Gen (MIM *601282) wurde im Zusammenhang mit Muskeldystrophie mit myofibrillären Defekten beschrieben, entweder mit oder ohne Kombination mit dermatologischen Symptomen (Epidermolysis bullosa, EBS-MD: MIM #226670; LGMD2Q: MIM #613723). Es kodiert Plectin, ein großes Protein, das der Quervernetzung der Intermediärfilamente in der Muskelfaser dient [115, 116]. *PLEC1* und *DNAJB6* werden jedoch nicht konsequent zu den MFM-Kandidatengen gezählt [93, 98] (siehe auch Anlage 3 und 4).

3.3.5 Hämophilie A und das *F8*-Gen

3.3.5.1 Hämophilie A und die Blutgerinnung

Hämophilie A und B (HA bzw. HB) werden X-chromosomal-rezessiv vererbt und machen zusammen mit der Von-Willebrand-Erkrankung 95 - 97 % der erblich bedingten Gerinnungsstörungen aus. Erkrankungen mit Defekten in anderen Gerinnungsfaktoren sind deutlich seltener [81, 117]. Die HA (MIM #306700) wird verursacht durch einen Mangel an Gerinnungsfaktor VIII (FVIII), wodurch die Hämostase (Abb. 2) aus dem Gleichgewicht gerät. Sie tritt mit einer Inzidenz von 1 : 5 000 männlichen Neugeborenen weitaus häufiger auf als die durch eine Faktor IX-Defizienz (FIX) verursachte HB (MIM #306900), die bei 1 : 25 000 - 30 000 Neugeborenen diagnostiziert wird [117-119]. Betroffene Patienten entwickeln je nach Schweregrad der Hämophilie mehr oder weniger starke und häufige Blutungen. Bei einem schweren Phänotyp treten diese oft spontan auf, bei leicht betroffenen Patienten dagegen nur nach Risikosituationen wie Verletzungen oder Operationen [120].

Eine häufige Komplikation v.a. bei schweren HA-Fällen ist die Entwicklung von Hemmkörpern als Reaktion auf eine therapeutische Substitution des fehlenden Gerinnungsfaktors [121]. Die Schweregrade der HA werden je nach gemessener Restaktivität des FVIII-Proteins im Plasma (FVIII:C) eingeteilt und treten unterschiedlich häufig auf. Etwa 50 % der HA-Patienten haben einen schweren Phänotyp mit einer Restaktivität von < 1 %, weitere 10 % der Fälle sind moderat (FVIII:C 1 - 5 %) und die restlichen 40 % mild (FVIII:C 5 - 40 %) [118, 122, 123].

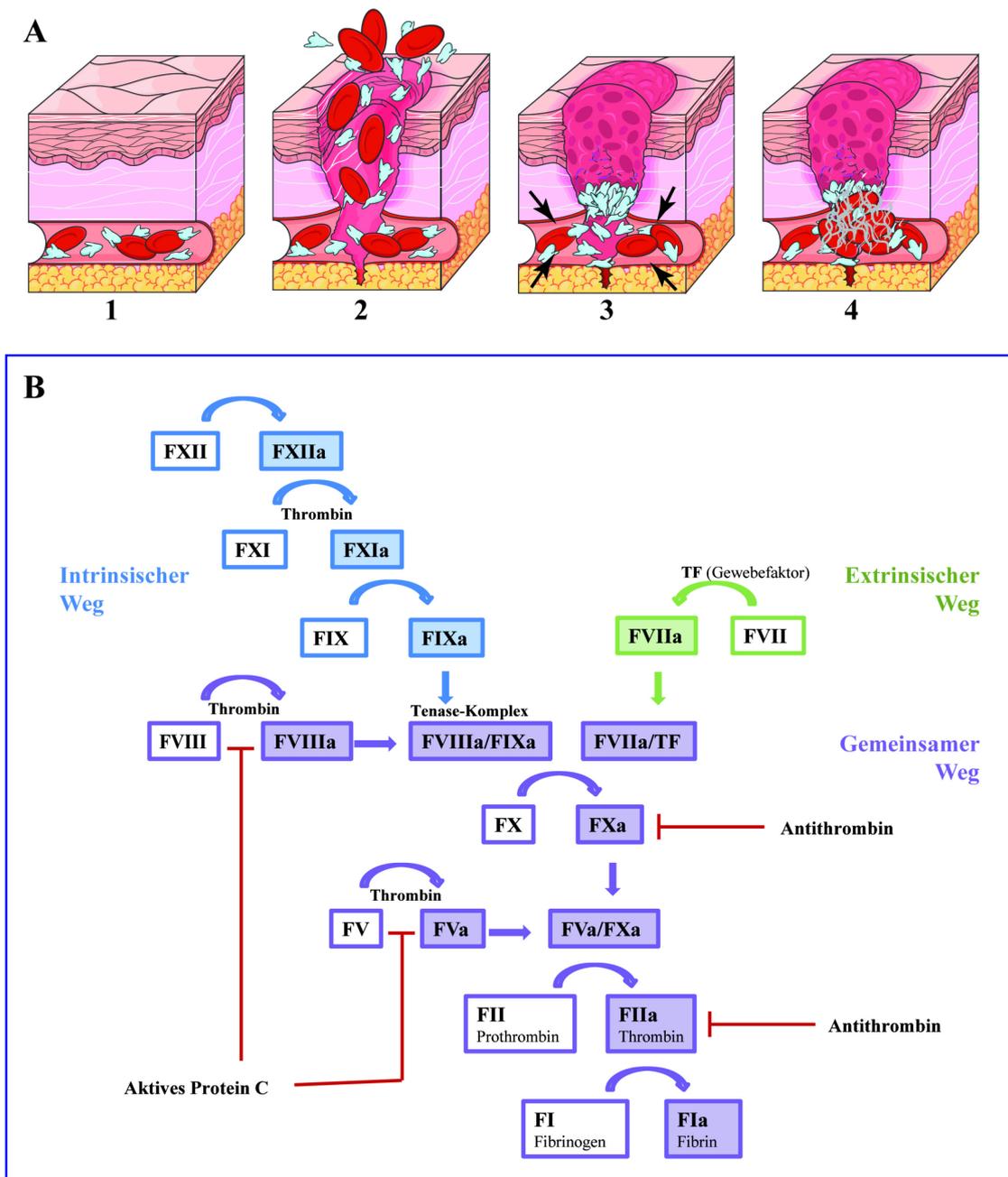


Abb. 2 Überblick über die zelluläre und plasmatische Hämostase. **(A)** Normaler Blutfluss (1), Austritt von Blut nach einer Verletzung (2). Im Rahmen der zellulären Hämostase aggregieren Thrombozyten an der verletzten Stelle und das Blutgefäß kontrahiert (Vasospasmus, angedeutet durch Pfeile; 3). Ein dichtes, stabilisierendes Fibrinnetz bildet sich am Blutgerinnsel (4). (Abbildung erstellt mit Servier Medical Art, Servier, Suresnes, Frankreich.) **(B)** Plasmatische Blutgerinnung mit intrinsischem und extrinsischem Weg und kaskadenartiger Aktivierung der Gerinnungsfaktoren (a = aktivierter Faktor). [124-126]

Das FVIII-Protein wird vor allem in Leber- und Nierenzellen synthetisiert und zirkuliert im Normalfall in inaktivierter Form gebunden an den Von-Willebrand-Faktor im Blut, [127, 128]. Es ist ein wichtiger Bestandteil der plasmatischen Blutstillung, die im Wesentlichen aus einer Reihe kaskadenartiger Aktivierungen von Proenzymen (Zymogenen) besteht, die letztendlich zu der Aktivierung von Thrombin führt (siehe Abb. 2 B). Dieses wird an mehreren Stellen als Aktivator sowie für die Spaltung von Fibrinogen in Fibrin benötigt, welches wiederum im Rahmen der zellulären Blutstillung (siehe Abb. 2 A) für die stabile Bildung eines Blutgerinnsels benötigt wird. FVIII dient dabei als Ko-Faktor und bildet mit FIX, welcher proteolytisch wirkt, den Tenase-Komplex, der wiederum den nächsten Faktor (FX) in der Kaskade aktiviert (Abb. 2 B) [124-126, 129].

Die Aufrechterhaltung der Hämostase ist essentiell, damit weder eine thrombophile noch eine hämophile Gerinnungsstörung entsteht, was komplexe Regulationsmechanismen erfordert. Wichtige Akteure sind dabei z.B. das schon erwähnte verstärkend wirkende Thrombin. Antithrombin hingegen hat eine hemmende Wirkung auf einige der Gerinnungsfaktoren, ebenso das aktivierte Protein C (siehe Abb. 2 B) [125, 129].

3.3.5.2 Das *F8*-Gen und das FVIII-Protein

Verursacht wird die HA durch verschiedene Mutationstypen im *F8*-Gen, welches auf dem langen Arm des X-Chromosoms telomernah lokalisiert ist (Xq28) und ca. 187 kb umfasst (Transkript Variante 1, ENST00000360256/NM_000132.3). Die 26 Exons variieren in ihrer Größe von 69 bis 3106 bp und die Introns von 207 bp bis 32,4 kb [130].

Intron 1 und 22 sind Hotspots für rekurrente Inversionen, die durch die homologe Rekombination von zwei bzw. drei homologen Kopien in- und außerhalb des Gens entstehen und einen schweren Phänotyp verursachen [131-133]. Die anderen Mutationen, im Wesentlichen Punktmutationen sowie kleine und große Deletionen und Insertionen treten über das gesamte *F8*-Gen verteilt auf [122].

In der molekulargenetischen Routinediagnostik wird daher stufenweise nach der ursächlichen Mutation gesucht. Nach dem Test auf die beiden Inversionen werden Promotor, Exons und unmittelbare Spleißstellen sequenziert und zuletzt eine MLPA (Multiplex ligation-dependent probe amplification) durchgeführt [134].

Das FVIII-Protein wird zunächst als inaktives Prä-Protein synthetisiert. Vor der Sekretion wird die sog. B-Domäne, welche allein durch Exon 14 (das größte Exon) kodiert wird, herausgespalten, sodass das entstehende noch inaktive Heterodimer gebunden an den Von-Willebrand-Faktor (VWF) im Blut zirkulieren kann. Bei Bedarf wird es schließlich durch Thrombin an mehreren Stellen gespalten und dadurch aktiviert (Abb. 2 B) [122, 128].

3.4 Zielsetzung und Inhalt der Arbeit

Ziel dieser Arbeit waren die Konzeptionierung, Etablierung, Validierung und Anwendung von Next Generation Sequencing in diversen Diagnostik- und Forschungsansätzen bei verschiedenen, meist heterogenen Krankheitsbildern und die Durchführung daraus resultierender Nachweisexperimente. Dabei sollten je nach Ziel und Größe der Projekte problemorientierte Lösungsansätze entwickelt werden, deren Größenmaßstab angemessen ist in Bezug auf vorhandene Ressourcen und Möglichkeiten in der Humangenetik Würzburg.

Die einzelnen bearbeiteten Projekte lassen sich dabei in Stufen mit immer größer werdendem NGS-Maßstab und –Umfang (siehe Abb. 1) einordnen: vom Screening eines einzelnen Gens oder sog. Panels mit mehreren Kandidatengenen für ein heterogenes Krankheitsbild über ein ganzes Exom bis hin zu der Analyse der kompletten genomischen Sequenz inklusive nicht-kodierender Bereiche. Dabei zeigt sich v.a. im Nachhinein betrachtet deutlich die Anpassung der Lösungsansätze an die sich unglaublich schnell weiterentwickelnden Möglichkeiten und Kapazitäten auf dem Gebiet der Anreicherungs- und Sequenziermethoden des NGS.

Projekt 1: Schnelles Screening eines einzelnen Kandidaten-Gens

Den Anfang bzw. die erste Stufe der in dieser Doktorarbeit bearbeiteten NGS-Projekte bildet die Sequenzierung eines einzelnen Gens in einer großen Gruppe an Patienten in einem einzigen Schritt. Das Projekt wurde schon vor dieser Doktorarbeit initiiert, als bei einem Patienten mit einer Hörstörung eine verdächtige Variante im *COL4A6*-Gen identifiziert wurde. Mit einem geeigneten NGS-Format sollte dann eine größere Gruppe an Hörstörungspatienten für dieses relativ große Gen gescreent werden, was schon im Rahmen der dieser Doktorarbeit vorangegangenen Diplomarbeit („*Multiplex-PCR als Probenvorbereitung für Next Generation Sequencing*“) durchgeführt wurde.

Resultierend aus diesen Ergebnissen sollten nun noch weitere Experimente zum Nachweis der Kausalität der entdeckten Variante durchgeführt werden, was letzten Endes zu der Veröffentlichung eines neuen Kandidatengens für X-chromosomale, nicht-syndromale Hörstörungen im *European Journal of Human Genetics* führte (siehe 5.1 und Anlage 1, [135]).

Projekt 2: Ein-Schritt-Analyse eines großen Gens

Die Komplettssequenzierung (der Exons) sehr großer Gene mittels klassischer Sanger-Sequenzierung ist eine mitunter sehr zeitintensive Aufgabe, vor allem wenn zusätzlich zu der Größe des Gens noch problematische Zielregionen z.B. aufgrund eines hohen GC-Gehaltes der Sequenz hinzukommen. Dieses Projekt zielte darauf ab, einen geeigneten, mit weniger Aufwand verbundenen NGS-Ansatz für die Sequenzierung der über hundert Exons des *RYR1*-Gens zu etablieren. Im Laufe dieses Projektes zeigte sich einerseits, dass nicht

jedes NGS-Format in gleichem Maß für jedes beliebige Gen geeignet ist und andererseits die Notwendigkeit einer schnellen Anpassung an die sich mit rasanter Geschwindigkeit weiterentwickelnden NGS-Technologien.

Zuerst wurde ein Ansatz nur für das *RYR1*-Gen mit einem kleinen PCR-basierten Panel gewählt. Da dies nicht optimale Ergebnisse lieferte, wurde es nicht in die Routinediagnostik übernommen.

In einem zweiten Ansatz sollten dann im Rahmen einer Kooperationsstudie mit der Uniklinik Lund, Schweden, *RYR1*, *CACNA1S* sowie weitere 62 Gene, die auch für MH und kongenitale Myopathien als Kandidatengene in Betracht gezogen wurden, in einem größeren NGS-Panel bei MH- bzw. CCD-Patienten analysiert werden. Die dabei erzielten Ergebnisse konnten schließlich in *Clinical Genetics* veröffentlicht werden (siehe 5.2 und Anlage 2, [136]).

Projekt 3: Screening mehrerer Kandidatengene für heterogene Krankheitsbilder

Ziel dieses Projektes war das Screening gleich mehrerer Kandidatengene für ein heterogenes Krankheitsbild mittels krankheits-spezifischer Panels. Dies sollte am Beispiel von zwei Ansätzen etabliert werden: ein kleines Panel für drei Gene, die mit drei Typen autosomal-rezessiver Gliedergürteldystrophie assoziiert werden (LGMD-Panel) sowie ein zweites für die Analyse von insgesamt acht Kandidatengenen für myofibrilläre Myopathien (MFM-Panel).

Im Rahmen einer Kooperationsstudie u.a. mit der Uniklinik Aachen konnten so in einer Kombination von NGS und eingehender klinisch-pathologischer Untersuchungen die kausalen Mutationen mehrerer MFM-Patienten identifiziert und näher charakterisiert werden. Die neu gewonnenen Erkenntnisse über unterschiedliche Ausprägungen von MFM sowie eine neue Mutation im *BAG3*-Gen wurden schließlich im *Orphanet Journal of Rare Diseases* publiziert (siehe 5.3 und Anlage 3, [137]).

Projekt 4: Exomanalyse

Nach der Sequenzierung lediglich einer Handvoll Kandidatengene ist der Schritt zum Umfang des nächsten Projekts, der Analyse eines ganzen Exoms, groß. Basis dieser Studie war die Auswertung der (extern durchgeführten) Sequenzierung des Exoms eines Patienten mit schwerer MFM sowie dreier Angehöriger. Resultierend aus dem Ergebnis der NGS-Daten sollten Nachweisexperimente durchgeführt werden, die plausible Beweise für die Kausalität der dabei gefundenen Variante liefern. Letztendlich konnte so die erste rezessive *MYOT*-Mutation, die MFM verursacht, in *Neurogenetics* veröffentlicht werden (siehe 5.4 und Anlage 4, [138]).

Projekt 5: Analyse der kompletten genomischen Sequenz eines Gens

Projekt 5.1: Mutationsstatistik und Diagnostik der Hämophilie A

Die Hämophilie A ist hinsichtlich der Aufklärungsquote eine dankbare genetisch bedingte Erkrankung, da es sich hier um eine monogen verursachte Erbkrankheit mit nur wenigen möglichen Differentialdiagnosen handelt. Diese Studie, welche in *Hämostaseologie* eingereicht wurde, sollte zunächst einen Überblick über das Mutationsspektrum der in der Würzburger Humangenetik analysierten HA-Patienten schaffen und besonders häufige und außergewöhnliche Mutationen hervorheben (siehe 5.5.1 und Anlage 5, [139]).

Projekt 5.2: Identifizierung tiefer intronischer Mutationen im F8-Gen mittels NGS

Trotz der hohen Aufklärungsquote in der Routinediagnostik der HA bleibt immer noch ein kleiner Prozentsatz der Fälle ungeklärt. Neueren Erkenntnissen zufolge stellen tiefe intronische Mutationen hier einen möglichen Ansatzpunkt dar. Da routinemäßig nur in den kodierenden Regionen des *F8*-Gens nach Mutationen gesucht wird, bot sich hier nun die nächste Stufe der DNA-Analyse mittels NGS an, die Sequenzierung der kompletten genomischen Sequenz eines Gens, d.h. kodierende und nicht-kodierende DNA-Bereiche. Im Rahmen dieses Projektes konnten so bei 15 ungelösten HA-Fällen mehrere verdächtige tiefe intronische Varianten identifiziert und mittels bioinformatischer *in silico*-Vorhersageprogramme als potentiell pathogen eingestuft werden. Die Studie wurde in *Thrombosis and Haemostasis* veröffentlicht (siehe 5.5.2 und Anlage 6, [140]).

Bei einem weiteren Patienten waren in Voruntersuchungen größere Kopienzahl-Variationen (copy number variations, CNVs) diagnostiziert worden. Über die Kompletsequenzierung der Introns des *F8*-Gens mittels des NGS-Ansatzes sollte getestet werden, ob sich diese Methode zur Bestimmung der CNV-Bruchpunkte eignet. Die Erkenntnisse hierzu wurden in Anlage 5 veröffentlicht (siehe 5.5.1 und Anlage 5, [139]).

Projekt 5.3: Minigen-Konstrukte zur Analyse tiefer intronischer Varianten im F8-Gen

Schließlich wurde nach einer Möglichkeit gesucht, für zumindest einen Teil der mittels NGS gefundenen tiefen intronischen Varianten den Verdacht auf Kausalität zu verifizieren. Als Methode wurde hierfür der *in vitro*-Nachweis über Minigen-Konstrukte gewählt. Mit diesem Ansatz sollte für zwei Kandidaten-Varianten der Effekt auf das Spleißverhalten der *F8*-mRNA untersucht werden. Die Ergebnisse wurden in *Thrombosis and Haemostasis* eingereicht (siehe 5.5.3 und Anlage 7, [141]).

4 METHODISCHE DETAILS

Im Folgenden sind die in dieser Arbeit angewendeten Methoden aufgelistet, je Methode sind die Verwendung in den einzelnen Projekten und ggf. detailliertere Informationen angegeben.

4.1 Polymerase-Kettenreaktion (PCR)

Projekt 1	Siehe Anlage 1 [135]	Projekt 4	Siehe Anlage 4 [138] bzw. Suppl. material
Projekt 2	Siehe Anlage 2 [136]	Projekt 5.2	Siehe Anlage 6 [140]
Projekt 3	Siehe Anlage 3 [137] bzw. unten	Projekt 5.3	Siehe Anlage 7 [141]

Details zu Projekt 3

Für PCRs im Rahmen von Nachsequenzierungen für das LGMD-Panel wurden die Reaktionen unter Standardbedingungen durchgeführt, wie auch in Anlage 3 [137] beschrieben, jedoch mit folgenden Spezifikationen: Für das Ansetzen der PCR-Reaktionen wurden die Primer für *ANO5*, *CAPN3* und *DYSF* aus der Routinediagnostik der Humangenetik Würzburg verwendet, welche unter Standardbedingungen mithilfe des Primerdesign-Programmes Primer3 v.0.4.0 entworfen [142] und bei der Firma Metabion (Planegg/Steinkirchen, Deutschland) bestellt wurden.

4.2 Marker- bzw. Haplotypanalyse

Projekt 1	Siehe Anlage 1 [135]
Projekt 4	Siehe Anlage 4 [138] bzw. Suppl. material
Projekt 5.2	Siehe Anlage 6 [140]

4.3 DNA-Sequenzierung nach Sanger

Projekt 1	Siehe Anlage 1 [135]	Projekt 4	Siehe Anlage 4 [138] bzw. Suppl. material
Projekt 2	Siehe Anlage 2 [136]	Projekt 5.2	Siehe Anlage 6 [140]
Projekt 3	Siehe Anlage 3 [137]	Projekt 5.3	Siehe Anlage 7 [141]

4.4 Probenanreicherung für Next Generation Sequencing

4.4.1 Access Array-System (Fluidigm)

- Projekt 1** Siehe Anlage 1 [135]
Projekt 2 Siehe unten
Projekt 3 Siehe Anlage 3 [137] bzw. unten

Details zu Projekt 2

Die Access Array-PCR zur Anreicherung der kodierenden Sequenzen des *RYR1*-Gens wurde gemäß Herstellerangaben (Fluidigm, South San Francisco, CA, USA) durchgeführt, wie auch in Anlage 1 [135] und 3 [137] beschrieben, jedoch mit folgenden Spezifikationen: Die Primer wurden nicht von der Firma Fluidigm entworfen und hergestellt, sondern es wurden die Primer der Routinediagnostik der Humangenetik Würzburg für die 106 Exons des *RYR1*-Gens (Transkript Variante 1, ENST00000359596/NM_000540.2) verwendet. Teilweise wurden für zu lange oder in den NGS-Daten anfangs schlecht abgedeckte Amplikons verschiedene neue Primerkombinationen getestet (Ex 3 - 5, 8 - 9, 13, 15 - 16, 21 - 22, 24, 28, 30, 32 - 34, 44, 46 - 49, 51 - 52, 55 - 56, 60, 62 - 63, 66, 68 - 69, 74 - 75, 77 - 78, 82, 86 - 87, 95, 103 - 104). Alle Primer wurden unter Standardbedingungen mittels Primer3 entworfen, von Metabion hergestellt und tragen die M13-Sequenz als Universal-Überhang an ihrem 5'-Ende (M13-Forward: TGTAACGACGGCCAGT; M13-Reverse: CAGGAACAGCTATGACC) anstelle der von Fluidigm verwendeten CS1- bzw. CS2-Universalsequenz. Exon 91 wurde in der endgültigen Version aufgrund seines problematisch hohen GC-Gehaltes von 76 % von Anreicherung und NGS ausgeschlossen, sodass *RYR1* in 88 Amplikons angereichert wurde. Um die amplifizierten Sequenzen patientenspezifisch zu markieren, wurden die Barcodes 51 - 58 von Fluidigm verwendet, jedoch wurden auch hier die CS1-/CS2-Sequenz durch die M13-Sequenz ersetzt und die Primer von Metabion hergestellt. Die Genbibliotheken wurden zweimal mit Agencourt AMPure XP Reagent Beads (Beckman Coulter, Krefeld, Deutschland) aufgereinigt.

Details zu Projekt 3

Die Probenanreicherung für das MFM-Panel wurde wie in Anlage 3 [137] beschrieben durchgeführt. Für das LGMD-Panel wurde ebenso bzw. gemäß Herstellerangaben vorgegangen, jedoch mit folgenden Spezifikationen: Von Fluidigm wurden 98 spezifische Primerpaare für die Amplifikation der insgesamt 101 Exons der drei Gene *CAPN3* (Transkript Variante 1, ENST00000397163/NM_000070.2), *DYSF* (Transkript Variante 8, ENST00000258104/NM_003494.3) und *ANO5* (Transkript Variante 1, ENST00000324559/NM_213599.2) unter Standardbedingungen entworfen und hergestellt. Um den Assay zu optimieren, wurden mit Primer3 für einige Amplikons unter Standardbedingungen neue Primer entworfen (*CAPN3* Ex 24; *DYSF* Ex 1, 12, 19, 46, 48) und von Metabion hergestellt,

sodass schließlich *CAPN3* in 22, *DYSF* in 52 und *ANO5* in 21 Amplikons angereichert wurde. Es wurden die Barcodeprimer 1 - 16 von Fluidigm verwendet.

4.4.2 SureSelect-System (Agilent)

Projekt 2 Siehe Anlage 2 [136]

Projekt 5.2 Siehe Anlage 6 [140]

4.4.3 TruSeq-System (Illumina)

Projekt 4 Siehe Anlage 4 [138] bzw. Suppl. material

4.5 Next Generation Sequencing

4.5.1 GS Junior (Roche)

Projekt 1 Siehe Anlage 1 [135]

Projekt 2 Siehe unten

Projekt 3 Siehe Anlage 3 [137]

Details zu Projekt 2

Die Sequenzierungen der mittels Access Array-PCR angereicherten Exons des *RYR1*-Gens wurden alle gemäß Herstellerangaben mit dem GS Junior (Roche Applied Science, Mannheim, Deutschland) durchgeführt, wie auch in Anlage 1 [135] und 3 [137] beschrieben.

4.5.2 MiSeq (Illumina)

Projekt 5.2 Siehe Anlage 6 [140]

4.5.3 HiSeq2000 (Illumina)

Projekt 2 Siehe Anlage 2 [136]

Projekt 4 Siehe Anlage 4 [138] bzw. Suppl. material

4.6 Molekulargenetische Diagnostikmethoden der HA-Routinediagnostik

Projekt 5.1 Siehe Anlage 5 [139]

4.7 *In vitro*-Spleißanalyse über Minigen-Konstrukte

Projekt 5.3 Siehe Anlage 7 [141]

5 ZUSAMMENFASSENDE DARSTELLUNG UND DISKUSSION DER PUBLIZIERTEN ERGEBNISSE

5.1 Schnelles Screening eines einzelnen Kandidaten-Gens

Untersuchungs-,„Objekt“ dieser Studie (Anlage 1, [135]), an der auch schon während meiner vorangehenden Diplomarbeit gearbeitet wurde, war eine ungarische Familie mit einer in drei Generationen auftretenden, nicht-syndromalen Hörstörung (Stammbaum siehe Anlage 1, Abb. 1 a). Es wurde ein X-chromosomal-rezessiver Erbgang angenommen, da die betroffenen Frauen einen im mittleren Erwachsenenalter auftretenden milden bis moderaten Phänotyp und die betroffenen Männer (inklusive des damals 11-jährigen Index-Patienten) eine kongenitale, beidseitige, schwere sensorineurale Hörstörung mit einer im MRT auffälligen Cochlea-Fehlbildung zeigten (Anlage 1, Abb. 2). Daher wurde schon vor Beginn der Diplomarbeit eine Sequenzierung des ganzen X-chromosomalen Exoms des Index-Patienten, seiner Mutter und eines betroffenen Cousins durchgeführt. Bei der Auswertung fiel nach dem Ausschließen von Polymorphismen, falsch-positiven NGS-Befunden (die sich durch eine Sanger-Sequenzierung nicht bestätigen ließen) und Varianten, die nicht auf das Vererbungsmuster passten, die Missense-Mutation c.1771G>A, p.Gly591Ser an der ersten Position von Exon 23 des *COL4A6*-Gens auf. Bioinformatische Vorhersageprogramme stufte die Variante durchgängig als potentiell pathogen ein. Über verschiedene Ansätze wurde nun versucht, die gefundene Variante als ursächliche Mutation für die Hörstörung der Familie zu bestätigen. Ein Teil der molekulargenetischen Nachweis-Experimente (DNA-Analytik) wurde dabei im Rahmen meiner Diplomarbeit und dieser Doktorarbeit durchgeführt.

Es konnte bestätigt werden, dass die fragliche Variante bei allen zur Verfügung stehenden Familienmitgliedern korrekt mit der Krankheit kosegregierte (Anlage 1, Abb. 1 a). Eine ursächliche Mutation in den bekannten, mit X-chromosomal-vererbten, nicht-syndromalen Hörstörungen assoziierten Genen *PRPS1* und *POU3F4* (DFNX1- und DFNX2-Loci) sowie dem DFNX3-Locus konnte über verschiedene Experimente ausgeschlossen werden (Sanger-Sequenzierung, MLPA, Markeranalyse).

Während der Diplomarbeit wurde ein NGS-Ansatz etabliert, mit dem schnell alle 46 Exons des *COL4A6*-Gens bei 96 weiteren Patienten mit nicht-syndromaler Hörstörung und ungeklärter genetischer Ursache analysiert werden sollten, um auf mögliche weitere Träger der gleichen oder einer anderen Mutation in diesem Gen schließen zu können. Als NGS-Technologie erwiesen sich eine Amplifikation mit dem Access Array-System (Fluidigm) und eine Sequenzierung mit dem GS Junior (Roche) als ideale und effektive Kombination

für die Anzahl an Exons und DNA-Proben. So konnten mit zwei Access Array-Chips mit jeweils 48 der 96 Patienten schnell alle Proben angereichert werden. Die erste Hälfte der Patienten wurde zunächst noch außer Haus sequenziert, die restlichen Proben in zwei Läufen am eigenen GS Junior. Die geforderte Mindest-Coverage von 15 wurde zumeist erfüllt. Als problematisch erwiesen sich 4 Exons, die bei vielen der Patienten mittels Sanger-Sequenzierung ergänzend analysiert werden mussten. Zwei der 96 Patienten konnten, vermutlich aufgrund einer schlechten Qualität der DNA-Proben oder technischer Probleme bei der Anreicherung, gar nicht ausgewertet werden. Wie sich bei anderen, im Rahmen der Diplom- und Doktorarbeit durchgeführten Access Array-Projekten gezeigt hat, lassen sich nach einem Neudesign der Primer problematischer Amplikons oft bessere Ergebnisse erzielen (siehe Projekt 2 und 3 bzw. Kapitel 5.2 und 5.3). Dies wäre sicherlich noch eine Optimierungsmöglichkeit für den *COL4A6*-Array, wenn eine weitere Anwendung geplant wäre, zumal hier nicht ein erhöhter GC-Gehalt der Zielregion (ein häufiger Grund für schwer zu amplifizierende Sequenzen [143]) das Problem sein konnte. Zusammengefasst konnte bei den analysierten Patienten zwar weder c.1771G>A noch eine andere potentiell pathogene Variante im *COL4A6*-Gen identifiziert werden. Die Variante konnte jedoch auch nicht bei der Analyse von insgesamt 407 X-Chromosomen anonymer gesunder Kontrollpersonen detektiert werden (Untersuchung mittels Sanger- oder Pyro-Sequenzierung), was es als unwahrscheinlich erscheinen lässt, dass es sich bei der Variante um einen Polymorphismus handelt. Weitere Analysen, deren Ergebnisse den Verdacht auf die Kausalität der *COL4A6*-Variante erhärteten, wurden von anderen Gruppen durchgeführt. Bioinformatische Analysen zeigten, dass sowohl Nukleotid als auch Aminosäure an der betroffenen Position hoch konserviert sind (Anlage 1, Abb. 3a). Außerdem wurde simuliert, dass der Aminosäureaustausch von Glycin nach Serin eine destabilisierende Wirkung auf die durch Glycine stabilisierte Quartärstruktur der Kollagen-Tripelhelix hat, zu der sich α -Kollagene funktionell zusammenlagern (siehe Anlage 1, Abb. 3 b) [144].

Mit Fibroblasten des Index-Patienten und seiner Mutter wurden mRNA-Analysen durchgeführt, um den bioinformatisch vorhergesagten Effekt der Variante auf das Spleißverhalten funktionell zu ermitteln. Fünf von fünf Algorithmen, die über die Software Alamut (Interactive Biosoftware, Rouen, Frankreich) abgefragt werden können (SplicesiteFinder-like [Alamut-spezifischer Algorithmus], Max-EntScan [145], NNSPLICE [146], GeneSplicer [147], Human Splicing Finder [12]), berechneten eine verminderte Wahrscheinlichkeit für die Nutzung der Original-Akzeptorspleißstelle (ass) in der veränderten Sequenz (Anlage 1, Tab. 1). Die mRNA-Analysen ergaben jedoch kein aberrantes Spleißmuster im Vergleich zur wildtypischen Sequenz. Dennoch wurde die Möglichkeit nicht ausgeschlossen, dass das Spleißverhalten im Gewebe des Innenohrs doch so ist, wie *in silico* vorhergesagt oder, dass nur ein kleiner Teil der Transkripte falsch gespleißt wird, der jedoch bei den durchgeführten mRNA-Analysen nicht nachweisbar war.

Es ist bekannt, dass *COL4A6* in vielen Basalmembranen (z.B. von Niere, Haut, Adipozyten), jedoch im Gegensatz zu *COL4A5* nicht in der Basalmembran der Nieren-Glomeruli

(glomerular basal membrane, GBM), exprimiert wird [148]. Durch Expressionsanalysen konnte gezeigt werden, dass zu *COL4A6* homologe Gene in membranösen und knöchernen Strukturen im Maus-Innenohr sowie während der Entwicklung der Ohranlagen im Zebrafisch-Embryo exprimiert werden (Anlage 1, Abb. 4). Diese Ergebnisse bekräftigen die Annahme, dass *COL4A6* eine Bedeutung für ein intaktes Gehörsystem hat und könnten eine Erklärung für die Cochlea-Malformation der Patienten sein. Durch die fehlende Expression in der GBM wurde bisher angenommen, dass Mutationen in *COL4A6* alleine (im Gegensatz zu *COL4A5*-Mutationen) nicht zu einem Alport-Syndrom passen. Bislang wurden daher auch nur Mutationen (Deletionen) beschrieben, die *beide* Gene betreffen und eine Form des Alport-Syndroms verursachen, die mit Weichteil-Tumoren (Leiomyomatose) einhergeht [54, 149]. Da jedoch bei keinem der betroffenen Familienmitglieder ein syndromales Erscheinungsbild mit Hinweis auf ein Alport-Syndrom auftrat und bei der Auswertung des X-Exoms keine Deletion im Bereich von *COL4A6* und *COL4A5* gefunden wurde, konnte mit dieser Studie die erste *COL4A6*-Mutation als Auslöser einer nicht-syndromalen Hörstörung beschrieben werden, was durch die beschriebenen Ergebnisse der klinischen, molekulargenetischen und bioinformatischen Versuche sowie Expressionsanalysen bekräftigt wird. Bei einem genetisch stark heterogenen Krankheitsbild wie den erblich bedingten Hörstörungen ist es nicht außergewöhnlich, dass ein weiteres Gen detektiert werden konnte, das bis jetzt (alleine) noch mit keiner anderen erblich bedingten Krankheit assoziiert werden konnte.

Der gewählte kleine NGS-Ansatz erwies sich als ideal für das schnelle Screening der 46 Exons des *COL4A6*-Gens bei einer großen Gruppe an Patienten. Neue Kandidatengene für genetisch heterogene Krankheitsbilder können auf diesem Weg natürlich nicht gefunden werden, für Fragestellungen wie die dieses Projektes oder auch für genetisch eindeutige (monogen oder nahezu monogen vererbte) Erkrankungen ist die Methode jedoch ein passender Ansatz.

5.2 Ein-Schritt-Analyse eines großen Gens

Wie bereits in Kapitel 3.3.2.2 beschrieben wurde, werden Mutationen im *RYR1*-Gen zu der häufigsten Ursache der Malignen Hyperthermie (MH) und Central Core Disease (CCD) gezählt [60]. In der Humangenetik Würzburg werden im Rahmen der MH- / CCD-Diagnostik meist zuerst die bekannten Hotspot-Exons und anschließend ggf. das komplette *RYR1*-Gen sequenziert, da letzteres mittels Sanger-Sequenzierung aufgrund der Gen-Größe mit 106 Exons ein aufwändiges Unterfangen darstellt. Daher war ein Ziel dieser Doktorarbeit einen geeigneten NGS-Ansatz zu testen, um die Komplettssequenzierung dieses großen Gens zu erleichtern. Angesichts der Anzahl an Exons wurde auch hier zunächst die Verbindung einer Anreicherung mit dem 48 x 48-Format des Access Array-Systems (Fluidigm) und einer 454-Sequenzierung mit dem GS Junior ausgewählt, was sich schon während der vorangegangenen Diplomarbeit bzw. Projekt 1 dieser Doktorarbeit als optimal aufeinander abgestimmte Kombination erwiesen hatte.

Zunächst wurden zwei Etablierungsrunden mit jeweils den gleichen acht Patienten mit insgesamt 24 bekannten Varianten durchgeführt. Dabei wurden, soweit möglich, die im Labor schon vorhandenen Primer der Sanger-Sequenzierungen verwendet, die die M13-Sequenz anstelle der CS1- / CS2-Sequenz als Bindungsstelle für plattform-spezifische Sequenzierprimer (GS Junior) tragen. Einige Primer wurden neu entworfen (siehe 4.4.1), um hinsichtlich der Amplikonlänge optimale Bedingungen für die gewählte NGS-Methode zu schaffen (ca. 300 - 650 bp), sodass insgesamt 91 Primerpaare eingesetzt wurden. In einem der beiden Läufe wurden alle Primerpaare einzeln (Singleplex) verwendet und daher für die Anreicherung je Lauf zwei Access Array-Chips mit 48 bzw. 43 Primerpaaren benötigt. Im zweiten Lauf wurden je zwei Primerpaare zusammen eingesetzt (Duplex) und nur ein Chip benötigt. Von den 24 bekannten Varianten konnten beim Singleplex-Lauf 19 (79 %) wiederentdeckt werden, eine weitere Variante war nur in einem von 20 Reads zu sehen und wurde daher unter Standardbedingungen (Filter für hetero- und homozygote Varianten: Frequenz > 20 %) aus der Variantenliste herausgefiltert. Bei dem Duplex-Lauf wurden nur 17 (71 %) der bekannten Varianten wiederentdeckt. Bei beiden Läufen konnte der Rest nicht detektiert werden, da das jeweils betroffene Exon eine Coverage von 0 in den NGS-Daten hatte.

Bei dem Singleplex- bzw. Duplex-Lauf waren insgesamt 21 % bzw. 22 % der Amplikons nicht ausreichend, sondern nur mit einer Coverage von < 15 abgedeckt und 11 % bzw. 18 % der Exons sogar mit weniger als fünf, meist gar keinen Reads sequenziert. Die Coverage-Verteilung schwankte generell erheblich über die Exons, sodass neben den regelmäßig ausgefallenen auch mehrere Exons mit > 100 Reads abgedeckt waren, wohingegen die Median-Coverage bei 25 bzw. 43 lag. Als Erklärung für die Ausfallquoten kamen mehrere Gründe in Betracht. Zunächst hatten die meisten der ausgefallenen Amplikons einen GC-Gehalt zwischen 60 % und 70 % und führten daher auch bei den Sanger-Sequenzierungen der Routinediagnostik gelegentlich zu Problemen. Allerdings gab es auch einzelne Exons mit einem

GC-Gehalt bis zu 67 %, die in den NGS-Daten ausreichend abgedeckt waren, sodass dies nicht alle Probleme erklärt. Ein weiteres mögliches Argument stellen unter anderem auch Amplikonlänge und Längen-Differenz zwischen allen Amplikons dar. Als optimale Länge für 454-Sequenzierungen wurden 200 - 600 bp empfohlen mit einer maximalen Differenz von 150 bp zwischen allen Amplikons, da sonst kleinere Fragmente leicht bevorzugt amplifiziert und bei der Sequenzierung überrepräsentiert werden können [150]. Eingesetzt wurden hier jedoch auch einzelne Primer für über 650 bp lange Fragmente und die Differenz betrug über 450 bp. Allerdings wurden auch bei anderen, im Labor etablierten und eingesetzten NGS-Ansätzen, die mittels GS Junior sequenziert wurden, die empfohlenen Werte überschritten, ohne, dass dabei so hohe Ausfälle zu verzeichnen waren.

Um das Panel für die weitere Verwendung zu optimieren, wurden mehrere Primerpaare neu entworfen (siehe 4.4.1), was teilweise zu einer Verbesserung führte. Exon 91, welches einen GC-Gehalt von $> 75 \%$ hat, wurde jedoch schließlich von der Anreicherung ausgeschlossen, da keines der getesteten Primerpaare Besserung brachte. Zuletzt wurden 88 Primerpaare für die Anreicherung der restlichen 105 Exons des *RYRI*-Gens eingesetzt. So wurden mit dieser Methode bei insgesamt 87 MH- und / oder CCD-Patienten, von denen bei 84 % schon vorab eine Sequenzierung der MH- / CCD-Hotspots ergebnislos durchgeführt worden war, alle Exons des *RYRI*-Gens analysiert. Die Proben wurden in drei Runden zu 48, 27 und 12 Patienten angereichert und in sechs Läufen am GS Junior mit je 12 - 14 bzw. 24 Patienten sequenziert. Auch hierbei fielen noch immer vier Exons auf, die bei keinem der Patienten auswertbar waren (Exon 13, 22, 48, 49), und mindestens sieben weitere, die in über der Hälfte der Läufe bei über 50 % der Patienten ausfielen, sodass viel kontroll-sequenziert werden musste. Testreaktionen mit einer klassischen PCR und verschiedenen Reaktionsansätzen zeigten, dass sich die betroffenen Amplikons nur unter individuellen Reaktionsbedingungen amplifizieren ließen, was bei dem recht starren Access Array-System nicht möglich ist. Dennoch konnte bei 39 der 87 Patienten eine auffällige Variante gefunden werden, von denen 25 in der Literatur oder Datenbanken als Mutationen gelistet sind und die restlichen mithilfe von Vorhersageprogrammen als wahrscheinlich kausal eingestuft werden konnten. Nur 21 der Varianten lagen dabei innerhalb der Hotspots, der Rest in anderen Exons, was die Kompletsequenzierung des *RYRI*-Gens als diagnostischen Ansatz bei MH und CCD unbedingt rechtfertigt.

Insgesamt betrachtet erscheint die gewählte NGS-Methode zwar hinsichtlich der Kapazität zunächst ein passender Ansatz für die Analyse der 106 *RYRI*-Exons zu sein. Allerdings zeigte sich in diesem Projekt als ungünstig, dass das Access Array-System eine recht unflexible Technologie ist, bei der keine spezifischen Reaktionsansätze und PCR-Bedingungen, z.B. für Amplikons mit hohem GC-Gehalt, eingesetzt werden können. *RYRI* besitzt mehrere knifflige Exons, die auch schon bei der Sanger-Sequenzierung Probleme bereiteten, daher half in diesem Projekt, im Gegensatz zu den Projekten 1 und 3 (Kapitel 5.1 und 5.3), das Entwerfen neuer Primer nur bedingt weiter und es entstand ein hoher Aufwand an Kontroll-Sequenzierungen mit klassischen Methoden. Dennoch konnte immerhin bei einem Drittel

der Patienten eine Mutation in *RYR1* detektiert und dabei bestätigt werden, dass auch außerhalb der Hotspot-Exons Mutationen auftreten. Auch der Einsatz im Labor schon vorhandener Primer mit M13-Anhang wurde erfolgreich getestet, sodass dies an die Möglichkeit der schnellen Etablierung weiterer NGS-Ansätze für andere im Labor etablierte Diagnostik-Bereiche denken lässt. Zuletzt wurde das Kombinieren von Primerpaaren auf einem Access Array-Chip getestet (Duplex), was generell funktioniert hat. Allerdings stellte sich das Finden passender Paare angesichts der Kriterien (z.B. ähnlicher GC-Gehalt, Zielregionen mindestens 5 kb voneinander entfernt, bekannte Problem-Exons bevorzugt einzeln amplifizieren) als äußerst komplex dar, was ein bekanntes Problem bei Multiplex-PCRs ist [29]. Daher wurde dieses Ziel nicht mehr weiter verfolgt, nachdem mehrere der Primer neu entworfen waren und so alles hätte neu kombiniert und getestet werden müssen. Angesichts des hohen Aufwands an Kontroll-Sequenzierungen wurde dieser etablierte NGS-Ansatz für *RYR1* nach der Anwendung für die 87 erwähnten Patienten nicht mehr weiter verwendet.

Stattdessen wurde in einer Kooperationsstudie mit der Uniklinik Lund (Schweden), ein größeres Projekt gestartet, im Rahmen dessen die Analyse von *RYR1* mit einer anderen NGS-Technologie getestet sowie nach weiteren Kandidatengen für MH-assoziierte Phänotypen gesucht werden sollte (Anlage 2, [136]). Dafür wurden insgesamt 64 potentielle Kandidatengene, die v.a. in den Kalzium-Stoffwechsel und strukturelle Funktionen der Skelettmuskulatur involviert sind (siehe Anlage 2, Supplementary table 1), für ein hybridisierungs-basiertes Panel (SureSelect-System) ausgewählt. Die DNAs von vier Patienten mit MH (NGS2, 4 und 5 männlich, NGS3 weiblich) und einer Patientin mit CCD (NGS1; siehe Anlage 2, Tab.1 und Supplementary table 2) wurden (außer Haus) mit diesem Panel angereichert und auf einem HiSeq sequenziert.

Die erhaltenen NGS-Daten zeigten eine Sensitivität der Sequenzierung von über 98 %, knapp 99 % der Exons waren mit mindestens 20 Reads abgedeckt und nur wenige Komplettausfälle einzelner Exons zu verzeichnen. Die Coverage schwankte auch hier recht stark über die Gene und Exons hinweg. Das *RYR1*-Gen ließ sich jedoch durchweg gut auswerten, lediglich in Exon 91 gab es kurze Bereiche, die nicht ausreichend abgedeckt waren. In Tab. 2 ist der Gesamt-Durchschnitt der Coverage für *RYR1* und das komplette Panel gezeigt.

Coverage	<i>RYR1</i>	Panel komplett
Min.	113	161
Max.	2005	2592
MW	933	1202

Tab. 2 Durchschnittliche Coverage für *RYR1* und das komplette MH- / CCD-Panel. Den Werten liegen die Coverage-Angaben aller fünf Patienten zugrunde. Min. = Minimum, Max. = Maximum, MW = Mittelwert.

Bei vier der fünf Patienten (NGS1 - 4) konnte in den NGS-Daten eine pathogene Mutation identifiziert werden. Tatsächlich war dies in allen vier Fällen eine bekannte, heterozygote

Mutation im *RYR1*-Gen [151-153]. Bei Patient NGS5 konnte überhaupt keine verdächtige Variante detektiert werden.

Bei jedem der fünf Patienten wurden auch zahlreiche (34 - 43) Polymorphismen detektiert. Die Patienten NGS 2 und 3 zeigten außerdem einige weitere auffällige, heterozygote Varianten, die mit bestimmten Muskelerkrankungen assoziiert und / oder von Vorhersageprogrammen als potentiell pathogen eingestuft wurden (siehe Anlage 2, Tab. 1 und 2). Allerdings passten hierbei teilweise die klinischen Merkmale der Patienten nicht zu denen der Literatur bzw. der assoziierten Erkrankung (*FKTN*, *KBTBD13*) oder die in Datenbanken gelistete Minor Allele Frequency (MAF) war zu hoch, sodass bei den betroffenen Varianten eher von einem Polymorphismus auszugehen ist (*FKTN*, *LAMA2*, *DMD*). Auch passte der Erbgang teilweise nicht, sodass die betroffenen Patienten „lediglich“ als Überträger hierfür einzustufen sind (*DMD*, *FKRP*). NGS3 trug noch je eine auffällige, heterozygote Variante in *ATP2A2*, *COL6A1* und dem riesigen *TTN*-Gen, deren Effekte unklar blieben. Das Auftreten der auffällig gehäuften, mit Muskelerkrankungen assoziierten Varianten von NGS2 und 3 ist vermutlich ein gutes Beispiel für die diskutierte Mutationslast, gemäß derer jeder von uns Anlagen-Träger seltener, rezessiver, mit Krankheiten assoziierter Varianten ist [154].

Da gemäß Literaturangaben 70 % der MH-Patienten eine Mutation im *RYR1*-Gen tragen [60], ist es nicht verwunderlich, dass vier der hier untersuchten fünf Patienten eine bekannte Mutation in diesem Gen tragen. Ob das Panel mit insgesamt 64 Genen ein sinnvoller Ansatz für eine routinemäßige Untersuchung von MH- und CCD-Patienten ist, kann nach der Analyse von lediglich fünf Patienten jedoch nicht beantwortet werden. Dazu wäre es sicherlich sinnvoll eine repräsentativere Kohorte aus MH-, CCD- und Patienten mit weiteren kongenitalen Myopathien mit bekanntem MH-Risiko zu analysieren.

Allerdings lässt sich auch schon mit den hier erzielten Ergebnissen zeigen, dass die gewählte NGS-Methode für das Ziel dieses Projektes [136] sehr gut geeignet ist, da die NGS-Daten eine hohe Sensitivität und Spezifität zeigten und das mit anderen Methoden meist schwierig zu analysierende *RYR1*-Gen erstaunlich gut auszuwerten war. Im Gegensatz zu einer Analyse dieses großen Gens über eine Sanger-Sequenzierung oder über den zuvor getesteten kleinen NGS-Ansatz (Access Array / GS Junior) waren hier (fast) keine Ausfälle zu verzeichnen und sogar das besonders knifflige und GC-reiche Exon 91 konnte analysiert werden.

Trotz der guten Ausbeute des kleinen NGS-Ansatzes (Mutationsfunde bei etwa einem Drittel der Patienten) war der dabei entstandene Kosten- und v.a. Zeitaufwand an Kontroll-Sequenzierungen so hoch, dass dieser kleine Ansatz kein Potential für einen Einsatz in der Routinediagnostik birgt und zu der Erkenntnis führte, dass das in anderen Projekten erfolgreich eingesetzte NGS-System trotz eines zunächst passend erscheinenden Formats nicht für jedes Gen ein geeigneter Ansatz ist. Das hinsichtlich der erzielten NGS-Daten vielversprechende größere Panel mit 64 Genen hingegen überzeugte zwar mit Qualität, Sensitivität und Spezifität der Daten (nicht nur des *RYR1*-Gens), zeigte allerdings auch, dass die Interpretation gefundener Varianten, von denen viele aufgrund von Datenbank- / Literaturrecherche und *in silico*-Vorhersagen zunächst relevant erscheinen, eine Herausforderung darstellt.

5.3 Screening mehrerer Kandidatengene für heterogene Krankheitsbilder

Die nächste Stufe nach der Analyse einzelner (großer) Gene über kleine NGS-Ansätze war ein kleines Panel für das Screening mehrerer Kandidatengene eines heterogenen Krankheitsbildes. In dieser Doktorarbeit wurde dies für zwei heterogene Muskelerkrankungen etabliert: Gliedergürteldystrophien (LGMDs) und myofibrilläre Myopathien (MFMs). Für beide Projekte wurde auch hier die Kombination von Access Array-PCR und GS Junior-Sequenzierung eingesetzt, da die Kapazität dieses NGS-Ansatzes ideal zu der Anzahl zu analysierender Exons passte.

5.3.1 LGMD-Panel

Zuerst wurde ein Panel für die drei in Zentraleuropa besonders häufigen autosomal-rezessiven LGMDs 2A, 2B und 2L bzw. die Gene *CAPN3*, *DYSF* und *ANO5* etabliert. Zusammen kommen die drei Gene auf 101 Exons mit mindestens 12 bp und maximal 309 bp Länge. Zunächst wurden 98 Primerpaare entworfen, die bei der Anreicherung auf zwei Access Array-Chips (Assay 1 und 2) mit 48 bzw. 50 Primerpaaren (d.h. zwei Paare jeweils als Duplex, die restlichen 94 als Singleplex eingesetzt) verteilt wurden. Bei einer Etablierungsrunde mit 18 Patienten mit bekannten Mutationen und SNPs in mindestens einem der drei Gene konnten 58 von 59 dieser bekannten Varianten (siehe Tab. 3) wiedergefunden werden, eine befand sich in einem bei der Sequenzierung ausgefallenen, d.h. nicht mit genügend Reads abgedeckten Exon. Abgesehen davon führte der Etablierungslauf zu einer Sensitivität von > 98 % und einer Spezifität von > 97 % des NGS-Panels.

Variantentyp	Anzahl	Wiedergefunden
Substitution	34	34
Deletion (1 - 16 bp)	11	11
Duplikation (1 - 7 bp)	12	11
Indel	del4ins3	1
	del5ins4	1

Tab. 3 Übersicht über die bekannten Varianten der Etablierungsrunde des LGMD-Panels. Bei den 18 LGMD-Patienten waren in den Genen *CAPN3*, *DYSF* und *ANO5* insgesamt 59 Varianten verschiedenen Typs bekannt, von denen alle wiedergefunden werden konnten, bis auf eine Duplikation von 7 bp, die in einem nicht-abgedeckten Bereich der NGS-Daten lag.

Nach der Etablierungsrunde wurde das Panel optimiert durch den Entwurf neuer Primer für problematische Amplikons, die auffällig oft ausfielen, d.h. bei der Sequenzierung häufig mit weniger als 15 Reads abgedeckt waren (siehe 4.4.1). Nach der erfolgreichen Optimierung wurde *CAPN3* in 22, *DYSF* in 52 und *ANO5* in 21 (d.h. insgesamt 95) Amplikons angereichert. Da Exon 46 des *DYSF*-Gens auch nach mehreren Neu-Entwürfen der Primer eher mäßig abgedeckt war, wurden die Primer hierfür doppelt, d.h. an zwei Positionen auf einem der beiden Assays eingesetzt, sodass die Kapazität des Access Array-Systems auf Seiten der Primer (maximal 48 Paare je Chip bei Singleplex) voll ausgenutzt wurde.

So konnte das optimierte LGMD-Panel schließlich erfolgreich validiert, akkreditiert und in der Diagnostik der Humangenetik Würzburg eingesetzt werden. Die häufige Mutation c.191dupA in einem A₇-Homopolymer in Exon 5 des *ANO5*-Gens [81] konnte aufgrund der bekannten Homopolymer-Problematik der 454-Sequenziergeräte bzw. der Pyrosequenzierungsreaktion [32-34] nicht mit Sicherheit in den NGS-Daten erkannt werden. Daher wurde dieses Exon standardmäßig parallel zum NGS mittels Sanger-Sequenzierung kontrolliert. Auch gelegentlich auftretende, einzelne schlecht abgedeckte Amplikons sowie Exons, in denen eine Mutation identifiziert werden konnte, wurden mit traditionellen Methoden kontrollsequenziert. Insgesamt konnten im Rahmen dieser Doktorarbeit bei 27 von 57 getesteten Patienten (47 %) die kausalen Mutationen identifiziert werden (homozygot oder compound heterozygot aufgrund des autosomal-rezessiven Erbgangs), falsch-positive NGS-Ergebnisse traten dabei nicht auf. In der Routinediagnostik wurden schließlich noch 3 weitere Runden mit insgesamt 26 Patienten durchgeführt. Da die LGMD-Diagnostik über das NGS-Panel jedoch dann aufgrund einer Änderung der Vorschriften der Kassenärztlichen Bundesvereinigung nicht mehr abgerechnet werden konnte (Aspekte zu dieser Problematik in Kapitel 5.6), musste das etablierte LGMD-Panel schließlich wieder eingestellt werden. Die Diagnostik läuft mittlerweile entweder wieder über die traditionelle Sanger-Sequenzierung oder über ein großes Muskelpanel, in dem insgesamt 37 (v.02) bzw. 65 (v.03) Kandidatengene für verschiedene erblich bedingte Muskelkrankheiten zusammengefasst sind (siehe 5.6).

5.3.2 MFM-Panel

Ein zweites kleines Panel wurde für die insgesamt 93 kodierenden Exons von acht Kandidatengenen für myofibrilläre Myopathien etabliert: *DES* (Exon 1 - 9), *CRYAB* (Exon 1 - 3), *MYOT* (Exon 2 - 10), *ZASP* (Exon 5 - 9 bzw. 4, 7, 8 und 10 der Transkripte in Skelett- bzw. Herzmuskulatur), *FLNC* (Exon 1 - 48), *BAG3* (Exon 1 - 4), *FHL1* (Exon 3 - 8) und *DNAJB6* (Exon 2 - 8). Für die Anreicherung mit dem Access Array-System wurden die kleineren Gene in einem Assay zusammengefasst (Assay 1), Assay 2 wurde mit 48 Primerpaaren für die 48 Exons des *FLNC*-Gens belegt.

Im ersten Lauf des Panels wurden neun Patienten analysiert. Zwecks Etablierung waren darunter fünf alte Fälle, deren bekannte Mutation wiedergefunden werden konnte. Mehrere Exons fielen auf, die bei (fast) allen Patienten nicht mit der geforderten Mindest-Coverage von 15 Reads abgedeckt waren: Exon 1 von *DES* und *BAG3*, Exon 10 von *MYOT* sowie die Exons 1, 8, 11, 12, 14 - 16, 22 und 29 von *FLNC*. Für Exon 1 von *DES* und *BAG3* wurden neue Primer entworfen, mit denen leider keine besseren Ergebnisse erzielt werden konnten. Es ist jedoch ein bekanntes Phänomen, dass die Anreicherung des meist sehr GC-reichen vorderen Abschnitts vieler Gene häufig kompliziert ist [155]; mit > 70 % bzw. > 68 % ist der GC-Gehalt des Exon 1 von *DES* und *BAG3* tatsächlich relativ hoch. Für Exon 10 in *MYOT* konnten keine geeigneten alternativen Primersequenzen gefunden werden. Da die weitere Verwendung von Assay 2 des Panels jedoch zwischenzeitlich infrage stand, wurde

für die Exons des *FLNC*-Gens kein Versuch unternommen, alternative Primer zu entwerfen. Die kritischen Exons der vier genannten Gene mussten daher bei den folgenden Läufen des MFM-Panels grundsätzlich mittels Sanger-Sequenzierung nachkontrolliert werden.

Insgesamt wurde das Panel bei 50 verschiedenen Patienten mit Verdacht auf MFM angewandt, je Lauf wurden dabei 9 - 18 Patienten analysiert. Da bei einigen Patienten teilweise schon in Voruntersuchungen das ein oder andere Kandidatengene ausgeschlossen werden konnte (Sanger-Sequenzierung in der Routinediagnostik der Humangenetik Würzburg), wurden nur bei 30 der 50 Patienten beide Teile des Panels durchgeführt, bei zwölf nur Assay 2 (*FLNC*) und bei acht weiteren nur Assay 1.

Wie auch in den vorigen Projekten wurden verdächtige Varianten, die z.B. durch *in silico*-Bewertungsprogramme wie PolyPhen-2 [14], MutationTaster [15] und SIFT [16] als potentiell pathogen eingestuft oder in der Literatur bzw. Datenbanken als Mutation beschrieben wurden, mittels Sanger-Sequenzierung nachkontrolliert und bestätigt. Dabei mussten insgesamt sieben falsch-positive NGS-Ergebnisse verzeichnet werden (siehe Tab. 4). Sechs dieser Varianten traten innerhalb eines NGS-Laufes auf (Lauf IV), die letzte in Lauf III.

Tab. 4 Falsch-positive NGS-Ergebnisse des MFM-Panels und Ergebnisse der SNPCheck-Analyse der Primer. Gezeigt sind Gen, Exon, cDNA-Koordinaten und Frequenz des falsch-positiven Ergebnisses und der betroffene Patient sowie das Ergebnis der SNPCheck-Analyse für Forward- und Reverse-Primer. Ex = Exon, For. = Forward-Primer, Rev. = Reverse-Primer, n.v. = nicht validiert.

Gen	Exon	cDNA-Koordinaten	Frequenz	Patient	SNPCheck	
					For.	Rev.
<i>DES</i>	Ex 3	c.688G>A	20,6 %	Pat. IV-7	1 x (n.v.)	1 x (n.v.)
	Ex 4	c.848C>T	22,5 %	Pat. IV-3	1 x (n.v.)	--
<i>MYOT</i>	Ex 6	c.754T>C	24,5 %	Pat. IV-4	--	--
<i>ZASP</i>	Ex 6	c.380T>C	21,7 %	Pat. IV-3	1 x (n.v.)	--
	Ex 6	c.548+10T>A	20,3 %	Pat. IV-4		
<i>FLNC</i>	Ex 1	c.225G>A	33,3 %	Pat. IV-6	--	--
	Ex 27	c.4644C>T	100 %	Pat. III-17	1 x (n.v.)	1 x (n.v.)

Alle betroffenen Stellen wurden beim NGS in beide Richtungen und mit guter Qualität sequenziert und die falsch-positiven Varianten traten balanciert auf, d.h. in Reads beider Richtungen. Ein Probenvertausch in Lauf IV konnte als Erklärung ausgeschlossen werden, da bei jedem der betroffenen Patienten mindestens eine weitere, individuelle NGS-Variante mittels Sanger-Sequenzierung korrekt bestätigt werden konnte. Um einen Allelausfall (allelic dropout) durch SNPs in der Primersequenz auszuschließen, wurden alle Primer (je Forward / F und Reverse / R) mit dem Online-Programm SNPCheck3 (<https://secure.ngsl.org.uk/SNPCheck/snpcheck.htm>) untersucht. Dabei wurde entweder keiner oder ein nicht-validierter SNP detektiert (ohne Angabe einer Minor Allele Frequency, MAF; siehe Tab. 4), allerdings wurden bei den Primern mit SNP für NGS und Kontrollsequenzierung jeweils die gleichen Primer verwendet, was einen Effekt dieser SNPs unwahrscheinlich macht. Die homozygot

detektierte Variante c.4644C>T in *FLNC*-Exon 27 konnte weder über eine Sanger-Sequenzierung mit den NGS- noch mit alternativen, im Labor vorhandenen Primern (keine SNPs) bestätigt werden. Ein Allelausfall aufgrund der detektierten SNPs in den Primersequenzen kann daher bei allen falsch-positiven Varianten ausgeschlossen werden.

Auch Pseudogene wurden als Ursache ausgeschlossen. Unter den betroffenen Genen ist zwar für *FLNC* ein Pseudogen bekannt (LOC392787 auf 7q32.1) und eine BLAST-Analyse [156] zeigte eine starke Homologie zwischen Pseudogen und Intron 45 bis 3'UTR von *FLNC* (c.7562-90_*100). Dies führte in den NGS-Daten gelegentlich zu (vermeintlich) falsch-positiven Varianten in *FLNC* (Tab. 5), wenn Sequenzen des aufgrund der Homologie mitamplifizierten und –sequenzierten Pseudogens irrtümlich in *FLNC* zugeordnet wurden. Die falsch-positiven *FLNC*-Varianten aus Lauf IV lagen jedoch außerhalb des homologen Bereiches und traten auch nicht in den Reads auf, die dem Pseudogen zuzuordnen sind.

Zuletzt handelte es sich bei den falsch-positiven Varianten auch nicht um kleine Insertionen oder Deletionen in Homopolymeren, in denen bei Pyrosequenzierungen (GS Junior) wegen ungenauer Signale und daraus resultierender Messfehler bekanntermaßen falsche Varianten angezeigt werden können [34]. Eine abschließende Erklärung für die Häufung der falsch-positiven Ergebnisse in Lauf IV konnte daher nicht gefunden werden.

Tab. 5 Sequenz-Unterschiede zwischen *FLNC* und Pseudogen. Gezeigt sind die betroffene Position (cDNA) im *FLNC*-Gen und das entsprechende Nukleotid im Pseudogen (A, C, T, G bzw. „-“, für ein im Pseudogen nicht vorhandenes Nukleotid).

<i>FLNC</i>	Pseudo- gen	<i>FLNC</i>	Pseudo- gen	<i>FLNC</i>	Pseudo- gen
c.7576T	A	c.7636T	A	c.7877G	A
c.7581C	-	c.7657C	T	c.8107G	-
c.7588A	G	c.7677T	C	c.8118C	T
c.7596T	C	c.7687T	G	c.8121T	C
c.7600G	T	c.7689T	C	c.*1A	G
c.7609G	A	c.7780+10A	G		

Etwa die Hälfte aller mit dem MFM-Panel analysierten Patienten wurde im Rahmen einer Kooperationsstudie mit der Neurologischen Klinik Aachen Teil eines Forschungsprojekts, im Laufe dessen die MFM von insgesamt 38 Index-Patienten durch eine genaue (para-)klinische (v.a. hinsichtlich multisystemischer Symptome), pathologische und genetische Untersuchung detailliert charakterisiert und die Eignung von NGS für die MFM-Diagnostik getestet werden sollte (Anlage 3, [137]). Im Rahmen der genetischen Untersuchung wurde bei 21 der Patienten in der Routinediagnostik mittels Sanger-Sequenzierung schon ein Großteil der MFM-Kandidatengene (teilweise nur die Hotspots) analysiert (Kohorte 1). Für die 17 restlichen Patienten wurden dann beide Teile des MFM-Panels angewandt (Kohorte 2) und bei 10 der Patienten aus Kohorte 1, bei denen bisher über die Sanger-Sequenzierung

keine Mutation gefunden werden konnte, nur Assay 2 (*FLNC* komplett). Exon 343 von *TTN* (ENST00000589042 / NM_001267550.1), welches im Panel nicht enthalten war, wurde schließlich noch bei 25 der 38 Patienten, bei denen in den anderen Genen keine potentiell pathogene Variante gefunden wurde, ebenfalls über eine Sanger-Sequenzierung analysiert.

Bei insgesamt 14 der Patienten konnte eine heterozygote Mutation in *DES*, *FLNC*, *MYOT*, *ZASP*, *CRYAB* oder *BAG3* identifiziert werden (siehe Anlage 3, Tab. 1), vier davon über das NGS-Panel (Patienten F1.III.1, F4.II.2, F12.III.3 und F13.1), die restlichen mittels Sanger-Sequenzierung. In *FHL1*, *DNAJB6* und *TTN* wurden keine Mutationen gefunden. Bis auf c.626C>A, p.Pro209Gln in Exon 3 von *BAG3* waren alle anderen Mutationen bekannt [102-104, 157-160]. Neun der 14 Patienten hatten eine positive Familienanamnese mit autosomal-dominantem Erbgang und in vier der Familien stand DNA-Material eines anderen erkrankten Familienmitglieds zur Verfügung, bei dem die Mutation korrekt bestätigt werden konnte (siehe Anlage 3, Abb. 1). Die anderen Fälle waren sporadisch, darunter auch die erwähnte, bisher unbekannte, *de novo* auftretende *BAG3*-Mutation, die bei den gesunden Eltern des betroffenen Patienten (F13.1) erwartungsgemäß nicht nachgewiesen werden konnte. An der gleichen Position ist, mit anderem Nukleotid- und Aminosäure-Austausch, auch die häufigste bekannte *BAG3*-Mutation c.626C>T, p.Pro209Leu lokalisiert [105].

Vor allem bei den Patienten, bei denen durch das NGS-Panel nicht nur die Hotspots, sondern alle Exons der Kandidatengene analysiert wurden, wurden neben den genannten Mutationen und zahlreichen Polymorphismen auch einige schwer klassifizierbare Varianten (unclassified variants, UVs) identifiziert, die zwar alle in der SNP-Datenbank dbSNP gelistet sind, allerdings mit einer MAF von < 1 % (siehe Anlage 3, Additional file 1). Vorhersageprogramme bewerteten diese Varianten meist widersprüchlich und eine Segregationsanalyse war zumeist nicht möglich, sodass der Effekt der UVs offen bleibt.

Die Patienten, bei denen eine Mutation identifiziert werden konnte, wurden an der Uniklinik Aachen eingehend klinisch untersucht, v.a. im Hinblick auf Erkrankungsalter, initiales Symptom und das Auftreten multisystemischer Merkmale wie eine Hörstörung oder Polyneuropathien (siehe Anlage 3, Tab. 1 und 2). Dabei fiel u.a. auf, dass bei Patienten mit der gleichen Mutation unterschiedlich ausgeprägte Phänotypen auftreten können, zum Teil sogar mit Differenzen innerhalb einer Familie wie z.B. bei Patienten und betroffenen Angehörigen mit der *DES*-Mutation c.1049G>C, p.Arg350Pro, die einen dominant-negativen Effekt hat und dazu führt, dass kein funktionelles Netzwerk aus Desmin-Intermediärfilamenten gebildet werden kann [157]. Außerdem zeigte Patient F13.1 mit der bisher unbekanntem *BAG3*-Mutation einen für *BAG3*opathien ungewöhnlich milden und spät auftretenden Krankheitsverlauf. Neben der genauen klinischen Beschreibung und teilweise der Definition neuer multisystemischer Phänotypen im Zusammenhang mit bekannten MFM-Mutationen (siehe Anlage 3, Abb. 2) konnte als Fazit dieser Studie bei 14 von 38 MFM-Patienten die kausale Mutation identifiziert werden, darunter eine neue im *BAG3*-Gen. Zwar wurden nur vier der Mutationen mittels NGS detektiert, allerdings zeigten die Ergebnisse aller über das NGS-Panel analysierten Patienten, dass eine genetische Analyse auch über die bekannten

Hotspots der Kandidatengene hinaus sinnvoll ist, da auch dort, v.a. im *FLNC*-Gen, viele Varianten gefunden wurden, die nur durch Studien wie diese besser charakterisiert und eingeschätzt werden können. Um einen pathogenen Effekt der UVs beweisen zu können, wären weitere, v.a. funktionelle Nachweis-Experimente sinnvoll, die jedoch im Rahmen dieser Studie nicht (mehr) möglich waren.

Nur bei einem der anderen, außerhalb der veröffentlichten Studie mit dem MFM-Panel analysierten Patienten konnte eine (wahrscheinlich) kausale Mutation in einem der Kandidatengene identifiziert werden. Die heterozygote Missense-Mutation c.1657G>A, p.Gly553Ser in Exon 10 des *FLNC*-Gen ist in dbSNP gelistet (rs201572079), jedoch mit einer MAF < 0,1 %. Sowohl betroffenes Nukleotid als auch die Aminosäure sind hoch konserviert und Vorhersageprogramme stuften die Variante durchgängig als potentiell pathogen ein.

Da jedoch insgesamt der größte Teil der mit dem MFM-Panel untersuchten Fälle ungeklärt bleibt, liegt die Vermutung nahe, dass es noch weitere Kandidatengene für MFM geben muss. Um diese zu finden, bietet sich z.B. ein großes Muskelpanel an mit weiteren Kandidatengenen für wichtige Struktur- und Stoffwechselproteine des Muskels oder eine Exom-Sequenzierung. Auch zeigte sich, dass eine extrem gute Phänotypisierung der Patienten (und damit generell eine enge Zusammenarbeit von Klinikern und Genetikern) vorab wichtig ist, da bei mindestens zwei der hier untersuchten Patienten die Verdachtsdiagnose MFM im Lauf der Studie korrigiert werden musste, da weitere klinische Untersuchungen eher auf eine andere Muskelerkrankung wie z.B. die FSHD oder LGMD hinwiesen.

Im Falle einer weiteren Anwendung des MFM-Panels, könnte es sicherlich noch optimiert werden, v.a. durch das Entwerfen neuer Primer für problematische *FLNC*-Exons. Auch eine Anreicherung mit einer klassischen PCR im 96 Well-Plattenformat mit individuellen Reaktionsansätzen z.B. für GC-reiche Amplikons wäre denkbar, wobei hier wieder der zeitliche Mehr-Aufwand kalkuliert werden müsste. Alternativ zur PCR-basierten Anreicherung bietet sich auch eine hybridisierungsbasierte Methode an, z.B. das SureSelect-System von Agilent, welches z.B. in Projekt 5.5.2 verwendet wurde und bei dem schwierige Zielregionen z.B. durch den Einsatz zusätzlicher Sonden leichter verstärkt werden können.

Abschließend lässt sich sagen, dass der für das LGMD- und das MFM-Panel gewählte NGS-Ansatz hinsichtlich der Kapazität ideal war, da in beiden Fällen das 48 x 48-Format des Access Array-Systems auf Seiten der Primer optimal ausgenutzt wurde. Das LGMD-Panel konnte sogar erfolgreich in der Routinediagnostik eingesetzt werden, wohingegen das MFM-Panel für eine Studie mit einer überschaubaren Anzahl Patienten zwar schnell etabliert war, allerdings für eine routinemäßige Verwendung weiter optimiert werden müsste, da mit der aktuellen Version Zeit- und Kosten-Aufwand an Nachsequenzierungen der beim NGS ausgefallenen Exons zu hoch sind. Auch wäre ein Wechsel zu einer NGS-Plattform, die nicht auf der Pyrosequenzierung basiert, denkbar, da so Mutationen in Homopolymeren leichter erkannt und falsch-positive Ergebnisse evtl. umgangen werden könnten.

5.4 Exomanalyse

Eine schwere MFM wurde auch bei dem Index-Patienten dieses Projekts diagnostiziert (Anlage 4, [138]). Die ersten Symptome progressiver Muskelschwäche traten bei ihm mit Ende dreißig auf. Seine Eltern und ein Bruder waren gesund, seine Schwester litt jedoch auch an einer schweren progressiven Myopathie, die sich bei ihr schon deutlich früher und stärker manifestierte. Um die ursächliche Mutation zu finden, wurde die Sequenzierung des kompletten Exoms der beiden Patienten (Index-Patient und Schwester), des gesunden Bruders und der gesunden Mutter durchgeführt. Die DNA des verstorbenen Vaters stand nicht zur Verfügung. Unter der Annahme, dass es sich um einen autosomal-rezessiven Erbgang handelt, wurden die Variantenlisten der Personen verglichen und gefiltert. Es wurde einerseits hinsichtlich eines echt rezessiven Erbgangs nach Varianten gesucht, die beide Patienten homozygot haben, die gesunde Mutter heterozygot und der gesunde Bruder entweder heterozygot oder gar nicht. Andererseits wurde auch für einen compound heterozygoten Erbgang gefiltert, d.h. nach einem Gen gesucht, in dem die beiden Patienten zwei auffällige Varianten heterozygot haben, von denen die gesunde Mutter jedoch nur eine und der gesunde Bruder eine oder keine aufweist. Die so erhaltenen Variantenlisten wurden weiter gefiltert hinsichtlich der Relevanz des Gens für den MFM-Phänotyp (Literaturrecherche und OMIM) und dem Vorhandensein von Pseudogenen (NCBI), bei dem ggf. einige Varianten aufgrund der falschen Zuordnung von Sequenzen direkt ausgeschlossen werden konnten.

Am auffälligsten war bei der Auswertung eine bei beiden Patienten homozygot und den Gesunden heterozygot auftretende Missense-Mutation in Exon 2 des *MYOT*-Gens (c.16C>G, p.Arg6Gly). *In silico*-Programme stuften die Variante durchgängig als potentiell pathogen ein. Die Variante konnte bei allen vier Personen mittels Sanger-Sequenzierung korrekt bestätigt werden und auch die gesunde Tochter der Patientin wurde als heterozygote Trägerin identifiziert (Anlage 4, Abb. 2 A). Das *MYOT*-Gen ist ein bekanntes Kandidatengen für MFM (MFM3, MIM #609200) sowie die phänotypisch oft überlappende LGMD1A (MIM #159000) und Spheroid-Body-Myopathie (MIM #182920). Exon 2 stellt einen bekannten Mutations-Hotspot im *MYOT*-Gen dar [102, 161, 162]. Da jedoch bisher nur heterozygote *MYOT*-Mutationen bzw. autosomal-dominante Erbgänge beschrieben wurden, wurde bei den beiden Patienten dieser Studie zunächst über eine MLPA eine Deletion bzw. der Verlust eines Allels ausgeschlossen. Über eine Markeranalyse konnte dann gezeigt werden, dass beide Patienten einen homozygoten Haplotyp für die durch die Marker abgedeckte Region haben (Anlage 4, Abb. 2 B), obwohl die Eltern (soweit bekannt) nicht konsanguin sind. Um einen Polymorphismus auszuschließen wurden 372 Allele gesunder Kontrollpersonen an der fraglichen Position c.16C sequenziert, sie alle zeigten jedoch nur das wildtypische Allel C.

Die Variante c.16C>G war weder in der Literatur noch in Datenbanken als Polymorphismus oder Mutation aufgeführt. Allerdings wurde eine Missense-Mutation einer LGMD1A-

Patientin beschrieben, die zwar das benachbarte Nukleotid, jedoch die gleiche Aminosäure betraf (c.17G>A, p.Arg6His, heterozygot), lediglich mit einem Austausch nach Histidin anstatt Glycin [163]. Sowohl das betroffene Nukleotid (c.16C) als auch die betroffene Aminosäure (p.Arg6) der in diesem Projekt beschriebenen neuen Variante sind hoch konserviert. Bestätigung für den pathogenen Effekt der Missense-Mutation fand sich auch in histopathologischen Untersuchungen einer Muskelbiopsie des Patienten. So konnten die für MFM pathologisch charakteristischen Zeichen myofibrillärer Degeneration ausgehend von den Z-Scheiben und Proteinaggregate in den Muskelfasern festgestellt werden (Anlage 4, Abb. 1 und Supplementary Fig. 1). Bei massenspektrometrischen Analysen der Aggregate zeigte Myotilin den höchsten Anteil unter den angereicherten bzw. aggregierten Proteinen (Anlage 4, Supplementary Tab. 1).

In der Summe bekräftigten und bestätigten alle Nachweis-Experimente die Annahme, dass es sich bei der gefundenen Variante c.16C>G um die erste beschriebene homozygote und somit autosomal-rezessive Missense-Mutation im *MYOT*-Gen handelt. Außerdem zeigte sich, dass NGS und Proteom-Analysen (Laser-Mikrodissektion plus Massenspektrometrie von Proteinaggregaten aus defekten Myofibrillen) eine gute Kombination in der MFM-Diagnostik sind. Proteom-Analysen eignen sich nicht nur als diagnostischer Ansatz bei histopathologisch auffälligen Erkrankungen wie der MFM und für eine detailliertere Charakterisierung des Phänotyps. Wie sich in diesem Projekt zeigte, sind sie gerade auch für große NGS-Ansätze wie eine Exom-Sequenzierung, bei der viele unbekannte, oft schwer zu klassifizierende Varianten gefunden werden (unclassified variants, UVs), eine gute Möglichkeit die Ergebnisse molekulargenetischer Diagnostik zu unterstützen bzw. zu bestätigen.

Ein häufig genannter Nutzen des WES ist, dass neue Kandidatengene entdeckt werden können. Auf diese Weise konnten sicherlich auch schon einige komplizierte alte Fälle nach mitunter jahrzehntelangem, mühsamen Durchsequenzieren einzelner Kandidatengene mit der Sanger-Methode endlich gelöst und das Wissen über genetisch heterogene Krankheitsbilder erweitert werden. Ein neues MFM-Gen wurde in diesem Projekt zwar nicht entdeckt, aber zumindest konnte eine neue Vererbungsform – die autosomal-rezessive Myotilinopathie – aufgedeckt werden. Ein großer Vorteil war hier auch, dass mehrere gesunde und kranke Familienmitglieder mittels WES analysiert werden konnten.

Für Exom-Sequenzierungen werden häufig Trio- bzw. Familienanalysen durchgeführt, da durch geschicktes Vergleichen und Filtern der Variantenlisten von betroffenen und nicht betroffenen Personen so die Variantenzahl oft erheblich eingegrenzt werden kann. Je nach (vermutetem) Erbgang kann z.B. bezüglich der Frequenz des veränderten Allels (z.B. 20 - 70 % für heterozygote, > 75 % für homo- oder hemizygoten Varianten) oder der Minor Allele Frequency (MAF) bekannter Varianten gefiltert werden. Da die Verteilung von Polymorphismen und pathogenen Mutationen in SNP- und Mutationsdatenbanken nicht immer korrekt ist und in beiden auch Varianten der jeweils anderen Kategorie gelistet sind, können bei der Suche nach der kausalen Mutation nicht grundsätzlich alle in SNP-Datenbanken bekannten Varianten aus den Listen aussortiert werden. Beim Setzen eines MAF-Filters zum Ein-

grenzen langer Variantenlisten gilt es verschiedene Aspekte zu beachten wie z.B. den Erbgang oder die Erkrankungsform. So kann die Heterozygoten-Frequenz rezessiv wirkender Mutationen in der gesunden Bevölkerung deutlich höher sein als bei dominant wirkenden Mutationen. Auch Erkrankungsalter und Schwere des Phänotyps sind hier ein Aspekt. So kann es vorkommen, dass z.B. in der Datenbank des 1 000-Genome- oder ähnlichen Projekten, bei denen Varianten (vermeintlich) gesunder Personen gelistet werden, auch pathogene Mutationen milder und daher nicht diagnostizierter Erkrankungsformen oder von erst im hohen Erwachsenenalter manifestierenden Krankheiten aufgeführt sind. Außerdem ist jedes Individuum Anlageträger zahlreicher Risikoallele und pathogener Mutationen mendelscher Erkrankungen, die über WES detektiert werden können [164]. Daher wird bei der Analyse von Exom-Sequenzierungen häufig ein „Auswertungs-Panel“ als Daten-Filter angewandt (im Gegensatz zu Anreicherungs-Panels wie in Kapitel 5.3, die schon vor der Sequenzierung als „Gen-Filter“ dienen), z.B. ein Auswertungs-Panel für Muskelerkrankungen. Dadurch werden nur mit Erkrankungen des Bewegungsapparates assoziierte Gene ausgewertet, um für die Fragestellung irrelevante Zufallsbefunde zu vermeiden.

5.5 Analyse der kompletten genomischen Sequenz eines Gens

5.5.1 Mutationsstatistik und Diagnostik der Hämophilie A

In der Humangenetik der Universität Würzburg wurden bis Ende 2014 in insgesamt 18 Jahren HA-Diagnostik 2 671 männliche Index-Patienten mit der Indikation HA untersucht. Im Rahmen der HA-Routinediagnostik wurde dabei Schritt für Schritt versucht, die genetische Ursache der Gerinnungsstörung zu finden. Diese stufenweise Diagnostik führte zu einer hohen Trefferquote und einer charakteristischen Mutationsverteilung. Die Statistik des Würzburger Patientenkollektivs (Abb. 3 und Anlage 5, Tab. 1, [139]) bietet angesichts der Größe der untersuchten Kohorte einen repräsentativen Überblick über das Mutationsspektrum im *F8*-Gen und ähnelt im Wesentlichen der Statistik anderer Studien mit größeren Patientengruppen sowie gen- bzw. krankheitsspezifischer Datenbanken [122, 165, 166]. Aktuell sind in der Factor VIII Variant Database über 2 000 verschiedene Mutationen im *F8*-Gen gelistet, die bei knapp 5 500 Patienten gefunden wurden (<http://www.factorviii-db.org/>; Stand 30.05.2015).

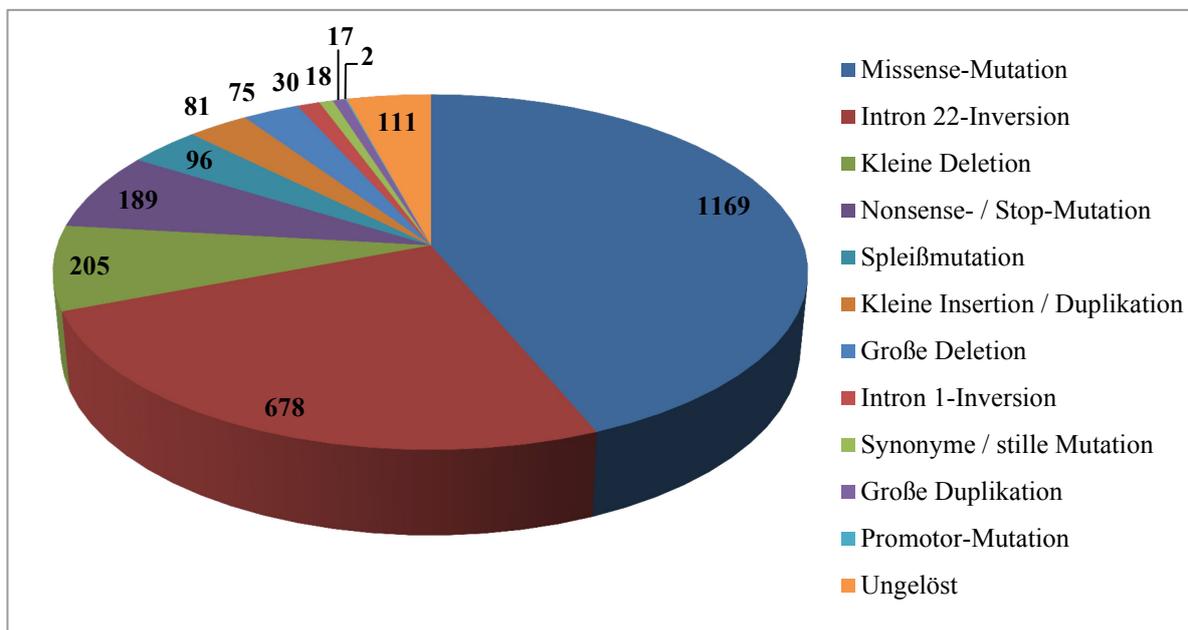


Abb. 3 Mutationsspektrum der in Würzburg untersuchten HA-Patienten. Anzahl der gefundenen Mutationstypen im *F8*-Gen der von Januar 1997 – Dezember 2014 untersuchten 2 671 männlichen Index-Patienten.

Die bekannten Mutationen verteilen sich im Allgemeinen über das gesamte *F8*-Gen, jedoch sind einige Exons und bestimmte Nukleotid-Positionen in der untersuchten Kohorte und auch in der Literatur als Hotspots für Punktmutationen aufgefallen. Abb. 4 zeigt die Verteilung der in Würzburg gefundenen Missense-Mutationen über die 26 Exons des *F8*-Gens. In

rot ist dabei in den Hotspot-Exons 11, 14 und 23 jeweils der Anteil der drei insgesamt am häufigsten entdeckten Missense-Mutationen markiert.

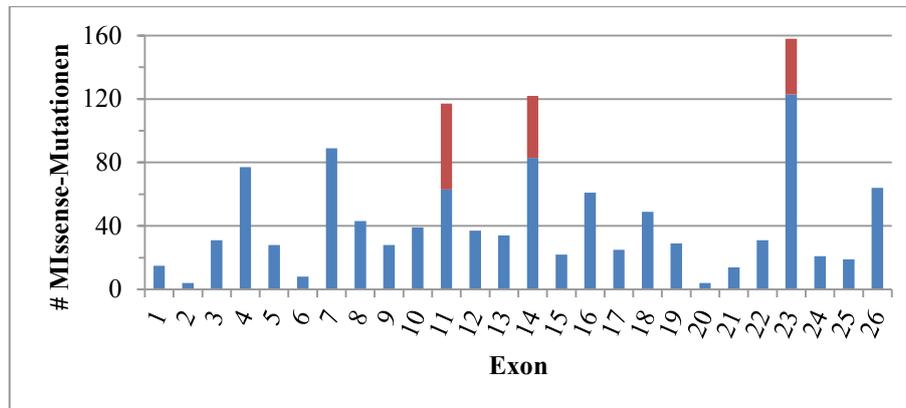


Abb. 4 Verteilung der bei insgesamt 1169 Patienten diagnostizierten Missense-Mutationen über die 26 Exons des *F8*-Gens. In Exon 11, 14 und 23 ist jeweils in rot der Anteil der insgesamt am häufigsten gefundenen Missense-Mutationen markiert: c.1636C>T in Exon 11 (54-mal gefunden), c.5096A>T in Exon 14 (39-mal), c.6506G>A in Exon 23 (35-mal).

Ein Großteil der Missense- und Nonsense-Mutationen trat an einigen der insgesamt 70 im *F8*-Gen vorkommenden CpG-Dinukleotide (Cytosin-Phosphat-Guanin) auf, meist in Arginin-Codons. Dies ist ein bekanntes Phänomen, da Cytosine in CpG-Inseln oft methyliert sind und daher leicht zu Thymin desaminiert werden können [167, 168]. Tab. 6 zeigt eine Übersicht über weitere Mutations-Hotspots, drei Adenin-Homopolymere (A_n) in Exon 14, an denen insgesamt 33 % der gefundenen kleinen Duplikationen oder Deletionen identifiziert wurden, die zu einer Verschiebung des Leserasters führen [122, 169, 170].

Tab. 6 Übersicht über die an drei Poly-As (A_n) gefundenen Mutationen. Angegeben ist die Häufigkeit (n) der jeweiligen Mutation in der untersuchten Patientenkohorte, mit einer Benennung der Mutation gemäß HGVS-Nomenklatur / Human Genome Variation Society [171].

Poly-A-Koordinaten	n
Mutation	
c.3629_3637, p.1210_1213 (A_9)	
c.3637dupA	15
c.3637delA	37
c.4372_4382, p.1458_1461 (A_8TA_2)	
c.4379dupA	17
c.4379delA	8
c.4380delT	5
c.4820_4825, p.1607_1609 (A_6)	
c.4825dupA	12
c.4825delA	1

Trotz der hohen Aufklärungsrate blieb die genetische Ursache von etwa 4 % der 2 671 HA-Fälle nach Durchführung der Routinediagnostik ungeklärt. Diskutierte Lösungen für unge-

löste Fälle sind z.B. Mutationen in den Genen für Interaktionspartner des FVIII-Proteins, die differentialdiagnostisch abgeklärt werden können, so z.B. das *VWF*-Gen (die milde Form 2N des Von-Willebrand-Syndroms, der Typ Normandy, ähnelt in ihrem Phänotyp einer milden HA) [172] und *LMAN1* oder *MCFD2* (Mutationen in diesen Genen stehen in Zusammenhang mit einer kombinierten FV / FVIII-Defizienz, F5F8D) [173, 174]. Auch komplexere genomische Veränderungen wurden diskutiert [134]. Epigenetische Veränderungen in regulatorischen Bereichen des *F8*-Gens konnten in einer Studie mit 80 Patienten nicht als Ursache einer HA bestätigt werden [175]. Am häufigsten wurden Mutationen in den Introns, meist fernab der regulären Spleißstellen (tiefe intronische Mutationen), als weitere Möglichkeit beschrieben, die zur Aktivierung kryptischer Spleißstellen und daher alternativen Transkripten führen [5, 176-180].

5.5.2 Identifizierung tiefer intronischer Mutationen im *F8*-Gen mittels NGS

Die in Kapitel 5.5.1 genannten tiefen intronischen Mutationen standen im Mittelpunkt dieses Projektes. Wie schon erwähnt, wurden sie in einigen Fällen als genetische Ursache meist milder HA-Erkrankungen beschrieben. In dieser Studie wurde nun ein Ansatz entwickelt, mittels NGS die komplette genomische Sequenz des *F8*-Gens bei 15 ungelösten HA-Fällen zu analysieren, um auf diesem Weg nach tiefen intronischen Mutationen suchen zu können. Vierzehn der Patienten hatten eine milde und einer (Pat. 8) eine moderate HA. Bei allen Patienten blieb die in Kapitel 5.5.1 (Anlage 5, [139]) beschriebene konventionelle Routinediagnostik ohne Befund, d.h. Mutationen in den kodierenden Bereichen, den unmittelbaren Spleißstellen und dem Promotor des *F8*-Gens sowie die beiden Inversionen (Intron 1 und Intron 22) konnten als genetische Ursache der HA ausgeschlossen werden. Auch Mutationen in den Hotspot-Exons 18 - 24 des *VWF*-Gens, die einen HA-ähnlichen Phänotyp verursachen können [122], wurden nicht gefunden (Anlage 6, [140]).

Ein weiterer Patient (Pat. 16) wurde in die Studie mit einbezogen, bei dem in Voruntersuchungen mittels MLPA größere CNVs diagnostiziert werden konnten: eine Duplikation der Exons 1 - 22 und eine Triplikation der Exons 23 - 25. Bei diesem Patienten sollte mittels NGS nach den genauen Bruchpunkten der CNVs gesucht werden (siehe Anlage 5, [139]).

Bei der Projektplanung wurden Repeat-Elemente vom Alu-Typ von der zu sequenzierenden Zielregion (*F8*-Gen) vorsorglich ausgeschlossen, da angenommen wurde, dass bei der Probenanreicherung genomweit vorhandene, meist stark homologe Alu-Kopien [181] mit vervielfältigt würden. Bei der Sequenzierung würde dies eine Überrepräsentierung der betroffenen Abschnitte mit jedoch unspezifisch zugeordneten Sequenzen mit sich bringen. Durch den Ausschluss der Alu-Elemente umfasste die Zielregion zwar nur ca. 95 % der genomischen Sequenz des *F8*-Gens, die ausgeschlossenen Bereiche wurden jedoch von beiden Seiten immerhin etwa 50 bp weit an-sequenziert. Dennoch führte das Vorhandensein zahlreicher anderer Repeat-Elemente im *F8*-Gen und generell dem ganzen menschlichen Genom [182] bei einem Alignment der NGS-Daten zunächst gegen die *F8*-Referenzsequenz zu einer sehr unausgeglichene Abdeckung (Coverage). Vor allem Abschnitte mit LINE /

L1- und LTR-Repeats waren bei der Sequenzierung stark überrepräsentiert. Durch ein Alignment gegen die gesamte X-chromosomale Referenzsequenz konnte dem Problem jedoch zum Teil Abhilfe verschafft werden. Insgesamt konnten nun etwa 66 % (X-Alignment) anstatt 48 % (F8-Alignment) der Sequenzen zugeordnet und spezifischer auf Repeat-Kopien im ganzen X-Chromosom verteilt werden, wodurch die Coverage in der Zielregion weitaus ausgeglichener war.

Bei der Auswertung der NGS-Daten der einzelnen Patienten wurde zunächst die Coverage-Verteilung über die Introns aller Patienten verglichen. Dabei fiel bei Patient 8 ein Bereich von etwa 9 - 10 kb in Intron 1 auf, der deletiert zu sein schien (siehe Anlage 6, Abb. 1). Die genauen Bruchpunkte konnten in den NGS-Daten nicht ermittelt werden, da beide innerhalb von Alu-Repeats lagen, die bei der Sequenzierung ausgeschlossen waren. Zudem waren zwischen den Grenzen des deletierten Bereiches einige kurze Abschnitte mit einer geringen Zahl Sequenzen abgedeckt, bei denen es sich um unspezifisch zugeordnete Sequenzen innerhalb von Repeat-Elementen sowie die Int1h-1-Region handelte, in der fälschlicherweise die Reads der stark homologen Kopie (Int1h-2) zugeordnet wurden, die etwa 140 kb weiter telomerwärts des *F8*-Gens liegt [133] und aufgrund der Homologie mit angereichert und sequenziert wurde. Sanger-Sequenzierungen von PCR-Produkten mit Primern nahe der Deletions-Bruchpunkte bestätigten die Deletion des kompletten Bereiches von 9 185 bp und die Bruchpunkte konnten in einem Motiv von 11 bp jeweils in der Mitte zweier stark homologer AluY-Repeats identifiziert werden (g.154,232,081_154,241,265 bzw. c.143+9420_144-4206del; siehe Anlage 6, Abb. 5).

Die schon bekannten CNVs von Patient 16 waren in den NGS-Daten gut erkennbar. Abb. 1 in Anlage 5 zeigt deutlich den Unterschied der Coverage zwischen dupliziertem (Exon 1 - 22) und tripliziertem Bereich (Exon 23 - 25) und der einfachen Dosis von Exon 26. Es ist erkennbar, dass die Bruchstellen tiefer in den betroffenen Introns liegen müssen, die genauen Bruchpunkte waren jedoch weder über die Coverage-Ansicht aus Abb. 1 in Anlage 5 noch direkt im Alignment identifizierbar. Es fielen weder Stellen im Alignment auf, an denen Coverage oder Qualität plötzlich einbrachen noch Sequenzen, die nur zum Teil zugeordnet werden konnten. Um die Bruchpunkte näher charakterisieren zu können, mussten demnach andere Methoden hinzugezogen werden, wie z.B. PCR und Sequenzierung mit geschickt gewählten Primerkombinationen. Schwierigkeiten stellte hierbei jedoch die Reihenfolge der zusätzlichen Exon-Kopien des Patienten dar, die weder aus den MLPA- noch den NGS-Daten hervorging und für die es zahlreiche Anordnungsmöglichkeiten gibt. So war die exakte Bruchpunktbestimmung genau dieses Patienten auch schon Teil einer anderen Studie der Humangenetik Würzburg, die der vorliegenden Arbeit voranging. Unter der Annahme, dass die Kopien bei dem Patienten tandemartig angeordnet sind, wurde mit verschiedenen Primerkombinationen versucht, mittels Long Range PCR (LR-PCR) und anschließender Sequenzierung die Bruchpunkte zu charakterisieren, was jedoch erfolglos blieb. Bekannt ist nur, dass die gleiche Kopienzahl-Veränderung bei einer weiblichen Verwandten des Patienten in heterozygoter Form auftrat [183]. Ein möglicher Ansatz, Kopie-

Reihenfolge und Bruchpunkte doch noch aufzuklären, wäre eine RNA-Analyse, die in diesem Fall jedoch nicht möglich war, da es sich um eine pseudonymisierte Studie handelte.

Das Vergleichen der Variantenlisten der 15 Patienten ohne bekannte Mutation ergab für jeden der 15 Patienten mindestens einen Treffer und insgesamt eine Liste von 23 verdächtigen Kandidatenvarianten (Single Nucleotide Variants, SNVs; siehe Tab. 7 und Anlage 6, Tab. 2). Darunter waren drei schon publizierte tiefe intronische Mutationen, sechs weitere Varianten waren in der SNP-Datenbank dbSNP [10] gelistet, allerdings mit einer globalen Minor Allele Frequency (MAF) von $< 1\%$ und nur 0 - 5 hemizygoten Allelträgern gemäß der Datenbank des 1 000-Genome-Projektes [184], was eine mögliche Pathogenität nicht von vorneherein ausschließt. Sechs der verdächtigen Varianten traten bei je zwei bzw. drei der 15 Patienten auf.

Der Effekt schon publizierter intronischer *F8*-Varianten wurde meist über die Analyse der mRNA aus dem Blut der Patienten untersucht [5, 177, 179, 180], was in dieser pseudonymisierten Studie, wie schon erwähnt, jedoch nicht möglich war. Daher wurde hier mithilfe verschiedener bioinformatischer Programme der potentielle Effekt der SNVs *in silico* bewertet. Die Combined Annotation Dependent Depletion-Analyse (CADD) fasst verschiedenste funktionelle Aspekte und Annotationen in einem Faktor, dem sog. C-score, zusammen, welcher einen Hinweis auf einen möglichen funktionellen Effekt eines SNV gibt [13]. Der C-score wurde für die 23 Kandidatenvarianten sowie eine Reihe bekannter Mutationen verschiedenen Typs im *F8*-Gen ermittelt, die Ergebnisse sind in Tab. 7 und Anlage 6, Abb.3 gezeigt (Abfrage des Scores über <http://cadd.gs.washington.edu/score>, 29.04.2014). Dabei war der Score für Stopp-Mutationen am höchsten, gefolgt von den Missense-, Spleiß- und stillen Mutationen und zuletzt Polymorphismen mit dem niedrigsten Score. Eine ähnliche Verteilung wurde auch von Kircher *et al.* für alle genomweit möglichen SNVs ermittelt [13].

Der C-score der publizierten tiefen intronischen *F8*-Mutationen lag zwischen 2,9 und 8,0. Der Wert für die 23 Kandidatenvarianten lag dabei in einem ähnlichen Rahmen (0,006 - 8,8), wobei die Scores der drei bekannten Mutationen mit 0,006, 0,458 und 1,79 eher im unteren Bereich lagen. Dies zeigt, dass trotz der generellen Tendenz, die sich aus der Studie von Kircher *et al.* [13] ableiten lässt (je höher der C-score, desto wahrscheinlicher die Schädlichkeit eines SNV), ein niedriger C-score die Pathogenität einer Variante nicht ausschließt und *in silico*-basierte Bewertungen zwar als Hinweis oder Tendenz, jedoch nicht als einziges Argument für die Bewertung einer Variante herangezogen werden sollten.

Tab. 7 Tiefe intronische Kandidatenvarianten der 15 HA-Patienten. Gegeben sind Intron-Nummer und cDNA-Koordinaten gemäß HGVS-Nomenklatur. Die drei bekannten Mutationen sind fett gedruckt. Für alle bekannten Varianten ist die Referenznummer gemäß dbSNP (rs-Nummer) bzw. HGMD (CS-Nummer) aufgeführt und ggf. die MAF gemäß 1 000-Genome-Projekt. Der vorhergesagte Effekt auf das Spleißverhalten ist gemäß Spleißvorhersage der Alamut-Software (Interactive Biosoftware, Rouen, Frankreich) angegeben, der C-score gemäß der Combined Annotation Dependent Depletion-Analyse (CADD) [13]. hemi = hemizygot, ass / dss = acceptor / donor splice site, ESE = exonic splice enhancer.

Intron	Variante	Patient	Referenz	Alamut Spleißvorhersage	C-score
	c.144-10758_144-10757insTATA	12	–	<i>de novo</i> ass	8.095
	c.144-5714C>T	10	–	<i>de novo</i> ass	0.476
1	c.144-3700C>T	8	<i>rs138903905</i> ; MAF: ALL 0.5 % (5x hemi), EUR 0.2 % (1x hemi)	<i>de novo</i> ass	1.858
	c.144-1259C>T	4	<i>rs184460929</i> ; MAF: ALL 0.2 % (2x hemi), EUR 0%	<i>de novo</i> ass	2.389
6	c.787+3221A>G	2, 9	–	<i>de novo</i> ass	1.290
7	c.1010-842G>A	3, 7, 14	<i>rs148954517</i> ; MAF: ALL 0.7 % (3x hemi), EUR 2.1 % (3x hemi)	<i>de novo</i> ass	2.088
9	c.1444-1833T>C	10	–	<i>de novo</i> ESE	8.749
	c.1537+325A>G	5	CS1211666; [180]	<i>de novo</i> dss	0.458
10	c.1538-18G>A	13	CS080683; [178]	<i>de novo</i> ass, normale ass abgeschwächt	1.790
12	c.1903+2536A>G	8	<i>rs150944728</i> ; MAF: ALL 0.4 % (3x hemi), EUR 0 %	<i>de novo</i> ass	0.380
	c.2114-6139C>T	10	–	<i>de novo</i> ass	0.130
13	c.2114-4382_2114-4381insG	3, 7, 14	<i>rs200992472</i> ; [5]; MAF: ALL 0.7 % (3x hemi), EUR 2.1 % (3x hemi)	<i>de novo</i> ass / dss	0.637
	c.2114-1681T>A	13	–	<i>de novo</i> dss	0.236
14	c.5219+293T>A	1	–	<i>de novo</i> dss	2.915
16	c.5587-98G>A	2, 9	–	<i>de novo</i> dss (und ass)	0.006
	c.5587-93C>T	6	CS117569; [5, 179]	<i>de novo</i> dss	0.006
18	c.5998+182A>G	8	–	<i>de novo</i> dss (und ass)	4.658
	c.5999-740T>C	4	–	kein Effekt	3.328
22	c.6429+2788T>C	12	<i>rs28430439</i> ; MAF: ALL 0.1 % (0x hemi), EUR 0%	<i>de novo</i> ass	6.540
	c.6430-4825T>C	8	–	<i>de novo</i> ESEs	0.312
25	c.6900+8517C>T	8	–	<i>de novo</i> dss (und ass)	3.711
	c.6901-725T>C	2, 9	–	<i>de novo</i> ass	2.285
3'UTR	c.*56G>T	11, 15	–	<i>de novo</i> dss (und ass)	5.270

Der potentielle Effekt der Varianten auf das Spleißverhalten wurde mittels verschiedener *in silico*-Vorhersageprogramme bewertet, die in der Alamut-Software (Interactive Biosoftware, Rouen, Frankreich) integriert sind und die Wahrscheinlichkeit einer Donor- oder Akzeptor-Spleißstelle (dss / ass) angeben sowie ESE-Motive (exonic splice enhancer) identifizieren. Für die meisten SNVs wurde die Aktivierung einer ass vorhergesagt (siehe Tab. 7 und Anlage 6, Abb. 2 und Supplementary Figure). Nicht immer handelte es sich dabei um eine *de novo*-Spleißstelle, in manchen Fällen wurde auch eine vorhandene, aber in gesunden Individuen nicht genutzte kryptische Spleißstelle durch die intronische Variante verstärkt und dadurch ihre Aktivierung bzw. Nutzung wahrscheinlicher. Die Folge der Aktivierung kryptischer Spleißstellen ist meist die Insertion intronischer Sequenzabschnitte in die mRNA, was oft zu einem vorzeitigen Stopp-Codon führt, wie es z.B. von Castaman *et al.* für die Mutation c.5587-93C>T in Intron 16 beschrieben wurde [179], die hier bei Patient 6 gefunden wurde. Das alternative verkürzte Transkript (bei verschiedenen tiefen intronischen Varianten auch mehrere Transkripte möglich) kann zu einem verkürzten Protein führen oder durch Nonsense-mediated mRNA Decay (NMD) direkt wieder abgebaut werden [185]. Je nachdem, wie stark eine kryptische Spleißstelle genutzt wird, kann dennoch in geringen Mengen normales *F8*-Transkript synthetisiert werden, wie z.B. Inaba *et al.* für die Mutation c.1537+325A>G gezeigt haben [180], die hier bei Patient 5 detektiert wurde. Dadurch erklärt sich auch der oft berichtete milde HA-Phänotyp von Patienten mit tiefer intronischer *F8*-Mutation [5, 178-180].

Zusammenfassend lässt sich sagen, dass sich der NGS-Ansatz als zweckdienliche und komfortable Methode bewiesen und den gestellten Anspruch, die Introns des *F8*-Gens nach Mutationen absuchen zu können, erfüllt hat. Effektivität zeigte die Methode einerseits dahingehend, dass bei allen 15 ungelösten HA-Fällen mindestens eine auffällige tiefe intronische Variante detektiert werden konnte, die zumindest *in silico* als potentiell pathogen eingestuft werden konnte. Dies zeigt, dass tiefe intronische Varianten ein vielversprechender Ansatzpunkt für die Klärung weiterer bisher ungelöster HA-Fälle sind. Effektiv ist der Ansatz auch dahingehend, dass nicht nur die Introns nach Mutationen abgesucht, sondern auch nahezu alle anderen HA verursachenden Mutationstypen entdeckt werden können: neben Punktmutationen in allen Bereichen des *F8*-Gens (Promotor, Exons, Introns, 5' und 3'UTR) können auch kleine und (hemizygot) große Deletionen und Insertionen bzw. Duplikationen aufgedeckt werden. Das macht bei den Würzburger Patienten immerhin über zwei Drittel der Fälle aus. Mittels NGS könnten daher die zeit- und kostenintensivere Sanger-Sequenzierung und MLPA durch eine einzige Methode ersetzt werden.

Die Grenzen sind jedoch auch klar: komplexere genomische Veränderungen wie die beiden Inversionen und auch die genauen Bruchpunkte von CNVs können dabei nicht ermittelt werden. Auch kommt NGS ebenso wenig wie die Sanger-Sequenzierung ohne traditionelle Methoden (z.B. mRNA-Analyse) zum funktionellen Nachweis der Pathogenität (tiefer) intronischer SNVs aus. *In silico*-Vorhersageprogramme liefern zwar schon den richtigen Hinweis auf die potentielle Schädlichkeit bzw. Auswirkung der entdeckten SNVs, scheitern

jedoch in einem kleinen Prozentsatz [186]. Sie können daher funktionelle Nachweisexperimente nicht ersetzen und sollten auf jeden Fall durchgeführt werden, ehe eine tiefe intronische Variante in einem diagnostischen Befund berichtet wird.

Der getestete NGS-Ansatz bietet sich aufgrund seiner Effektivität und Praktikabilität generell gut für den Einsatz in der HA-Routinediagnostik an. Es besteht jedoch auch Potential für Optimierungen. Die überrepräsentierten Bereiche hauptsächlich in LINE / L1- und LTR-Repeats, in denen die Coverage deutlich über dem Durchschnitt lag, könnten z.B. ebenso wie die Alu-Elemente von der Zielregion ausgeschlossen werden, um dadurch insgesamt eine höhere Qualität und ausgeglichene Coverage in den spezifischen Bereichen erzielen zu können. Dadurch entgingen zwar noch mehr Teile des *F8*-Gens der Analyse, allerdings war die Qualität der Sequenzen in den betroffenen Bereichen ohnehin so schlecht, dass eine (Punkt-) Mutation hier nicht weiter aufgefallen wäre.

Im Rahmen dieser Arbeit wurde das Design der Hybridisierungssonden tatsächlich schon für eine zweite Sequenzierungs-Runde optimiert, dabei wurde die Zielregion durch Ausschluss besonders problematischer Repeat-Regionen auf 81,5 % des *F8*-Gens reduziert, was jedoch mit einem deutlichen Zugewinn bezüglich Coverage-Verteilung und Qualität in den spezifisch abgedeckten Bereichen einherging. Diese zweite Runde hatte einen vielversprechenden Ansatz: Es wurden 16 neue HA-Patienten ausgewählt, von denen jeweils zwei bis vier die gleiche Missense-Mutation im *F8*-Gen, jedoch einen unterschiedlich schweren Phänotyp hatten. Die Hypothese war, dass die Patienten mit schwererem Phänotyp zusätzlich zu der bekannten Missense-Mutation noch eine tiefe intronische Variante haben, die in der Routinediagnostik nicht erkannt werden konnte und die z.B. das ohnehin veränderte FVIII-Protein oder seine Synthese so modifiziert, dass es schließlich zu der schweren Hämophilie kommt. Bei der Auswertung der NGS-Daten fielen jedoch weder CNVs noch verdächtige SNVs in den Introns der Patienten auf, die von Vorhersageprogrammen als potentiell pathogen eingestuft wurden. Eine mögliche Erklärung für den milderen Phänotyp einiger Patienten wäre z.B. eine zusätzliche, protektive Mutation in einem der gerinnungsfördernden Faktoren, die der gerinnungshemmenden *F8*-Mutation entgegenwirkt und diese gewissermaßen bis zu einem gewissen Grad „antagonisiert“ [187]. Eine *VWF*-Mutation konnte auch bei diesen Patienten zumindest in den Exons 18 - 24 ausgeschlossen werden, jedoch wurden auch Mutationen in den Exons 17 und 25 - 27 in Zusammenhang mit einem Von-Willebrand-Syndrom mit hämophilie-ähnlichem Phänotyp beschrieben [188]. Eine *F5F8D* schien eher unwahrscheinlich, da bei den Patienten keine *FV*-Defizienz bekannt war. Da es sich jedoch auch hier um eine pseudonymisierte Studie handelte und die zur Verfügung stehenden Gerinnungswerte (FVIII:C) extern erhoben wurden, ist nicht gesichert, dass ausschließlich eine FVIII-Defizienz bei den Patienten vorliegt oder ob der bekannte FVIII:C-Wert auf einer oder mehreren Messungen beruht. Daher sollte eine Mutation in einem der anderen Gerinnungsfaktoren oder auch den Genen *LMAN1* oder *MCFD2* nicht grundsätzlich ausgeschlossen werden. Auch eine weiter vom *F8*-Gen entfernte Mutation in einem Bereich mit bisher unbekannter regulatorischer Funktion für das *F8*-Gen wäre denkbar.

5.5.3 Minigen-Konstrukte zur Analyse tiefer intronischer Varianten im *F8*-Gen

Da mRNA-Analysen aus Patientenproben nicht möglich waren, wurde nach einer Alternative gesucht, den Effekt der in Kapitel 5.5.2 (Anlage 6, [140]) gefundenen tiefen intronischen Varianten funktionell zu überprüfen. Daher wurde mithilfe sog. Minigen-Konstrukte ein System zur *in vitro*-Spleißanalyse entwickelt, mit dem zumindest zwei der SNVs aus der NGS-Studie untersucht wurden (Anlage 7, [141]). Dazu wurde die zu analysierende Sequenz kloniert, in HEK-Zellen (Human Embryonic Kidney) transfiziert und schließlich die mRNA extrahiert und analysiert. Als Varianten wurden c.5587-98G>A in Intron 16 (Minigen 1) ausgewählt, gefunden bei Patient 2 und 9, sowie c.5998+182A>G in Intron 18 (Minigen 2), gefunden bei Patient 8 (Aufbau der Minigen-Konstrukte siehe Anlage 7, Abb.1 A). Die von Castaman *et al.* beschriebene Mutation c.5587-93C>T [179], welche 5 bp neben der Minigen 1-Variante liegt und hier bei Patient 6 gefunden wurde, wurde als Positiv-Kontrolle für Minigen 1 eingesetzt, außerdem für beide Ansätze eine untransfizierte Kontrolle sowie eine mit wildtypischer Sequenz.

Fünf von fünf Vorhersageprogrammen berechneten für c.5587-98G>A die Verstärkung einer im Wildtyp nicht genutzten kryptischen Spleißstelle (dss). Für c.5998+182A>G zeigten vier von fünf Programmen eine hohe Wahrscheinlichkeit für eine *de novo* dss (siehe Abb. 5 A und C).

Für die Positiv-Kontrolle c.5587-93C>T konnten Castaman *et al.* zeigen, dass die *in silico* vorhergesagte *de novo* dss (siehe Abb. 5 B) bei c.5587-95 tatsächlich genutzt wird. Zusätzlich wird eine kryptische ass an der Position c.5586+137 genutzt, sodass das dazwischen liegende Stück von 56 bp in die mRNA inseriert wird. Durch die Insertion entsteht eine Verschiebung des Leserasters und nach 24 bp ein Stopp-Codon (TGA), was zu einem verkürzten Transkript führt [179]. Dies konnte mit den Minigen-Versuchen bestätigt werden. Allerdings konnten Castaman *et al.* in der RNA direkt aus Patientenblut sowie über ein ähnliches *in vitro*-System neben der aberranten auch wildtypische *F8*-mRNA nachweisen [179], was in dieser Arbeit über das *in vitro*-System nicht der Fall war (siehe Anlage 7, Abb. 1 B).

Der gleiche Mechanismus wie für c.5587-93C>T konnte mit den hier durchgeführten Experimenten auch für die Minigen 1-Variante c.5587-98G>A nachgewiesen werden. Die Sequenzierung der PCR-Produkte der cDNA zeigte einheitlich die Insertion intronischer Sequenzen in die mRNA (siehe Anlage 7, Abb.1 B und C). Die laut *in silico*-Programmen verstärkte dss bei c.5587-102 (siehe Abb. 5) und die gleiche kryptische ass wie bei der Positiv-Kontrolle (c.5586+137) wurden aktiviert. Dadurch wurden 49 bp als kryptisches Exon gespleißt, in dem an der gleichen Stelle wie bei der Kontrolle ein vorzeitiger Abbruch entsteht (siehe Anlage 7, Abb. 1 C), was die Pathogenität der Variante c.5587-98G>A sehr wahrscheinlich macht. Allerdings ist trotz dieser *in vitro*-Ergebnisse nicht auszuschließen, dass auch bei dieser Mutation (ähnlich wie bei der Positiv-Kontrolle) *in vivo* zusätzlich zu der veränderten mRNA eine wildtypische synthetisiert wird, die jedoch im *in vitro*-System nicht nachweisbar war.

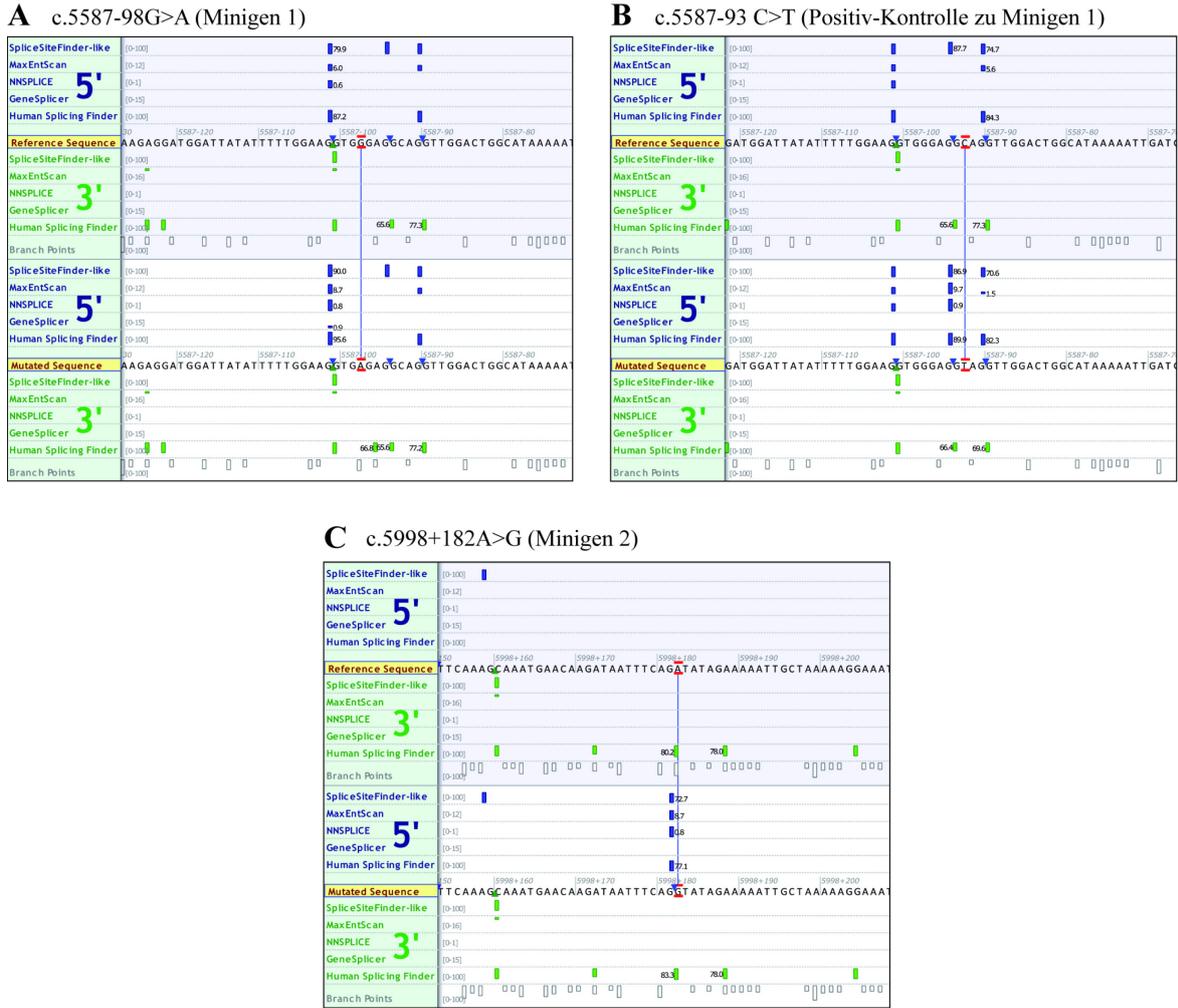


Abb. 5 Splice Prediction nach Alamut (Interactive Biosoftware, Rouen, Frankreich). (A) Für c.5587-98G>A sagen fünf von fünf Programmen die Verstärkung einer im Wildtyp nicht genutzten kryptischen Spleißstelle (dss) an der Position c.5587-102 vorher. (B) Für c.5587-93C>T wird eine *de novo* dss bei c.5587-95 berechnet. (C) Auch für c.5998+182A>G wird mit vier von fünf Programmen eine *de novo* dss vorhergesagt.

Bei Minigen 2 konnte weder über Gelelektrophorese noch über Sequenzierung ein Unterschied zwischen Wildtyp und verändertem Konstrukt detektiert werden (siehe Abb. 6 / Anlage 7, Abb. 1 B), was die Pathogenität der Variante c.5998+182A>G unwahrscheinlich macht.

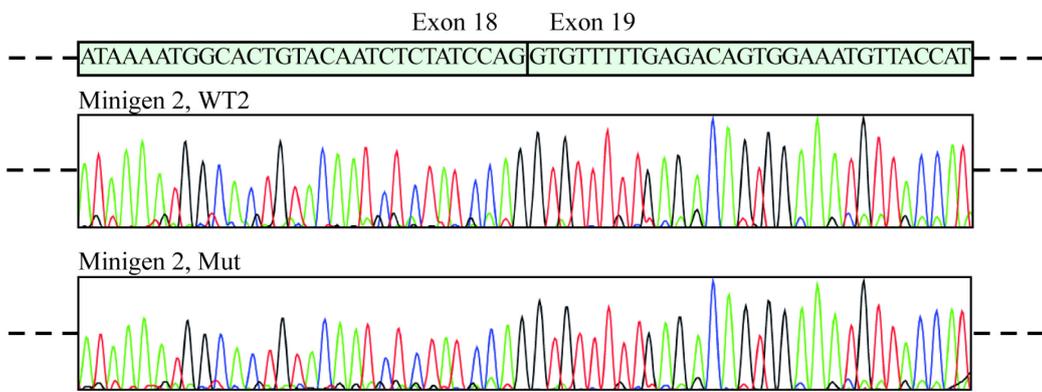


Abb. 6 Sequenzierungs-Ergebnis von Minigen 2. Die Sequenz des Konstrukts mit wildtypischer Sequenz (WT2, oben) und Mutante mit c.5998+182A>G (Mut, unten) sind identisch, beide Konstrukte wurden daher gleich gespleißt.

Insgesamt hat sich der Weg über Minigen-Konstrukte zum Nachweis der Pathogenität der getesteten tiefen intronischen Varianten als zuverlässige und effektive Alternative zu mRNA-Analysen aus Patientenblut bewiesen. Die Verlässlichkeit der Methode lässt sich durch verschiedene Aspekte begründen: zunächst konnte *in vitro* für die Positiv-Kontrolle der gleiche alternative Spleißmechanismus wie von Castaman *et al.* [179] nachgewiesen werden. Außerdem wurde die wildtypische Sequenz in den Konstrukten genauso wie erwartet gespleißt, d.h. an den natürlichen Exon-Intron-Grenzen, an denen auch *in vivo* gespalten wird. Die Effektivität des Ansatzes zeigte sich darin, dass für eine der beiden getesteten Varianten *in vitro* ein abnormales Spleißverhalten gezeigt werden konnte, das genau der *in silico*-Vorhersage entsprach und den gleichen Mechanismus auslöste wie die bekannte Positiv-Kontrolle.

Allerdings konnte über das *in vitro*-System nicht wie bei Castaman *et al.* [179] zusätzlich zu der veränderten auch normal gespleißte mRNA bei den Konstrukten mit Mutation detektiert werden. Eine mögliche Erklärung hierfür kann einerseits sein, dass in dem *in vitro*-System natürlich nicht die gleichen Bedingungen wie *in vivo* herrschen, angefangen bei der unmittelbaren Umgebung der Mutation (im Minigen-Konstrukt war nur ein kleiner Teil des *F8*-Gens enthalten) über die regulatorischen Abschnitte (der verwendete Vektor besitzt einen recht starken Promotor aus dem Cytomegalievirus, kurz CMV, dessen Aktivität mit der von *in vivo* agierenden Promotor- und anderen regulatorischen Bereichen nicht direkt vergleichbar ist) bis hin zu den künstlichen Bedingungen einer Zellkultur. Eine Erklärung wäre allerdings auch, dass das kryptische Exon trotz der gleichen Mutation bei manchen Patienten stärker und bei manchen weniger genutzt wird, sodass unterschiedliche Mengen an normalem Transkript entstehen können.

Intron 16 (mit 286 bp das drittkleinste der *F8*-Introns) bzw. die Region um die Mutationen c.5587-93C>T und c.5587-98G>A kann aufgrund der Ergebnisse als Hotspot für tiefe intronische *F8*-Mutationen betrachtet werden. Zusammen mit den Patienten dieser Doktorarbeit [140, 141], denen von Castaman *et al.* [179] und von Pezeshkpoor *et al.* (hier wurde die Mutation c.5587-93C>T auch gefunden, jedoch nicht nach HGVS-Nomenklatur benannt, sondern c.5586+194C>T) [5], wurden dadurch sieben nicht-verwandte Patienten mit einer der beiden Mutationen beschrieben.

Das Ergebnis für Minigen 2 bestätigt die Vorsichtsmaßnahme, dass Vorhersageprogramme *in silico*-Methoden bleiben und funktionelle Analysen nicht ersetzen können. Die laut *in silico*-Bewertung vielversprechende Variante c.5998+182A>G führte den *in vitro*-Ergebnissen zufolge nicht zu der vorhergesagten Aktivierung einer starken *de novo* dss. Allerdings konnten in den NGS-Daten des betroffenen Patienten noch vier weitere verdächtige intronische Varianten und eine große Deletion von > 9 kb in Intron 1 identifiziert werden [140]. Zudem ist auch eine Möglichkeit, dass die Variante c.5998+182A>G nur zu so geringen Mengen an verändertem *F8*-Transkript führt, dass diese über das *in vitro*-System nicht nachweisbar sind, ähnlich wie im Minigen 1-Konstrukt nicht, wie bei Castaman *et al.* [179], auch wildtypisches Transkript nachgewiesen werden konnte. Daher bleibt die Vermu-

tung, dass der moderate Phänotyp dieses Patienten v.a. durch die Deletion aber auch durch einen Effekt seiner tiefen intronischen Varianten verursacht wird.

Der getestete Minigen-Versuch bietet sich auch für die Analyse weiterer tiefer intronischer Varianten an und ist eine effektive Alternative, wenn mRNA-Analysen direkt aus Patienten-Blut nicht möglich sind. Probleme können jedoch entstehen bei zu großen Introns, die nicht oder nur schwierig klonierbar sind. Ein möglicher Ansatz, um die Problematik zu großer Introns bzw. zu tiefer intronischer Varianten zu umgehen, wäre z.B. statt eines Konstruktes aus insgesamt vier Exons und drei Introns nur das betroffene Intron und die beiden benachbarten Exons zu klonieren (die in dieser Studie klonierten Abschnitte umfassten ca. 3,3 kb bzw. 3,5 kb). Auch könnte man künstliche Introns synthetisieren, die neben der Mutation nur die wichtigen, zum Spleißen benötigten Elemente enthalten (Konsensussequenzen der *ass* und *dss*, Verzweigungspunkt und Polypyrimidinbereich [189]), jedoch nicht die volle Länge des natürlichen Introns umfassen. Allerdings ist hierbei fraglich, ob die Ergebnisse, die in dieser konstruierten Umgebung mit künstlich aneinandergereihten Spleißsignalen erzielt werden, auf eine natürliche Umgebung, in der z.B. eine kryptische *ass* in passender Nähe zu einer *de novo* *dss* liegt, unmittelbar übertragbar sind.

Das Fazit des Hämophilie-Projektes dieser Doktorarbeit, welches die Sequenzierung der *F8*-Introns und daraus resultierende funktionelle Nachweisexperimente umfasste, ist jedenfalls, dass insgesamt ein Drittel (5 von 15) der ungelösten HA-Fälle geklärt werden konnte. Patient 5, 6 und 13 sind jeweils Träger einer schon bekannten tiefen intronischen Mutation [178-180] und für Patient 2 und 9 konnte die Pathogenität der gefundenen Variante in dieser Arbeit *in vitro* bestätigt werden.

Innerhalb dieser Doktorarbeit wurde das Hämophilie-Projekt unter den etablierten NGS-Ansätzen in der größten NGS-Kategorie „Sequenzierung der genomischen Sequenz“ eingeordnet. Es handelte sich hierbei zwar nicht um die Sequenzierung eines kompletten Genoms (Whole Genome Sequencing / WGS), sondern nur der kompletten Sequenz eines einzelnen Gens. Allerdings werden trotzdem die Vor- und Nachteile der Sequenzierung nicht-kodierender DNA-Sequenzen deutlich. Neben den Aspekten, die auch schon in Kapitel 5.4 im Zusammenhang mit Exom-Sequenzierungen genannt wurden (z.B. neue Kandidatengene; Polymorphismen und Mutationen; Zufallsbefunde; geschicktes Filtern), kommen bei einer Genom-Sequenzierung die schwer zu beurteilenden Varianten in den riesigen nicht-kodierenden Bereichen (immerhin nahezu 99 % [4]) des Genoms hinzu. Um hier die Pathogenität einer Variante nachweisen zu können, sind oft umfangreiche Nachweis-Experimente nötig wie Expressionsanalysen, Versuche im Tiermodell oder die Sequenzierung unzähliger Kontrollen, wie verschiedene Projekte dieser Doktorarbeit verdeutlichen. Zu wenig ist bisher über die nicht-kodierenden Bereiche des menschlichen Genoms bekannt, über vom Ziellokus weit entfernte Bereiche mit regulatorischer Funktion und über die natürliche genetische Varianz. Allerdings kann auch nur über die Durchführung von WES- und WGS-Studien neues Wissen hierzu gewonnen werden.

5.6 Fazit und Ausblick

An den Projekten bzw. dem Umfang der einzelnen NGS-Projekte dieser Doktorarbeit wird deutlich, wie sich die Dimensionen des NGS allein in den letzten drei bis vier Jahren mit rasantem Tempo weiterentwickelt haben, wie neue, schnellere Technologien mit immer höherem Durchsatz entwickelt wurden und die ersten NGS-Technologien schon wieder veraltet sind. So soll z.B. der Support für die 454-Technologien (454 / Roche) Mitte 2016 schon wieder komplett eingestellt werden [1], dafür investiert Roche jetzt in Sequenziertechnologien der dritten Generation, in ein System für Einzelmolekül-Sequenzierungen über das sog. „Sequencing by Expansion“-Prinzip (Homepage von 454 Life Sciences: <http://www.454.com> > Resources & Support > News > 6/26/2014). Auch in der Humangenetik Würzburg wird der GS Junior kaum noch genutzt, momentan lediglich für spezielle Ansätze wie „deep bisulfite sequencing“ für epigenetische Fragestellungen. In der Molekulargenetik werden häufig große Panels eingesetzt, so sind z.B. die Gene aller im Laufe dieser Doktorarbeit etablierten kleinen Muskelpanels (das komplizierte und große *RYR1*-Gen sowie die Gene des LGMD- und MFM-Panels), die über das Access Array-System angereichert und am GS Junior sequenziert wurden, mittlerweile zusammen mit vielen weiteren „Muskelgenen“ in einem großen Muskelpanel mit 37 (v.02) bzw. 65 (v.03) Kandidatengenereint, bei dem hybridisierungs-basiert angereichert und mit Illumina-Geräten sequenziert werden kann. Das bedeutet, dass die während der ersten Hälfte der Doktorarbeit etablierten, zu dem Zeitpunkt optimal ausgewählten und teilweise erfolgreich in der Diagnostik eingesetzten Panels schon während der zweiten Hälfte der Doktorarbeit eingestellt wurden und im Prinzip schon veraltet sind, wenn man bedenkt, welche Kapazitäten und Möglichkeiten größere Ansätze wie das Muskelpanel bieten, die zum jetzigen Zeitpunkt „State of the Art“-Methoden darstellen.

Ein Problem bei dem Einsatz solcher Panels (bzw. NGS im Allgemeinen) in der Routinediagnostik stellt in Deutschland (zum aktuellen Zeitpunkt) das kassenärztliche Abrechnungssystem dar, da NGS nicht als Regelleistung bei den gesetzlichen Krankenkassen gilt und nicht im EBM (Einheitlicher Bewertungsmaßstab) der Kassenärztlichen Bundesvereinigung mit eigener Abrechnungsziffer aufgeführt wird. Ausnahmen können lediglich beantragte Einzelgenehmigungen sein, für die z.B. der Berufsverband Deutscher Humangenetiker e.V. (BVDH) eine Liste mit Gründen für den Einsatz von NGS entworfen hat, die den Krankenkassen bei der Einschätzung von Einzelanträgen helfen sollen (Homepage des BVDH: <http://www.bvdh.de/> > Leistungsabrechnung > NGS-Hochdurchsatzverfahren, Stand 30.05.2015: „Die Kommission [...] hat deshalb [...] Indikationen für den Einsatz der NGS erarbeitet. Da die NGS aber keine im EBM aufgeführte Regelleistung ist, soll sie Ärzten und Krankenversicherungen Kriterien zur Beurteilung von Einzelfallanträgen an die Hand geben.“). Für einige Gene bzw. Krankheiten, für die eine sog. Komplexziffer vorhanden ist, die keine bestimmte Methode vorschreibt, kann direkt NGS angewandt werden.

Andernfalls gilt jedoch für eine Diagnostik mittels DNA-Sequenzierung die Ziffer 11322 mit der hinsichtlich der anzuwendenden Methode eindeutigen Beschreibung „*Nachweis oder Ausschluss einer krankheitsrelevanten oder krankheitsauslösenden genomischen Mutation mittels Sequenzierung menschlicher DNA nach der Kettenabbruchmethode nach Sanger*“ (Homepage der Kassenärztlichen Bundesvereinigung / KBV: <http://www.kbv.de/html/ebm.php>, Stand 30.05.2015).

Die Abrechnungsproblematik ist jedoch keineswegs der einzige Aspekt des NGS, der Fragen aufwirft und weiterer Entwicklungen (wie der schon lang diskutierten Komplexziffer für Next Generation Sequencing) und Richtlinien bedarf. Weitere Aspekte, die auch schon an verschiedenen Stellen dieser Doktorarbeit genannt wurden, sind u.a. die langfristige Datenspeicherung, der Datenschutz und genetische Diskriminierung, Zufallsbefunde und andere ethische Gesichtspunkte. Auch kann NGS zum jetzigen Zeitpunkt z.B. (noch) nicht chromosomale und zytogenetische Techniken zum Nachweis komplexer genomischer Veränderungen ersetzen.

Nichtsdestotrotz ist NGS eine äußerst innovative Technik, deren Vorzüge nicht nur die in dieser Arbeit diskutierten Aspekte hinsichtlich der Erforschung und Diagnostik erblich bedingter Erkrankungen, der personalisierten Medizin und des menschlichen Genoms im Allgemeinen sind. Auch in anderen Bereichen, z.B. mit evolutionären Fragestellungen, der *de novo*-Sequenzierung des Genoms unerforschter Organismen, der Grundlagenforschung biologischer Vorgänge oder der Analyse von Transkriptomen (RNA-Sequenzierung), ist Next Generation Sequencing mittlerweile eine “State of the Art”-Methode.

6 LITERATURVERZEICHNIS

1. **van Dijk EL, Auger H, Jaszczyszyn Y, et al.** *Ten years of next-generation sequencing technology.* Trends Genet **2014**; 30(9): 418-426.
2. **Ng SB, Turner EH, Robertson PD, et al.** *Targeted capture and massively parallel sequencing of 12 human exomes.* Nature **2009**; 461(7261): 272-276.
3. **Goldstein DB, Cavalleri GL.** *Genomics: understanding human diversity.* Nature **2005**; 437(7063): 1241-1242.
4. **Frebourg T.** *The challenge for the next generation of medical geneticists.* Hum Mutat **2014**; 35(8): 909-911.
5. **Pezeshkpoor B, Zimmer N, Marquardt N, et al.** *Deep intronic 'mutations' cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA.* J Thromb Haemost **2013**; 11(9): 1679-1687.
6. **Akman HO, Kakhlon O, Coku J, et al.** *Deep Intronic GBE1 Mutation in Manifesting Heterozygous Patients With Adult Polyglucosan Body Disease.* JAMA Neurol **2015**; 72(4): 441-445.
7. **Anczukow O, Buisson M, Leone M, et al.** *BRCA2 deep intronic mutation causing activation of a cryptic exon: opening toward a new preventive therapeutic strategy.* Clin Cancer Res **2012**; 18(18): 4903-4909.
8. **Webb TR, Parfitt DA, Gardner JC, et al.** *Deep intronic mutation in OFDI, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23).* Hum Mol Genet **2012**; 21(16): 3647-3654.
9. **Cooper DN, Krawczak M.** *Human Gene Mutation Database.* Hum Genet **1996**; 98(5): 629.
10. **Sherry ST, Ward MH, Kholodov M, et al.** *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res **2001**; 29(1): 308-311.
11. **Amberger JS, Bocchini CA, Schiettecatte F, et al.** *OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders.* Nucleic Acids Res **2015**; 43(Database issue): D789-798.
12. **Desmet FO, Hamroun D, Lalande M, et al.** *Human Splicing Finder: an online bioinformatics tool to predict splicing signals.* Nucleic Acids Res **2009**; 37(9): e67.
13. **Kircher M, Witten DM, Jain P, et al.** *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet **2014**; 46(3): 310-315.
14. **Adzhubei IA, Schmidt S, Peshkin L, et al.** *A method and server for predicting damaging missense mutations.* Nat Methods **2010**; 7(4): 248-249.

15. **Schwarz JM, Cooper DN, Schuelke M, et al.** *MutationTaster2: mutation prediction for the deep-sequencing age.* Nat Methods **2014**; 11(4): 361-362.
16. **Ng PC, Henikoff S.** *Predicting deleterious amino acid substitutions.* Genome Res **2001**; 11(5): 863-874.
17. **International Human Genome Sequencing Consortium.** *Finishing the euchromatic sequence of the human genome.* Nature **2004**; 431(7011): 931-945.
18. **Hood L, Rowen L.** *The Human Genome Project: big science transforms biology and medicine.* Genome Med **2013**; 5(9): 79.
19. **Ball MP, Bobe JR, Chou MF, et al.** *Harvard Personal Genome Project: lessons from participatory public research.* Genome Med **2014**; 6(2): 10.
20. **Nelson B.** *A fast track to personalized medicine?: whole-genome sequencing may lead to better cancer therapies.* Cancer Cytopathol **2012**; 120(5): 283-284.
21. **Esplin ED, Oei L, Snyder MP.** *Personalized sequencing and the future of medicine: discovery, diagnosis and defeat of disease.* Pharmacogenomics **2014**; 15(14): 1771-1790.
22. **Collins FS, Green ED, Guttmacher AE, et al.** *A vision for the future of genomics research.* Nature **2003**; 422(6934): 835-847.
23. **Bao R, Huang L, Andrade J, et al.** *Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing.* Cancer Inform **2014**; 13(Suppl 2): 67-82.
24. **van El CG, Cornel MC, Borry P, et al.** *Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics.* Eur J Hum Genet **2013**; 21(6): 580-584.
25. **Green RC, Lautenbach D, McGuire AL.** *GINA, genetic discrimination, and genomic medicine.* N Engl J Med **2015**; 372(5): 397-399.
26. **Solomon BD, Hadley DW, Pineda-Alvarez DE, et al.** *Incidental medical information in whole-exome sequencing.* Pediatrics **2012**; 129(6): e1605-1611.
27. **Rigter T, Henneman L, Kristoffersson U, et al.** *Reflecting on earlier experiences with unsolicited findings: points to consider for next-generation sequencing and informed consent in diagnostics.* Hum Mutat **2013**; 34(10): 1322-1328.
28. **Tewhey R, Warner JB, Nakano M, et al.** *Microdroplet-based PCR enrichment for large-scale targeted sequencing.* Nat Biotechnol **2009**; 27(11): 1025-1031.
29. **Mamanova L, Coffey AJ, Scott CE, et al.** *Target-enrichment strategies for next-generation sequencing.* Nat Methods **2010**; 7(2): 111-118.
30. **Hollants S, Redeker EJ, Matthijs G.** *Microfluidic amplification as a tool for massive parallel sequencing of the familial hypercholesterolemia genes.* Clin Chem **2012**; 58(4): 717-724.
31. **Bodi K, Perera AG, Adams PS, et al.** *Comparison of commercially available target enrichment methods for next-generation sequencing.* J Biomol Tech **2013**; 24(2): 73-86.

32. **Margulies M, Egholm M, Altman WE, et al.** *Genome sequencing in microfabricated high-density picolitre reactors.* Nature **2005**; 437(7057): 376-380.
33. **Ronaghi M.** *Pyrosequencing sheds light on DNA sequencing.* Genome Res **2001**; 11(1): 3-11.
34. **Metzker ML.** *Sequencing technologies - the next generation.* Nat Rev Genet **2010**; 11(1): 31-46.
35. **Rothberg JM, Hinz W, Rearick TM, et al.** *An integrated semiconductor device enabling non-optical genome sequencing.* Nature **2011**; 475(7356): 348-352.
36. **Valouev A, Ichikawa J, Tonthat T, et al.** *A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning.* Genome Res **2008**; 18(7): 1051-1063.
37. **Mardis ER.** *Next-generation DNA sequencing methods.* Annu Rev Genomics Hum Genet **2008**; 9: 387-402.
38. **Liu L, Li Y, Li S, et al.** *Comparison of next-generation sequencing systems.* J Biomed Biotechnol **2012**; 2012: 251364.
39. **Schadt EE, Turner S, Kasarskis A.** *A window into third-generation sequencing.* Hum Mol Genet **2010**; 19(R2): R227-240.
40. **Eid J, Fehr A, Gray J, et al.** *Real-time DNA sequencing from single polymerase molecules.* Science **2009**; 323(5910): 133-138.
41. **Tekin M, Arnos KS, Pandya A.** *Advances in hereditary deafness.* Lancet **2001**; 358(9287): 1082-1090.
42. **Hilgert N, Smith RJ, Van Camp G.** *Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics?* Mutat Res **2009**; 681(2-3): 189-196.
43. **Morton CC, Nance WE.** *Newborn hearing screening-a silent revolution.* N Engl J Med **2006**; 354(20): 2151-2164.
44. **Petersen MB, Willems PJ.** *Non-syndromic, autosomal-recessive deafness.* Clin Genet **2006**; 69(5): 371-392.
45. **Bayazit YA, Yilmaz M.** *An overview of hereditary hearing loss.* ORL J Otorhinolaryngol Relat Spec **2006**; 68(2): 57-63.
46. **Parker M, Bitner-Glindzicz M.** *Genetic investigations in childhood deafness.* Arch Dis Child **2015**; 100(3): 271-278.
47. **Mahdieh N, Rabbani B.** *Statistical study of 35delG mutation of GJB2 gene: a meta-analysis of carrier frequency.* Int J Audiol **2009**; 48(6): 363-370.
48. **Gasparini P, Rabionet R, Barbujani G, et al.** *High carrier frequency of the 35delG deafness mutation in European populations. Genetic Analysis Consortium of GJB2 35delG.* Eur J Hum Genet **2000**; 8(1): 19-23.

49. **Morell RJ, Kim HJ, Hood LJ, et al.** *Mutations in the connexin 26 gene (GJB2) among Ashkenazi Jews with nonsyndromic recessive deafness.* N Engl J Med **1998**; 339(21): 1500-1505.
50. **Abe S, Usami S, Shinkawa H, et al.** *Prevalent connexin 26 gene (GJB2) mutations in Japanese.* J Med Genet **2000**; 37(1): 41-43.
51. **Harrison M, Roush J, Wallace J.** *Trends in age of identification and intervention in infants with hearing loss.* Ear Hear **2003**; 24(1): 89-95.
52. **Pimperton H, Kennedy CR.** *The impact of early identification of permanent childhood hearing impairment on speech and language outcomes.* Arch Dis Child **2012**; 97(7): 648-653.
53. **Vona B, Muller T, Nanda I, et al.** *Targeted next-generation sequencing of deafness genes in hearing-impaired individuals uncovers informative mutations.* Genet Med **2014**; 16(12): 945-953.
54. **Zhou J, Mochizuki T, Smeets H, et al.** *Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in inherited smooth muscle tumors.* Science **1993**; 261(5125): 1167-1169.
55. **Barker DF, Hostikka SL, Zhou J, et al.** *Identification of mutations in the COL4A5 collagen gene in Alport syndrome.* Science **1990**; 248(4960): 1224-1227.
56. **Oohashi T, Sugimoto M, Mattei MG, et al.** *Identification of a new collagen IV chain, alpha 6(IV), by cDNA isolation and assignment of the gene to chromosome Xq22, which is the same locus for COL4A5.* J Biol Chem **1994**; 269(10): 7520-7526.
57. **Sugimoto M, Oohashi T, Ninomiya Y.** *The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5(IV) and alpha 6(IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters.* Proc Natl Acad Sci U S A **1994**; 91(24): 11679-11683.
58. **Wappler F.** *Malignant hyperthermia.* Eur J Anaesthesiol **2001**; 18(10): 632-652.
59. **Rosenberg H, Davis M, James D, et al.** *Malignant hyperthermia.* Orphanet J Rare Dis **2007**; 2: 21.
60. **Bandschapp O, Girard T.** *Malignant hyperthermia.* Swiss Med Wkly **2012**; 142: w13652.
61. **Fiege M, Wappler F.** *Malignant hyperthermia - update 2002.* Anasthesiol Intensivmed Notfallmed Schmerzther **2003**; 38(7): 478-482.
62. **Glahn KP, Ellis FR, Halsall PJ, et al.** *Recognizing and managing a malignant hyperthermia crisis: guidelines from the European Malignant Hyperthermia Group.* Br J Anaesth **2010**; 105(4): 417-420.
63. **Britt BA.** *Combined anesthetic- and stress-induced malignant hyperthermia in two offspring of malignant hyperthermic-susceptible parents.* Anesth Analg **1988**; 67(4): 393-399.
64. **Tobin JR, Jason DR, Challa VR, et al.** *Malignant hyperthermia and apparent heat stroke.* JAMA **2001**; 286(2): 168-169.

65. **Nishio H, Sato T, Fukunishi S, et al.** *Identification of malignant hyperthermia-susceptible ryanodine receptor type 1 gene (RYR1) mutations in a child who died in a car after exposure to a high environmental temperature.* Leg Med (Tokyo) **2009**; 11(3): 142-143.
66. **Quinlivan RM, Muller CR, Davis M, et al.** *Central core disease: clinical, pathological, and genetic features.* Arch Dis Child **2003**; 88(12): 1051-1055.
67. **Mathews KD, Moore SA.** *Multiminicore myopathy, central core disease, malignant hyperthermia susceptibility, and RYR1 mutations: one disease with many faces?* Arch Neurol **2004**; 61(1): 27-29.
68. **Ording H, Brancadoro V, Cozzolino S, et al.** *In vitro contracture test for diagnosis of malignant hyperthermia following the protocol of the European MH Group: results of testing patients surviving fulminant MH and unrelated low-risk subjects. The European Malignant Hyperthermia Group.* Acta Anaesthesiol Scand **1997**; 41(8): 955-966.
69. **The European Malignant Hyperpyrexia Group.** *A protocol for the investigation of malignant hyperpyrexia (MH) susceptibility.* Br J Anaesth **1984**; 56(11): 1267-1269.
70. **Urwyler A, Deufel T, McCarthy T, et al.** *Guidelines for molecular genetic detection of susceptibility to malignant hyperthermia.* Br J Anaesth **2001**; 86(2): 283-287.
71. **Brandom BW.** *Genetics of malignant hyperthermia.* ScientificWorldJournal **2006**; 6: 1722-1730.
72. **Protasi F.** *Structural interaction between RYRs and DHPRs in calcium release units of cardiac and skeletal muscle cells.* Front Biosci **2002**; 7: d650-658.
73. **MacLennan DH, Duff C, Zorzato F, et al.** *Ryanodine receptor gene is a candidate for predisposition to malignant hyperthermia.* Nature **1990**; 343(6258): 559-561.
74. **Phillips MS, Fujii J, Khanna VK, et al.** *The structural organization of the human skeletal muscle ryanodine receptor (RYR1) gene.* Genomics **1996**; 34(1): 24-41.
75. **Monnier N, Procaccio V, Stieglitz P, et al.** *Malignant-hyperthermia susceptibility is associated with a mutation of the alpha 1-subunit of the human dihydropyridine-sensitive L-type voltage-dependent calcium-channel receptor in skeletal muscle.* Am J Hum Genet **1997**; 60(6): 1316-1325.
76. **Moslehi R, Langlois S, Yam I, et al.** *Linkage of malignant hyperthermia and hyperkalemic periodic paralysis to the adult skeletal muscle sodium channel (SCN4A) gene in a large pedigree.* Am J Med Genet **1998**; 76(1): 21-27.
77. **Mathews KD, Moore SA.** *Limb-girdle muscular dystrophy.* Curr Neurol Neurosci Rep **2003**; 3(1): 78-85.
78. **Nigro V, Aurino S, Piluso G.** *Limb girdle muscular dystrophies: update on genetic diagnosis and therapeutic approaches.* Curr Opin Neurol **2011**; 24(5): 429-436.
79. **Nigro V, Savarese M.** *Genetic basis of limb-girdle muscular dystrophies: the 2014 update.* Acta Myol **2014**; 33(1): 1-12.

80. **Bushby K.** *Diagnosis and management of the limb girdle muscular dystrophies.* Pract Neurol **2009**; 9(6): 314-323.
81. **Hicks D, Sarkozy A, Muelas N, et al.** *A founder mutation in Anoctamin 5 is a major cause of limb-girdle muscular dystrophy.* Brain **2011**; 134(Pt 1): 171-182.
82. **Witting N, Duno M, Petri H, et al.** *Anoctamin 5 muscular dystrophy in Denmark: prevalence, genotypes, phenotypes, cardiac findings, and muscle protein expression.* J Neurol **2013**; 260(8): 2084-2093.
83. **Lo HP, Cooper ST, Evesson FJ, et al.** *Limb-girdle muscular dystrophy: diagnostic evaluation, frequency and clues to pathogenesis.* Neuromuscul Disord **2008**; 18(1): 34-44.
84. **Richard I, Broux O, Allamand V, et al.** *Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A.* Cell **1995**; 81(1): 27-40.
85. **Liu J, Aoki M, Illa I, et al.** *Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy.* Nat Genet **1998**; 20(1): 31-36.
86. **Stehlikova K, Skalova D, Zidkova J, et al.** *Autosomal recessive limb-girdle muscular dystrophies in the Czech Republic.* BMC Neurol **2014**; 14: 154.
87. **Canki-Klain N, Milic A, Kovac B, et al.** *Prevalence of the 550delA mutation in calpainopathy (LGMD 2A) in Croatia.* Am J Med Genet A **2004**; 125A(2): 152-156.
88. **Hanisch F, Muller CR, Grimm D, et al.** *Frequency of calpain-3 c.550delA mutation in limb girdle muscular dystrophy type 2 and isolated hyperCKemia in German patients.* Clin Neuropathol **2007**; 26(4): 157-163.
89. **Bashir R, Britton S, Strachan T, et al.** *A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B.* Nat Genet **1998**; 20(1): 37-42.
90. **Bolduc V, Marlow G, Boycott KM, et al.** *Recessive mutations in the putative calcium-activated chloride channel Anoctamin 5 cause proximal LGMD2L and distal MMD3 muscular dystrophies.* Am J Hum Genet **2010**; 86(2): 213-221.
91. **Olive M, Odgerel Z, Martinez A, et al.** *Clinical and myopathological evaluation of early- and late-onset subtypes of myofibrillar myopathy.* Neuromuscul Disord **2011**; 21(8): 533-542.
92. **Forrest KM, Al-Sarraj S, Sewry C, et al.** *Infantile onset myofibrillar myopathy due to recessive CRYAB mutations.* Neuromuscul Disord **2011**; 21(1): 37-40.
93. **Olive M, Kley RA, Goldfarb LG.** *Myofibrillar myopathies: new developments.* Curr Opin Neurol **2013**; 26(5): 527-535.
94. **Selcen D.** *Myofibrillar myopathies.* Neuromuscul Disord **2011**; 21(3): 161-171.
95. **De Bleecker JL, Engel AG, Ertl BB.** *Myofibrillar myopathy with abnormal foci of desmin positivity. II. Immunocytochemical analysis reveals accumulation of multiple other proteins.* J Neuropathol Exp Neurol **1996**; 55(5): 563-577.

96. **Nakano S, Engel AG, Waclawik AJ, et al.** *Myofibrillar myopathy with abnormal foci of desmin positivity. I. Light and electron microscopy analysis of 10 cases.* J Neuropathol Exp Neurol **1996**; 55(5): 549-562.
97. **Claeys KG, Fardeau M, Schroder R, et al.** *Electron microscopy in myofibrillar myopathies reveals clues to the mutated gene.* Neuromuscul Disord **2008**; 18(8): 656-666.
98. **Schroder R, Schoser B.** *Myofibrillar myopathies: a clinical and myopathological guide.* Brain Pathol **2009**; 19(3): 483-492.
99. **Goldfarb LG, Park KY, Cervenakova L, et al.** *Missense mutations in desmin associated with familial cardiac and skeletal myopathy.* Nat Genet **1998**; 19(4): 402-403.
100. **Viegas-Pequignot E, Li ZL, Dutrillaux B, et al.** *Assignment of human desmin gene to band 2q35 by nonradioactive in situ hybridization.* Hum Genet **1989**; 83(1): 33-36.
101. **Vicart P, Caron A, Guicheney P, et al.** *A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy.* Nat Genet **1998**; 20(1): 92-95.
102. **Selcen D, Engel AG.** *Mutations in myotilin cause myofibrillar myopathy.* Neurology **2004**; 62(8): 1363-1371.
103. **Selcen D, Engel AG.** *Mutations in ZASP define a novel form of muscular dystrophy in humans.* Ann Neurol **2005**; 57(2): 269-276.
104. **Vorgerd M, van der Ven PF, Bruchertseifer V, et al.** *A mutation in the dimerization domain of filamin c causes a novel type of autosomal dominant myofibrillar myopathy.* Am J Hum Genet **2005**; 77(2): 297-304.
105. **Selcen D, Muntoni F, Burton BK, et al.** *Mutation in BAG3 causes severe dominant childhood muscular dystrophy.* Ann Neurol **2009**; 65(1): 83-89.
106. **Selcen D, Bromberg MB, Chin SS, et al.** *Reducing bodies and myofibrillar myopathy features in FHL1 muscular dystrophy.* Neurology **2011**; 77(22): 1951-1959.
107. **Ohlsson M, Hedberg C, Bradvik B, et al.** *Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin.* Brain **2012**; 135(Pt 6): 1682-1694.
108. **Stein KC, Bengoechea R, Harms MB, et al.** *Myopathy-causing mutations in an HSP40 chaperone disrupt processing of specific client conformers.* J Biol Chem **2014**; 289(30): 21120-21130.
109. **Shalaby S, Mitsuhashi H, Matsuda C, et al.** *Defective myotilin homodimerization caused by a novel mutation in MYOT exon 9 in the first Japanese limb girdle muscular dystrophy 1A patient.* J Neuropathol Exp Neurol **2009**; 68(6): 701-707.
110. **van der Flier A, Sonnenberg A.** *Structural and functional aspects of filamins.* Biochim Biophys Acta **2001**; 1538(2-3): 99-117.
111. **Furst DO, Goldfarb LG, Kley RA, et al.** *Filamin C-related myopathies: pathology and mechanisms.* Acta Neuropathol **2013**; 125(1): 33-46.

112. **Homma S, Iwasaki M, Shelton GD, et al.** *BAG3 deficiency results in fulminant myopathy and early lethality.* Am J Pathol **2006**; 169(3): 761-773.
113. **Cowling BS, Cottle DL, Wilding BR, et al.** *Four and a half LIM protein 1 gene mutations cause four distinct human myopathies: a comprehensive review of the clinical, histological and pathological features.* Neuromuscul Disord **2011**; 21(4): 237-251.
114. **Bang ML, Centner T, Fornoff F, et al.** *The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system.* Circ Res **2001**; 89(11): 1065-1072.
115. **Gundesli H, Talim B, Korkusuz P, et al.** *Mutation in exon 1f of PLEC, leading to disruption of plectin isoform 1f, causes autosomal-recessive limb-girdle muscular dystrophy.* Am J Hum Genet **2010**; 87(6): 834-841.
116. **Selcen D, Juel VC, Hobson-Webb LD, et al.** *Myasthenic syndrome caused by plectinopathy.* Neurology **2011**; 76(4): 327-336.
117. **Peyvandi F, Jayandharan G, Chandy M, et al.** *Genetic diagnosis of haemophilia and other inherited bleeding disorders.* Haemophilia **2006**; 12 Suppl 3: 82-89.
118. **Bolton-Maggs PH, Pasi KJ.** *Haemophilias A and B.* Lancet **2003**; 361(9371): 1801-1809.
119. **Belvini D, Salviato R, Radossi P, et al.** *Molecular genotyping of the Italian cohort of patients with hemophilia B.* Haematologica **2005**; 90(5): 635-642.
120. **Mannucci PM, Tuddenham EG.** *The hemophilias - from royal genes to gene therapy.* N Engl J Med **2001**; 344(23): 1773-1779.
121. **Oldenburg J, Schroder J, Brackmann HH, et al.** *Environmental and genetic factors influencing inhibitor development.* Semin Hematol **2004**; 41(1 Suppl 1): 82-88.
122. **Oldenburg J, El-Maarri O.** *New insight into the molecular basis of hemophilia A.* Int J Hematol **2006**; 83(2): 96-102.
123. **Bogdanova N, Markoff A, Eisert R, et al.** *Spectrum of molecular defects and mutation detection rate in patients with mild and moderate hemophilia A.* Hum Mutat **2007**; 28(1): 54-60.
124. **Clemetson KJ.** *Platelets and primary haemostasis.* Thromb Res **2012**; 129(3): 220-224.
125. **Palta S, Saroa R, Palta A.** *Overview of the coagulation system.* Indian J Anaesth **2014**; 58(5): 515-523.
126. **Bowen DJ.** *Haemophilia A and haemophilia B: molecular insights.* Mol Pathol **2002**; 55(2): 127-144.
127. **Hollestelle MJ, Thinnes T, Crain K, et al.** *Tissue distribution of factor VIII gene expression in vivo - a closer look.* Thromb Haemost **2001**; 86(3): 855-861.
128. **Lenting PJ, van Mourik JA, Mertens K.** *The life cycle of coagulation factor VIII in view of its structure and function.* Blood **1998**; 92(11): 3983-3996.

129. **Esmon CT.** *Regulation of blood coagulation.* *Biochim Biophys Acta* **2000**; 1477(1-2): 349-360.
130. **Gitschier J, Wood WI, Goralka TM, et al.** *Characterization of the human factor VIII gene.* *Nature* **1984**; 312(5992): 326-330.
131. **Lakich D, Kazazian HH, Jr., Antonarakis SE, et al.** *Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A.* *Nat Genet* **1993**; 5(3): 236-241.
132. **Naylor JA, Green PM, Rizza CR, et al.** *Analysis of factor VIII mRNA reveals defects in everyone of 28 haemophilia A patients.* *Hum Mol Genet* **1993**; 2(1): 11-17.
133. **Bagnall RD, Waseem N, Green PM, et al.** *Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A.* *Blood* **2002**; 99(1): 168-174.
134. **Pezeshkpoor B, Pavlova A, Oldenburg J, et al.** *F8 genetic analysis strategies when standard approaches fail.* *Hamostaseologie* **2014**; 34(2): 167-173.
135. **Rost S, Bach E, Neuner C, et al.** *Novel form of X-linked nonsyndromic hearing loss with cochlear malformation caused by a mutation in the type IV collagen gene COL4A6.* *Eur J Hum Genet* **2014**; 22(2): 208-215.
136. **Broman M, Kleinschnitz I, Bach JE, et al.** *Next-generation DNA sequencing of a Swedish malignant hyperthermia cohort.* *Clin Genet* **2015**; 88(4): 381-385.
137. **Semmler AL, Sacconi S, Bach JE, et al.** *Unusual multisystemic involvement and a novel BAG3 mutation revealed by NGS screening in a large cohort of myofibrillar myopathies.* *Orphanet J Rare Dis* **2014**; 9(1): 121.
138. **Schessl J, Bach E, Rost S, et al.** *Novel recessive myotilin mutation causes severe myofibrillar myopathy.* *Neurogenetics* **2014**; 15(3): 151-156.
139. **Bach JE, Oldenburg J, Muller CR, et al.** *Mutational spectrum and identification of deep intronic variants by next generation sequencing in the factor VIII gene of haemophilia A patients.* *Hamostaseologie* **2015**; Accepted for publication(doi.org/10.5482/HAMO-15-05-0017).
140. **Bach JE, Wolf B, Oldenburg J, et al.** *Identification of deep intronic variants in 15 haemophilia A patients by next generation sequencing of the whole factor VIII gene.* *Thromb Haemost* **2015**; 114(4): 757-767.
141. **Bach JE, Muller CR, Rost S.** *Mini-gene assays confirm the splicing effect of deep intronic variants in the factor VIII gene.* *Thromb Haemost* **2016**; 115(1): 222-224.
142. **Untergasser A, Cutcutache I, Koressaar T, et al.** *Primer3 - new capabilities and interfaces.* *Nucleic Acids Res* **2012**; 40(15): e115.
143. **Varadaraj K, Skinner DM.** *Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases.* *Gene* **1994**; 140(1): 1-5.
144. **Ricard-Blum S.** *The collagen family.* *Cold Spring Harb Perspect Biol* **2011**; 3(1): a004978.

145. **Yeo G, Burge CB.** *Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals.* J Comput Biol **2004**; 11(2-3): 377-394.
146. **Reese MG, Eckman FH, Kulp D, et al.** *Improved splice site detection in Genie.* J Comput Biol **1997**; 4(3): 311-323.
147. **Pertea M, Lin X, Salzberg SL.** *GeneSplicer: a new computational method for splice site prediction.* Nucleic Acids Res **2001**; 29(5): 1185-1190.
148. **Peissel B, Geng L, Kalluri R, et al.** *Comparative distribution of the alpha 1(IV), alpha 5(IV), and alpha 6(IV) collagen chains in normal human adult and fetal tissues and in kidneys from X-linked Alport syndrome patients.* J Clin Invest **1995**; 96(4): 1948-1957.
149. **Miner JH.** *Alport syndrome with diffuse leiomyomatosis. When and when not?* Am J Pathol **1999**; 154(6): 1633-1635.
150. **Williams R, Peisajovich SG, Miller OJ, et al.** *Amplification of complex gene libraries by emulsion PCR.* Nat Methods **2006**; 3(7): 545-550.
151. **Amburgey K, Bailey A, Hwang JH, et al.** *Genotype-phenotype correlations in recessive RYR1-related myopathies.* Orphanet J Rare Dis **2013**; 8: 117.
152. **Carpenter D, Robinson RL, Quinell RJ, et al.** *Genetic variation in RYR1 and malignant hyperthermia phenotypes.* Br J Anaesth **2009**; 103(4): 538-548.
153. **McWilliams S, Nelson T, Sudo RT, et al.** *Novel skeletal muscle ryanodine receptor mutation in a large Brazilian family with malignant hyperthermia.* Clin Genet **2002**; 62(1): 80-83.
154. **Bell CJ, Dinwiddie DL, Miller NA, et al.** *Carrier testing for severe childhood recessive diseases by next-generation sequencing.* Sci Transl Med **2011**; 3(65): 65ra64.
155. **Li LY, Li Q, Yu YH, et al.** *A primer design strategy for PCR amplification of GC-rich DNA sequences.* Clin Biochem **2011**; 44(8-9): 692-698.
156. **Altschul SF, Gish W, Miller W, et al.** *Basic local alignment search tool.* J Mol Biol **1990**; 215(3): 403-410.
157. **Bar H, Fischer D, Goudeau B, et al.** *Pathogenic effects of a novel heterozygous R350P desmin mutation on the assembly of desmin intermediate filaments in vivo and in vitro.* Hum Mol Genet **2005**; 14(10): 1251-1260.
158. **Sharma S, Mucke N, Katus HA, et al.** *Disease mutations in the "head" domain of the extra-sarcomeric protein desmin distinctly alter its assembly and network-forming properties.* J Mol Med (Berl) **2009**; 87(12): 1207-1219.
159. **Vrabie A, Goldfarb LG, Shatunov A, et al.** *The enlarging spectrum of desminopathies: new morphological findings, eastward geographic spread, novel exon 3 desmin mutation.* Acta Neuropathol **2005**; 109(4): 411-417.
160. **Shatunov A, Olive M, Odgerel Z, et al.** *In-frame deletion in the seventh immunoglobulin-like repeat of filamin C in a family with myofibrillar myopathy.* Eur J Hum Genet **2009**; 17(5): 656-663.

161. **Hauser MA, Horrigan SK, Salmikangas P, et al.** *Myotilin is mutated in limb girdle muscular dystrophy IA.* Hum Mol Genet **2000**; 9(14): 2141-2147.
162. **Foroud T, Pankratz N, Batchman AP, et al.** *A mutation in myotilin causes spheroid body myopathy.* Neurology **2005**; 65(12): 1936-1940.
163. **Reilich P, Krause S, Schramm N, et al.** *A novel mutation in the myotilin gene (MYOT) causes a severe form of limb girdle muscular dystrophy IA (LGMD1A).* J Neurol **2011**; 258(8): 1437-1444.
164. **Tabor HK, Auer PL, Jamal SM, et al.** *Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: implications for the return of incidental results.* Am J Hum Genet **2014**; 95(2): 183-193.
165. **Green PM, Bagnall RD, Waseem NH, et al.** *Haemophilia A mutations in the UK: results of screening one-third of the population.* Br J Haematol **2008**; 143(1): 115-128.
166. **Becker J, Schwaab R, Moller-Taube A, et al.** *Characterization of the factor VIII defect in 147 patients with sporadic hemophilia A: family studies indicate a mutation type-dependent sex ratio of mutation frequencies.* Am J Hum Genet **1996**; 58(4): 657-670.
167. **Cooper DN.** *Eukaryotic DNA methylation.* Hum Genet **1983**; 64(4): 315-333.
168. **Pattinson JK, Millar DS, McVey JH, et al.** *The molecular genetic analysis of hemophilia A: a directed search strategy for the detection of point mutations in the human factor VIII gene.* Blood **1990**; 76(11): 2242-2248.
169. **Margaglione M, Castaman G, Morfini M, et al.** *The Italian AICE-Genetics hemophilia A database: results and correlation with clinical phenotype.* Haematologica **2008**; 93(5): 722-728.
170. **Young M, Inaba H, Hoyer LW, et al.** *Partial correction of a severe molecular defect in hemophilia A, because of errors during expression of the factor VIII gene.* Am J Hum Genet **1997**; 60(3): 565-573.
171. **den Dunnen JT, Antonarakis SE.** *Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion.* Hum Mutat **2000**; 15(1): 7-12.
172. **Tuley EA, Gaucher C, Jorieux S, et al.** *Expression of von Willebrand factor "Normandy": an autosomal mutation that mimics hemophilia A.* Proc Natl Acad Sci U S A **1991**; 88(14): 6377-6381.
173. **Nichols WC, Seligsohn U, Zivelin A, et al.** *Mutations in the ER-Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII.* Cell **1998**; 93(1): 61-70.
174. **Zhang B, Cunningham MA, Nichols WC, et al.** *Bleeding due to disruption of a cargo-specific ER-to-Golgi transport complex.* Nat Genet **2003**; 34(2): 220-225.
175. **Zimmermann MA, Hansmann T, Haaf T, et al.** *Methylation analysis of the promoter region and intron 1 of the factor VIII gene in haemophilia A patients.* Hamostaseologie **2013**; 33 Suppl 1: S46-49.

176. **Youssoufian H, Kazazian HH, Jr., Patel A, et al.** *Mild hemophilia A associated with a cryptic donor splice site mutation in intron 4 of the factor VIII gene.* *Genomics* **1988**; 2(1): 32-36.
177. **Bagnall RD, Waseem NH, Green PM, et al.** *Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients.* *Br J Haematol* **1999**; 107(4): 766-771.
178. **Santacroce R, Santoro R, Sessa F, et al.** *Screening of mutations of hemophilia A in 40 Italian patients: a novel G-to-A mutation in intron 10 of the F8 gene as a putative cause of mild hemophilia A in southern Italy.* *Blood Coagul Fibrinolysis* **2008**; 19(3): 197-202.
179. **Castaman G, Giacomelli SH, Mancuso ME, et al.** *Deep intronic variations may cause mild hemophilia A.* *J Thromb Haemost* **2011**; 9(8): 1541-1548.
180. **Inaba H, Koyama T, Shinozawa K, et al.** *Identification and characterization of an adenine to guanine transition within intron 10 of the factor VIII gene as a causative mutation in a patient with mild haemophilia A.* *Haemophilia* **2013**; 19(1): 100-105.
181. **Deininger PL, Batzer MA.** *Alu repeats and human disease.* *Mol Genet Metab* **1999**; 67(3): 183-193.
182. **Cordaux R, Batzer MA.** *The impact of retrotransposons on human genome evolution.* *Nat Rev Genet* **2009**; 10(10): 691-703.
183. **Zimmermann MA, Oldenburg J, Muller CR, et al.** *Characterization of duplication breakpoints in the factor VIII gene.* *J Thromb Haemost* **2010**; 8(12): 2696-2704.
184. **Genomes Project C, Abecasis GR, Auton A, et al.** *An integrated map of genetic variation from 1,092 human genomes.* *Nature* **2012**; 491(7422): 56-65.
185. **McGlinicy NJ, Smith CW.** *Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?* *Trends Biochem Sci* **2008**; 33(8): 385-393.
186. **Zimmermann MA, Gehrig A, Oldenburg J, et al.** *Analysis of F8 mRNA in haemophilia A patients with silent mutations or presumptive splice site mutations.* *Haemophilia* **2013**; 19(2): 310-317.
187. **Tizzano EF, Cornet M, Domenech M, et al.** *Modifier genes in haemophilia: their expansion in the human genome.* *Haemophilia* **2002**; 8(3): 250-254.
188. **Laffan MA, Lester W, O'Donnell JS, et al.** *The diagnosis and management of von Willebrand disease: a United Kingdom Haemophilia Centre Doctors Organization guideline approved by the British Committee for Standards in Haematology.* *Br J Haematol* **2014**; 167(4): 453-465.
189. **Black DL.** *Mechanisms of alternative pre-messenger RNA splicing.* *Annu Rev Biochem* **2003**; 72: 291-336.

7 ABKÜRZUNGSVERZEICHNIS

A	Adenin
AD	Autosomal-dominant
<i>ANO5</i>	Anoctamin 5; MIM *608662
AR	Autosomal-rezessiv
Ass	Acceptor splice site
<i>ATP2A2</i>	ATPase, Ca(2+)-Transportin, slow-twitch 2; MIM *108740
<i>BAG3</i>	BCL2-associated athanogene 3; MIM *603883
Bp	Basenpaare
C	Cytosin
<i>CACNA1S</i>	Calcium channel, voltage-dependent, L type, alpha 1S subunit; MIM *114208
<i>CAPN3</i>	Calpain 3; MIM *114240
CCD	Central Core Disease; MIM #117000
CK	Creatin-Kinase
CNV	Kopienzahl-Variation (copy number variation)
<i>COL4A5</i>	Kollagen, Typ IV, alpha 5; MIM *303630
<i>COL4A6</i>	Kollagen, Typ IV, alpha 6; MIM *303631
<i>COL6A1</i>	Kollagen, Typ VI, alpha 1; MIM *120220
<i>CRYAB</i>	Alpha-B Crystallin; MIM *123590
<i>DES</i>	Desmin; MIM *125660
<i>DMD</i>	Dystrophin; MIM *300377
DMD / BMB	Muskeldystrophie Typ Duchenne / Becker; MIM #310200 / 300376
DNA	Desoxyribonukleinsäure, cDNA: complementary DNA
<i>DNAJB6</i>	DnaJ (Hsp40) Homolog, subfamily B, member 6; MIM *611332
Dss	Donor splice site
<i>DYSF</i>	Dysferlin; MIM *603009
Ex	Exon
F	Faktor (z.B. FVIII, FIX)
<i>F8</i>	Coagulation Factor VIII; MIM *300841
<i>FHL1</i>	Four and a half LIM domains 1; MIM *300163
<i>FKRP</i>	Fukutin-related protein; MIM *606596

<i>FKTN</i>	Fukutin; MIM *607440
<i>FLNC</i>	Filamin C; MIM *102565
FSHD	Fazioscapulohumerale Dystrophie; MIM #158900
FVIII:C	FactorVIII clotting activity
G	Guanin
Gb	Gigabasen
GBM	Basalmembran der Nieren-Glomeruli (glomerular basal membrane)
<i>GJB2</i>	Gap junction protein, beta 2; MIM *121011
HA	Hämophilie A; MIM #306700
HB	Hämophilie B; MIM #306900
HEK	Embryonale Nierenzellen (human embryonic kidney)
Int	Intron
IVCT	<i>in vitro</i> -Kontrakturtest
Kb	Kilobasen
<i>KBTBD13</i>	Kelch repeat and BTB (POZ) domain containing 13; MIM *613727
<i>LAMA2</i>	Laminin, alpha 2; MIM *156225
LGMD	Gliedergürteldystrophie (limb-girdle muscular dystrophy)
LINE	Long interspersed nuclear element
<i>LMAN1</i>	Lectin, mannose-binding, 1; MIM *601567
LTR	Long terminal repeat
MAF	Minor Allele Frequency
Mb	Megabasen
<i>MCFD2</i>	Multiple coagulation factor deficiency 2; MIM *607788
MFM	Myofibrilläre Myopathie
MH	Maligne Hyperthermie; MHS: Malignant Hyperthermia Susceptibility
MLPA	Multiplex ligation-dependent probe amplification
Mut.	Mutation
<i>MYOT</i>	Myotilin; MIM *604103
NGS	Next Generation Sequencing
Pat.	Patient
PCR	Polymerase-Kettenreaktion (polymerase chain reaction)
<i>PLEC1</i>	Plectin 1; MIM *601282
<i>POU3F4</i>	POU domain, class 3, transcription factor 4; MIM *300039
<i>PRPS1</i>	Phosphoribosyl pyrophosphate synthetase 1; MIM *311850
RNA	Ribonukleinsäure; mRNA: messenger-RNA

<i>RYR1</i>	Ryanodin Rezeptor 1; MIM *180901
SNP	Einzelnukleotid-Polymorphismus (Single Nucleotide Polymorphism)
SNV	Einzelnukleotid-Variante (Single Nucleotide Variant)
SR	Sarkoplasmatisches Reticulum
T	Thymin
TF	Gewebefaktor (Tissue Factor)
<i>TTN</i>	Titin; MIM *188840
UTR	Untranslatierte Region
UV	Unklassifizierte Variante
<i>VWF</i>	Von Willebrand factor; MIM *613160
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WT	Wildtyp
XR	X-chromosomal-rezessiv
<i>ZASP</i>	LIM domain binding 3; MIM *605906 (Synonym <i>LDB3</i>)

8 VERWENDETE DATENBANKEN UND ONLINE-TOOLS

Datenbank / Online-Tool	Link
1 000 Genomes Browser	http://browser.1000genomes.org/Homo_sapiens/Info/Index
BLAST (Basic Local Alignment Search Tool)	http://blast.ncbi.nlm.nih.gov/Blast.cgi
CADD (Combined Annotation Dependent Depletion)	http://cadd.gs.washington.edu/score
dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP/
Ensembl	http://www.ensembl.org/index.html
ESEfinder	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home
FVIII Variant Database	http://www.factorviii-db.org/
GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml
Hereditary Hearing Loss Homepage	http://hereditaryhearingloss.org/
HGMD (Human Gene Mutation Database)	https://portal.biobase-international.com/cgi-bin/portal/login.cgi
Human Splicing Finder	http://www.umd.be/HSF/
Max-EntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
MutationTaster	http://www.mutationtaster.org/
NCBI	http://www.ncbi.nlm.nih.gov/
NNSPLICE	http://www.fruitfly.org/seq_tools/splice.html
OMIM (Online Mendelian Inheritance in Man)	http://omim.org/entry/
PolyPhen-2 (Polymorphism Phenotyping v2)	http://genetics.bwh.harvard.edu/pph2/
RepeatMasker	http://repeatmasker.org/cgi-bin/WEBRepeatMasker
RESCUE-ESE	http://genes.mit.edu/burgelab/rescue-ese/
SIFT (Sorting Intolerant from Tolerant)	http://sift.jcvi.org/
SNPCheck3	https://secure.ngrl.org.uk/SNPCheck/snpcheck.htm

9 LISTE DER VORGELEGTEN PUBLIKATIONEN

Anlage 1:

Simone Rost, **Elisa Bach**, Cordula Neuner, Indrajit Nanda, Sandra Dysek, Reginald E Bittner, Alexander Keller, Oliver Bartsch, Robert Mlynski, Thomas Haaf, Clemens R Müller and Erdmute Kunstmann

Novel form of X-linked nonsyndromic hearing loss with cochlear malformation caused by a mutation in the type IV collagen gene COL4A6

Eur J Hum Genet 2014; 22(2): 208-215. doi: 10.1038/ejhg.2013.108

Anlage 2:

M. Broman, I. Kleinschnitz, **J.E. Bach**, S.Rost, G. Islander and C.R. Müller

Next-generation DNA sequencing of a Swedish malignant hyperthermia cohort

Clin Genet. 2015 Oct; 88(4):381-5. doi: 10.1111/cge.12508

Anlage 3:

Anna-Lena Semmler, Sabrina Sacconi, **J Elisa Bach**, Claus Liebe, Jan Bürmann, Rudolf A Kley, Andreas Ferbert, Roland Anderheiden, Peter Van den Bergh, Jean-Jacques Martin, Peter De Jonghe, Eva Neuen-Jacob, Oliver Müller, Marcus Deschauer, Markus Bergmann, J Michael Schröder, Matthias Vorgerd, Jörg B Schulz, Joachim Weis, Wolfram Kress and Kristl G Claeys

Unusual multisystemic involvement and a novel BAG3 mutation revealed by NGS screening in a large cohort of myofibrillar myopathies

Orphanet J Rare Dis. 2014; 9(1):121. doi: 10.1186/s13023-014-0121-9

Anlage 4:

Joachim Schessl, **Elisa Bach**, Simone Rost, Sarah Feldkirchner, Christiana Kubny, Stefan Müller, Franz-Georg Hanisch, Wolfram Kress, Benedikt Schoser

Novel recessive myotilin mutation causes severe myofibrillar myopathy

Neurogenetics. 2014, 15(3):151-6. doi: 10.1007/s10048-014-0410-4

Anlage 5:

Bach JE, Oldenburg J, Müller CR, Rost S

Mutational spectrum and identification of deep intronic variants by next generation sequencing in the factor VIII gene of haemophilia A patients

Hämostaseologie 2015 [Accepted for publication]. doi.org/10.5482/HAMO-15-05-0017

Anlage 6:

Bach JE, Wolf B, Oldenburg J, Müller CR, Rost S

Identification of deep intronic variants in 15 hemophilia A patients by Next Generation Sequencing of the whole factor VIII gene

Thromb Haemost. 2015 Oct;114(4):757-67. doi: 10.1160/TH14-12-1011.

Anlage 7:

Bach JE, Müller CR, Rost S

Mini-gene assays confirm the splicing effect of deep intronic variants in the factor VIII gene

Thromb Haemost. 2015 Dec 22;115(1):222-4. doi: 10.1160/TH15-05-0399

10 SONDERDRUCKE DER VORGELEGTEN PUBLIKATIONEN

Anlage 1

Simone Rost, **Elisa Bach**, Cordula Neuner, Indrajit Nanda, Sandra Dysek, Reginald E Bittner, Alexander Keller, Oliver Bartsch, Robert Mlynski, Thomas Haaf, Clemens R Müller and Erdmute Kunstmann

Novel form of X-linked nonsyndromic hearing loss with cochlear malformation caused by a mutation in the type IV collagen gene COL4A6

Eur J Hum Genet 2014; 22(2): 208-215. doi: 10.1038/ejhg.2013.108

Novel form of X-linked nonsyndromic hearing loss with cochlear malformation caused by a mutation in the type IV collagen gene *COL4A6*

Running title: X-linked hearing loss caused by *COL4A6* mutation

Simone Rost^{1*}, Elisa Bach¹, Cordula Neuner¹, Indrajit Nanda¹, Sandra Dysek², Reginald E. Bittner², Alexander Keller³, Oliver Bartsch⁴, Robert Mlynski⁵, Thomas Haaf^{1*}, Clemens R. Müller¹, Erdmute Kunstmann¹

¹Department of Human Genetics, University of Würzburg, Würzburg, Germany

²Neuromuscular Research Department, Center for Anatomy and Cell Biology, Medical University of Vienna, Vienna, Austria

³DNA Analytics Core Facility and Department of Animal Ecology and Tropical Biology, University of Würzburg, Würzburg, Germany

⁴Institute of Human Genetics, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany

⁵Department of Oto-Rhino-Laryngology, Plastic, Aesthetic and Reconstructive Head and Neck Surgery, Comprehensive Hearing Center University of Würzburg, Würzburg, Germany

*Correspondence: Simone Rost or Thomas Haaf, Department of Human Genetics, University of Würzburg, Biocenter, Am Hubland, 97074 Würzburg, Tel: +49 931 31-84095 or -88738; Fax: +49 931 31-84069; Email: simone.rost@biozentrum.uni-wuerzburg.de or thomas.haaf@uni-wuerzburg.de

ABSTRACT

Hereditary hearing loss is the most common human sensorineural disorder. Genetic causes are highly heterogeneous with mutations detected in more than 40 genes associated with nonsyndromic hearing loss to date. While autosomal recessive and autosomal dominant inheritance is prevalent, X-linked forms of nonsyndromic hearing impairment are extremely rare. Here, we present a Hungarian three-generation family with X-linked nonsyndromic congenital hearing loss and the underlying genetic defect.

Next generation sequencing and subsequent segregation analysis detected a missense mutation (c.1771G>A, p.Gly591Ser) in the type IV collagen gene *COL4A6* in all affected family members. Bioinformatic analysis and expression studies support this substitution as being causative. *COL4A6* encodes the alpha-6 chain of type IV collagen of basal membranes which forms a heterotrimer with two alpha-5 chains encoded by *COL4A5*. While mutations in *COL4A5* and contiguous X-chromosomal deletions involving *COL4A5* and *COL4A6* are associated with X-linked Alport syndrome, a nephropathy associated with deafness and cataract, mutations in *COL4A6* alone have not been related to any hereditary disease so far. Moreover, our index patient and other affected family members show normal renal and ocular function, which is not consistent with Alport syndrome but with a nonsyndromic type of hearing loss. In situ hybridization and immunostaining demonstrated expression of the *COL4A6* homologues in the otic vesicle of the zebrafish and in the murine inner ear, supporting its role in normal ear development and function.

In conclusion, our results suggest *COL4A6* as being the fourth gene associated with X-linked nonsyndromic hearing loss.

Keywords: next generation sequencing; X-linked, nonsyndromic hearing loss; type IV collagen; *COL4A6*

INTRODUCTION

Childhood hearing loss is one of the most common and most heterogeneous diseases worldwide. In addition to various environmental causes, e.g. intrauterine, peri- or postnatal infections, numerous genetic defects leading to hearing impairment are known. Hereditary hearing loss is classified into two main forms: syndromic and nonsyndromic. Syndromic hearing loss includes autosomal dominant (e.g. Waardenburg syndrome, MIM 193500, branchio-oto-renal syndrome, MIM 113650 and Stickler syndrome, MIM 108300), autosomal recessive (e.g. Usher syndrome, MIM 276900, Pendred syndrome, MIM 274600 and Jervell-Lange-Nielsen syndrome, MIM 220400), and X-linked forms (e.g. Alport syndrome, MIM 301050).¹ Each of these syndromes is caused by mutations in more than one gene.^{1,2} Syndromic hearing loss can also occur in a number of mitochondrial diseases.³ Similarly, nonsyndromic hearing loss comprises autosomal dominant (DNFA), autosomal recessive (DFNB) and X-linked (DFNX or DFN) subtypes as well as some mitochondrial forms (Hereditary Hearing Loss Homepage, <http://hereditaryhearingloss.org/>). Mutations in the *GJB2* gene (MIM 121011) encoding connexin 26 (CX26)⁴ are responsible for most cases of autosomal recessive nonsyndromic hearing loss, accounting for more than 30% of affected children in Western Europe.⁵ More than 40 different genes have been identified for autosomal recessive deafness (Hereditary Hearing Loss Homepage), while X-linked nonsyndromic hearing loss is much rarer, accounting for 1-3% of hereditary deafness.¹ Only three genes have been associated with X-linked nonsyndromic hearing loss until now: *PRPS1* (MIM 311850; DFNX1, formerly DFN2) on Xq22 which encodes phosphoribosyl pyrophosphate synthetase 1 (ref. 6), *POU3F4* (MIM 300039; DFNX2, formerly DFN3) on Xq21, encoding

a member of a transcription factor family which contains a POU domain⁷ and the recently identified gene for the small muscle protein, X-linked (*SMPX*; MIM 300226) on Xp22 accounting for DFNX4 (formerly DFN6).^{8,9} An additional locus has been mapped to the short arm of the X chromosome (DFNX3, formerly DFN4) overlapping with the *DMD* locus on Xp21.2.^{10,11}

Here, we present data on a Hungarian family in which only males suffer from severe congenital hearing loss while females are not or only mildly to moderately affected. Based on the highly suggestive X-linked recessive inheritance and after ruling out the known X-linked genes *PRPS1* and *POU3F4*, next generation sequencing of the X chromosomal exome was performed on three family members in order to identify the underlying genetic defect. Segregation, bioinformatic and expression analyses identified a missense mutation in the *COL4A6* gene underlying the hearing impairment in this family.

MATERIALS AND METHODS

Karyotyping and SNP array

To verify the presence of structural changes, conventional cytogenetic analysis was performed on chromosomes prepared from 72 hrs PHA stimulated peripheral blood lymphocytes of the index patient using the standard procedure. Metaphase chromosomes were then evaluated by Giemsa-Trypsin-Giemsa (GTG) banding at about the 500 band level according to the International System for Human Cytogenetic Nomenclature (ISCN, 2009).

In order to ascertain genomic alterations, a genome-wide SNP array was performed using the HumanCytoSNP-12 BeadChip (Illumina Inc., San Diego, Calif., USA) encompassing 300,000 SNPs dispersed across the genome with an average distance of 10 kb between markers. In brief, 200 ng of genomic DNA from the index patient, mother and cousin was hybridized to the BeadChip according to manufacturer's instruction. After hybridization, the BeadChip was scanned with the Illumina BeadArray Reader and the data were analysed by examining signal intensity (log R ratio) and allelic composition (BAF) with GenomeStudio v2010.1 and cnvPartition v3.1.6 software. A minimum of a 5 probe cut-off value was used to define a copy number change.

Marker and MLPA analysis

CA-repeat markers located in or adjacent to the dystrophin gene (*DMD*) on Xp21.2 (5'DYS-I, 5'-7n4, STR-45, STR-49, 3'-19n8, DXS992; according to Leiden Muscular Dystrophy pages, http://www.dmd.nl/ca_dmd.html) were examined in all available family members using fluorescence-labeled primers and standard PCR conditions. Fragment analyses were performed on an ABI 3130xl capillary sequencer (Applied Biosystems, Carlsbad, CA, USA) according to standard protocols. Data were evaluated using GeneMapper v.4.0 (Applied Biosystems, Carlsbad, CA, USA) and haplotypes were constructed by Cyrillic v.2.1 (CyrillicSoftware, Oxfordshire, UK).

DNA samples of the index patient and his mother were analysed by multiplex ligation-dependent probe amplification (MLPA) for quantitative genomic changes in the *POU3F4* gene using the P163-C1 probe mix (MRC Holland, Amsterdam, The Netherlands) according to manufacturer's instructions. MLPA PCR products were separated on an ABI 3130xl automatic sequencer, pre-analysed by GeneMapper v.4.0 and finally evaluated by the Excel-based software Coffalyser v.5.2 (MRC Holland, Amsterdam, The Netherlands).

DNA sequencing analysis

Sanger sequencing was performed for the coding exons of the X-linked genes *PRPS1* and *POU3F4* in the index patient on an ABI 3130xl capillary sequencer using standard conditions. Sequence data were analysed by Gensearch v.3.8 (PhenoSystems, Lillois, Belgium) and automatically aligned to reference sequences from GenBank (NCBI, <http://www.ncbi.nlm.nih.gov/genbank/>).

The whole exome of the X chromosome was sequenced in two affected males (Fig. 1, III.3 and III.5) and an obligatory carrier female (II.2) by next generation sequencing (NGS) on a HiSeq2000 (Illumina, San Diego, CA, USA) after target enrichment with the SureSelect Human Chromosome X Exome Kit (Agilent Technologies, Santa Clara, CA, USA) according to manufacturer's instructions. Sequence reads were mapped using the Burrows-Wheeler Aligner (BWA)¹² after trimming and sorting of the raw data. Processed data were aligned to the X chromosome sequence of the human reference sequence hg19 (GRCh37) and analyzed using the software GensearchNGS (PhenoSystems, Lillois, Belgium), which enables variant detection based on VarScan.¹³ Detected variants of the three examined patients were automatically compared in a table and additional filtering of the variants was performed considering quality (Phred quality score > 20), coverage (> 20x), variant frequency (> 0.8 for male patients, > 0.3 and < 0.7 for the female carrier), intragenic location (elimination of extragenic variants) and biochemical effect of the variant (severity > synonymous variants).

Candidate variants were verified by Sanger sequencing using standard protocols and their segregation was studied within the family.

DNA samples of 93 healthy Hungarian women as well as 91 male and 65 female blood donors from Central Europe were screened for the presence of the c.1771 variant in the *COL4A6* gene either by Sanger sequencing or by pyrosequencing on a PyroMark Q96 MD system (Qiagen GmbH, Hilden, Germany) according to manufacturer's conditions.

The *COL4A6* gene was sequenced in 94 unrelated male and two female patients suffering from hearing loss, following target enrichment of all 46 exons (including the two alternative exons: 1A of NM_001847.2 and 1B of NM_033641.2) and parts of the 5' and 3'UTRs with the Access Array System (Fluidigm Europe B.V., Amsterdam, The Netherlands) according to the instructions provided by Fluidigm. Amplicon sequencing was performed on a Genome Sequencer 454 FLX System (Roche Applied Science, Mannheim, Germany) after amplification by emulsion PCR using reagents and conditions provided by Roche Applied Science. Sequence data were processed and analyzed using the software GensearchNGS as described above.

Bioinformatic analysis

Prediction of the effect of amino acid substitutions on protein function was done by the bioinformatic tools SIFT (<http://sift.jcvi.org/>), PolyPhen-2 (Polymorphism Phenotyping v2, <http://genetics.bwh.harvard.edu/pph2/index.shtml>) and Mutation Taster (<http://www.mutationtaster.org/>) which are included in Alamut version 2.2 (Interactive Bio-software, Rouen, France). Splice site prediction was also performed using Alamut which comprises five different prediction tools: SplicesiteFinder-like (Alamut-specific tool), MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html), NN-SPLICE (http://www.fruitfly.org/seq_tools/splice.html), GeneSplicer (<http://www.cbcb.umd.edu/software/genesplicer/>) and Human Splicing Finder (<http://www.umd.be/HSF/>). The alignment tool MultAlin (<http://multalin.toulouse.inra.fr/multalin/>)¹⁴ was used for multiple sequence

alignments of COL4A6 protein sequences from different species (mammals, fish and reptiles) and of the six alpha chains ($\alpha 1$ - $\alpha 6$) of human type IV collagens.

Bioinformatic analysis for protein stability and structure prediction was performed using software THE BuScr 1.06 (<http://structbio.biochem.dal.ca/jrailey/THeBuScr.html>), SCWRL 4.0 (<http://dunbrack.fccc.edu/Software.php>) and UCSF Chimera 1.5 (<http://www.cgl.ucsf.edu/chimera/>).

COL4A6 mRNA analysis

Fibroblasts from skin biopsies of the index patient and his mother were grown in Dulbecco's Modified Eagle Medium (DMEM; Gibco, Life Technologies GmbH, Darmstadt, Germany) supplemented with 15% Fetal Bovine Serum (FBS; Gibco). After harvesting the cultured cells with trypsin/EDTA solution (Gibco), total mRNA was isolated using RNeasy Mini Kit (Qiagen). Complementary DNA was synthesized by reverse transcriptase (Superscript II, Invitrogen, Life Technologies GmbH, Darmstadt, Germany). Amplification of exons 22 to 24 of *COL4A6* cDNA was done using specific primers located in the adjacent exons (21F: 5'-AAAGGAGCCAGAGGAGATCG-3', 25R: 5'-GAGGCCTCGAGACCCTTTAG-3'). Obtained fragments were sequenced on an ABI 3130xl capillary sequencer using standard conditions.

Expression studies

We performed an immunohistochemical staining of the mouse inner ear (C57Bl/10 male mouse, 12 weeks of age) with a COL4A6 specific antibody (rabbit polyclonal antibody M137, Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) following published procedures.¹⁵ Antibody specificity was checked by omitting the first antibody (data not shown).

Whole-mount *in situ* hybridization according to established protocols¹⁶ was done to assess the *col4a6* expression pattern in wild-type developing embryos of the zebrafish. A 620 bp fragment corresponding to the coding region (nt 1487 - nt 2106) adjacent to the 3'UTR of the *col4a6* mRNA (XM_001921595.2) was isolated from cDNA of whole embryos (26-28 hours post fertilization, hpf) and cloned into pCRII TOPO vector (Invitrogen). After linearization of the plasmid both sense and antisense digoxigenin labelled probes were generated through *in vitro* transcription using Sp6 and T7 RNA polymerase. The control sense probe did not reveal a detectable signal. The riboprobe sequence specificity was checked by BLAST analysis against the entire zebrafish genome which showed exclusive full length homology to the *col4a6* on chromosome 7.

RESULTS

Clinical data

The family examined in the present study, consists of 17 available members from three generations with 9 males, five of which suffering from bilateral severe sensorineural hearing loss while three females are healthy and five females suffer from mild to moderate hearing impairment. The pedigree of the non-consanguineous family is shown in Fig. 1A. The index patient (III.3) did not pass the standardized hearing screening during pediatric examination at the age of three months. Brain stem audiometry revealed bilateral severe hearing loss with unaided hearing thresholds of 80 dB right side and 80 to 100 dB left side (data not shown). Lacking any obvious postnatal diseases or treatments with potential impact on

hearing, the hearing loss was considered congenital and the patient received bilateral hearing aids. He developed speech in two to three word sentences at the age of three years but articulation was insufficient. At the age of three years, magnetic resonance imaging (MRI) of the cranium and high resolution computed tomography (HRCT) of the temporal bone showed bilateral malformation of the cochlea with incomplete separation of the cochlea from the internal auditory canal (Fig. 2A). No further malformation of the temporal bone, labyrinth and cerebello-pontine angle was detected. The patient received his first cochlear implant (CI) on the left side at the age of four years. Losing his usable residual hearing at the right side by the age of eight (Fig. 1C) he underwent cochlear implantation of the second, right side for bilateral rehabilitation. Intraoperative, a gusher phenomenon (profuse escape of cerebrospinal fluid after opening of the cochlea) was observed in both ears. Although the index patient is fully focused on spoken language communication, his hearing perception is reduced in situations of background noise. His performance in speech perception is well within the range of patients with inner ear malformation.¹⁷ No further clinical manifestations were obvious. In particular, renal laboratory parameters, abdominal ultrasound and ophthalmological examination were normal.

Three further male family members (II.3, III.5 and III.6) also suffer from severe hearing loss due to malformation of the cochlea as evidenced by HRCT and MRI imaging (Fig. 1B and 2B+C). Language skills of these affected males are severely restricted indicating pre-lingual onset of hearing loss and lack of appropriate treatment with hearing aids and CIs. Patient II.1 showed normal development in early childhood until he suffered from encephalitis at the age of one year and presumably developed deaf-mutism and severe mental retardation as a consequence of this infection. Computed tomography of his cochlea showed no malformation but signs of a post-inflammation status supporting the assumption that hearing loss is not inherited but acquired.

Four female family members (I.1, II.4, II.5 and II.6) developed mild to moderate hearing impairment during the third and fourth decade of their lives, respectively. Female II.4 (aged 55) suffers from mixed hearing loss and uses conventional hearing aids since approx. 20 years (Fig. 1B). Female II.6 (aged 45) uses a hearing aid since ten years due to a conductive hearing loss on the left side (Fig. 1B). Reliable data for the reason of hearing loss in patients I.1 and II.5 are missing. Female III.10 shows normal thresholds of her right ear and suffers from low grade conductive hearing loss on the left side (Fig. 1B). The 46-year-old mother (II.2) of the index patient shows no hearing impairment to date and computed tomography of her cochlea indicated no pathological findings (Fig. 2D).

Since only male family members are affected by severe sensorineural hearing loss already in early childhood, nonsyndromic X-chromosomal recessive hearing loss was most likely in this family.

Informed consent for genetic analyses was given by all ascertained family members.

Mapping and sequencing data

The pedigree of the family fits best to X-chromosomal inheritance of the hearing loss. Because of the malformation of the cochlea *POU3F4* was the most promising candidate gene. Sanger sequencing and MLPA analysis of *POU3F4* including the promoter region revealed no mutation. No distinct chromosomal rearrangements were detected by standard cytogenetic analysis using GTG banding ruling out a translocation disrupting the *POU3F4* gene. Large copy number variations (CNVs) were excluded by SNP array analysis.

Further, the X-linked gene *PRPS1* was excluded in the index patient by Sanger sequencing and linkage to the *DFNX3* locus on Xp21.2 was ruled out by analysis of markers located in or adjacent to the *DMD* gene (data not shown).

Sequencing of the X-chromosomal exome of the index patient (III.3), his mother (II.2) and his cousin (III.5) revealed 86 million passed filter reads (100 bp single reads) and an average coverage between 87 and 103 for the complete X chromosomal exome. After alignment against the X chromosomal reference sequence of hg19 we obtained 8 to 10 million reads per patient. About 10.000 exonic variants were scanned for each patient after quality filtering (Phred >20 and coverage >20x). Due to additional filtering of variant frequencies and mutation type as well as comparison of the detected variants in the female carrier and both male patients we obtained a list of 31 shared variants in 23 different genes. About half of these variants could be excluded because they were annotated SNPs. Of the 15 remaining variants (in eight genes), six could not be confirmed by Sanger sequencing and hence represent NGS technical artefacts or can be explained by mis-aligned NGS reads due to existing pseudogenes. Segregation analyses of the remaining nine variants (in four genes: *NHS*, MIM 300457; *SLC6A8*, MIM 300036; *BCAP31*, MIM 300398; and *COL4A6*, MIM 303631) revealed that only a missense mutation in exon 23 of the gene encoding the alpha-6 chain of collagen type IV (*COL4A6*, NM_001847.2) perfectly co-segregated with the disease in the family (Fig. 1A). The transition c.1771G>A resulting in the amino acid substitution p.Gly591Ser was present in all affected male patients in a hemizygous form except for patient II.1 who had suffered from an infantile encephalitis. All three obligatory female carriers and the slightly affected female family members carried the mutation in a heterozygous form. The substitution was absent from all unaffected male family members. Linkage analysis of the detected *COL4A6* variant revealed a LOD score of 2.36.

None out of 407 examined X chromosomes of male and female anonymous blood donors from Hungary and Central Europe showed the c.1771G>A transition in *COL4A6*, thus excluding a common polymorphism. Moreover, the mutation has not been described in any public SNP database.

Amplicon sequencing of *COL4A6* in 96 patients suffering from hearing loss (in whom mutations in *GJB2* had been excluded) could not detect the candidate mutation c.1771G>A or any other causal mutation for hearing loss in the *COL4A6* gene.

Bioinformatic data

Bioinformatic tools that predict the effect of amino acid substitutions on protein function classified the missense mutation c.1771G>A, p.Gly591Ser as pathogenic (SIFT: deleterious, score: 0.00; PolyPhen: probably damaging, score: 1.00; Mutation Taster: disease causing, p-value: 0.983). According to Alamut, both, nucleotide c.1771G and amino acid Glycine-591 are highly conserved. Multiple sequence alignments of *COL4A6* protein sequences from different species (Fig. 3A) and of the six alpha chains of human type IV collagens using MultAlin also showed high evolutionary conservation for Glycine-591 which is substituted by serine in the affected members of the Hungarian family.

Five different splice site prediction tools which are included in Alamut predicted minor to moderate effects on mRNA splicing in the mutated sequence compared to the wild-type sequence ranging from -3.9% to -36.5% (mean of all five tools: -18.9%) for predicted change in acceptor splice site usage (Table 1).

Bioinformatic analysis for protein stability and structure prediction using the described software showed striking differences in the collagen triple helix stability and structure of the region containing the p.Gly591Ser variant compared to the wild-type (Fig. 3B).

prediction tool	scale	wild-type 3' splice site pred.	Gly591Ser 3' splice site pred.	percentage of predicted change
SpliceSiteFinder-like	0-100	80.6	76.7	-4.8 %
MaxEntScan	0-12	6.6	4.8	-27.3 %
NNSPLICE	0-1	0.9	0.7	-22.2 %
GeneSplicer	0-15	5.2	3.3	-36.5 %
Human Splicing Finder	0-100	80.2	77.1	-3.9 %

Table 1: Splice site prediction (pred.) of five different software tools which are integrated in Alamut (Interactive Biosoftware) showing differences in acceptor splice site usage of intron 22 / exon 23 for the mutated sequence (Gly591Ser) compared to the wild-type COL4A6 gene.

COL4A6 mRNA and expression data

Since the identified *COL4A6* missense mutation affects the first nucleotide of exon 23 and splice site prediction software indicated a moderate effect on splicing probability caused by the mutation, we studied the *COL4A6* mRNA from fibroblasts of the index patient and his mother. Except for the c.1771G>A substitution, no difference in size, sequence or quantity compared to the *COL4A6* cDNA from control fibroblasts was detected (data not shown).

Immunohistochemical staining of the mouse inner ear showed apart from multiple structures where a weak and somewhat diffuse staining was observed, that most pronounced immune signals of COL4A6 were visible at membranous and osseous structures at the *stria vascularis* of the spiral ligament (Fig. 4A) and at higher magnification, in a subgroup of ganglia cells of the *ganglion spirale* (Fig. 4B).

Whole-mount *in situ* hybridization was performed to assess the *col4a6* expression pattern in wildtype developing zebrafish embryos (Fig. 4C). Around the 15-16 somite stage, the *col4a6* transcript was clearly visible in the dorsal midbrain and notochord with a weak expression in the otic vesicle. At the 18 somite stage, expression in the otic vesicle became more prominent, while expression in the dorsal midbrain region persisted (not shown). Furthermore, faint expression was seen in the hind brain and pharyngeal arch. During the early pharyngeal period (28 hpf), in addition to strong expression in the otic vesicle, high expression was noted in the mid hindbrain boundary and pharyngeal arches (not shown). At the same time, the expression in the notochord and dorsal midbrain diminished and a low level of expression appeared in the neural tube. After 42 hpf expression in the ear and pharyngeal arches remained strong, but additional expression occurred in the swim bladder (Fig. 4C), along with weak expression in the lateral line. A similar pattern of expression was also observed at 96 hpf, which did not differ remarkably from the pattern described in 42 hpf.

DISCUSSION

X-chromosomal exome sequencing revealed a missense mutation in *COL4A6* encoding the alpha-6 chain of type IV collagen which co-segregated in all affected members of a Hungarian family suffering from congenital nonsyndromic hearing loss. Bioinformatic analyses, expression studies and clinical data collectively provided evidence for the p.Gly591Ser variant underlying the hearing impairment in the three-generation family and for the type IV-collagen playing an important role in cochlea development. Although the missense mutation affects the first nucleotide of exon 23, aberrant splicing could not be demonstrated by *COL4A6* mRNA analysis from fibroblasts of the index patient. Since splice site prediction tools indicated low to moderate effects on splicing for the detected mutation, minor missplicing is considered possible, eventually resulting in an unstable product that escapes detection by PCR and subsequent sequencing. In addition, we cannot exclude abnormal splicing in the target tissue (inner ear), which was not accessible to analysis.

All members of the collagen superfamily display the characteristic triple helical structure of three alpha chains which is stabilized by the presence of glycine as every third residue and a high content of proline and hydroxyproline.¹⁸ Thus, mutation of a glycine residue within the triple helix structure such as p.Gly591Ser in the alpha-6 chain can be expected to exert a negative effect on both helical partners (two alpha-5 chains), destabilizing the whole collagen molecule which could be shown by bioinformatic collagen stability and structure prediction (Fig. 3B). This is furthermore strongly supported by the fact that other genetic diseases of collagens, e. g. osteogenesis imperfecta (*COL1A1*, MIM 120150; *COL1A2*, MIM 120160) or Ehlers-Danlos syndrome (*COL1A1*; *COL1A2*; *COL3A1*, MIM 120180; *COL5A1*, MIM 120215; *COL5A2*, MIM 120190) are predominantly caused by mutations of glycine residues in the corresponding collagen genes.¹⁹

The great diversity in the collagen superfamily is not only due to the existence of 28 different family members but also to the presence of several molecular isoforms of the triple helices formed either by homo- or heterotrimers.¹⁸ Type IV collagens are network-forming collagens which are made up of two molecular isoforms e.g. two alpha-5 chains combined with one alpha-6 chain [$\alpha 5(\text{IV})_2\alpha 6(\text{IV})$].²⁰ The genes *COL4A5* (MIM 303630) and *COL4A6* coding for the alpha-5 and alpha-6 chains of type IV collagens are located close to each other in a head-to-head orientation on chromosome Xq22.²¹ The transcription of both genes is controlled by a common bidirectional promoter element but separate promoters for *COL4A5* and *COL4A6* allow also for tissue-specific gene expression.²² All type IV collagens are components of basement membranes: $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$ are expressed ubiquitously and the remaining alpha chains are only present in specific tissues. *COL4A5* is expressed in basement membranes of the kidney, skin, trachea, eye and neuromuscular junction, *COL4A6* is also present in these tissues except for the renal glomerular basement membrane.²² Consistently, the index patient of the Hungarian family showed normal renal function. *COL4A1-A5* could be localized in different structures of the murine cochlea by immunohistochemistry, however *COL4A6* was not included in this investigation.¹⁵ In the present study, distinct expression of *COL4A6* in the *stria vascularis* and the *ganglion spirale* of the mouse inner ear were shown by immunohistochemical staining (Fig. 4A and B). The current zebrafish genome (build Zv9) shows a single *COL4A6* ortholog on chromosome 7 (DRE7). Zebrafish *col4a6* is organized head-to-head with another type IV collagen gene, *col4a5*, on DRE7, demonstrating synteny conservation with human Xq22. Whole-mount *in situ* hybridization in wild-type developing zebrafish embryos showed dynamic expression of zebrafish *col4a6* in the ear which is characterized by relatively low transcript levels in the otic vesicle during early developmental stages, followed by a more pronounced expression at later stages (42-

96 hpf; Fig. 4C). These data provide further evidence for the involvement of *COL4A6* in the hearing process.

Mutations in the human *COL4A3* and *COL4A4* genes are associated with autosomal recessive Alport syndrome (<20% of cases) whereas the more frequent X-linked form of Alport syndrome (>80%) is caused by mutations in *COL4A5*.²³ In Alport syndrome, renal basement membranes are affected resulting in glomerulonephropathy and renal failure frequently accompanied by variable hearing impairment and ocular anomalies (MIM 301050 and 203780).²³ A subtype of X-linked Alport syndrome associated with diffuse leiomyomatosis is caused by a large deletion involving the two contiguous genes *COL4A5* and *COL4A6* (MIM 308940). This subtype is frequently combined with cataract and deafness.^{24,25} Mutations in *COL4A6* alone have not been associated with Alport syndrome or any other hereditary disorder so far. Since renal problems have not been reported in any of the affected members of the Hungarian family in the present study, there is no evidence for Alport syndrome. Our results rather suggest that missense mutations in the *COL4A6* gene cause severe nonsyndromic hearing impairment in males. The female carriers in the family show variable phenotypes ranging from normal to mild or moderate hearing capabilities. Female II.2 showed normal hearing at age 46 (Fig. 1B). In contrast, females I.1, II.4 and II.6 developed hearing impairment in their third or fourth decade of life requiring hearing aids. This variability in severity and onset in these female carriers might be explained by skewed X-inactivation resulting in a higher or lower expression ratio of the mutated and wild-type *COL4A6* protein encoded by the two X-chromosomal gene copies. Since there are no CT images available of the three affected carriers, no statement about the presence of cochlear malformations in female carriers can be made except for the healthy mother of the index patient who is a proven carrier but shows no malformation of her cochlea. The hearing impairment in females II.5 (left side only) and III.10 (Fig. 1B) could be of mixed origin since there is chronic otitis of the middle ear reported (more details not available).

Interestingly, the *COL4A6* gene is located only approx. 500 kb apart from the *DFNX1/PRPS1* locus. Mutations in the coding regions of *PRPS1* had been excluded in our family by Sanger sequencing. Most likely, the two genes cannot be discriminated by linkage analysis in pedigrees and, therefore, may not have been recognized before as separate loci. Microdeletions upstream of *POU3F4* which is located about 24.6 Mb apart from *COL4A6* have been described repeatedly as causative for the most common form of X-linked non-syndromic deafness associated with inner ear malformations.^{26,27} Both MLPA analysis comprising six probes in the critical upstream region and SNP array analysis did not provide any indication for CNVs in the *POU3F4* region. Additional Sanger sequencing of the coding region of *POU3F4* excluded this locus to be linked to hearing loss in the index patient.

In conclusion, our study provides evidence to suggest *COL4A6* as the fourth gene for X-linked hearing loss. In all affected males of the Hungarian family the p.Gly591Ser mutation segregates with severe hearing impairment and cochlear malformation while further known X-linked deafness loci have been excluded. The alpha-6 chain of collagen type IV is part of the basement membrane of the inner ear and may play an important role in cochlea development. We propose that the p.Gly591Ser mutation is responsible for the observed malformation of the cochlea in the affected males.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The authors are deeply grateful to all family members for their participation in this study and to Dr. Jörg Schröder, Prof. Wafaa Shehata-Dieler and to Prof. Annerose Keilmann for clinical support. We thank Prof. Tiemo Grimm for calculation of the LOD scores, Dr. Eberhard Schneider for inspecting radiographs of patients with hearing loss and Dr. Wolfram Kress for constructive discussions. We appreciate the donation of control samples by Dr. Veronika Karcagi and Dr. Ulrich Zechner. We also thank Prof. Manfred Scharl for his support with the whole mount in situ experiment in zebrafish embryos. Additionally, we would like to thank Simone Günther, Birgit Halliger-Keller, Jens Gräf and Mohammed Uddin for technical assistance.

This work was supported by the German Research Foundation.

REFERENCES

1. Bayazit YA, Yilmaz M: An overview of hereditary hearing loss. *ORL J Otorhinolaryngol Relat Spec* 2006; **68**: 57-63.
2. Yan D, Liu XZ: Genetics and pathological mechanisms of Usher syndrome. *J Hum Genet* 2010; **55**: 327-335.
3. Kokotas H, Petersen MB, Willems PJ: Mitochondrial deafness. *Clin Genet* 2007; **71**: 379-391.
4. Kelsell DP, Dunlop J, Stevens HP *et al*: Connexin 26 mutations in hereditary nonsyndromic sensorineural deafness. *Nature* 1997; **387**: 80-83.
5. Tekin M, Arnos KS, Pandya A: Advances in hereditary deafness. *Lancet* 2001; **358**: 1082-1090.
6. Liu X, Han D, Li J *et al*: Loss-of-function mutations in the PRPS1 gene cause a type of nonsyndromic X-linked sensorineural deafness, DFN2. *Am J Hum Genet* 2010; **86**: 65-71.
7. de Kok YJ, van der Maarel SM, Bitner-Glindzicz M *et al*: Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. *Science* 1995; **267**: 685-688.
8. Huebner AK, Gandia M, Frommolt P *et al*: Nonsense mutations in SMPX, encoding a protein responsive to physical force, result in X-chromosomal hearing loss. *Am J Hum Genet* 2011; **88**: 621-627.
9. Schraders M, Haas SA, Weegerink NJ *et al*: Next-generation sequencing identifies mutations of SMPX, which encodes the small muscle protein, X-linked, as a cause of progressive hearing impairment. *Am J Hum Genet* 2011; **88**: 628-634.
10. Lalwani AK, Brister JR, Fex J *et al*: A new nonsyndromic X-linked sensorineural hearing impairment linked to Xp21.2. *Am J Hum Genet* 1994; **55**: 685-694.
11. Pfister MH, Apaydin F, Turan O *et al*: A second family with nonsyndromic sensorineural hearing loss linked to Xp21.2: refinement of the DFN4 locus within DMD. *Genomics* 1998; **53**: 377-382.
12. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 2009; **25**: 1754-1760.
13. Koboldt DC, Chen K, Wylie T *et al*: VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 2009; **25**: 2283-2285.

14. Corpet F: Multiple sequence alignment with hierarchical clustering. *Nucl Acids Res* 1988; **16**: 10881-10890.
15. Cosgrove D, Samuelson G, Pinnt J: Immunohistochemical localization of basement membrane collagens and associated proteins in the murine cochlea. *Hear Res* 1996; **97**: 54-65.
16. Winkler C, Schafer M, Duschl J, Scharl M, Volff JN: Functional divergence of two zebrafish midkine growth factors following fish-specific gene duplication. *Genome Res* 2003; **13**: 1067-1081.
17. Buchman CA, Copeland BJ, Yu KK, Brown CJ, Carrasco VN, Pillsbury HC: Cochlear implantation in children with congenital inner ear malformations. *Laryngoscope* 2004; **114**: 309-316.
18. Ricard-Blum S: The collagen family. *Cold Spring Harb Perspect Biol* 2011; **3**: a004978.
19. Dalgleish R: The Human Collagen Mutation Database 1998. *Nucleic Acids Res* 1998; **26**: 253-255.
20. Borza DB, Bondar O, Ninomiya Y *et al*: The NC1 domain of collagen IV encodes a novel network composed of the alpha 1, alpha 2, alpha 5, and alpha 6 chains in smooth muscle basement membranes. *J Biol Chem* 2001; **276**: 28532-28540.
21. Sugimoto M, Oohashi T, Ninomiya Y: The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5(IV) and alpha 6(IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. *Proc Natl Acad Sci USA* 1994; **91**: 11679-11683.
22. Segal Y, Zhuang L, Rondeau E, Sraer JD, Zhou J: Regulation of the paired type IV collagen genes COL4A5 and COL4A6. Role of the proximal promoter region. *J Biol Chem* 2001; **276**: 11791-11797.
23. Kashtan CE: Alport syndrome: an inherited disorder of renal, ocular, and cochlear basement membranes. *Medicine* 1999; **78**: 338-360.
24. Cochat P, Guibaud P, Garcia-Torres R, Roussel B, Guarner V, Larbre F: Diffuse leiomyomatosis in Alport syndrome. *J Pediat* 1988; **113**: 339-343.
25. Anker MC, Arnemann J, Neumann K, Ahrens P, Schmidt H, Konig R: Alport syndrome with diffuse leiomyomatosis. *Am J Med Genet* 2003; **119A**: 381-385.
26. de Kok YJ, Vossenaar ER, Cremers CW *et al*: Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Hum Mol Genet* 1996; **5**: 1229-1235.
27. Song MH, Lee HK, Choi JY, Kim S, Bok J, Kim UK: Clinical evaluation of DFN3 patients with deletions in the POU3F4 locus and detection of carrier female using MLPA. *Clin Genet* 2010; **78**: 524-532.

TITLES AND LEGENDS TO FIGURES

Figure 1

A) Pedigree of the family. Black squares indicate affected males with severe hereditary hearing loss, the gray square (II.1) indicates a male with acquired deafness, circles with black dots show mutation carriers who are healthy to date (unshaded) or who are affected mildly to moderately (shaded). The detected nucleotides at position c.1771 of the *COL4A6* gene are given for every family member: G or GG means hemi- or homozygous for the wild-type sequence, GA means c.1771G>A heterozygous and A denotes the mutation in hemizygous form.

B) Representative pure tone audiometry of eight family members. The females II.2 and III.4 show regular thresholds at their current age. The females II.6 (at 41 years) and III.10 (at 7 years) show normal thresholds of their right ears and suffer from low grade conductive hearing loss of maximum 25 dB at 1kHz (II.6) and a maximum of 20 dB between 0.5 and 1kHz (III.10) on their left side. Female II.4 (at present) expresses bilateral moderate mixed hearing loss of maximum 35 dB (right) and 40 dB (left) at 1kHz and an additional conductive component of maximum 25 dB (right) at 2kHz and 30 dB (left) at 250 Hz and 1kHz. Thresholds of patients III.3 (at 8 years), III.5 (at 24 years) and III.6 (at 12 years) at their latest follow up showing only residual hearing in single frequencies (red – right side, blue – left side, circles and crosses – air conduction, arrow heads – bone conduction, vertical arrows – frequency not heard).

C) Behavior audiometry under best aided conditions (bilateral hearing aids) of the index patient III.3 at the age of three years prior to cochlea implantation of the left ear (dashed lines) and at the age of eight years before cochlear implantation of the right ear. On the left: Speech audiometry “Mainzer Kindertest” (solid diamonds) and “Freiburger monosyllables” (solid boxes), on the right: Functional gain using hearing aids measured with warble thresholds (red – right ear, blue – left ear).

Figure 2

High resolution imaging of the temporal bone of the index patient (III.3) at the age of three years, of his affected cousins (III.5 and III.6) and his healthy mother (II.2). Both sides are shown in each case.

A-C) Axial high resolution computed tomography (HRCT) scans showing the isolated malformed cochlea with incomplete partition of the cochlea and incomplete separation from the internal auditory canal (asterisks). The vestibular labyrinth is normal. The mastoid cavity and the middle ear ossicles appear regular (white arrow indicating malleo-incudial complex).

D) Representative images from a HRCT scan of the temporal bone in the mother of the index patient demonstrating a regular cochlea.

Figure 3

A) Multiple sequence alignment of 80 amino acids encompassing glycine-591 of the human *COL4A6* protein and orthologous proteins from other species using the software tool Multalin (<http://multalin.toulouse.inra.fr/multalin/>). Glycine-591 is perfectly conserved (black vertical bar; note that amino acid 591 is shifted by 19 positions to 610 in the alignment due

to insertions in COL4A6 of the other species; red letters: high consensus, blue: low consensus, black: neutral).

B) Structural comparison of wild-type COL4A6 (upper part) with the Gly591Ser variant (lower part). Left subfigures illustrate estimated melting temperatures in degrees Celsius along the COL4A6 amino acid sequence. The respective mutation site is highlighted with an arrow. The Gly591Ser mutation seriously affects the melting temperature of the natural model.

In the three-dimensional models (right subfigures) the COL4A6 chain is displayed in blue, whereas the white ribbons are the corresponding COL4A5 chains in the heterotrimerical structure. Atoms and bonds are colored in red in case they clash or have unfavorable contacts with atoms of the other helices (van der Waals overlap greater than 0.6 angstroms). The pairings of clashing atoms are indicated by yellow lines. In the wild-type COL4A6, at position 591 Gly has no side-chain at all and yields a clash-free conformation of the three chains. By contrast, the mutation to Ser produces an atypical large side-chain structure at the same position. In the model of the triple-helical collagen structure, this side-chain has several hard clashes that reduce the stability of the interchain-conformation and likely triggers disarrangement of the quaternary structure.

Figure 4

COL4A6 expression in the mouse inner ear and in the zebrafish embryo.

A) Thin sections of a de-mineralised murine inner ear were incubated with rabbit polyclonal Col4A6 antibody M137. COL4A6 is strongly expressed at membranous and osseous structures at the *stria vascularis* of the spiral ligament (black arrow) which were more intensely stained as compared to the basilar membrane underlying the organ of Corti (small black arrows). GS: *ganglion spirale*, OC: organ of Corti, RM: Reissner's membrane, SV: *stria vascularis*.

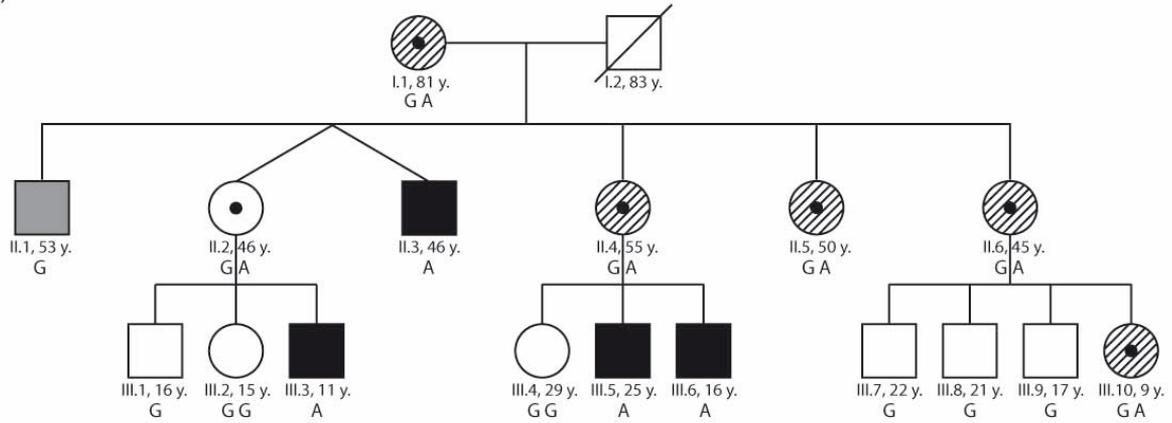
B) At higher magnification, a very distinct and pronounced reactivity was seen in a subgroup of ganglia cells of the *ganglion spirale* (white arrows).

C) Whole-mount *in situ* hybridization on zebrafish embryos of different developmental stages (e: ear, hpf: hours post fertilization, ot: otic vesicle, pa: pharyngeal arch, sb: swim bladder).

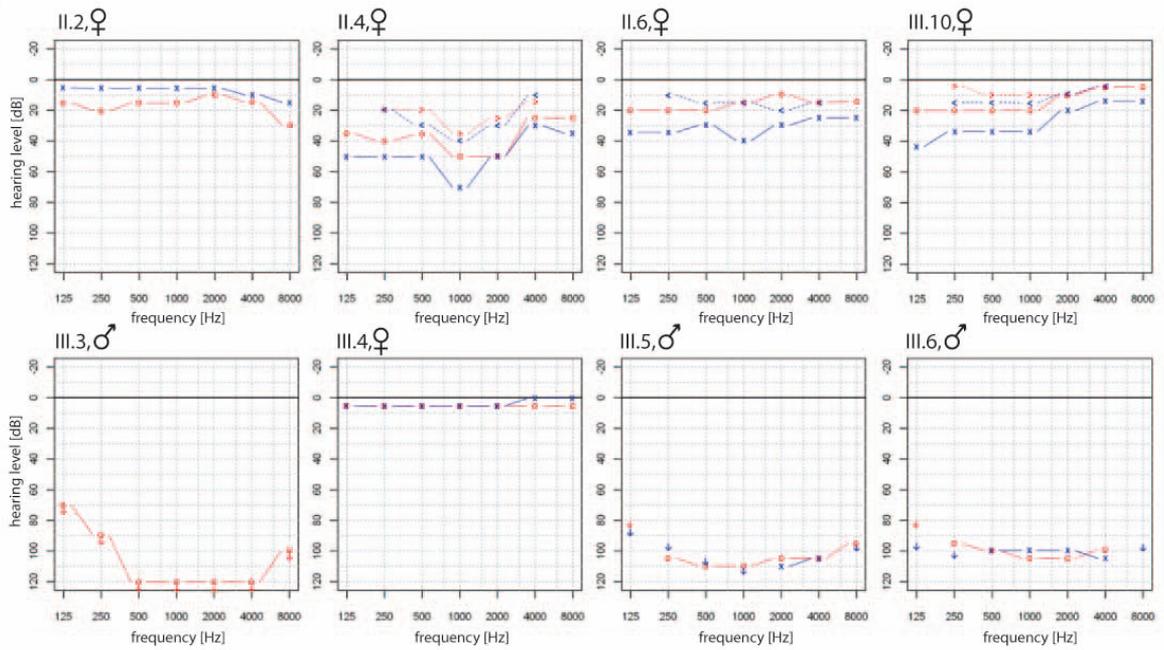
FIGURES

Figure 1

A)



B)



C)

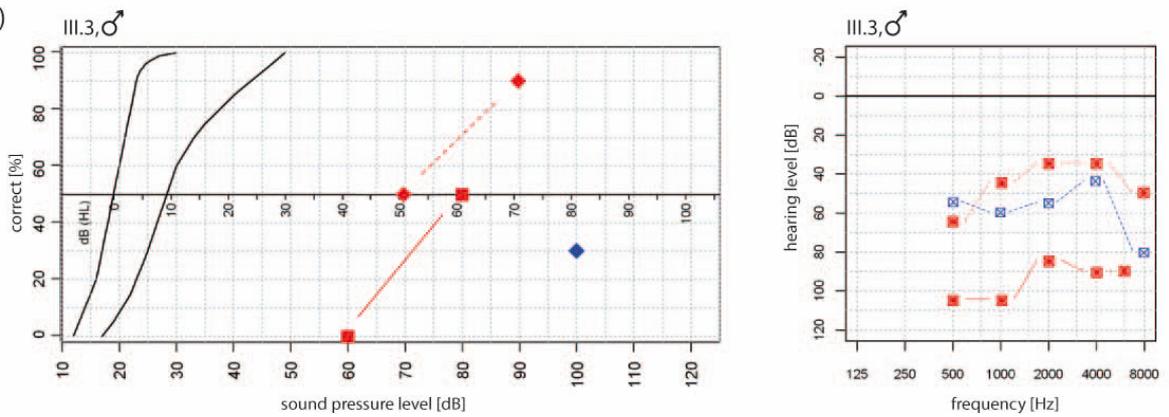
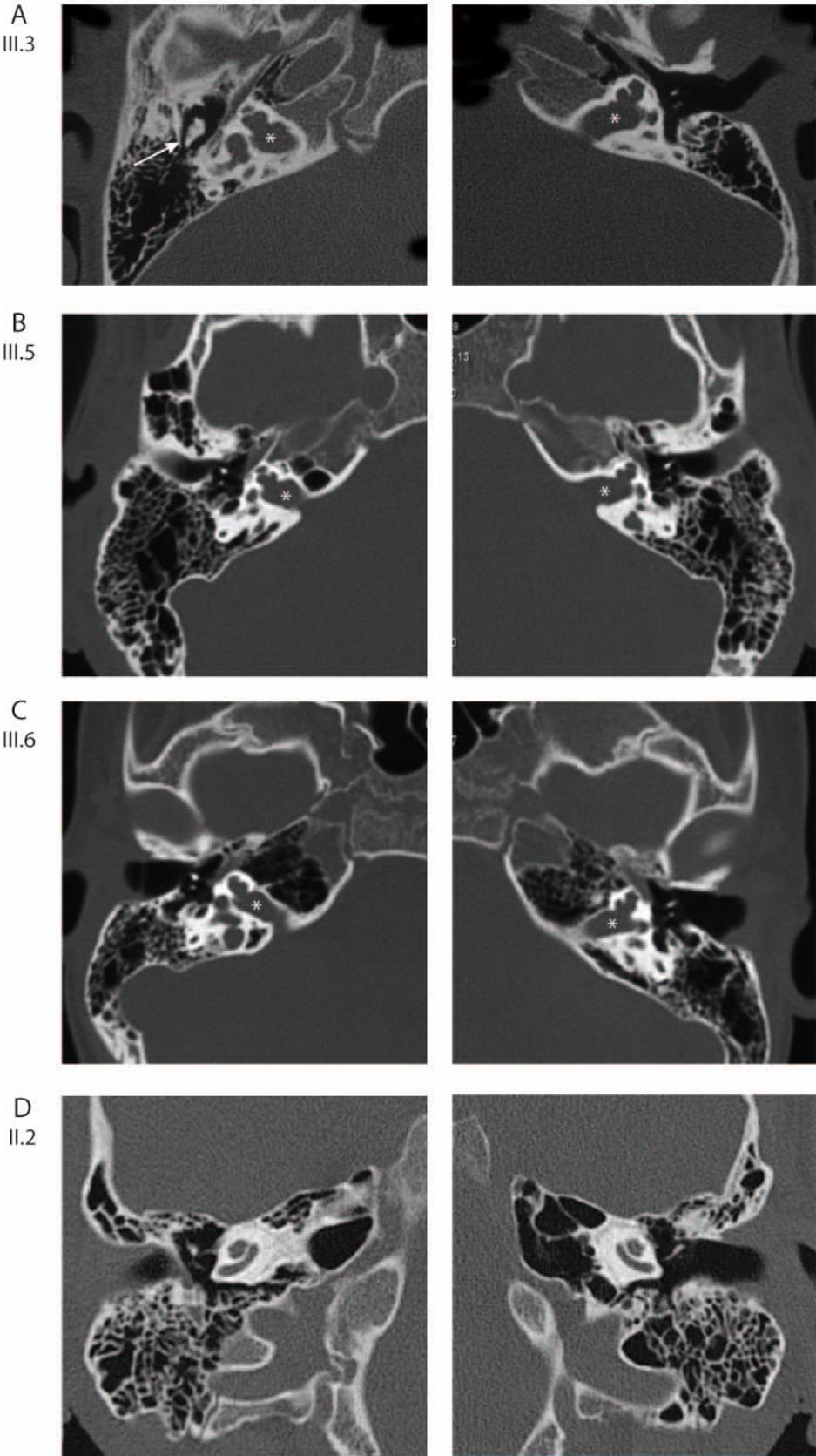


Figure 2



Anlage 2

M. Broman, I. Kleinschnitz, **J.E. Bach**, S.Rost, G. Islander and C.R. Müller

Next-generation DNA sequencing of a Swedish malignant hyperthermia cohort

Clin Genet. 2015 Oct; 88(4):381-5. doi: 10.1111/cge.12508



Short Report

Next-generation DNA sequencing of a Swedish malignant hyperthermia cohort

Broman M., Kleinschnitz I., Bach J.E., Rost S., Islander G., Müller C.R.
Next-generation DNA sequencing of a Swedish malignant hyperthermia cohort.
Clin Genet 2014. © John Wiley & Sons A/S. Published by John Wiley & Sons Ltd, 2014

Malignant hyperthermia (MH)-related mutations have been identified in the ryanodine receptor type 1 gene (*RYR1*) and in the dihydropyridine gene (*CACNA1S*), but about half of the patients do not have causative mutations in these genes. We wanted to study the contribution of other muscle genes to the *RYR1* phenotypes. We designed a gene panel for sequence enrichment targeting 64 genes of proteins involved in the homeostasis of the striated muscle cell. Next-generation sequencing (NGS) resulted in >50,000 sequence variants which were further analyzed by software filtering criteria to identify causative variants. In four of five patients we identified previously reported *RYR1* mutations while the fifth patient did not show any candidate variant in any of the genes investigated. In two patients pathogenic variants were found in other genes known to cause a muscle disorders. All but one patient carried likely benign rare polymorphisms. The NGS technique proved convenient in identifying variants in the *RYR1*. However, with a clinically variable phenotype-like MH, the pre-selection of genes poses problems in variant interpretation.

Conflict of interest

The authors have declared no conflicting interests.

**M. Broman^{a*}, I. Kleinschnitz^{b*},
J.E. Bach^b, S. Rost^b,
G. Islander^a and C.R. Müller^b**

^aDepartment of Perioperative and Intensive Care, Skåne University Hospital, Lund, Sweden and ^bInstitute of Human Genetics, Biocentre, University of Würzburg, Germany

*These authors contributed equally to this work.

Key words: malignant hyperthermia – muscle – mutation – next-generation sequencing

Corresponding author: Marcus Broman, Department of Perioperative and Intensive Care, Getingevägen 4, 22241 Lund Skåne University Hospital, Lund, Sweden.
Tel.: +4646171539;
email: marcus.broman@skane.se

Received 3 August 2014, revised and accepted for publication 22 September 2014

The malignant hyperthermia susceptibility (MHS) is a pharmacogenetic disorder in which halogenated inhalational anesthetics cause a dysregulation of intracellular $[Ca^{2+}]$ of the striated muscle cell inducing life threatening hypermetabolism and muscle rigidity. Outside anesthesia MH subjects vary from physically fit to myopathic patients with central core disease (CCD) and other muscle pathologies (1). There are a few reports on lethal MH-like reactions unrelated to anesthesia (2).

Mutations in the *RYR1* have been identified in about half of MH cases and a few more in *CACNA1S* (3, 4). However, the analysis of these two very large genes was incomplete in many studies. Recently, 32 genes were sequenced by NGS technique in two MHS cases (5) and a whole exome analysis was carried out in four MHS families (6), the first study ending up in identifying mutations in *RYR1*, and the latter in both *RYR1* and *CACNA1S*. In this pilot project, we selected *RYR1* and

CACNA1S and another 62 genes coding for proteins involved in Ca^{2+} homeostasis or other key functions of striated muscle to explore the power of NGS in variant detection.

Patients and methods

After approval from the Regional Ethical Board, Lund, Sweden and written consent from all patients, DNA was extracted from peripheral blood according to standard procedures from four novel Swedish MHS patients and one CCD patient (Table 1) who had not undergone genetic testing before.

The set included genes coding for proteins:

- i. in Ca^{2+} signaling (ASPH proteins, the SERCAs, dihydropyridine receptor, *RYR1*, calsequestrin),
- ii. in the contraction process (actin, troponin, tropomyosin),

Broman et al.

Table 1. Clinical and genetic details of the patients studied. Genetic variants are listed in the last column. Causative mutations are in bold type and candidate mutations in regular type. Detailed descriptions of the patients' clinical histories are given in the Appendix S1

Patient	Clinical diagnosis	In vitro muscle contracture test (IVCT) results MH status and max contractions or muscle biopsy	Larach score (10)	Physical disability	Found mutation: GENE and variant causative mutation (bold) candidate mutation (regular)
NGS1	Central core disease	Cores in muscle biopsy	n.a.	Wheelchair-bound, scoliosis, respiratory impairment	<i>RYR1</i> p.Thr4920Ile
NGS2	Malignant hyperthermia	MHS caff 4.4 g haloth 2.2 g	45	None observed	<i>FKTN</i> p.Gly125Ser <i>LAMA2</i> p.Arg1844Ser <i>RYR1</i> p.Arg2355Trp <i>KBTD13</i> p.Leu440Pro
NGS3	Malignant hyperthermia	MHS caff 3.3 g haloth 6.2 g Cores in muscle biopsy	63	None observed	<i>DMD</i> p.His2921Arg <i>FKRP</i> p.Ala114Gly <i>ATP2A2</i> intronic c.-10C>G <i>RYR1</i> p.Arg2435His <i>COL5A1</i> p.Asp636Asn <i>TTN</i> p.Arg18637Ser <i>RYR1</i> p.Arg2435His
NGS4	Malignant hyperthermia	MHS caff 5.3 g haloth 3.4 g	39	None observed	<i>RYR1</i> p.Arg2435His
NGS5	Malignant hyperthermia	MHS caff 0.4 g haloth 0.0 g	30	None observed	No variants

MHS, malignant hyperthermia susceptibility.

- iii. for Na⁺, K⁺, Ca²⁺ ion channels,
- iv. representing the immune system (interleukines 2 and 6),
- v. in the cytoskeleton (collagens),
- vi. for multifunctional enzymes/communication processes (junctophilins, laminin, integrin).

A detailed description is given in the Table S1, Supporting Information.

Probes were designed for the SureSelect Target Enrichment (Agilent, Böblingen, Germany) to cover the coding/splice consensus sequences. Sequencing was carried out on an Illumina HiSeq2000 (San Diego, CA, USA).

Of the targeted genomic area, 98.7% were covered by >20 reads. Analysis of variants was performed by alignment to the human reference genome (GRCh37, hg19) using GenSearch Software (PhenoSystems, Braine le Chateau, Belgium). The initial output of variants was >50,000 which were filtered by using the parameters: occurrence >20% of the forward/reverse reads AND balance >0 between forward/reverse reads AND minimum coverage >20 reads/base AND minor allele frequency (MAF) <3% in Caucasians in order to eliminate the majority of annotated single nucleotide polymorphisms (SNPs) (in databases dbSNP, ESP and 1000Genomes). Because the transmission of MH is autosomal dominant, we expected *heterozygous* variants to be covered by >20% of reads for the given position. This resulted in a secondary output of 203 variants (35–49 per patient),

as shown in Table 2, which were manually evaluated with prediction tools included in the software package ALAMUT (Interactive Biosciences, Rouen, France) for their impact on the amino acid sequence (SIFT, Mutation Taster, PolyPhen-2) and on splicing probabilities (SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer, Human Splicing Finder). Their annotation in genetic databases (dbSNP, ESP, 1000 genomes, HGMD) was checked for allele frequencies and known association to diseases. The data were not analyzed for copy number variation.

Variants were classified into three groups: (i) previously known/reported pathogenic, (ii) previously unknown but predicted pathogenic by all prediction tools, (iii) polymorphisms if at least one prediction tool predicted 'no pathogenicity' or previously published without functional relevance.

The variants in groups (i) and (ii) were confirmed by Sanger sequencing.

Results in *RYR1* or *CACNA1S*

In four of five patients we identified heterozygous pathogenic mutations in *RYR1* (Table 2). Patient NGS1 showed the mutation c.14759C>T/p.Thr4920Ile reported once before in association with CCD (7). Patient NGS2 showed: c.7063C>T/p.Arg2355Trp, whereas patients NGS3 and NGS4 both showed the

NGS DNA analysis of Swedish MH patients

Table 2. Variants identified and their classification with respect to reported or predicted pathogenicity

Patient	Variants after software analysis (n = 203)	Published known variants which are reported pathogenic (n = 9)	Previously unknown variants, but predicted likely pathogenic by predictions tools (n = 3)	Predicted SNPs by software analysis (n = 10) or database entries (n = 151)
NGS1	35	<i>RYR1</i> c.14759C>T p.Thr4920Ile rs118192155 Heterozygous Associated previously to malignant hyperthermia and central core disease		34
NGS2	42	<i>RFTN</i> c.373G>A p.Gly125Ser rs3400667 Heterozygous Associated previously to muscular dystrophy-dystroglycanopathy <i>LAMA2</i> c.5530C>A p.Arg1844Ser rs5617362 Heterozygous Associated previously to congenital muscle dystrophy 1 <i>RYR1</i> c.7063C>T p.Arg2355Trp rs183922803 Heterozygous Associated previously to malignant hyperthermia	<i>KBTD13</i> c.1319T>C p.Leu440Pro Heterozygous The variant is previously unknown	36
NGS3	49	<i>DMD</i> c.9762A>G p.His2921Arg rs1800279 Heterozygous Associated previously to Becker muscle dystrophy <i>FKBP</i> c.341C>G p.Ala114Gly rs143793528 Heterozygous Associated previously to congenital muscle dystrophy <i>ATP2A2</i> g.110719585C>G intronic variant c.-10C>G rs201929640 Heterozygous Associated previously to reduced mRNA expression <i>RYR1</i> c.7304G>A p.Arg2435His rs29933396 Heterozygous Associated previously to malignant hyperthermia	<i>COL6A1</i> c.1908G>A p.Asp636Asn Heterozygous The variant is previously unknown <i>TTN</i> c.55909C>A p.Arg18637Ser Heterozygous The variant is previously unknown	43
NGS4	42	<i>RYR1</i> c.7304G>A p.Arg2435His rs29933396 Heterozygous Associated previously to malignant hyperthermia		41
NGS5	35	No variants	No variants	35

NGS, next-generation sequencing; SNPs, single nucleotide polymorphisms.

Broman et al.

mutation c.7304G>A/p.Arg2435His. These latter mutations have been repeatedly associated with MH and/or CCD and have been found to impact on RYR1 function (8–11).

No variants were found in the *CACNA1S*. Of note, patient NGS5 did not show any candidate variant in any gene investigated.

Results in other genes

NGS2 harbored the heterozygous variant c.373G>A/p.Gly125Ser in the *FKTN* coding for fukutin, a protein with multiple functions in the Golgi apparatus. The variant has previously been found in a patient with autosomal recessive Walker–Warburg syndrome (OMIM #613152, #253800 and #611588), together with a polyA signal deletion (12) but based on a MAF of 1.1–8.7% it has to be considered a polymorphism (rs34006675). NGS2 has no clinical sign of Walker–Warburg and therefore the variant is probably unrelated to the phenotype. Another known variant; c.5530C>A/p.Arg1844Ser was found heterozygous in the *LAMA2*, coding for the laminin alpha 2 protein, which is part of the basement membrane. Variants in *LAMA2* have been associated to congenital merosin-deficient muscle dystrophy (OMIM #607855) (13). With a MAF of 1.7%, this variant is a likely polymorphism (rs56173620). Finally, an unknown variant; c.1319T>C/p.Leu440Pro was found in *KBTBD13* coding for the 'Kelch repeat and BTB domain containing protein 13', which has overall regulatory cellular functions. Variants in this gene have been found in the autosomal dominant nemaline myopathy, type 6 NEM6 (OMIM #609273). Nemaline bodies are the key morphological feature of NEM muscle but have also been reported in *RYR1*-related congenital muscular dystrophies (14). However the actual patient has no detectable nemaline bodies in muscle biopsy.

Patient NGS3 showed a known polymorphism; c.8762A>G/p.His2921Arg in the large *DMD* coding for dystrophin, a cytoskeleton protein of muscle. Mutations are associated with Duchenne or Becker muscular dystrophies and the actual variant has earlier been reported in a Becker patient (15). NGS3 is female and could be a carrier of BMD. However, the variant has been observed at 1–3% frequency in >10,000 control samples (rs1800279). Another known variant; c.341C>G/p.Ala114Gly (rs143793528) was found in the *FKRP* coding for the fukutin-related protein, which has similar properties as fukutin in glycosylation and has been reported previously in severe congenital forms (16). NGS3 could be a carrier of this condition. Mutations in *FKRP* can give rise to recessive muscular dystrophies of variable onset and severity (OMIM #607155, #606612, #613153). A known variation was also identified in the promoter region of the *ATP2A2* (g.110719519C>G; rs201929640). Similar variants nearby, although not this particular one, have been related to lower mRNA expression (17). In our context, *ATP2A2* is of great interest because it encodes the intracellular calcium ion pump SERCA which, after a

muscle contraction, replenishes the intracellular calcium stores in the sarcoplasmic reticulum. To our knowledge, *ATP2A2* variants have not yet been associated to the MH phenotype. Finally, two unknown variants of predicted likely pathogenicity were identified; one in *COL6A1*; c.1906G>A/p.Asp636Asn and another in *TTN*; c.5589G>T/p.Arg1863Ser. *COL6A1* codes for the collagen 6-alpha-1-subunit, which maintains cell integrity in striated muscle tissue. Mutations in *COL6A1* are associated with Bethlem or Ulrich muscular dystrophies, whereas *TTN* codes for the huge protein titin, the main structural organizer of the Z-disk.

Discussion

Because about half of the MH patients do not show mutations in *RYR1* nor in *CACNA1S* the power of next-generation sequencing (NGS) technologies has prompted us to design a NGS-panel of genes coding for proteins involved in the Ca²⁺ metabolism or in other key functions of the striated muscle cell for a wider analytic approach.

Four of the five patients were found to carry *bona fide* mutations in *RYR1* indicating that *RYR1* is the main mutation target for MH. Of note is that patient NGS5 without any mutation in any gene studied showed weaker IVCT results and a lower Larach score than patients NGS2–4. In two patients, sequence variants in other genes were identified, which have previously been related to other muscle disorders. All but one patient carried likely benign polymorphisms in different genes with MAF <3%.

Our panel design focused on genes known to be related to muscle dysfunction. Such a preselection poses a problem of variant interpretation: because many muscle disorders result in calcium dysregulation and eventually anesthetic complications, it is hard to decide if these concomitant variants outside *RYR1* have an impact on the MH event. Recently, for example, MH-like reactions have been reported in glycosylation disorders (18). It is conceivable that some mutations require a certain constellation of background genetic variation in order to display an altered function. In fact, the few variants identified here by screening of only 64 genes are a small part of the 'background of the genome' on which individual mutations exert their effect.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

References

- Hopkins PM. Malignant hyperthermia: advances in clinical management and diagnosis. *Br J Anaesth* 2000; 85: 118–128.
- Brandon B, Muldoon S. Unexpected MH deaths without exposure to inhalation anesthetics in pediatric patients. *Pediatr Anesth* 2013; 23: 851–854.
- Robinson R, Carpenter D, Shaw MA, Halsall J, Hopkins P. Mutations in RYR1 in malignant hyperthermia and central core disease. *Hum Mutat* 2006; 27: 977–989.

NGS DNA analysis of Swedish MH patients

4. Carpenter D, Ringrose C, Leo V et al. The role of CACNA1S in predisposition to malignant hyperthermia. *BMC Med Genet* 2009; 10: 104.
5. Schiemann AH, Dirnholt EM, Pollock N, Stowell KM. Sequence capture and massively parallel sequencing to detect mutations associated with malignant hyperthermia. *Br J Anaesth* 2013; 1: 122–127.
6. Kim JH, Jarvik GP, Browning BL et al. Exome sequencing reveals novel rare variants in the ryanodine receptor and calcium channel genes in malignant hyperthermia families. *Anesthesiology* 2013; 119: 1054–1065.
7. Amburgey K, Bailey A, Hwang JH et al. Genotype-phenotype correlations in recessive RYR1-related myopathies. *Orphanet J Rare Dis* 2013; 8: 117.
8. Carpenter D, Robinson RL, Quinnell RJ et al. Genetic variation in RYR1 and malignant hyperthermia phenotypes. *Br J Anaesth* 2009; 103: 538–548.
9. Brandom BW, Bina S, Wong CA et al. Ryanodine receptor type 1 gene variants in the malignant hyperthermia-susceptible population of the United States. *Anesth Analg* 2013; 116: 1078–1086.
10. Avila G, Dirksen RT. Functional effects of central core disease mutations in the cytoplasmic region of the skeletal muscle ryanodine receptor. *J Gen Physiol* 2001; 118: 277–290.
11. Tong J, McCarthy TV, MacLennan DH. Measurement of resting cytosolic Ca^{2+} concentrations and Ca^{2+} store size in HEK-293 cells transfected with malignant hyperthermia or central core disease mutant Ca^{2+} release channels. *J Biol Chem* 1999; 274: 693–702.
12. Cotarelo RP, Valero MC, Prados B et al. Two new patients bearing mutations in the fukutin gene confirm the relevance of this gene in Walker-Warburg syndrome. *Clin Genet* 2008; 73: 139–145.
13. Guicheney P, Vignier N, Zhang X et al. PCR based mutation screening of the laminin alpha2 chain gene (LAMA2): application to prenatal diagnosis and search for founder effects in congenital muscular dystrophy. *J Med Genet* 1998; 35: 211–217.
14. Kondo E, Nishimura T, Kosho T et al. Recessive RYR1 mutations in a patient with severe congenital nemaline myopathy with ophthalmoplegia identified through massively parallel sequencing. *Am J Med Genet A* 2012; 4: 772–778.
15. Lenk U, Hanke R, Speer A. Carrier detection in DMD families with point mutations, using PCR-SSCP and direct sequencing. *Neuromuscul Disord* 1994; 4: 411–418.
16. Brockington M, Blake DJ, Pandini P et al. Mutations in the Fukutin-related protein gene (FKRP) cause a form of Congenital Muscular Dystrophy with secondary laminin alpha2 deficiency and abnormal glycosylation of alpha-dystroglycan. *Am J Hum Genet* 2001; 69: 1198–1209.
17. Korosec B, Glavac D, Metka V, Metka RG. ATP2A3 gene is involved in cancer susceptibility. *Cancer Genet Cytogenet* 2009; 188: 88–94.
18. Tegtmeyer LC, Rust S, van Scherpenzeel M et al. Multiple phenotypes in phosphoglucomutase 1 deficiency. *N Engl J Med* 2014; 6: 533–542.
19. Larach MG, Localio AR, Allen GC et al. A clinical grading scale to predict malignant hyperthermia susceptibility. *Anesthesiology* 1994; 80: 771–779.

Supplementary table 1

Supplementary Table. Genes included in the NGS analysis: Gene name, chromosomal localization, protein name and function and associated genetic disorders (if any). Column 4 indicates the coverage (in %) of the targeted regions (exons plus 20 bp of flanking intronic sequences).

Gene name	Chr.	Start (GRCh37/hg19) end (GRCh37/hg19)	Percentage of bases of the gene covered with minimum reads of >20, average of all five patients	Protein name	Function	Protein expressed in	Known associated muscle pathology (or other disease)
<i>ACTA1</i>	1	229566992 - 229569843	100	Actin	Contraction	Striated muscle cell	Congenital myopathy with cores Nemaline myopathy 3 with excess thin filaments fibre type 1
<i>ASPH</i>	8	62413115 - 62627199	97,3	Set of proteins in complexes with Calsequestrin, Triadin and RYR1	Regulates Ca release from sarcoplasmic reticulum	Striated muscle cell	No defined disease entity
<i>ATP2A1</i>	16	28889809 - 28915830	100	SERCA	Intracellular Ca pump which pumps Ca from sarcoplasm to sarcoplasmic reticulum	Striated muscle cell	Brody's disease
<i>ATP2A2</i>	12	28889809 - 110788998	100	SERCA	Intracellular Ca pump which pumps Ca from sarcoplasm to sarcoplasmic reticulum	Striated muscle cell	Darier-White disease (skin)
<i>ATP2A3</i>	17	28889809 - 3867758	98,8	SERCA	Intracellular Ca pump which pumps Ca from sarcoplasm to sarcoplasmic reticulum	Striated muscle cell	No defined disease entity
<i>BIN1</i>	2	28889809 - 127864903	100	Unknown	Participates in apoptosis	Probably striated muscle cell	Centronuclear myopathy type 2
<i>CACNA1S</i>	1	28889809 - 201081694	100	Dihydropyridine receptor	Ca channel	Striated muscle cell	Malignant hyperthermia susceptibility Hypokalemic periodic paralysis type 1 Thyreotoxic periodic paralysis
<i>CALM1</i>	14	28889809 - 90874619	98,8	Calmodulin	Ca binding protein	Striated muscle cell	Cardiac ventricular tachycardia
<i>CALR</i>	19	28889809 - 13055304	100	Calreticulin	Ca binding protein	Striated muscle cell	No defined disease entity
<i>CASQ1</i>	1	28889809 - 160171676	100	Calsequestrin	Ca binding protein	Striated muscle cell	No defined disease entity
<i>CD38</i>	4	28889809 - 15850706	100	Multifunctional ectoenzyme	Participates in Ca signalling	Striated muscle cell and most cells	No defined disease entity
<i>CFL2</i>	14	28889809 - 35184029	100	Cofilin 2	Regulation of actin-filament dynamics	Striated muscle cell	Nemaline myopathy 7
<i>CLCN1</i>	7	28889809 - 143049097	100	Chloride channel 1	Cl channel	Striated muscle cell	Myotonia congenita
<i>CNTN1</i>	12	28889809 - 41466214	100	Contactin 1	Cell adhesion in Neuronal membrane	Striated muscle cell and most cells	Compton North congenital myopathy
<i>COL6A1</i>	21	28889809 - 47424963	100	Collagen type 4 alpha 1	Maintaining integrity in tissue	Striated muscle cell and most cells	Bethlem myopathy Ulrich muscular dystrophy
<i>COL6A2</i>	21	28889809 - 47551763	100	Collagen type 4 alpha 2	Maintaining integrity in tissue	Striated muscle cell and most cells	Bethlem myopathy Ulrich muscular dystrophy
<i>COL6A3</i>	2	28889809 - 238322850	100	Collagen type 4 alpha 3	Maintaining integrity in tissue	Striated muscle cell and most cells	Bethlem myopathy Ulrich muscular dystrophy

<i>CPT2</i>	1	28889809 53679869	100	Carnithine palmitoyl-transferase 2	Mitochondrial metabolism	Striated muscle cell and most cells	Myopathy due to CPT2 deficiency adult/infantile/neonatal
<i>DMD</i>	X	28889809 33357726	100	Dystrophin	Cytoskeleton Function in muscle cells	Striated muscle cell	Duchenne muscle dystrophy Becker muscle dystrophy
<i>DNM2</i>	19	28889809 10942586	99,9	Dynamin 2	Multiple functions Binds to actin and acts as cytoskeleton	Striated muscle cell and most cells	Centronuclear myopathy Charcot-Marie-Tooth disease
<i>FBN3</i>	19	28889809 8212385	100	Fibrillin 3	Extracellular microfibrils	Unknown Fetal tissue Connective tissue	Probably involved in Weill-Marchesani syndrome
<i>FHL1</i>	X	28889809 135293518	100	LIM protein SLIMMER	Unknown	Striated muscle cell and most cells	X linked myopathies; Emery Dreyfuss muscular dystrophy type 6 Scapulo peronal muscular dystrophy
<i>FKBP1A</i>	20	28889809 1373816	95,6	FK506 binding protein	Coordinates RYR1 Complex formation During Ca release	Striated muscle cell	No defined disease entity
<i>FKRP</i>	19	28889809 47261832	100	Fukutin related protein	Multiple functions in Golgi apparatus	Striated muscle cell and most cells	Muscular dystrophy-dystroglycanopathy Limb girdle dystrophy type 2I
<i>FKTN</i>	9	28889809 10840399	100	Fukutin	Multiple functions in Golgi apparatus	Striated muscle cell and most cells	Muscular dystrophy-dystroglycanopathy Walker Warburg syndrome Fukuyama muscle dystrophy Limb girdle dystrophy type 2H
<i>HOMER1</i>	5	28889809 78809700	100	Homer homolog 1	Dendritic protein, Ca balance in neuronal cells	Neurons and probably striated muscle cells	No defined disease entity
<i>HOMER2</i>	15	28889809 83621476	97,3	Homer homolog 2	Dendritic protein, Ca balance in neuronal cells	Neurons and probably striated muscle cells	No defined disease entity
<i>HOMER3</i>	19	28889809 19052041	99,9	Homer homolog 3	Dendritic protein, Ca balance in neuronal cells	Neurons and probably striated muscle cells	No defined disease entity
<i>HSPG2</i>	1	28889809 22263750	99,3	Heparan proteoglykan 2	Multiple functions, interacts with several signaling intracellular proteins	Striated muscle cell and most cells	Schwarz Jampel syndrome 1 Dyssegmental dysplasia
<i>IL1R1</i>	2	28889809 102796334	100	Interleukine 1 receptor 1	Immune system	Striated muscle cell and most cells	Dysregulation of IL1R seen in several diseases
<i>IL6R</i>	1	28889809 154441926	99,8	Interleukine 6 receptor	Immune system	Striated muscle cell and most cells	Dysregulation of IL6R seen in several diseases
<i>ITGA7</i>	12	28889809 56106089	100	Integrin alpha 7	Multiple functions Binds to multiple intracellular proteins	Striated muscle cell and most cells	Congenital muscular dystrophy due to integrin alpha 7 deficiency
<i>ITPR1</i>	3	28889809 4889524	100	Inositol 1,4,5 tri- phosphate receptor 1	Ca release from endoplasmic Reticulum	Striated muscle cell and most cells	Spinocerebellar ataxia type 15
<i>ITPR2</i>	12	28889809 26986131	100	Inositol 1,4,5 tri- phosphate receptor 2	Unknown	Unknown	No defined disease entity
<i>ITPR3</i>	6	28889809 33664351	100	Inositol 1,4,5 tri- phosphate receptor 2	Unknown	Unknown	No defined disease entity
<i>JPH1</i>	8	28889809 75233562	100	Junctophilin 1	Mediates interaction between ion channels in endoplasmatic reticulum membrane	Striated muscle cell and most cells	No defined disease entity
<i>JPH3</i>	16	28889809 87721762	100	Junctophilin 3	Mediates interaction between ion channels in endoplasmatic reticulum membrane and cell surface	Striated muscle cell and most cells	Huntington like disease 2
<i>JPH4</i>	14	28889809 24048009	99,8	Junctophilin 4	Mediates interaction Between ion channels in endoplasmatic reticulum membrane and cell surface	Striated muscle cell and most cells	No defined disease entity

<i>JSRP1</i>	19	28889809 - 2256422	58,7	Junctional sarcoplasmic reticulum protein 1	Interacts in electro-contraction coupling with dihydropyridine receptor and calsequestrin	Striated muscle cell	No defined disease entity
<i>KBTBD13</i>	15	28889809 - 65372276	100	Kelch repeat and BTB domain containing protein 13	Multiple regulatory functions in the cell	Striated muscle cell and most cells	Nemaline myopathy type 6
<i>KCNJ2</i>	17	28889809 - 68176185	100	Potassium inwardly rectifying channel	K channel	Striated muscle cell and cardiac muscle cells	Andersen syndrome Familial atrial fibrillation Short QT syndrome
<i>LAMA2</i>	6	28889809 - 129837711	99,6	Laminin alpha 2	Component of basement membrane	Striated muscle cell and most cells	Congenital merosin deficient muscle dystrophy Congenital laminin alpha 2 deficient muscle dystrophy
<i>LARGE</i>	22	28889809 - 34316416	100	LARGE glycotransferase	Involved in synthesis of sugar chains	Striated muscle cell and most cells	Muscular dystrophy-dystroglycanopathy MDC1D
<i>MEGF10</i>	5	28889809 - 126796914	100	Multiple EGF factor domains 10	Growth factor	Striated muscle cell and most cells	Early onset myopathy, arreflexia, respiratory distress and dysphagia
<i>MTM1</i>	X	28889809 - 149841616	100	Myotubularin 1	Muscle cell differentiation	Striated muscle cell	X linked myotubular myopathy
<i>MYH7</i>	14	28889809 - 23903496	100	Myosin heavy chain	Participates in contractile apparatus	Striated muscle cell and cardiac muscle cells	Familial hypertrophic cardiomyopathy Familial dilated cardiomyopathy Lainig distal myopathy Myosin storage myopathy Scapulo peroneal syndrome myopathic type
<i>NEB</i>	2	28889809 - 152591001	100	Nebulin	Cytoskeletal matrix	Striated muscle cell	Nemaline myopathy type 2
<i>ORAI1</i>	12	28889809 - 122079946	98,5	Calcium release activated calcium modulator	Restores Ca balance in T lymphocytes	T lymphocytes and probably striated muscle cells	Immune dysfunction with T cell inactivation
<i>POMGN1</i>	1	46654353 - 46685977	100	Protein O-linked mannose beta 1,2 N-acetylglucosaminyltransferase	Processes in the Golgi apparatus	Striated muscle cell and most cells	Muscular dystrophy-dystroglycanopathy type A3, B3, C3
<i>POMT1</i>	9	134378289 - 134399193	100	Protein O-mannosyl transferase 1	Processes in endoplasmatic reticulum	Striated muscle cell and most cells	Muscular dystrophy-dystroglycanopathy type A3, B3, C3
<i>POMT2</i>	14	77741299 - 77787225	100	Protein O-mannosyl transferase 2	Processes in endoplasmatic reticulum	Striated muscle cell and most cells	Muscular dystrophy-dystroglycanopathy type A3, B3, C3 Walker Warburg syndrome
<i>RYR1</i>	19	38924340 - 39078204	99,3	Ryanodine receptor 1	Ca channel	Striated muscle cell	Malignant hyperthermia susceptibility Central Core disease Congenital myopathies
<i>S100A13</i>	1	153591275 - 153606568	100	Calcium binding protein S100	Participates in Ca regulation in the cell	Striated muscle cell and most cells	No defined disease entity
<i>SCN4A</i>	17	62015914 - 62050278	100	Sodium channel alfa subunit	Na channel	Striated muscle cell	Myotonia congenita Hyperkaleamic paralysis type 2 Hypokaleamic paralysis type 2 Paramyotonia congenita Acetazolamide responsive myasthenic syndrome
<i>SEPN1</i>	1	26126667 - 26144713	90,3	Selenoprotein 1	Participates in oxidation/reduction processes with cell protecting function Binds ca	Striated muscle cell and most cells	Multiminicore disease Rigid spine muscular dystrophy 1 Congenital myopathy with fibre type disproportion
<i>STIM1</i>	11	3876933 - 4114440	100	Transmembrane protein that mediates Ca flux	Ca channel	Striated muscle cell and most cells	Associated to certain tumors Associated to distinct neuromuscular phenotypes, tubular aggregate myopathy, TAM
<i>SYPL2</i>	1	110009100 - 110024764	89,0	Synaptophysin	Unknown	Unknown	No defined disease entity
<i>TNNT1</i>	19	55644161 - 55660606	94,3	Troponin T skeletal muscle	Regulates contraction in skeletal muscle cell together with RYR1 and Ca	Striated muscle cell	Nemaline myopathy Amish type
<i>TPM2</i>	9	35681990 - 35690053	100	Tropomyosin 2	Involved in contractile system by binding to actin in muscle cell	Striated muscle cell	Arthrogryposis multiplex congenita 1, 2B Nemaline myopathy
<i>TPM3</i>	1	154127780 - 154164609	100	Tropomyosin 3	Involved in contractile system by binding to actin in muscle cell	Striated muscle cell	CAP myopathy Congenital myopathy with fibre type disproportion Nemaline myopathy 1
<i>TRDN</i>	6	123537484 - 123958238	100	Triadin	Participates with RYR1 in Ca handling in electro chemical coupling process	Striated muscle cell	Polymorphisms can be markers for IgA nephritis
<i>TRPC1</i>	3	142443266 - 142526730	100	Nonselective membrane channel permeable to Ca and other cations	Unspecified Ca transmembrane transport	Striated muscle cell	No defined disease entity
<i>TRPC3</i>	4	122800183 - 122872909	100	Nonselective membrane channel permeable to Ca and other cations	Unspecified Ca transmembrane transport	Striated muscle cell	No defined disease entity
<i>TTN</i>	2	179390716 - 179672150	100	Titin	Adhesion template for contractile machinery in muscle cells	Striated muscle cell	Associated to a wide range of neuromuscular phenotypes, hereditary myopathy with early respiratory failure, HMERF, early-onset myopathies with cores, cardiac involvement Limb girdle muscular dystrophy 2J Subgroups of centronuclear myopathy

Supplementary table 2

Supplementary material. Clinical histories of the patients. Age is at the time of the malignant hyperthermia crisis.

NGS1

A girl who was floppy infant at birth. Never been able to walk. Central core disease was diagnosed at age of two. She had major scoliosis surgery several times performed with trigger free anaesthesia without complications. She is now grown up with severe myopathic symptoms, wheelchair-bound and she suffers respiratory impairment. She is not IVC tested. Histopathological investigation revealed central cores in almost all muscle fibers. There is no other family member with muscle symptoms.

NGS2

Forty-six year old previously healthy and physically fit male practicing endurance sports. Surgery: osteosynthesis of tibial fracture. Anaesthesia was induced with propofol, fentanyl and suxamethonium, which induced masseter muscle spasm. In spite of this anaesthesia was continued with isoflurane. Generalised muscle rigidity, hypercapnia, fever up to 38,1 C and acidosis developed within a short time. Postoperatively the patient had cola coloured urine and a more than 100 fold increase in creatine kinase as a sign of rhabdomyolysis. Histopathological investigation revealed slight variation in fibre diameter, mainly in type 1 fibres. The findings were very discrete and not consistent with any specific diagnosis. No sequelae after the MH reaction remained.

NGS3

Seven year old girl. Surgery: arthroscopy due to knee pain. Anaesthesia was induced and maintained with isoflurane and nitrous oxide. Thirty minutes after anesthesia induction, she developed tachycardia of 180, masseter spasm, hyperventilation, flushed skin and the temperature increased to 38,9 C in 10 minutes. Blood gas analysis revealed metabolic acidosis. Surgery and anaesthesia was stopped and she was hyperventilated with 100% oxygene and cooled with ice and iv fluids. Dantrolene was never given. Postoperatively she developed a 200 fold increase in creatine kinase. Histopathological investigation revealed abundantly occurring cores in type 1 fibres. Findings congruent with Central core disease. She is now a grown up physically healthy female without muscle symptoms or disability.

NGS4

Twenty-seven year old male with ulcerous colitis. Surgery: cholecotomy. Anaesthesia was induced with propofol, fentanyl, suxamethonium and maintained with enflurane and nitrous oxide. During anaesthesia he developed hypercapnia, fever and tachychardia. He responded promptly on Dantrolene treatment. He did not develop any muscular symptoms. Histopathological investigation revealed discrete signs of myopathy; degeneration and regeneration processes in some muscle fibres, but no diagnosis could be established. No sequelae after the MH reaction remained.

NGS5

Forty year old man with hyperparathyroidism and insulin dependent diabetes. Surgery: parathyroidectomy. Anaesthesia was induced with propofol, fentanyl and maintained with sevoflurane. He developed hypercapnia and fever up to 40 C. Postoperatively myoglobine was more than doubled. Histopathological investigation was normal. No sequelae after the MH reaction remained.

Anlage 3

Anna-Lena Semmler, Sabrina Sacconi, **J Elisa Bach**, Claus Liebe, Jan Bürmann, Rudolf A Kley, Andreas Ferbert, Roland Anderheiden, Peter Van den Bergh, Jean-Jacques Martin, Peter De Jonghe, Eva Neuen-Jacob, Oliver Müller, Marcus Deschauer, Markus Bergmann, J Michael Schröder, Matthias Vorgerd, Jörg B Schulz, Joachim Weis, Wolfram Kress and Kristl G Claeys

Unusual multisystemic involvement and a novel BAG3 mutation revealed by NGS screening in a large cohort of myofibrillar myopathies

Orphanet J Rare Dis. 2014; 9(1):121. doi: 10.1186/s13023-014-0121-9

Semmler et al. *Orphanet Journal of Rare Diseases* 2014, **9**:121
<http://www.orphandis.com/content/9/1/121>



RESEARCH

Open Access

Unusual multisystemic involvement and a novel *BAG3* mutation revealed by NGS screening in a large cohort of myofibrillar myopathies

Anna-Lena Semmler^{1,2}, Sabrina Sacconi³, J Elisa Bach⁴, Claus Liebe^{1,2}, Jan Bürmann⁵, Rudolf A Kley⁶, Andreas Ferbert⁷, Roland Anderheiden⁸, Peter Van den Bergh⁹, Jean-Jacques Martin¹⁰, Peter De Jonghe^{10,11,12}, Eva Neuen-Jacob¹³, Oliver Müller^{14,15}, Marcus Deschauer¹⁶, Markus Bergmann¹⁷, J Michael Schröder², Matthias Vorgerd⁶, Jörg B Schulz^{1,18}, Joachim Weis^{2,18}, Wolfram Kress⁴ and Kristl G Claeys^{1,2,18*}

Abstract

Background: Myofibrillar myopathies (MFM) are a group of phenotypically and genetically heterogeneous neuromuscular disorders, which are characterized by protein aggregations in muscle fibres and can be associated with multisystemic involvement.

Methods: We screened a large cohort of 38 index patients with MFM for mutations in the nine thus far known causative genes using Sanger and next generation sequencing (NGS). We studied the clinical and histopathological characteristics in 38 index patients and five additional relatives (n = 43) and particularly focused on the associated multisystemic symptoms.

Results: We identified 14 heterozygous mutations (diagnostic yield of 37%), among them the novel p.Pro209Gln mutation in the *BAG3* gene, which was associated with onset in adulthood, a mild phenotype and an axonal sensorimotor polyneuropathy, in the absence of giant axons at the nerve biopsy. We revealed several novel clinical phenotypes and unusual multisystemic presentations with previously described mutations: hearing impairment with a *FLNC* mutation, dysphonia with a mutation in *DES* and the first patient with a *FLNC* mutation presenting respiratory insufficiency as the initial symptom. Moreover, we described for the first time respiratory insufficiency occurring in a patient with the p.Gly154Ser mutation in *CRYAB*. Interestingly, we detected a polyneuropathy in 28% of the MFM patients, including a *BAG3* and a *MYOT* case, and hearing impairment in 13%, including one patient with a *FLNC* mutation and two with mutations in the *DES* gene. In four index patients with a mutation in one of the MFM genes, typical histological findings were only identified at the ultrastructural level (29%).

Conclusions: We conclude that extraskelletal symptoms frequently occur in MFM, particularly cardiac and respiratory involvement, polyneuropathy and/or deafness. *BAG3* mutations should be considered even in cases with a mild phenotype or an adult onset. We identified a genetic defect in one of the known genes in less than half of the MFM patients, indicating that more causative genes are still to be found. Next generation sequencing techniques should be helpful in achieving this aim.

Keywords: MFM, Next generation sequencing, bcl-2 associated athanogene protein 3, Protein aggregation, Hearing impairment, Polyneuropathy

* Correspondence: kclaeys@ukaachende

¹Department of Neurology, RWTH Aachen University, Aachen, Germany

²Institute of Neuropathology, RWTH Aachen University, Aachen, Germany

Full list of author information is available at the end of the article



© 2014 Semmler et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Myofibrillar myopathies (MFM) are a group of phenotypically and genetically heterogeneous neuromuscular disorders. The morphological hallmark of MFM is the presence of protein aggregations in muscle fibres, a focal myofibrillar disorganisation starting at the Z-disk and an ectopic expression of proteins, such as desmin [1-6].

Muscle biopsies of MFM patients reveal protein aggregates that are dark blue or purple on the modified Gomori trichrome (mGT) and pink on the haematoxylin and eosin (HE) stains. The aggregations are lacking oxidative enzyme activity on the nicotinamide adenine dinucleotide tetrazolium reductase (NADH-TR) stains. Other light microscopic findings in MFM are rimmed and non-rimmed vacuoles and sometimes cytoplasmic bodies. At the ultrastructural level, the combination of Z-disk streaming and distinct types of protein accumulations is characteristic for MFM and can even give hints towards the mutated gene [7].

To date, mutations in nine genes are known to cause MFM: desmin (*DES*) [8], α B-crystallin (*CRYAB*) [9], myotilin (*MYOT*, also *TTID*) [10], filamin C (*FLNC*) [11], Z-band alternatively spliced PDZ motif-containing protein (*ZASP*, also *LDB3*) [12], four and a half LIM domain protein 1 (*FHL1*) [13], bcl-2 associated athanogene protein 3 (*BAG3*) [14], dnaJ homolog subfamily B member 6 (*DNAJB6*) [15] and titin (*TTN*) [16]. Mutations in exon 343 of the A-band region of *TTN* were discovered to cause MFM only very recently [16]. Inheritance in MFM is autosomal dominant in most patients, autosomal recessive in rare cases and X-linked in patients with mutations in *FHL1*. A broad inter- and intrafamilial variability of phenotypes can be seen in MFM. The age at onset of MFM depends on the underlying gene, mutation and inheritance pattern, but usually the disease starts in adulthood [17,18]. Nevertheless also childhood onsets have been described [8,13,14,19-21]. The initial symptom in MFM most frequently is skeletal muscle weakness, other muscular symptoms such as atrophy, hypertrophy, muscle rigidity, contractures, pain or cramps can be present [13,22-26]. Also, the respiratory muscles can be affected in MFM, resulting in a restrictive respiratory insufficiency, which has been reported for most MFM subtypes [4,8,11,14,27-31].

Furthermore, MFM can be associated with distinct multisystemic symptoms, the most frequent being cardiac involvement, ranging from conduction defects and arrhythmias to cardiomyopathies and sudden cardiac death [4,9,15,26,29,30,32-38]. The smooth muscles can also be affected in MFM, presenting as chronic diarrhoea or enteric hypomotility [37,39-42], or contributing to bulbar symptoms like swallowing difficulties [27,28,31,33,37,43-45]. Further symptoms associated with MFM exceed the muscles, such as polyneuropathy (PNP) [10-12,14,46],

cataracts [9,31,33], gynaecomastia [29,47], and hearing impairment [39,40,48,49].

Serum creatine kinase levels (CK) in MFM patients are normal or only slightly elevated. Electromyography (EMG) reveals myopathic or mixed patterns, but also (pseudo-) myotonic discharges can appear [17,50]. Magnetic resonance imaging (MRI) of the muscles plays an increasing role in the differential diagnosis of muscular disorders and of MFM in particular [51,52].

In the present study, we screened a large cohort of 38 unrelated index patients with MFM for mutations in the nine causative genes, employing next generation sequencing (NGS) and traditional Sanger sequencing. We identified a mutation in 14 index patients, including one novel *BAG3* mutation. We performed a detailed clinicopathological characterisation in the index patients and five affected relatives and particularly focused on the associated multisystemic symptoms. We highlight new phenotypical findings in MFM and correlate our data with the literature.

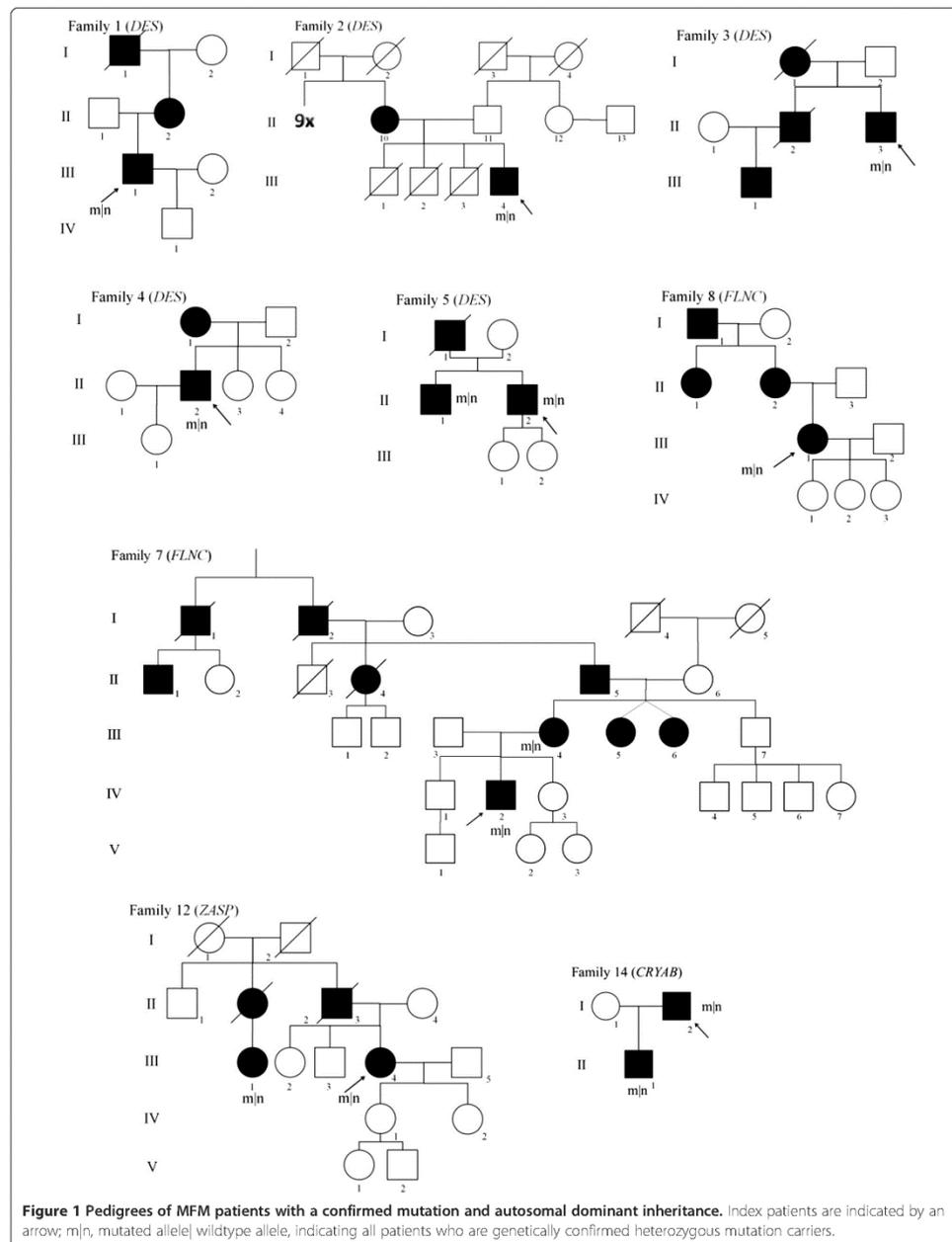
Patients and methods

Patients

In this study, we included 43 patients with MFM belonging to 38 unrelated families (Figure 1). We selected the MFM cases from the muscle and nerve biopsy archive at the Institute of Neuropathology, and from the Neuromuscular Clinic at the Department of Neurology, at the RWTH Aachen University Hospital (Aachen, Germany). The biopsy of patient F10.1 was also included in the study of Joshi et al. [53]. Additionally, eight index patients/biopsies with MFM were provided by referring centres. The study was performed according to the Declaration of Helsinki and was approved by the ethical committee of the RWTH Aachen University. Written informed consent was obtained from all patients.

Clinical and paraclinical evaluation

In all 43 patients, we performed a detailed neurological examination, including hearing and swallowing tests, and a blood analysis of the creatine kinase (CK) level. Nerve conduction velocity (NCV) and electromyography (EMG) studies, heart (electrocardiogram (ECG), echocardiography, long-term ECG) and lung examinations (spirometry/bodyplethysmography) were performed in the majority of patients. In addition, an ophthalmological exam was undertaken in patients with complaints of ocular involvement. An EMG was done in all patients with an identified gene defect, except for patients F5.II.2 and F14.1, a diagnostic muscle MRI of the lower extremities in four patients (F1.III.1, F2.III.4, F10.1, F12.III.3) and a whole-body muscle MRI in one (F7.III.4) (1.5 Tesla, Philips, Intera, Best – the Netherlands).



Light microscopy (LM) and electron microscopy (EM) of muscle and nerve biopsies

Open muscle biopsies had been obtained from the 38 index patients for diagnostic purposes after written informed consent. The muscle tissue sample was divided into unfixed (cryostat), formaldehyde-fixed (paraffin) and glutaraldehyde-fixed (EM) sections and further prepared according to routine protocols. Standard enzyme histochemical and immunohistochemical stains were applied [54]. For electron microscopy (EM) of muscle tissue, which was performed in 31 out of 38 probands, we used a CM10 transmission electron microscope (Philips, Amsterdam, the Netherlands). In addition, seven index patients had simultaneously a diagnostic sural nerve biopsy taken, among them patient F9.1 and F13.1 with a confirmed mutation and a polyneuropathy. The nerve biopsies were processed following standard procedures [55].

Genetic analyses

Peripheral blood samples and a written informed consent for genetic analysis were available from the 38 index patients, as well as from the affected mother of patient F7.IV.2, the affected brother of patient F5.II.2, the cousin of patient F12.III.4 and the son of patient F14.1. Genomic DNA was extracted using standard procedures. We used two different techniques for genetic analyses in our cohort: Sanger sequencing [56] and next generation sequencing (NGS) [57].

Sanger sequencing of the MFM genes

We had started screening a group of 21 MFM index patients with conventional Sanger sequencing before the NGS panel for MFM was established. We analysed *DES*, *CRYAB*, *MYOT* (exon 2 and 9), *FLNC* (exon 48), *ZASP* (exon 4–10), *FHL1*, *BAG3* and *DNAJB6*. We performed PCR using standard protocols and used the BigDye Terminator v1.1 mix (Applied Biosystems) for cycle sequencing reactions according to the instructions in the kit. Sequence data were generated with an ABI 3130xl capillary sequencer (Applied Biosystems, Carlsbad, CA, USA) using standard protocols and analysed by GenSearch v.4.0.8 (PhenoSystems, Lillois, Belgium).

NGS panel diagnostics in MFM

In a NGS approach, the coding exons of all causative genes with the exception of the most recently identified *TTN* gene were sequenced in the additional 17 index patients included in our study: *DES* (exon 1–9, NM_001927.3), *MYOT* (exon 2–10, NM_006790.2), *ZASP* (exons 5–9 of NM_001080116.1 and exons 4, 7, 8 and 10 of NM_007078.2), *FHL1* (exon 3–8, NM_001159702.2), *BAG3* (exon 1–4, NM_004281.3), *DNAJB6* (exon 2–8,

NM_005494.2), *CRYAB* (exon 1–3, NM_001885.1) and *FLNC* (exon 1–48, NM_001458.3). Furthermore, in 10 index patients of our first cohort in which Sanger sequencing was initially performed (see above) and in which no mutations were identified, we subsequently analysed the complete *FLNC* gene (exon 1–48) applying NGS. Target enrichment was performed with the Access Array System (Fluidigm Corporation, South San Francisco, USA) followed by emulsion PCR and sequence analyses on a GS Junior using kits and instructions by Fluidigm and 454 Life Sciences (Branford, USA). Sequence data were analysed with GenSearch NGS v.1.4 (PhenoSystems, Lillois, Belgium) automatically aligning data to the reference sequences mentioned above. Variants within exons and 20 bp of the flanking intronic sequences were detected with VarScan (<http://varscan.sourceforge.net/>) using following filters: coverage >15, frequency >20% for heterozygous or >75% for homo- and hemizygous variants. Candidate variants were verified by Sanger sequencing on an ABI 3130xl automated capillary sequencer using standard protocols, and their effect predicted with bioinformatic tools like PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/index.shtml>), SIFT (<http://sift.jevl.org/>) and MutationTaster (<http://www.mutationtaster.org/>) provided by Alamut (Interactive Biosoftware, Rouen, France).

Sanger sequencing of exon 343 of *TTN*

Finally, we performed Sanger sequencing to analyse exon 343 of *TTN* in 25 of our index patients, in whom an underlying gene defect had thus far not been identified (21 patients) or only an unknown variant was revealed (4 patients). PCR, sequencing reaction and data analysis of exon 343 of the *TTN* gene (NM_001267550.1) have been performed using standard protocols as described above.

Results

Genetic results in our MFM cohort (n = 43): known and novel pathogenic mutations

In 14 index patients, we identified a heterozygous pathogenic mutation in one of the nine genes causing MFM (Table 1), corresponding to a diagnostic yield of 37% in our MFM cohort. In ten patients we found the underlying gene defect using the initial Sanger sequencing study and in four patients by directly employing NGS panel diagnostics. Nine out of the 14 index patients had a positive family history corresponding with an autosomal dominant inheritance (for pedigrees see Figure 1), whereas the others occurred sporadically (Patient F6.1 with a *DES* mutation, Patient F13.1 with a mutation in *BAG3* and all three patients, F9.1, F10.1 and F11.1, with a *MYOT* mutation, Table 1). We identified six heterozygous mutations in *DES*, four of these were the Arg350Pro

Table 1 Clinical characteristics of myofibrillar myopathy patients with an identified mutation

Pat.	MF	FH	Gene	Exon	Mutation	AAO	AAE	First symptom	Weakness					Multisystemic symptoms				CK	
									UL prox. dist.	LL prox. dist.	axial	scapular winging	dys-plagia	dys-phonias	respiratory involvement	cardiac involvement	PNP		hearing imp.
F1.III.1	M	+	DES	6	Arg350Pro	39	47	weakness LL	+	+	-	-	-	-	-	-	-	-	1031
F2.III.4	M	+	DES	6	Arg350Pro	42	74	weakness LL	-	+	+	-	-	-	-	-	-	-	2690
F3.II.3	M	+	DES	6	Arg350Pro	48	59	dyspnoea	+	+	+	-	-	-	VC 45%NV	TAA intermittent, LV hypertrophy	-	+	2044
F4.II.2	M	+	DES	6	Arg350Pro	36	42	weakness LL	+	+	-	-	+	-	-	-	-	-	1159
F5.II.2	M	+	DES	1	Ser2Ile	60	72	syncope	+	+	+	-	-	-	-	3°AVB, PM	-	-	nl
F6.1	F	-	DES	3	Glu245Asp	51	62	weakness LL	-	+	+	-	-	-	-	DCM, AF, 3°AVB, PM, TI	-	-	nl
F7.IV.2	M	+	FLNC	18	Val930_Thr930del	28	40	weakness LL	-	+	-	-	+	-	-	-	-	-	629
F7.III.4	F	+	FLNC	18	Val930_Thr930del	53	60	weakness LL	-	+	-	-	-	-	-	-	-	-	388
F8.III.3	F	+	FLNC	48	Trp2710X	53	59	dyspnoea	+	+	-	+	-	-	VC 35%NV	-	-	-	500
F9.1	M	-	MYOT	2	Ser60Phe	67	71	dyspnoea	-	-	+	-	-	-	-	LBBB, HF, AF, 3°AVB intermittent, PM	ad/s/m	-	206
F10.1	M	-	MYOT	2	Ser60Phe	63	67	myalgia LL	-	-	-	-	-	-	-	-	-	-	589
F11.1	M	-	MYOT	2	Ser65Phe	55	60	weakness LL	+	+	+	-	-	-	mild	AF, PM, ICD, DCM, HTX,	-	-	608
F12.III.3	F	+	ZASP	6	Ala165Val	46	57	walking difficulties	-	+	-	-	-	-	-	-	-	-	182
F13.1	M	-	BAG3	3	Pro209Gln	34	43	weakness LL	+	+	+	-	-	-	-	-	a/s/m	-	1050
F14.1	M	+	CRYAB	3	Gly154Ser	69	69	rhabdomyolysis	-	-	-	-	-	-	VC 65%	-	-	-	2000

MF, male/female; FH, family history; AAO, age at onset; AAE, age at examination; CK, level of creatine kinase in serum in U/L (normal value < 174); UL, upper limb; LL, lower limb; prox, proximal; dist, distal; PNP, polyneuropathy; Hearing imp., hearing impairment; VC, vital capacity; NC, nocturnal ventilation; TAA, tachyarrhythmia absoluta; LV, left ventricle; 3° AVB, complete atrioventricular conduction block; nl, normal; DCM, dilated cardiomyopathy; AF, atrial fibrillation; PM, pace maker; TI, tricuspid insufficiency; LBBB, left bundle branch block; HF, heart failure; ad/s/m, axonal demyelinating/sensorimotor; ICD, implantable cardioverter defibrillator; HTX, heart transplantation; a/s/m, axonal/sensorimotor; for abbreviations of genes see text.

mutation, two different mutations in *FLNC*, and three *MYOT* mutations, including two times the Ser60Phe mutation in exon 2. One mutation each was found in *ZASP*, *CRYAB* and in *BAG3* (Table 1, Figure 1). The mutation in *BAG3* represents a novel mutation, whereas the other mutations have been described previously.

The c.626C>A, p.Pro209Gln mutation in exon 3 of *BAG3* (Patient F13.1, Table 1) is so far not known in the databases of the NHLBI exome sequencing project (ESP) or the 1000 genome project and was predicted to be disease causing by all used prediction programmes. Further arguments for pathogenicity came from familial segregation analysis, in which both healthy parents did not harbour the mutation, indicating a *de novo* mutation.

Genetic results in our MFM cohort (n = 43): polymorphisms and unclassified variants

With means of NGS panel diagnostics we found a large number of variants, especially in *FLNC*. The suspicious variants with a minor allele frequency beneath 1% are listed in Additional file 1. We assessed these variants with various databases and prediction programmes. Especially the pathogenicity of the c.6595G>A, p.Gly2199Arg mutation in exon 40 of *FLNC* (patient F20.1, Additional file 1) finally remained unclear. This genetic variant has only been counted once by the NHLBI exome sequencing project (ESP) and was predicted to be benign by PolyPhen-2, but disease causing by Mutation Taster. The glycine residue at position c.6595 is highly conserved, as well as the nucleotide at this position. Unfixed muscle tissue of this patient was not available anymore and no new muscle biopsy could be obtained, so that additional proteomic analyses for possible further clarification could not be performed. Moreover, family members were not available for segregation analysis to further clarify the pathogenicity.

Clinical findings in MFM patients with an identified gene defect (n = 14)

The 14 index patients with an identified mutation included eleven males and three females (Table 1). The age at onset ranged from 28 to 69 years (mean 49 years). The most common initial symptom was weakness in the legs (64%). In three patients the first symptom was dyspnoea (21%) and in one syncope (7%), indicating a respiratory or cardiac onset of the disease. At the time of examination, the disease persisted already up to 32 years (mean disease duration 9 years). Muscle weakness was distributed in the distal and proximal lower limbs, combined with weakness in the proximal upper limbs in six of the 14 patients. Patients F3.II.3 and F6.1 were wheelchair bound since the age of 59 and 60 years respectively, both after 11 years of disease progression. Scapular winging was evident in 29% of the patients, 14% showed additional

axial weakness and 14% a ptosis without ophthalmoparesis or facial weakness (Table 1). In 64% of the patients, atrophy of the distal lower limbs was present, among them the two patients with an additional PNP. EMG revealed a myopathic pattern in 58% and a mixed pattern in 42%, in three of the patients with a mixed pattern also pseudomyotonic discharges were present. The serum CK levels varied from normal to 16 fold elevated (mean: 5.5 N) (Table 1). Only patient F14.1 did not show permanent skeletal muscle weakness.

Multisystemic symptoms in the whole MFM cohort (n = 43)

In general, 16% of the patients included in the study (n = 43) had respiratory involvement and 60% presented one or more multisystemic symptom(s) (Table 2). The most striking finding was the large number of MFM patients with mainly a sensorimotor axonal-demyelinating polyneuropathy (12/43 or 28%), which occurred as frequently as cardiac disease (Table 2). In 50% (n = 6) of the MFM patients with a polyneuropathy, no other possible cause was present. However, another 33% (n = 4) suffered from diabetes mellitus and 17% (n = 2) had taken neurotoxic medications, as possible (additional) causes of a polyneuropathy. The PNP was diagnosed after 0–6 years of disease duration (mean: 1.7 years). 14% of the MFM patients suffered from hearing impairment, with a symptom onset between 16 and 80 years of age. Half of these patients used hearing devices. We did not detect diarrhoea, intestinal malabsorption or pseudoobstruction in our cohort.

Multisystemic symptoms in MFM patients with an identified mutation (n = 14 and affected relatives)

In the group of index patients with a pathogenic mutation (n = 14) 64% showed at least one multisystemic symptom (Figure 1, Tables 1 and 2). Patient F3.II.3 harbours the Arg350Pro mutation in *DES* and presented with exertional dyspnoea as initial symptom at 48 years of age. Subsequently, he developed dyspnoea at rest, had a vital capacity (VC) of 45% of the theoretical value and needed non-invasive continuous positive airway pressure (CPAP) ventilation at night. Furthermore, he presented an intermittent tachyarrhythmia absoluta and a hypertrophic left ventricle at echocardiography. He also suffered from bilateral hypacusis and used a hearing aid on the right side since the age of 52 years.

Patient F4.II.2 carried the same Arg350Pro mutation in *DES* and recognised swallowing problems after six years of disease progression. He also presented a bilateral hypacusis since the age of 38 years.

Only the affected mothers of two more patients with the Arg350Pro mutation in *DES* (patient F1.III.1; F2.III.4) showed multisystemic symptoms (respiratory insufficiency,

Table 2 Multisystemic symptoms of all included MFM patients (n = 43)

Symptom		Frequency	
Respiratory symptoms (6)	restricted vital capacity	6**	(1x <i>DES</i> , 1x <i>FLNC</i> , 1x <i>MYOT</i> , 1x <i>CRYAB</i>)
	ventilation assistance	1	
	nocturnal ventilation	2	(1x <i>DES</i> , 1x <i>FLNC</i>)
Cardiac symptoms (12)	atrial fibrillation	8	(1x <i>DES</i> , 2x <i>MYOT</i>)
	tachyarrhythmia absoluta	1	(1x <i>DES</i>)
	ventricular tachycardia	1	
	bradycardia	1	
	ventricular extrasystoles	1	
	3° AV block	3*	(2x <i>DES</i> , 1x <i>MYOT</i>)
	LBBB	2	(1x <i>MYOT</i>)
	RBBB	1	
	bifascicular block	1	
	tricuspid insufficiency	2	(1x <i>DES</i>)
	aortic/mitral/tricuspid insufficiency	1	
	heart failure	3	(1x <i>MYOT</i>)
	LV hypertrophy	3	
	DCM	2	(1x <i>DES</i> , 1x <i>MYOT</i>)
pace maker	6	(2x <i>DES</i> , 2x <i>MYOT</i>)	
ICD	1	(1x <i>MYOT</i>)	
heart transplantation	1	(1x <i>MYOT</i>)	
sudden cardiac death	0		
Polyneuropathy (12)	axonal	3	(1x <i>BAG3</i>)
	demyelinating	0	
	axonal-demyelinating	9	(1x <i>MYOT</i>)
	sensory	0	
	sensory + autonomic	1	
	motor	1	
Bulbar symptoms (8)	sensorimotor	10	(1x <i>MYOT</i> , 1x <i>BAG3</i>)
	dysphagia	5	(1x <i>DES</i>)
	dysphonia	2	(1x <i>DES</i>)
Hearing impairment		6	(2x <i>DES</i> , 1x <i>FLNC</i>)
Gynaecomastia		1	

One patient can present more than one multisystemic symptom; between brackets, the number of cases with an identified mutation in whom the symptom was present; **, symptom was the onset symptom in two cases; 3° AV block, complete atrioventricular conduction block; *, symptom was the onset symptom in one case; LBBB, left bundle branch block; RBBB, right bundle branch block; LV, left ventricle; DCM, dilated cardiomyopathy; ICD, implantable cardioverter defibrillator.

cardiac involvement with a pace maker implantation, polyneuropathy).

Patient F6.1 carried the Glu245Asp mutation in *DES* and exhibited cardiac symptoms: a dilated cardiomyopathy was diagnosed approximately ten years after disease onset and a chronic atrial fibrillation and complete atrioventricular block occurred, necessitating pace maker implantation.

The patient carrying the Ser21le mutation in *DES* (patient F5.II.2) presented with syncopes as the first symptom at the age of 60 years, due to a complete

atrioventricular conduction block leading to pace maker implantation. At the age of 65 years, he developed mild distal lower leg weakness and dysphonia. Other causes of dysphonia were excluded. Interestingly, the patient's father also had dysphonia and a pacemaker implantation at the age of 58 years. The patient's brother, carrying the same mutation, received a pacemaker at the age of 45 years and had mild distal involvement of the lower legs, but no dysphonia.

Patient F8.III.3 with the Trp2710X mutation in *FLNC* developed respiratory problems three years before

weakness in the extremities occurred. After five years of disease progression, the patient was dependent on non-invasive nocturnal ventilation. VC decreased to 35%.

The affected mother of the index patient harbouring the Val930_Thr933del mutation in *FLNC* (patient F7.III.4) presented with bilateral hearing impairment and hearing aids at the age of 60 years.

Patient F9.1 carrying the Ser60Phe mutation in *MYOT* developed exertional dyspnoea due to heart failure at the age of 64 years as the first symptom. A left bundle branch block, an intermittent atrioventricular conduction block and atrial fibrillation were diagnosed, and a pace maker was implanted. The patient also suffered from coronary disease. Four years later, a sensorimotor axonal-demyelinating polyneuropathy was diagnosed.

In Patient F11.1 with a Ser55Phe mutation in *MYOT*, cardiac symptoms began shortly after skeletal weakness became apparent at the age of 55 years. He had a permanent atrial fibrillation with tachyarrhythmia absoluta and received an implantable cardioverter defibrillator (ICD). In addition, he had several myocardial infarctions due

to coronary disease. One year later, a dilated cardiomyopathy was diagnosed and heart transplantation was performed.

Patient F14.1 carries the p.Gly154Ser mutation in *CRYAB* and presented at the age of 69 years with several episodes of rhabdomyolysis. The patient also complained about exercise intolerance and fatigue. The neurological examination was normal, but a respiratory involvement with a VC of 65% in lying position was diagnosed. His 42-year-old son, carrying the same mutation, was asymptomatic and only showed a moderate serum CK elevation.

MRI findings in MFM patients with an identified mutation
Muscle MRI findings are summarised and compared to the literature in Table 3.

Histological findings of the MFM cohort

In the 14 index patients with an identified mutation we consistently found an increased variability of fibre diameter (100%), frequent nuclear bags (64%), internal nuclei (86%), vacuoles (64%), rimmed vacuoles (50%) and

Table 3 MRI findings in five MFM patients with an identified mutation

Gene	Fischer et al. 2008				Current study				
	<i>DES</i>	<i>FLNC</i>	<i>MYOT</i>	<i>ZASP</i> *	<i>DES</i>		<i>FLNC</i>	<i>MYOT</i>	<i>ZASP</i>
					F1.III.1	F2.III.4	F7.III.4	F10.1	F12.III.3
Pelvic level									
Gluteus maximus	2,3	1,3	2	3	3	np	1	np	np
Gluteus medius	1,8	1,7	2,8	1	1	np	1	np	np
Gluteus minimus	1,8	1,7	3,3	3	3	np	1	np	np
Mid thigh level									
Vastus lateralis	1,2	1,2	1,6	2	3	1	2	1	1
Vastus intermedius	1,2	2,8	2,8	3	4	1	3	1	1
Vastus medialis	1,3	2,3	2,8	3	4	1	3	1	1
Rectus femoris	1,3	0,4	0,1	2	1	1	2	1	1
Sartorius	2,8	0,7	1,5	1	4	3	1	2	1
Gracilis	2,2	0,6	0,4	1	3	3	1	1	1
Biceps femoris	1,5	3,3	2,9	4	1	1	3	3	3
Adductor magnus	1,7	3,2	3,2	4	3	1	3	3	2
Semitendinosus	3,1	1,8	1,2	3	3	3	3	2	1
Semimembranosus	1,5	3,3	2,8	4	2	1	3	2	3
Mid leg level									
Tibialis anterior	1,9	2,8	3,1	3	3	2	0	1	3
Peroneal muscle group	3	2,7	2,8	3	2	4	1	1	2
Medial gastrocnemius	2	3	3,4	3	2	3	3	4	4
Lateral gastrocnemius	2,1	0,9	2,1	3	2	3	1	3	4
Soleus	2,2	3,6	3,9	4	4	2	3	3	4

5-point scale according to Fischer et al. 2008 [51]; stage 0 is referred to a normal muscle appearance; stage 1 means traces of increased signal intensity on the T1-weighted MR sequences; stage 2 refers to increased signal intensity on MRI in less than 50% of the muscle; stage 3 means an increased signal intensity on MRI in more than 50% of the muscle; stage 4 indicates an increased signal intensity on MRI in the entire muscle (end-stage disease); *, *ZASP* classification is based on Figure four in the study of Fischer et al. 2008; np, not performed; for abbreviations of genes see text.

increased endomysial connective tissue (79%). In 43% of the biopsies, cytoplasmic bodies and core-like lesions on the NADH-TR staining were revealed. Myofibrillar disorganisation was seen in all cases, but characteristic protein aggregations in the cytoplasm of mGT stained fibres were found in only 10 of the 14 cases (71%) (Patient F11.1: in the vastus medialis muscle biopsy protein aggregations were seen, but not in the simultaneously obtained tibialis anterior muscle biopsy). So 29% only showed typical MFM findings at the ultrastructural level (EM was performed in 11 of the 14 cases). Z-disk streaming and myofibrillar disorganisation were seen in 100%, granulofilamentous material in 64%. Abnormal mitochondria (64%), tubulofilamentous accumulations (29%) and cytoplasmic bodies (36%) were other frequent findings. Immunohistochemistry with antibodies directed against desmin showed immunoreactivity in all 8 performed cases.

In the group of 24 patients, in whom the underlying genetic defect had not been identified, we made similar observations. Characteristic protein aggregations in the cytoplasm of mGT stained fibres were found in 76% of the cases, myofibrillar disorganisation in 82%.

In ten of the 12 MFM patients additionally presenting a polyneuropathy, we observed neurogenic muscle atrophy, which found expression in angular shaped muscle fibres and fibre type grouping.

Clinicopathological phenotype of the patient with a novel *BAG3* mutation

Patient F13.1 harbours the novel Pro209Gln mutation in *BAG3* (Table 1). He presented first symptoms of distal lower limb weakness and symmetrical calve atrophy at the age of 34 years. The skeletal muscle weakness spread to the proximal lower limbs and finally to the proximal upper limbs with scapular winging after 2 more years of disease progression. Moreover, he developed an axonal sensorimotor polyneuropathy, as multisystemic symptom of MFM. No other obvious cause for the PNP could be found. The PNP was first diagnosed at the age of 39 years and finds expression in decreased vibration sense and ataxic, clumsy gait as well as an increased sensitivity for cold. EMG showed a mixed pattern and the maximum CK level was 1050 U/L (Table 1). The muscle biopsy showed desmin positive protein deposits, vacuoles and core-like lesions as well as some necrotic fibres. At the ultrastructural level, tubulofilamentous accumulations, lobulated nuclei and glycogen accumulations were seen, in addition to typical MFM findings like Z-disk streaming and the accumulation of granulofilamentous material. We did not observe regenerating fibres or apoptotic nuclei. The ultrastructural study of nerve tissue revealed no giant axons.

Discussion

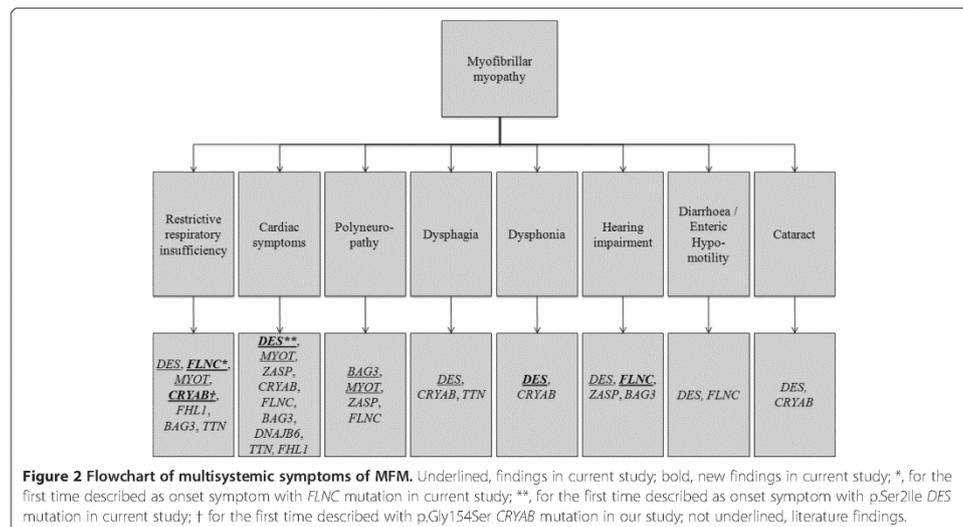
In a large cohort of MFM patients, we identified heterozygous mutations in 14 of 38 index patients (diagnostic yield of 37%), including the novel p.Pro209Gln mutation in exon 3 of *BAG3*, using Sanger and next generation sequencing (NGS) of the nine thus far known causative genes of MFM. Furthermore, we discovered new phenotypes associated with previously described mutations, such as hearing impairment with a *FLNC* mutation, dysphonia with a mutation in *DES* and the first patient with a *FLNC* mutation presenting respiratory insufficiency as the onset symptom. Interestingly, we detected a polyneuropathy in more than a quarter of the MFM patients, including a *BAG3* and a *MYOT* case (Figure 2).

Methodological issue: Sanger versus NGS sequencing

In ten of our patients, a mutation in one of the MFM-causative genes was identified using conventional Sanger sequencing. By employing NGS, we additionally found four mutations, among them the novel *BAG3* mutation, as well as a lot of variants, especially in *FLNC*, beyond the mutational hotspot in exon 48. The mass of variants that are found using NGS and their sometimes difficult interpretation can currently still be considered as a drawback of this technique. However, knowledge concerning the variants will increase in the following years as NGS will become more widely available and databases more complete. Advantages of NGS are reduced costs and decreased time to diagnosis. Furthermore, a specific and often unreliable genetic testing sequence is made redundant, especially in diseases with overlapping morphological and/or clinical phenotypes, such as MFM. The difficulty in interpretation of certain variants is illustrated by the unknown variant p.Gly2199Arg in exon 40 of *FLNC* of patient F20.1 (Additional file 1).

Novel mutation identified in *BAG3* and associated phenotype

The novel *BAG3* mutation is located at the amino acid position p.209 and proline is replaced by glutamine. At the same position the p.Pro209Leu mutation is situated, which is the most common *BAG3* mutation associated with MFM [14,36,46,48]. In all previously described MFM cases with the p.Pro209Leu mutation, the disease started within the first or second decade and was hallmarked by rapidly progressive muscle weakness, severe cardiac and respiratory involvement and early death of the patients in some cases. In nine of the eleven reported cases an axonal polyneuropathy was additionally found. Our patient's (Patient F13.1, Table 1) mild phenotype with adult-onset skeletal muscle weakness and the absence of cardiac or respiratory involvement clearly differs from the other *BAG3*-associated MFM cases, solely his axonal sensorimotor PNP is in concordance. Only one other



adult-onset patient with *BAG3*-related MFM has thus far been described: a woman harbouring the p.261-265 RAASP deletion in *BAG3*, who showed involvement of the orbicularis oculi muscles and the lower limb muscles, as well as photophobia and diplopia, starting at the age of 60 years. Her muscle biopsy revealed typical MFM findings [58].

In our patient's nerve and muscle biopsy no giant axons could be found, which were described before in *BAG3*-associated MFM cases [36,46,48]. Muscle biopsy findings were typical for MFM. We observed some lobulated nuclei at electron microscopy, corresponding to the findings of Selcen et al. who also described abnormal nuclei up to apoptotic changes in one MFM patient with a *BAG3* mutation [14].

Multisystemic symptoms in MFM, unusual phenotypes

We identified new phenotypes considering the multisystemic symptoms of our patients in association with certain mutations in the MFM genes (Figure 2; for review of the literature see Additional file 2). We reported hearing impairment for the first time in patients with the p.Arg350Pro *DES* mutation and for the first time in combination with a *FLNC* mutation (p.Val930_Thr933del). Also three other patients of our cohort (Table 2) presented with hearing problems, in one case (unidentified gene defect) the complaints began at the age of 16 years. Hearing impairment as a multisystemic symptom of MFM was described by Kraya et al. in two patients with a *ZASP* mutation, in a patient harbouring an unknown

variant in *BAG3* and in one autosomal recessive *DES* case, as a congenital feature [39,40,48,49].

We moreover present two new onset symptoms: syncopes due to a conduction block associated with the p.Ser2Ile mutation in *DES* and dyspnoea due to respiratory insufficiency for the first time with a *FLNC* mutation. In the patient with the *DES* mutation mild skeletal muscle involvement appeared only after several years. The two reported cases with the same mutation both showed skeletal muscle involvement at baseline [59], so this is the first described pure cardiac onset case associated with this mutation. Moreover dysphonia was noted in the proband and his father. So far, only dysphagia and nasal voice have been described with *DES* mutations [27], dysphonia has only been described with *CRYAB* mutations [31].

Respiratory insufficiency has been described in association with the p.Trp2710X mutation in *FLNC*, and frequently was the cause of death, but never the initial symptom [11,30,60]. It was not reported as an initial symptom in patients with other *FLNC* mutations, either. Also in patient F14.1 respiratory involvement was diagnosed, which has not been reported before with the p.Gly154Ser mutation in *CRYAB* [47,61].

The high frequency of polyneuropathy in our MFM cohort was striking (Table 2). We found PNP in combination with a novel *BAG3* and a known *MYOT* (p.Ser60Phe) mutation (Table 1). Both, *BAG3* and *MYOT* mutations, have been reported in MFM patients who also presented a PNP [10,14,46]. Nine additional index patients

of our cohort showed evidence for a polyneuropathy, but no mutation was found in these cases. Thus, there probably are other genes which cause a combined phenotype of MFM and polyneuropathy.

The most common cardiac symptoms in our cohort were arrhythmias (83%), followed by conduction defects (50%). This is in accordance with a 10-year longitudinal study in *DES* patients, in which conduction defects were the most common cardiac manifestation [59].

Taking the whole cohort into account, five patients complained about swallowing problems (one carried a *DES* p.Arg350Pro mutation).

The occurrence of multisystemic symptoms might be explained by the expression profiles of the proteins encoded by the MFM genes. However, the expression of proteins remains partially unknown, e.g. *FLNC* expression in the vestibulocochlear nerve.

Muscle MRI findings compared to the literature

For detailed values of the different muscles compared to the literature see Table 3 [51]. Patient F1.III.1 and patient F2.III.4 harbour the same mutation in *DES* (Table 1, Figure 1). Patient F2.III.4 showed an equal pattern of muscle involvement as described before even after 32 years of disease duration. However, the pattern seen in patient F1.III.1 is different, especially his gluteus minimus, quadriceps, and soleus muscle were far more involved than in previously reported cases. Patient F7.III.4 carries a *FLNC* mutation and showed a relatively typical pattern of muscle involvement in the MRI, only the severe involvement of the rectus femoris and semitendinosus and the spared tibialis anterior muscle was rather unusual. In the patient with a *MYOT* mutation (patient F10.1) we would have expected the involvement of the tibialis anterior muscle, because of the severely involved posterior compartment, as well as the vastus medialis and intermedius muscles which are described to be one of the most affected in the literature and are only mildly involved here. In the patient with a *ZASP* mutation (patient F.12.III.3) we found a rather mild involvement at muscle MRI, but as was described before biceps femoris and semimembranosus muscles were most severely affected.

Conclusion

We conclude that multisystemic involvement frequently occurs in MFM. Most important are the cardiac and respiratory symptoms, but also polyneuropathies, hearing loss and bulbar symptoms can occur. *BAG3* should be included in the genetic workup of MFM patients, even in cases with an adult onset and a mild phenotype. In 29% of all biopsies no aggregations were found at the light microscopic level, however, typical ultrastructural findings were present, underlining the importance of

electron microscopy in the diagnosis of MFM. In contrast, 76% of the patients/biopsies without a genetic identification showed distinct protein deposits in the cytoplasm of the muscle cells. This high rate and our percentage of identified mutations (37%) lead to the conclusion that more causative genes for MFM are still to be found. NGS might be helpful in achieving this aim.

Additional files

Additional file 1: Non-pathogenic or unclassified variants identified by means of NGS panel diagnostics in the MFM-causing genes.

Additional file 2: Review of multisystemic symptoms described in literature.

Competing interests

The authors declare that they have no conflicts of interest.

Authors' contributions

AS carried out the acquisition and interpretation of clinical, histopathological and imaging data, participated in the molecular genetic and histopathological studies and drafted the manuscript. SS was involved in collecting and analysing clinical and histopathological data and revised the manuscript. JEB carried out molecular genetic studies and drafted parts of the manuscript. CL participated in acquisition and interpretation of clinical and histopathological data and revised the manuscript. JB was involved in collecting and analysing clinical data and revised the manuscript. BAK was involved in collecting and analysing clinical data and revised the manuscript. AF participated in collecting and analysing clinical and imaging data and revised the manuscript. RA was involved in collecting and analysing clinical data and revised the manuscript. PvdB was involved in collecting and analysing clinical and histopathological data and revised the manuscript. JJM participated in the acquisition and interpretation of clinical and histopathological data and revised the manuscript. PDJ was involved in collecting clinical data and revised the manuscript. ENJ participated in collecting and analysing of histopathological data and revised the manuscript. OM participated in the acquisition and interpretation of clinical data and revised the manuscript. MD participated in collecting and analysing of clinical and imaging data and revised the manuscript. MB participated in collecting and interpretation of clinical data and revised the manuscript. JMS participated in the acquisition and interpretation of histopathological data and revised the manuscript. MV participated in collecting and analysing of clinical and histopathological data and revised the manuscript. JBS participated in the acquisition and interpretation of clinical data and revised the manuscript. JW participated in the acquisition and interpretation of histopathological data and revised the manuscript. WK carried out the molecular genetic studies and drafted and revised the manuscript. KGC conceived, coordinated and supervised the study, participated in the collection and analysing of the clinical, histopathological, imaging and genetic data, and corrected the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to the personnel of the Institute of Neuropathology and the Department of Neurology at the RWTH Aachen University hospital, Aachen, Germany, for the technical support. We are thankful to Birgit Halliger-Keller of the Department of Human Genetics at the University of Würzburg, Würzburg, Germany, and Philipp Helbig, for their technical support, as well as to Pauline Lahaut at the Nice University Hospital for her help in collecting patient's data.

Author details

¹Department of Neurology, RWTH Aachen University, Aachen, Germany. ²Institute of Neuropathology, RWTH Aachen University, Aachen, Germany. ³Centre de Référence des Maladies Neuro-musculaires, Nice Hospital and UMR CNRS6543, Nice University, Nice, France. ⁴Department of Human Genetics, University of Würzburg, Würzburg, Germany. ⁵Department of

Neurology, Saarland University, Homburg/Saar, Germany. ⁶Department of Neurology, Neuromuscular Center Ruhrgebiet, University Hospital Bergmannsheil, Ruhr-University Bochum, Bochum, Germany. ⁷Department of Neurology, Klinikum Kassel, Kassel, Germany. ⁸Department of Neurology, Klinikum Merzig, Merzig, Germany. ⁹Department of Neurology, Neuromuscular Reference Center, University Hospital Saint-Luc, Brussel, Belgium. ¹⁰Institute Born-Bunge, University of Antwerpen, Antwerpen, Belgium. ¹¹Neurogenetics Group, VIB-Department of Molecular Genetics, University of Antwerpen, Antwerpen, Belgium. ¹²Department of Neurology, University Hospital of Antwerpen, Antwerpen, Belgium. ¹³Institute of Neurophysiology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany. ¹⁴Department of Cardiology, Angiology and Pneumology, University Hospital Heidelberg, Heidelberg, Germany. ¹⁵DZHK (German Centre for Cardiovascular Research), partner site Heidelberg, Heidelberg, Germany. ¹⁶Department of Neurology, Martin-Luther-University Halle-Wittenberg, Halle, Germany. ¹⁷Institute of Neurophysiology, Klinikum Bremen-Mitte, Bremen, Germany. ¹⁸JARA - Translational Brain Medicine, Jülich and Aachen, Germany.

Received: 26 March 2014 Accepted: 21 July 2014
Published: 1 August 2014

References

- De Blecker JL, Engel AG, Ertl BB: **Myofibrillar myopathy with abnormal foci of desmin positivity. II. Immunocytochemical analysis reveals accumulation of multiple other proteins.** *J Neuropathol Exp Neurol* 1996, **55**:563-577.
- Nakano S, Engel AG, Wadawik AJ, Emslie-Smith AM, Busis NA: **Myofibrillar myopathy with abnormal foci of desmin positivity. I. Light and electron microscopy analysis of 10 cases.** *J Neuropathol Exp Neurol* 1996, **55**:549-562.
- Selcen D, Ohno K, Engel AG: **Myofibrillar myopathy: clinical, morphological and genetic studies in 63 patients.** *Brain* 2004, **127**:439-451.
- Olive M, Goldfarb LG, Shatunov A, Fischer D, Ferrer I: **Myofibrillar myopathy: refining the clinical and morphological phenotype.** *Brain* 2005, **128**:2315-2326.
- Schroder R, Viaboe A, Goebel HH: **Primary desminopathies.** *J Cell Mol Med* 2007, **11**:416-426.
- Claeys KG, van der Ven PF, Behin A, Stojkovic T, Eymard B, Dubourg O, Lafont P, Faulkner G, Richard P, Vicart P, Romero NB, Stoltenberg G, Udd B, Fardeau M, Voit T, Furst DG: **Differential involvement of sarcomeric proteins in myofibrillar myopathies: a morphological and immunohistochemical study.** *Acta Neuropathol* 2009, **117**:293-307.
- Claeys KG, Fardeau M, Schroder R, Suominen T, Tolksdorf K, Behin A, Dubourg O, Eymard B, Naissonobe T, Stojkovic T, Faulkner G, Richard P, Vicart P, Udd B, Voit T, Stoltenberg G: **Electron microscopy in myofibrillar myopathies reveals clues to the mutated gene.** *Neuromuscul Disord* 2008, **18**:656-666.
- Goldfarb LG, Park KY, Genenakova L, Gorothova S, Lee HS, Vasconcelos O, Nagle JW, Semino-Mora C, Shatunov A, Dalakas MC: **Missense mutations in desmin associated with familial cardiac and skeletal myopathy.** *Nat Genet* 1998, **19**:402-403.
- Vicart P, Caron A, Guicheney P, Li Z, Prevost MC, Faure A, Chateau D, Chapon F, Tome F, Dupret JM, Paulin D, Fardeau M: **A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy.** *Nat Genet* 1998, **20**:92-95.
- Selcen D, Engel AG: **Mutations in myotilin cause myofibrillar myopathy.** *Neurology* 2004, **62**:1363-1371.
- Vorgerd M, van der Ven PF, Bruchertsefer V, Lowe T, Kley RA, Schroder R, Lochmuller H, Himmel M, Roehler K, Furst DG, Huebner A: **A mutation in the dimerization domain of filamin c causes a novel type of autosomal dominant myofibrillar myopathy.** *Am J Hum Genet* 2005, **77**:297-304.
- Selcen D, Engel AG: **Mutations in ZASP define a novel form of muscular dystrophy in humans.** *Ann Neurol* 2005, **57**:269-276.
- Selcen D, Bromberg MB, Chin SS, Engel AG: **Reducing bodies and myofibrillar myopathy features in FHL1 muscular dystrophy.** *Neurology* 2011, **77**:1951-1959.
- Selcen D, Muntani F, Burton BK, Pegoraro E, Sewry C, Bite AV, Engel AG: **Mutation in BAG3 causes severe dominant childhood muscular dystrophy.** *Ann Neurol* 2009, **65**:83-89.
- Sato T, Hayashi YK, Oya Y, Kondo T, Sugie K, Kaneda D, Houzen H, Yabe I, Sasaki H, Noguchi S, Nonaka I, Osawa M, Nishino I: **DNAJB6 myopathy in an Asian cohort and cytoplasmic/nuclear inclusions.** *Neuromuscul Disord* 2013, **23**:269-276.
- Ohlsson M, Hedberg C, Bradvik B, Lindberg C, Tajsharhi H, Danielsson G, Melberg A, Udd B, Malmsson T, Oldfors A: **Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin.** *Brain* 2012, **135**:1682-1694.
- Schroder R, Schoser B: **Myofibrillar myopathies: a clinical and myopathological guide.** *Brain Pathol* 2009, **19**:483-492.
- Olive M, Odegerel Z, Martinez A, Pozo JJ, Bragado FG, Zabalza RJ, Jerico I, Gonzalez-Mera I, Shatunov A, Lee HS, Armstrong J, Marafé E, Arroyo MR, Pascual-Cabot J, Navarro C, Paradas C, Huera M, Marquez F, Rivas EG, Pou A, Ferrer I, Goldfarb LG: **Clinical and myopathological evaluation of early- and late-onset subtypes of myofibrillar myopathy.** *Neuromuscul Disord* 2011, **21**:533-542.
- Forrest KM, Al-Sarraf S, Sewry C, Buk S, Tan SW, Pitt M, Durward A, McDougall M, Irving M, Hanna MG, Matthews E, Sarkozy A, Hudson J, Baresi B, Bushby K, Jungbluth H, Waige E: **Infantile onset myofibrillar myopathy due to recessive CRYAB mutations.** *Neuromuscul Disord* 2011, **21**:37-40.
- McLaughlin HM, Kelly MA, Hawley PP, Damas BT, Funke B, Pickler J: **Compound heterozygosity of predicted loss-of-function DES variants in a family with recessive desminopathy.** *BMC Med Genet* 2013, **14**:68.
- Claeys KG, Fardeau M: **Myofibrillar myopathies.** *Handb Clin Neurol* 2013, **113**:1337-1342.
- Olive M, Goldfarb L, Moreno D, Lafont E, Dagvadovaj A, Samboughin N, Martinez-Matos JA, Martinez F, Alia J, Ferrero E, Vicart P, Ferrer I: **Desmin-related myopathy: clinical, electrophysiological, radiological, neuropathological and genetic studies.** *J Neurol Sci* 2004, **219**:125-137.
- Gamez J, Armstrong J, Shatunov A, Selva-O Callaghan A, Dominguez-Coronoz R, Ortega A, Goldfarb L, Ferrer I, Olive M: **Generalized muscle pseudo-hypertrophy and stiffness associated with the myotilin Ser55Phe mutation: a novel myotilinopathy phenotype?** *J Neurol Sci* 2009, **277**:167-171.
- Vattemi G, Neri M, Piffer S, Vicart P, Guaiardi F, Marini M, Guglielmi V, Filosto M, Tonin P, Ferlini A, Temelieri G: **Clinical, morphological and genetic studies in a cohort of 21 patients with myofibrillar myopathy.** *Acta Myol* 2011, **30**:121-126.
- Palma J, Penttila S, Huovinen S, Haapasalo H, Udd B: **An unusual phenotype of late-onset desminopathy.** *Neuromuscul Disord* 2013, **23**:922-923.
- Pfeffer G, Baresi R, Wilson U, Hardy SA, Griffin H, Hudson J, Elliott HR, Bamesh AV, Radunovic A, Winer JB, Valdyja S, Ramon A, Busby M, Farugio ME, Ming A, Everett C, Emsley HC, Horvath R, Straub V, Bushby K, Lochmuller H, Chimney PF, Sarkozy A: **Titin founder mutation is a common cause of myofibrillar myopathy with early respiratory failure.** *J Neurol Neurosurg Psychiatry* 2014, **85**:331-338.
- Dagvadovaj A, Olive M, Urtizbera JA, Halle M, Shatunov A, Bonnemant C, Park KY, Goebel HH, Ferrer I, Vicart P, Dalakas MC, Goldfarb LG: **A series of West European patients with severe cardiac and skeletal myopathy associated with a de novo R406W mutation in desmin.** *J Neurol* 2004, **251**:143-149.
- Selcen D, Engel AG: **Myofibrillar myopathy caused by novel dominant negative alpha B-crystallin mutations.** *Ann Neurol* 2003, **54**:804-810.
- Walter MC, Reilich P, Huebner A, Fischer D, Schroder R, Vorgerd M, Kress W, Born C, Schoser BG, Krause KH, Klutzny U, Bulst S, Frey JR, Lochmuller H: **Scapuloperoneal syndrome type Kaeser and a wide phenotypic spectrum of adult-onset, dominant myopathies are associated with the desmin mutation R350P.** *Brain* 2007, **130**:1485-1496.
- Kley RA, Hellenbroich Y, van der Ven PF, Furst DG, Huebner A, Bruchertsefer V, Peters SA, Heyer GM, Kirschner J, Schroder R, Fischer D, Muller K, Tolksdorf K, Eger K, Gemming A, Biedner T, Reum C, Walter MC, Lochmuller H, Ketelsen UP, Vorgerd M: **Clinical and morphological phenotype of the filamin myopathy: a study of 31 German patients.** *Brain* 2007, **130**:3250-3264.
- Sacconi S, Feasson L, Antoine JC, Pechoux C, Bernard R, Cabo AM, Casarin A, Salvati L, Desnuelle C, Urtizbera A: **A novel CRYAB mutation resulting in multisystemic disease.** *Neuromuscul Disord* 2012, **22**:66-72.
- Pruszczyk P, Koscielna-Pruszczyk A, Shatunov A, Goudeau B, Draminska A, Takeda K, Samboughin N, Vicart P, Strelkov SV, Goldfarb LG, Kaminska A: **Restrictive cardiomyopathy with atrioventricular conduction block resulting from a desmin mutation.** *Int J Cardiol* 2007, **117**:244-253.

33. Olive M, Armstrong J, Miralles F, Pou A, Fardeau M, Gonzalez L, Martinez F, Fischer D, Martinez Matos JA, Shatunov A, Goldfarb L, Ferrer I: **Phenotypic patterns of desminopathy associated with three novel mutations in the desmin gene.** *Neuromuscul Disord* 2007, **17**:443–450.
34. Yuri T, Miki K, Tsukamoto R, Shinde A, Kusaka H, Tsubura A: **Autopsy case of desminopathy involving skeletal and cardiac muscle.** *Pathol Int* 2007, **57**:32–36.
35. Strach K, Sommer T, Grohe C, Meyer C, Fischer D, Walter MC, Vorgerd M, Reilich P, Bar H, Reimann J, Reuner U, Gemring A, Goebel HH, Lochmuller H, Wintersparger B, Schroder R: **Clinical, genetic, and cardiac magnetic resonance imaging findings in primary desminopathies.** *Neuromuscul Disord* 2008, **18**:475–482.
36. Odgerel Z, Sarkozy A, Lee HS, McKenna C, Rankin J, Straub V, Lochmuller H, Paola F, D'Amico A, Bertini E, Bushby K, Goldfarb LG: **Inheritance patterns and phenotypic features of myofibrillar myopathy associated with a BAG3 mutation.** *Neuromuscul Disord* 2010, **20**:438–442.
37. Hong D, Wang Z, Zhang W, Xi J, Lu J, Luan X, Yuan Y: **A series of Chinese patients with desminopathy associated with six novel and one reported mutations in the desmin gene.** *Neuropathol Appl Neurobiol* 2011, **37**:257–270.
38. Hedberg C, Melberg A, Kuhl A, Jenne D, Oldfors A: **Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy 7 is caused by a DES mutation.** *Eur J Hum Genet* 2012, **20**:984–985.
39. Ariza A, Coll J, Fernandez-Figueras MT, Lopez MD, Mate JL, Garcia O, Fernandez-Vasalo A, Navas-Palacios JJ: **Desmin myopathy: a multisystem disorder involving skeletal, cardiac, and smooth muscle.** *Hum Pathol* 1995, **26**:1032–1037.
40. Munoz-Marmol AM, Strasser G, Isamat M, Coulombe PA, Yang Y, Roca X, Vela E, Mate JL, Coll J, Fernandez-Figueras MT, Navas-Palacios JJ, Ariza A, Fuchs E: **A dysfunctional desmin mutation in a patient with severe generalized myopathy.** *Proc Natl Acad Sci U S A* 1998, **95**:11312–11317.
41. Piniol-Ripoll G, Shatunov A, Cabello A, Larode P, de la Puerta I, Pelegrin J, Ramos FJ, Olive M, Goldfarb LG: **Severe infantile-onset cardiomyopathy associated with a homozygous deletion in desmin.** *Neuromuscul Disord* 2009, **19**:418–422.
42. Luan X, Hong D, Zhang W, Wang Z, Yuan Y: **A novel heterozygous deletion-insertion mutation (2695–2712 del/GTTTGT ins) in exon 18 of the filamin C gene causes filaminopathy in a large Chinese family.** *Neuromuscul Disord* 2010, **20**:390–396.
43. Dalakas MC, Dagvadorj A, Goudeau B, Park KY, Takeda K, Simon-Casteras M, Vasconcelos O, Sambuughin N, Shatunov A, Nagle JW, Sivakumar K, Vicart P, Goldfarb LG: **Progressive skeletal myopathy, a phenotypic variant of desmin myopathy associated with desmin mutations.** *Neuromuscul Disord* 2003, **13**:252–258.
44. Henderson M, De Waele L, Hudson J, Eagle M, Sewry C, Marsh J, Chariton R, He L, Blakely EL, Horrocks J, Stewart W, Taylor RW, Longman C, Bushby K, Barresi R: **Recessive desmin-null muscular dystrophy with central nuclei and mitochondrial abnormalities.** *Acta Neuropathol* 2013, **125**:917–919.
45. Izumi R, Niihori T, Aoki Y, Suzuki N, Kato M, Warita H, Takahashi T, Tateyama M, Nagashima T, Funayama R, Abe K, Nakayama K, Aoki M, Matsubara Y: **Exome sequencing identifies a novel TTN mutation in a family with hereditary myopathy with early respiratory failure.** *J Hum Genet* 2013, **58**:259–266.
46. Jaffer F, Murphy SM, Scoto M, Healy E, Rossor AM, Brandner S, Phadke R, Selcen D, Jungbluth H, Muntoni F, Reilly MM: **BAG3 mutations: another cause of giant axonal neuropathy.** *J Peripher Nerv Syst* 2012, **17**:210–216.
47. Reilich P, Schoser B, Schramm N, Krause S, Schessl J, Kress W, Muller-Hocker J, Walter MC, Lochmuller H: **The p.G154S mutation of the alpha-B crystallin gene (CRYAB) causes late-onset distal myopathy.** *Neuromuscul Disord* 2010, **20**:255–259.
48. Lee HC, Cherk SW, Chan SK, Wong S, Tong TW, Ho WS, Chan AY, Lee KC, Mak CM: **BAG3-related myofibrillar myopathy in a Chinese family.** *Clin Genet* 2012, **81**:394–398.
49. Kraya T, Kress W, Stoevesant D, Deschauer M, Zierz S: **[Myofibrillar myopathy due to the ZASP mutation Ala147Thr: two cases with exclusively distal leg involvement].** *Nervenarzt* 2013, **84**:209–213.
50. Hanisch F, Kraya T, Kornhuber M, Zierz S: **Diagnostic impact of myotonic discharges in myofibrillar myopathies.** *Muscle Nerve* 2013, **47**:845–848.
51. Fischer D, Kley RA, Strach K, Meyer C, Sommer T, Eger K, Rolfs A, Meyer W, Pou A, Pradas J, Heyer CM, Grossmann A, Huebner A, Kress W, Reimann J, Schroder R, Eymard B, Fardeau M, Udd B, Goldfarb L, Vorgerd M, Olive M: **Distinct muscle imaging patterns in myofibrillar myopathies.** *Neurology* 2008, **71**:758–765.
52. Wattjes MP, Kley RA, Fischer D: **Neuromuscular imaging in inherited muscle diseases.** *Eur Radiol* 2010, **20**:2447–2460.
53. Joshi PR, Hauburger A, Kley R, Claeys KG, Schneider I, Kress W, Stoltenberg G, Weis J, Vorgerd M, Deschauer M, Hanisch F: **Mitochondrial abnormalities in myofibrillar myopathies.** *Clin Neuropathol* 2014, **33**:134–142.
54. Dubowitz V, Sewry C: *Muscle biopsy a practical approach*. 3rd edition. Philadelphia: Elsevier; 2007.
55. Weis J, Brandner S, Lammens M, Sommer C, Vallat JM: **Processing of nerve biopsies: a practical guide for neuropathologists.** *Clin Neuropathol* 2012, **31**:7–23.
56. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocum PM, Smith M: **Nucleotide sequence of bacteriophage phi X174 DNA.** *Nature* 1977, **265**:687–695.
57. Vasli N, Laporte J: **Impacts of massively parallel sequencing for genetic diagnosis of neuromuscular disorders.** *Acta Neuropathol* 2013, **125**:173–185.
58. Sato T, Hayashi YK, Keduka E, Noguchi S, Osawa M, Nonaka I, Nishino I: **Novel BAG3 mutations in myofibrillar myopathy patients [abstract].** *Neuromuscul Disord* 2011, **21**:740.
59. Wahbi K, Behin A, Charon P, Dunand M, Richard P, Meune C, Vicart P, Laforet P, Stojkovic T, Becane HM, Kuntzer T, Duboc D: **High cardiovascular morbidity and mortality in myofibrillar myopathies due to DES gene mutations: a 10-year longitudinal study.** *Neuromuscul Disord* 2012, **22**:211–218.
60. Kley RA, Serdaroglu-Offlazer P, Leber Y, Odgerel Z, van der Ven PF, Olive M, Ferrer I, Onipe A, Mihaylov M, Bilbao JM, Lee HS, Hohfeld J, Djinnovic-Carugo K, Kong K, Tegenthoff M, Peters SA, Stenzel W, Vorgerd M, Goldfarb LG, Furst DO: **Pathophysiology of protein aggregation and extended phenotyping in filaminopathy.** *Brain* 2012, **135**:2642–2660.
61. Pilotto A, Marziliano N, Pasotti M, Grasso M, Costante AM, Arbustini E: **alphaB-crystallin mutation in dilated cardiomyopathies: low prevalence in a consecutive series of 200 unrelated probands.** *Biochem Biophys Res Commun* 2006, **346**:1115–1117.

doi:10.1186/s13023-014-0121-9

Cite this article as: Semmler et al.: Unusual multisystemic involvement and a novel BAG3 mutation revealed by NGS screening in a large cohort of myofibrillar myopathies. *Orphanet Journal of Rare Diseases* 2014 **9**:121.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Additional file 1

Additional file 1: Non-pathogenic or unclassified variants identified by means of NGS

panel diagnostics in the MFM-causing genes.

Pat.	Gene	Exon	Variant	SNP database	Predicted as	Conservation status	NHLBI	1000GP
F13.1	<i>ZASP</i>	4	c.664G>A, p.Ala222Thr*	rs139922045, MAF <0.1 %	predicted inconsistently	AA highly, Nt weakly	4/8596 (European American population)	not found
F15.1	<i>FLNC</i>	21	c.3721C>T, p.Arg1241Cys	rs146953558, MAF 0,95%	benign	AA highly, Nt weakly	81/8433 (European American population)	5/753 (European population)
F17.1	<i>FLNC</i>	23	c.4022G>A, p.Arg1341Gln	rs149641783, MAF 0,18%	predicted inconsistently	AA highly	15/8425 (European American population)	2/756 (European population)
F19.1	<i>FLNC</i>	34	c.5578C>T, p.Arg1860Cys	rs181067717, MAF 0,62%	probably/possibly damaging	AA highly, Nt moderately	52/8320 (European American population)	3/755 (European population)
F20.1	<i>FLNC</i>	40	c.6595G>A, p.Gly2199Arg	rs368977589, MAF 0,012%	predicted inconsistently	AA and Nt highly	1/8357 (European American population)	not found
F23.1	<i>TTN</i>	343	c.95297C>T, p.Ser31766Phe	rs191484894, MAF 0.2 %	disease causing	AA and Nt highly	16/8334 (European American population)	2/756 (European population)

Predicted as, here the results of the prediction programs PolyPhen-2 and Mutation Taster are shown; Conservation status, conservation of the changed amino acid (AA) and nucleotide (Nt) using data of Alamut; NHLBI, frequency of the variant according to the database of NHLBI exome sequencing project (ESP); 1000GP, frequency of the variant according to the database of the 1000 genome project. *, the disease did not segregate in the family; MAF, minor allele frequency; predicted inconsistently, the mutation is predicted to be disease causing by one programme and predicted to be benign by the other; -, not performed; for abbreviations of genes see text.

Additional file 2

Additional file 2.pdf – Review of multisystemic symptoms described in literature.

Gene/ Symptom	DES	CRYAB	FLNC	MYOT	ZASP	FHL-1	BAG3	DNAJB6	TTN
Cardiac involvement	AF: Arias et al. 2006 Arrhythmias: [34, 41] Conduction defects: [8, 27, 37, 39, 40] Cao et al. 2012 MI/TI: Dalakas et al. 2000 HCM: [33] DCM: [20] RCM: [32] Vrabie et al. 2005 Arbustini et al. 2006 ARVC: [38] HTx: Klauke et al. 2010 SCD: [22, 29, 35]	[9, 31, 47, 61] Iganaki et al. 2006 van der Smagt et al. 2014	[30, 42, 60] Tasca et al. 2012 Avila-Smirnov et al. 2010	[4, 10]	[12, 49] Vatta et al. 2003	Binder et al. 2012	[14, 36, 46, 48, 58]	[15]	[26]
Respiratory involvement	[8, 20, 27, 32, 33, 34, 35, 37, 39, 40, 44] Dagvadorj et al. 2003 Dalakas et al. 2000	[19, 28, 31] Del Bigio et al. 2011	[11, 30, 60] Tasca et al. 2012	[4]		Feldkirchner et al. 2013	[14, 36, 46, 48, 58]		[16, 26, 45] Edström et al. 1990
Poly-neuropathy			[11]	[10]	[12]	[13]	[14, 36, 46, 48]		
Dysphagia	[8, 20, 22, 27, 29, 33, 37, 39, 40, 43, 44] Dagvadorj et al. 2003	[9, 28, 31, 47]							[45]
Dysarthria	[22, 27, 33]	[31]	Tasca et al. 2012	[4]			[14]		[45]
Cataract	[22, 33]	[9, 31] van der Smagt et al. 2014							
Diarrhoe	[37, 39, 40, 41]		[42, 60]						
Scleroderma		[28]							
Hearing impairment	[39, 40]				[49]				
Gynaecomastia	[29, 37]	[47]							
Lipodystrophy						Feldkirchner et al. 2013			
Cerebellar ataxia			Tasca et al. 2012						

AF, atrial fibrillation; TI/MI, tricuspid or mitral insufficiency; HCM, hypertrophic cardiomyopathy; DCM, dilated cardiomyopathy; RCM, restrictive cardiomyopathy; ARVC, arrhythmogenic right ventricular cardiomyopathy; HTx, heart transplantation; SCD, sudden cardiac death, for abbreviations of genes see text.

References:

- Arias M, Pardo J, Blanco-Arias P, Sobrido M, Arias S, Dapena D, Carracedo A, Goldfarb LG, Navarro C: **Distinct phenotypic features and gender-specific disease manifestations in a Spanish family with desmin L370P mutation.** *Neuromuscul Disord* 2006, **16**:498-503.
- Cao L, Hong D, Zhu M, Li X, Wan H, Hong K: **A novel heterozygous deletion-insertion mutation in the desmin gene causes complete atrioventricular block and mild myopathy.** *Clin Neuropathol* 2013, **32**:9-15.
- Dalakas MC, Park K, Semino-Mora C, Lee HS, Sivakumar K, Goldfarb LG: **Desmin myopathy, a skeletal myopathy with cardiomyopathy caused by mutations in the desmin gene.** *NEngl J Med* 2000, **342**:770-780.
- Vrabie A, Goldfarb LG, Shatunov A, Nagele A, Fritz P, Kaczmarek I, Goebel HH: **The enlarging spectrum of desminopathies: new morphological findings, eastward geographic spread, novel exon 3 desmin mutation.** *Acta Neuropathol* 2005, **109**:411-417.
- Arbustini E, Pasotti M, Pilotto A, Pellegrini C, Grasso M, Previtati S, Repetto A, Bellini O, Azan G, Scaffino M, Campana C, Piccolo G, Viganò M, Tavazzi L: **Desmin accumulation restrictive cardiomyopathy and atrioventricular block associated with desmin gene defects.** *Eur J Heart Fail* 2006, **8**:477-483.
- Klauke B, Kossmann S, Gaertner A, Brand K, Stork I, Brodehl A, Dieding M, Walhorn V, Anselmetti D, Gerdes D, Bohms B, Schulz U, zu Knyphausen E, Vorgerd M, Gummert J, Milting H: **De novo desmin-mutation N116S is associated with arrhythmogenic right ventricular cardiomyopathy.** *Hum Mol Genet* 2010, **19**:4595-4607.
- Dagvadorj A, Goudeau B, Hilton-Jones D, Blancato JK, Shatunov A, Simon-Casteras M, Squier W, Nagle JW, Goldfarb LG, Vicart P: **Respiratory insufficiency in desminopathy patients caused by introduction of proline residues in desmin c-terminal α -helical segment.** *Muscle Nerve* 2003, **27**:669-675.
- Dagvadorj A, Olivé M, Urtizberera JA, Halle M, Shatunov A, Bönnemann C, Park KY, Goebel HH, Ferrer I, Vicart P, Dalakas MC, Goldfarb LG: **A series of West European patients with severe cardiac and skeletal myopathy associated with a de novo R406W mutation in desmin.** *J Neurol* 2004, **251**:143-149.
- Iganaki N, Hayashi T, Arimura T, Koga Y, Takahashi M, Shibata H, Teraoka K, Chikamori T, Yamashina A, Kimura A: **Alpha B-crystallin mutation in dilated cardiomyopathy.** *Biochem Biophys Res Commun* 2006, **342**:379-386.
- van der Smagt JJ, Vink A, Kirkels JH, Nelen M, ter Heide H, Molenschot MMC, Weger RA, Schellekens PAW, Hoogendijk J, Dooijes D: **Congenital posterior pole cataract and adult onset dilating cardiomyopathy. Expanding the phenotype of α B-crystallinopathies.** *Clin Genet* 2014, **85**:381-385.
- Del Bigio MR, Chudley AE, Sarnat HB, Campbell C, Goobie S, Chodirker BN, Selcen D: **Infantile Muscular Dystrophy in Canadian Aboriginals is an α BCrystallinopathy.** *Ann Neurol* 2011, **69**:866-871.
- Tasca G, Odgerel Z, Monforte M, Aurino S, Clarke NF, Waddell LB, Udd B, Ricci E, Goldfarb LG: **Novel FLNC mutation in a patient with myofibrillar myopathy in combination with late-onset cerebellar ataxia.** *Muscle Nerve* 2012, **46**:275-282.
- Avila-Smirnov D, Béhin A, Gueneau L, Claeys K, Beuvin M, Goudeau B, Richard P, Yaou RB, Romero NB, Mathis S, Voit T, Eymard B, Gil R, Fardeau M, Bonne G: **A novel missense FLNC mutation causes arrhythmia and late onset myofibrillar myopathy with particular histopathology features [abstract].** *Neuromuscul Disord* 2010, **20**:596-680.
- Vatta M, Mohapatra B, Jimenez S, Sanchez X, Faulkner G, Perles Z, Sinagra G, Lin JH, Vu TM, Zhou Q, Bowles KR, Di Lenarda A, Schimmenti L, Fox M, Chrisco MA, Murphy RT, McKenna W, Elliott P, Bowles NE, Chen J, Valle J, Towbin JA: **Mutations in *Cypher/ZASP* in Patients with dilated cardiomyopathy and left ventricular non-compaction.** *J Am Coll Cardiol* 2003, **42**:2014-2027.
- Binder JS, Weidemann F, Schoser B, Niemann M, Machann W, Beer M, Plank G, Schmidt A, Bisping E, Poparie I, Lafer I, Stojakovic, Quasthoff S, Vincent JB, Riemmueller R, Speicher MR, Berghold A, Pieske B, Windpassinger C: **Spongy hypertrophic Cardiomyopathy in patients with mutations in the Four-and-a-Half LIM Domain 1 Gene.** *Circ Cardiovasc Genet* 2012, **5**:490-502.
- Feldkirchner S, Walter MC, Müller S, Kubny C, Krause S, Kress W, Hanisch FG, Schoser B, Schessl J: **Proteomic characterization of aggregate components in an intrafamilial variable FHL1-associated myopathy.** *Neuromuscul Disord* 2013, **23**:418-426.
- Edström L, Thornell LE, Albo J, Landin S, Samuelsson M: **Myopathy with respiratory failure and typical myofibrillar lesions.** *J Neurol Sci* 1990, **96**:211-228.

Anlage 4

Joachim Schessl, **Elisa Bach**, Simone Rost, Sarah Feldkirchner, Christiana Kubny, Stefan Müller, Franz-Georg Hanisch, Wolfram Kress, Benedikt Schoser

Novel recessive myotilin mutation causes severe myofibrillar myopathy

Neurogenetics. 2014, 15(3):151-6. doi: 10.1007/s10048-014-0410-4

Neurogenetics (2014) 15:151–156
DOI 10.1007/s10048-014-0410-4

SHORT COMMUNICATION

Novel recessive myotilin mutation causes severe myofibrillar myopathy

Joachim Schessl · Elisa Bach · Simone Rost · Sarah Feldkirchner · Christiana Kubny · Stefan Müller · Franz-Georg Hanisch · Wolfram Kress · Benedikt Schoser

Received: 12 May 2014 / Accepted: 2 June 2014 / Published online: 14 June 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract We identified the first homozygous and hence recessive mutation in the myotilin gene (*MYOT*) in a family affected by a severe myofibrillar myopathy (MFM). MFM is a rare, progressive and devastating disease of human skeletal muscle with distinct histopathological pattern of protein aggregates and myofibrillar degeneration. So far, only heterozygous missense mutations in *MYOT* have been associated with autosomal dominant myofibrillar myopathy, limb-girdle muscular dystrophy type 1A and distal myopathy. Myotilin itself is highly expressed in skeletal and cardiac muscle and is localized at the Z-disc and therefore interacts in sarcomere assembly. We performed whole-exome sequencing in a German family clinically diagnosed with MFM and identified a homozygous mutation in exon 2, c.16C>G (p.Arg6Gly). Using laser microdissection followed by quantitative mass spectrometry, we identified the myotilin protein as one component showing the highest increased abundance in the aggregates in the index patient. We suggest that the combined approach has a high potential as a new tool for the confirmation of unclassified variants which are found in whole-exome sequencing approaches.

Keywords Myotilin · Myofibrillar myopathy · Recessive muscle disease · Whole-exome sequencing

Introduction

Myofibrillar myopathies (MFM) are progressive and devastating neuromuscular diseases of human skeletal muscle which often lead to severe muscle weakness and even to premature death. There are distinct histopathological features in the muscle biopsy like desmin-positive protein aggregates and degeneration of myofibrils [1–5]. While part of the MFM are caused by mutations in genes encoding sarcomeric and extra-sarcomeric proteins (desmin, filamin C, FHL1, plectin, valosin-containing protein (VCP), Z-band alternatively spliced PDZ-motif-containing protein (ZASP), BAG 3, alpha B-crystallin and myotilin), more than 50 % of cases are caused by still unresolved gene defects [1].

Myotilin is a 57-kDa protein with 10 exons encoded by the myotilin gene (*MYOT*, OMIM# 609200) on chromosome 5q31.2. It is highly expressed in the skeletal muscle and less in the cardiac muscle, and localizes to the Z-disc, and therefore is important in sarcomeric assembly and actin stabilization [6]. Myotilin contains a serine-rich region, two immunoglobulin-like- and a carboxy-terminal domain [6, 7]. There are several Z-disc-associated binding partners like actinin, filamin C and ZASP [6, 8, 9]. Only heterozygous missense mutations in *MYOT* have been reported so far in MFM, limb-girdle muscular dystrophy type 1A and distal myopathy [3, 10–17]. Except for one mutation c.1214G>A (p.Arg405Lys) in exon 9, almost all other previously reported *MYOT* mutations are located in exon 2 with c.179C>T (p.Ser60Cys) being the most common one.

In this report, we describe the first recessive acting homozygous mutation in the myotilin gene which affects a family with severe myofibrillar myopathy by performing whole-

Electronic supplementary material The online version of this article (doi:10.1007/s10048-014-0410-4) contains supplementary material, which is available to authorized users.

J. Schessl (✉) · S. Feldkirchner · C. Kubny · B. Schoser
Friedrich-Baur-Institute, Department of Neurology,
Ludwig-Maximilians University of Munich, Ziemssenstr. 1 A,
80336 Munich, Germany
e-mail: jschessl@med.lmu.de

E. Bach · S. Rost · W. Kress
Department of Human Genetics, Julius-Maximilian University,
Würzburg, Germany

S. Müller · F.-G. Hanisch
Central Bioanalytics, Center for Molecular Medicine Cologne,
University of Cologne, Köln, Germany

exome sequencing and a proteomic approach. Using laser microdissection, we were able to confirm the mutated protein identified via mutational analysis, due to its prominent abundance in the aggregates in the muscle biopsy of the patient.

Results

Case report

This 44-year-old male patient noticed first symptoms at the age of 38 years with progressive muscle weakness of the proximal legs with difficulties climbing stairs and rising from a chair. Muscle strength testing (MRC) demonstrated 4/5 paresis of the quadriceps femoris, iliopsoas, hamstring and gastrocnemius muscles. The upper extremities were not affected, and he did not report any swallowing problems or ocular deficits. Sensory examination and mental status were normal. CK levels were elevated up to 1,143 U/l (normal <190 U/l). The family history revealed that the parents and one brother were reported healthy. One sister was bedridden at age 35 due to a severe and progressive myopathy (Fig. 2b). Electromyography describes myopathic changes in the leg muscles without spontaneous activities. The nerve conduction studies were normal. An MRI of the upper leg muscles showed fatty changes in the vastus muscles. Cardiac investigations revealed sinus arrhythmia but no functional or structural involvement. The muscle biopsy taken from the medial head of the gastrocnemius muscle showed a severe degenerative myopathy with abundant fibres with pathological protein aggregates (Fig. 1a, b, c). This protein aggregates were positively stained with anti-desmin and anti-alpha B crystalline antibodies (Suppl. Fig. 1A,B) Electron microscopy revealed predominantly myofibrillar destruction (Fig. 1f, g).

Whole-exome sequencing, multiplex ligation-dependent probe amplification and marker analysis

Sequencing of the whole exome of four family members (both affected siblings, their healthy mother and brother) revealed 14.2 to 24.8 million passed chastity filter read pairs per sample. Prior to specific filtering, about 150,000 to 200,000 variants (SNPs and indels) were detected in each analysed exome. After quality filtering, all variants of both affected siblings were compared for shared variants using GensearchNGS. A total of 50 unknown variants were concurrently present in the index patient and his affected sister in a homozygous form. The most conspicuous homozygous variant was present in the myotilin gene (*MYOT*) in which mutations are known to be associated with myofibrillar myopathy (MFM) and limb-girdle muscular dystrophy type 1A (LGMD1A). The missense mutation c.16C>G, p.Arg6Gly in exon 2 of *MYOT* was confirmed for both patients by

Sanger sequencing. It could also be detected in the mother, the healthy brother as well as in the healthy daughter of the female patient but in a heterozygous form classifying them as carriers (Fig. 2a). The father was not available for examination because he had already died. Algorithms that predict the effect of amino acid substitutions on protein function classified the missense mutation c.16C>G, p.Arg6Gly as pathogenic (SIFT: deleterious, score 0.00; PolyPhen-2: probably damaging, score 1.00; Mutation Taster: disease causing, *p* value 1.0). According to Alamut (Interactive Biosoftware), both nucleotide c.16C and amino acid arginine-6 are highly conserved (considering 11 species). Furthermore, the substitution has neither been reported yet as a polymorphism in dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) nor as a causative mutation in any available disease mutation database (for example, LOVD: Leiden Muscular Dystrophy pages or HGMD: Human Gene Mutation Database). None out of 186 examined anonymous blood donors from Central Europe showed the c.16C>G transition in *MYOT*. However, an adjacent nucleotide exchange (c.17G>A) resulting in an amino acid substitution of the same arginine at position 6 (p.Arg6His) has been reported as disease causing in HGMD. Since the c.16C>G mutation was found in a homozygous form in both patients, we first performed multiplex ligation-dependent probe amplification (MLPA) analysis for the detection of potential deletions in the *MYOT* gene and subsequently microsatellite analyses in order to proof the region adjacent to *MYOT* for the presence of homozygosity. No deletions were detected in the *MYOT* gene, and both patients were homozygous for all 10 markers tested while the healthy mother and brother as well as the daughter of the affected female were heterozygous (Fig. 2b).

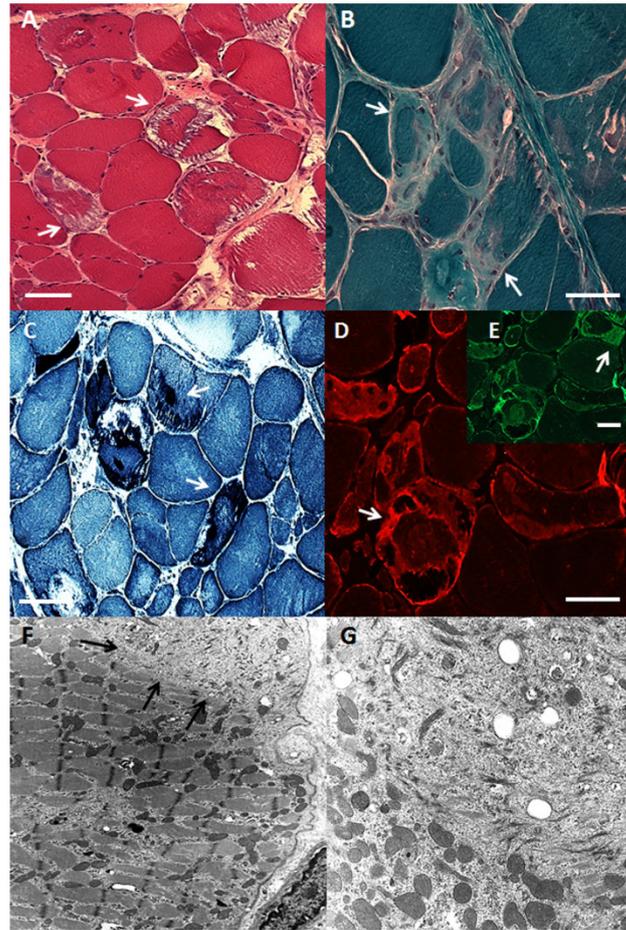
Myotilin expression is increased in patient muscle

Immunoblot analysis showed that the myotilin expression levels in the skeletal muscle from the patient were increased compared to control muscles (Suppl. Fig. 1C).

Myotilin level is increased in aggregates

For quantitative analysis of proteins in pathological protein aggregates, we performed iTRAQ-labelling combined with MALDI-TOF-TOF mass spectrometry of laser microdissected fibre material [18]. We compared the quantitative protein content of precisely excised plaque areas and morphologically normal fibres of the patient. The mass spectrometric results suggested that there was at least a 2.22-fold increased abundance of myotilin in the aggregates compared to controls (the regulation/accumulation factors measured in the iTRAQ technology are generally attenuated compared to those determined by label-free approaches). Known aggregate markers like filamin C (1.50-fold), desmin (2.31-fold) and α -B crystallin

Fig. 1 Muscle pathology of recessive myotilinopathy. Muscle biopsy taken from the gastrocnemius shows a moderate degenerative myopathy with some subsarcolemmal vacuolisations and opaque subsarcolemmal masses in H&E stains (*arrows*) (a). Trichrome-stained sections (b) reveal subsarcolemmal and intrasarcoplasmic aggregate material (*arrows*). NADH stain shows the abundance of the aggregate throughout the biopsy (*arrows*) (c). Immunohistochemistry with monoclonal antibodies against desmin (d) shows positive staining in core-like lesions and some subsarcolemmal accumulation (*arrows*). Identical aggregates are positive-stained by alpha-B crystalline (E, Suppl. Figure 1). Bars in (a–e) adjusted to 45 μ m. Electron microscopic analysis shows Z-line disorganisation and Z-line streaming. Moreover, subsarcolemmal myofibrillar destruction without classic granulofilamentous material was found in an irregular outer sarcomeric orientation (*arrows*) (f, g). Magnification factor b $\times 8,800$; c $\times 10,500$



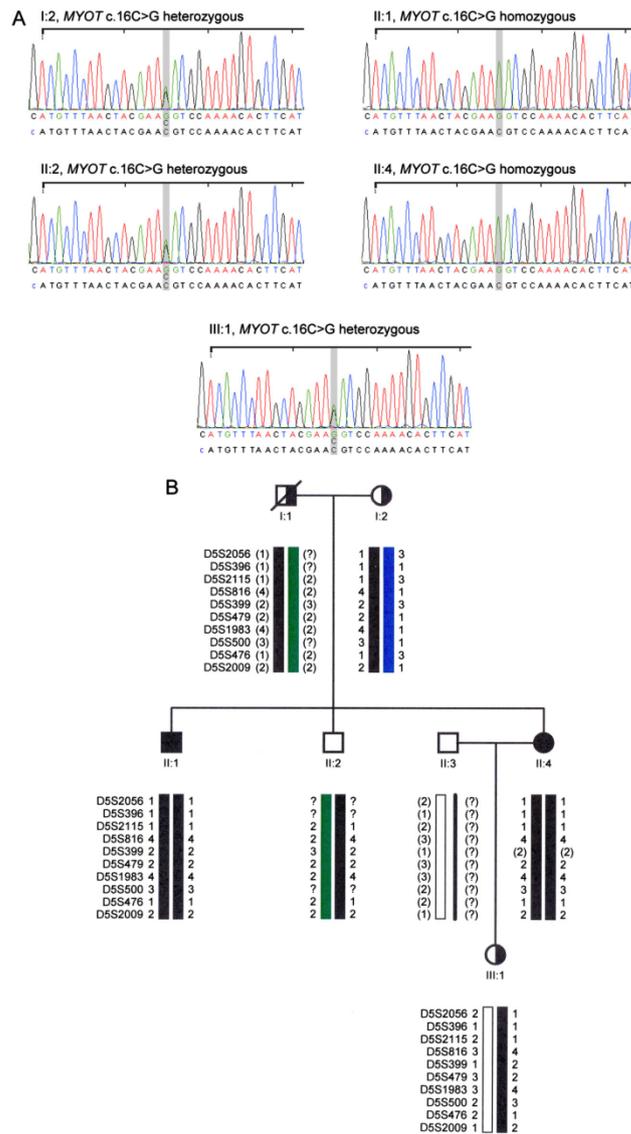
(2.53-fold) were increased in aggregates compared to morphologically normal fibres of the patient (Suppl. Tab. 1).

Discussion

We describe the first homozygous mutation in the myotilin gene leading to a novel, autosomal recessive subtype of myofibrillar myopathy (MFM). Performing whole-exome sequencing, followed by Sanger sequencing, we identified the c.16C>G (p.Arg6Gly) mutation in exon 2 affecting a German family with severe MFm. This mutation has not been reported so far and was not found in 372 chromosomes of healthy

blood donors, thus excluding a common polymorphism. All bioinformatics tools applied classified the detected substitution as pathogenic. A large deletion in the *MYOT* gene could be ruled out by MLPA analysis while microsatellite analysis clearly showed homozygosity for the region adjacent to *MYOT* in both patients, indicating a shared haplotype of the parents probably inherited by a common ancestor. Although consanguinity was not known for the parents, they originate from the same geographic area. An Arg-to-His substitution at the same amino acid position as the newly discovered mutation has been reported in a Turkish family with LGMD1A by Reilich et al., additionally supporting the causality of the detected p.Arg6Gly substitution in the present study [15]. The recently published mutation p.Arg6His was found in the

Fig. 2 Sequence analyses and pedigree of the family with haplotypes. **a** Sequence analyses of exon 2 of the *MYOT* gene in five family members. *Black letters* (below the chromatograms) indicate the wild-type sequences; *coloured letters* represent the corresponding text sequences of the chromatograms. The c.16C>G mutation is marked by *grey bars* in each chromatogram. **b** Pedigree of the family with haplotypes of the microsatellite analyses showing 10 different markers adjacent to the *MYOT* gene. *Numbers in brackets* indicate alleles which were reconstructed for the family members which were not available for analysis (I:1 and II:3); *question marks* represent alleles which could not be reconstructed or failed in the analysis



patients in a heterozygous form causing LGMD1A with late onset, rapid progression, loss of ambulation and respiratory failure. Since both mutations are detected at the same amino acid position in *MYOT*, a similar phenotype might be expected in the patients of both families. The difference in the zygosity status (heterozygous p.Arg6His in the Turkish family versus

homozygous p.Arg6Gly in the German family) could explain the more severe MFM phenotype of the patients in the present study. In detail, the two affected patients of the German family showed a severe and progressive muscle weakness and atrophy and elevated CK levels. The index patient's sister was bedridden since the age of 35 years. There was striking

intrafamilial variability of the clinical presentation with different onset of the disease and progression of the course of the muscle weakness. Muscle pathology presented a severe degenerative myopathy with abundant pathological protein aggregates. These histopathological features together with myofibrillar disorganization and intracellular accumulation of proteins associated to the Z-disc, including myotilin, are hallmarks of the homogeneous group of myofibrillar myopathies [1, 2, 4, 5]. Additional different proteins like cytoskeletal proteins, intermediate filaments, proteasome-related proteins and chaperones accumulate in these aggregates. Myotilin mutations have been associated with MFM, but only missense mutations in exon 2, most commonly the c.179C>T (p.Ser60Cys) mutation, has been reported so far [11, 14, 17, 19]. Missense mutations also have been identified to be causative for limb-girdle muscular dystrophy type 1A (LGMD1A) and distal myopathy phenotype [10, 12–16, 20]. The only *MYOT* mutation not found in exon 2, the c.1214G>A (p.Arg405Lys) mutation, located in exon 9 in the second immunoglobulin-like domain of myotilin was described to consequence in LGMD1A [20]. Interestingly, this p.Arg405Lys mutant results in a defective myotilin homodimerization and decreases the α -actinin interaction [20]. The detailed composition of the pathological protein aggregates in MFM is widely unknown. In a family with MFM caused by a mutation in the *FHL1* gene, the mutated protein identified via molecular diagnostics was confirmed by the proteomic approach due to its prominent abundance in the aggregates compared to the rest of the fibres and to morphologically normal fibres of the patients [21]. Also in patients with known filamin C mutations, filamin C was detected as highly abundant protein in the aggregates in these patients [22]. To understand the aggregate composition in the muscle biopsy of the family, we isolated the aggregates by laser microdissection and analysed the samples by proteomic downstream analysis. We identified myotilin in the aggregates with one of the highest abundance, being about 2.22-fold higher than in normal fibres of the patient. Western blot confirmed the high myotilin level in the muscle of the patient compared to healthy normal control muscle. Using this proteomic approach, we were able to confirm the mutation in the myotilin gene by its highly increased abundance in the aggregates of the patients' biopsy. This approach is not only an important diagnostic feature for the analysis of protein contents of pathological aggregates in muscle diseases but also an additional proof of mutational results. Moreover, this tool may help to identify new hereditary transmissions of mutations of known genes and therefore helps to understand pedigrees of families with rare diseases.

In conclusion, we identified the first recessive and homozygous mutation in the myotilin gene causing MFM by whole-exome sequencing and confirmed the mutational data by proteomic approach. This approach may help to understand

the protein composition of aggregates in MFM and protein contents in other myopathies and therefore helps to stratify unclear results from expensive whole-exome sequencing by identifying highly abundant proteins.

Acknowledgments We would like to thank the patients reported here for their participation and encouragement. We thank Ursula Klutzny and Maria Schmuck for the technical assistance. We also would like to thank Dr. Michael A. Hauser (Duke University Medical Center, Durham, NC, USA) for the myotilin antibody. J.S., F.G.H., W.K. and B.S. are supported by the German Research Association (DFG, FOR1228) and are members of the German network on muscular dystrophies (MD-NET funded by BMBF, Bonn, Germany; www.md-net.org). MD-NET is a partner of TREAT-NMD (EC, 6th FP, proposal no. 036825; www.treat-nmd.eu).

Disclosure The authors have reported no conflicts of interest.

References

- Schroder R, Schoser B (2009) Myofibrillar myopathies: a clinical and myopathological guide. *Brain Pathol* 19(3):483–492
- Selcen D (2011) Myofibrillar myopathies. *Neuromuscul Disord* 21(3):161–171
- Selcen D (2010) Myofibrillar myopathies. *Curr Opin Neurol* 23(5):477–481. doi:10.1097/WCO.0b013e32833d38b0
- Selcen D, Engel AG (2011) Myofibrillar myopathies. *Handb Clin Neurol* 101:143–154
- Selcen D, Ohno K, Engel AG (2004) Myofibrillar myopathy: clinical, morphological and genetic studies in 63 patients. *Brain* 127(Pt 2):439–451. doi:10.1093/brain/awh052
- Salmikangas P, Mykkanen OM, Gronholm M, Heiska L, Kere J, Carpen O (1999) Myotilin, a novel sarcomeric protein with two Ig-like domains, is encoded by a candidate gene for limb-girdle muscular dystrophy. *Hum Mol Genet* 8(7):1329–1336
- Salmikangas P, van der Ven PF, Lalowski M, Taivainen A, Zhao F, Suila H, Schroder R, Lappalainen P, Furst DO, Carpen O (2003) Myotilin, the limb-girdle muscular dystrophy 1A (LGMD1A) protein, cross-links actin filaments and controls sarcomere assembly. *Hum Mol Genet* 12(2):189–203
- Gontier Y, Taivainen A, Fontao L, Sonnenberg A, van der Flier A, Carpen O, Faulkner G, Borradori L (2005) The Z-disc proteins myotilin and FATZ-1 interact with each other and are connected to the sarcolemma via muscle-specific filamins. *J Cell Sci* 118(Pt 16):3739–3749
- van der Ven PF, Wiesner S, Salmikangas P, Auerbach D, Himmel M, Kempa S, Hayess K, Pacholsky D, Taivainen A, Schroder R, Carpen O, Furst DO (2000) Indications for a novel muscular dystrophy pathway. gamma-filamin, the muscle-specific filamin isoform, interacts with myotilin. *J Cell Biol* 151(2):235–248
- Berciano J, Gallardo E, Dominguez-Perles R, Garcia A, Garcia-Barredo R, Combarros O, Infante J, Illa I (2008) Autosomal-dominant distal myopathy with a myotilin S55F mutation: sorting out the phenotype. *J Neurol Neurosurg Psychiatry* 79(2):205–208
- Foroud T, Pankratz N, Batchman AP, Pauciulo MW, Vidal R, Miravalle L, Goebel HH, Cushman LJ, Azzarelli B, Horak H, Farlow M, Nichols WC (2005) A mutation in myotilin causes spheroid body myopathy. *Neurology* 65(12):1936–1940. doi:10.1212/01.wnl.0000188872.28149.9a
- Hauser MA, Conde CB, Kowalijow V, Zeppa G, Taratuto AL, Torian UM, Vance J, Pericak-Vance MA, Speer MC, Rosa AL (2002) Myotilin mutation found in second pedigree with LGMD1A. *Am J Hum Genet* 71(6):1428–1432

13. Hauser MA, Horrigan SK, Salmikangas P, Torian UM, Viles KD, Dancel R, Tim RW, Taiivainen A, Bartoloni L, Gilchrist JM, Stajich JM, Gaskell PC, Gilbert JR, Vance JM, Pericak-Vance MA, Carpen O, Westbrook CA, Speer MC (2000) Myotilin is mutated in limb girdle muscular dystrophy 1A. *Hum Mol Genet* 9(14):2141–2147
14. Olive M, Goldfarb LG, Shatunov A, Fischer D, Ferrer I (2005) Myotilinopathy: refining the clinical and myopathological phenotype. *Brain* 128(Pt 10):2315–2326
15. Reilich P, Krause S, Schramm N, Klutzny U, Bulst S, Zehetmayer B, Schneiderat P, Walter MC, Schoser B, Lochmuller H (2011) A novel mutation in the myotilin gene (MYOT) causes a severe form of limb girdle muscular dystrophy 1A (LGMD1A). *J Neurol* 258(8):1437–1444. doi:10.1007/s00415-011-5953-9
16. Penisson-Besnier I, Talvinen K, Dumez C, Vihola A, Dubas F, Fardeau M, Hackman P, Carpen O, Udd B (2006) Myotilinopathy in a family with late onset myopathy. *Neuromuscul Disord* 16(7):427–431
17. Vattermi G, Neri M, Piffer S, Vicart P, Gualandi F, Marini M, Guglielmi V, Filosto M, Tonin P, Ferlini A, Tomelleri G (2011) Clinical, morphological and genetic studies in a cohort of 21 patients with myofibrillar myopathy. *Acta Myol* 30(2):121–126
18. Feldkirchner S, Schessl J, Muller S, Schoser B, Hanisch FG (2012) Patient-specific protein aggregates in myofibrillar myopathies: laser microdissection and differential proteomics for identification of plaque components. *Proteomics* 12(23–24):3598–3609. doi:10.1002/pmic.201100559
19. Selcen D, Engel AG (2004) Mutations in myotilin cause myofibrillar myopathy. *Neurology* 62(8):1363–1371
20. Shalaby S, Mitsuhashi H, Matsuda C, Minami N, Noguchi S, Nonaka I, Nishino I, Hayashi YK (2009) Defective myotilin homodimerization caused by a novel mutation in MYOT exon 9 in the first Japanese limb girdle muscular dystrophy 1A patient. *J Neuropathol Exp Neurol* 68(6):701–707. doi:10.1097/NEN.0b013e3181a7f703
21. Feldkirchner S, Walter MC, Muller S, Kubny C, Krause S, Kress W, Hanisch FG, Schoser B, Schessl J (2013) Proteomic characterization of aggregate components in an intrafamilial variable FHL1-associated myopathy. *Neuromuscul Disord* 23:418–426
22. Kley RA, Maerkens A, Leber Y, Theis V, Schreiner A, van der Ven PF, Uszkoreit J, Stephan C, Eulitz S, Euler N, Kirschner J, Muller K, Meyer HE, Tegenthoff M, Furst DO, Vorgerd M, Muller T, Marcus K (2013) A combined laser microdissection and mass spectrometry approach reveals new disease relevant proteins accumulating in aggregates of filaminopathy patients. *Mol Cell Proteomics* 12(1):215–227

Supplementary Methods

Methods:

Clinical assessment

All blood and tissue samples were collected under Ludwig-Maximilians University Institutional Review Board protocols. Patients and family members were recruited after giving informed, written consent. Full neurological examination, including nerve conduction velocity (NCV) testing and electromyography (EMG), as well as pulmonary function tests and cardiac examination were performed. Serum creatine kinase (CK) levels were measured.

Muscle histology, immunohistochemistry and electron microscopy.

Standard histology, immunocytochemistry and electron microscopy protocols were used on muscle specimens taken from the medial head of gastrocnemius muscle from the patient as previously reported.[1-3]

Protein isolation and Myotilin Western blotting

Frozen muscle tissues from the patient and from one unaffected male individual were cut in slices of 10 μ m. 25 slices of each muscle biopsy specimen were collected for analysis. After lysis with Lysis-Buffer and homogenisation using a pestle performing additional heating steps (5 min at 95 C), cell lysate were centrifuged for 15 min at 14,000 rpm. Total cell lysate was frozen and stored at -20 C until use. Western blot analysis was performed using standard SDS-PAGE, immunoblotting, and detection procedures. For immunodetection anti-myotilin (generated in Dr. Michael A. Hauser's laboratory, Duke University Medical Center, Durham) as a primary antibody (diluted 1:7000) and HRP-conjugated secondary antibody (rabbit anti-mouse HRP (Dako #P0260, Glostrup, Denmark, diluted 1:10,000) were used.

Whole-exome and Sanger sequencing

The whole exome was sequenced in both affected siblings, their healthy mother and their healthy brother by next generation sequencing (NGS) on a HiSeq2000 (Illumina, San Diego, CA, USA) after target enrichment using the 62Mb TruSeq Exome Enrichment Kit (Illumina) according to manufacturer's instructions. Sequence reads (100bp paired-end reads) were mapped using the Burrows-Wheeler Aligner (BWA; <http://bio-bwa.sourceforge.net/>) after trimming and sorting of the raw data. Processed data were aligned to all chromosomes of the human reference sequence hg19 (GRCh37) and analyzed using the software Gensearch-NGS (PhenoSystems, Lillois, Belgium), which enables variant detection based on VarScan (<http://varscan.sourceforge.net/>). Detected variants of both patients were compared after filtering of the variants considering quality (Phred quality score > 20), coverage (> 5x), variant frequency (> 20% for hetero- and homozygous or > 80% for only homozygous variants), known/unknown status and biochemical effect of the variants (with highest priority on indels, non-synonymous, stop-gain and splice-site variants). Candidate variants were in-

spected using different bioinformatics tools provided by Alamut (Interactive Biosoftware) and verified by Sanger sequencing on an ABI 3130xl capillary sequencer (Applied Biosystems, Carlsbad, CA, USA) according to standard protocols. DNA samples of 186 healthy blood donors from Central Europe were screened for the presence of the detected MYOT gene variant by Sanger sequencing.

MLPA and marker analysis

DNA samples of the four family members and the daughter of the affected woman were analysed by multiplex ligation-dependent probe amplification (MLPA) for quantitative genomic changes in the MYOT gene using the P048-B2 probe mix (MRC Holland, Amsterdam, The Netherlands) according to manufacturer's instructions. MLPA PCR products were separated on an ABI 3130xl automatic sequencer (Applied Biosystems), pre-analysed by GeneMapper v.4.0 and finally evaluated by the Excel-based software Coffalyser v.5.2 (MRC Holland, Amsterdam, The Netherlands). Microsatellite markers located adjacent to the MYOT gene on 5q31.2 were chosen from the deCODE high resolution genetic map (<http://www.ncbi.nlm.nih.gov/mapview/>), primer sequences were taken from UniSTS (<http://www.ncbi.nlm.nih.gov/unists/>) and examined in five available family members using fluorescence-labeled primers and standard PCR conditions. Fragment analyses were performed on an ABI 3130xl capillary sequencer (Applied Biosystems) according to standard protocols. Data were evaluated using GeneMapper v.4.0 (Applied Biosystems) and haplotypes were constructed by Cyrillic v.2.1 (CyrillicSoftware, Oxfordshire, UK).

Laser microdissection and iTraQ-labeling combined with mass spectrometry

Detailed protocols of the used methods and techniques were recently published.[4]

Web Resources

UCSC Genome Browser: <http://genome.ucsc.edu>

ExpASy: <http://expasy.org/sprot/>

<http://www.ncbi.nlm.nih.gov/>

BWA: <http://bio-bwa.sourceforge.net/>

VarScan: <http://varscan.sourceforge.net/>

NCBI map viewer: <http://www.ncbi.nlm.nih.gov/mapview/>

UniSTS: <http://www.ncbi.nlm.nih.gov/unists/>

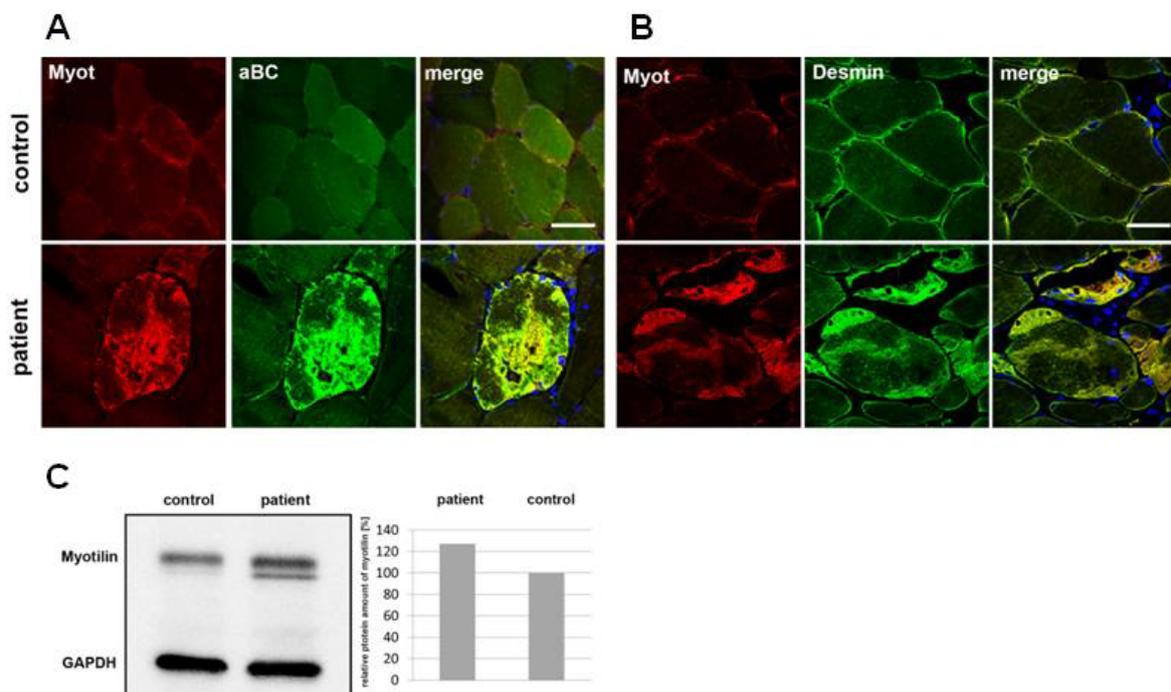
LOVD: <http://www.lovd.nl/3.0/home>

HGMD: <http://www.hgmd.org/>

References Methods Part

1. Dubowitz V, Sewry, C.A. (2007.). In: Muscle biopsy: a practical approach. . 3 edn. Saunders/Elsevier, Edinburgh,
2. Engel AG, Franzini-Armstrong, C. (2003) Myology. 3 edn. McGraw-Hill, New York
3. Windpassinger C, Schoser B, Straub V, Hochmeister S, Noor A, Lohberger B, Farra N, Petek E, Schwarzbraun T, Ofner L, Loscher WN, Wagner K, Lochmuller H, Vincent JB, Quasthoff S (2008) An X-linked myopathy with postural muscle atrophy and generalized hypertrophy, termed XMPMA, is caused by mutations in FHL1. *Am J Hum Genet* 82 (1):88-99. doi:S0002-9297(07)00010-9
4. Feldkirchner S, Schessl J, Muller S, Schoser B, Hanisch FG (2012) Patient-specific protein aggregates in myofibrillar myopathies: laser microdissection and differential proteomics for identification of plaque components. *Proteomics* 12 (23-24):3598-3609. doi:10.1002/pmic.201100559

Supplementary Figure 1



Suppl. Figure 1: Immunohistochemistry and Western Blot analysis.

The protein aggregates were positively stained with anti-myotilin (A,B), anti-alpha B crystalline (A), and anti-desmin (B) antibodies. (C) Immunoblot analysis showed that the myotilin expression levels in skeletal muscle from the patient was increased compared to control muscles.

Supplementary Table 1**Quantitative proteomics in Myotilin-associated myopathy**

PROTEIN		REGULATION FACTOR (AG/NK)
Protein	Name	patient
CRYAB	Alpha-crystallin B chain	2,5
SQSTM1	Sequestosome-1	2,4
RTN2	Reticulon-2	2,4
DESM	Desmin	2,3
MYOT	Myotilin	2,2
ACTC	Actin, alpha cardiac muscle 1	2,1
FLNA	Filamin-A	1,9
XIRP2	Xin actin-binding repeat-containing protein 2	1,7
XIRP1	Xin actin-binding repeat-containing protein 1	1,6
HSPB1	Heat shock protein beta-1	1,6
FLNC	Filamin-C	1,5
HSPB7	Heat shock protein beta-7	1,5

Suppl. Table 1: Proteomics (iTRAQ analysis): Quantitative proteomic data generated by differential analysis of precisely dissected aggregates (AG) vs. aggregate-free cells (NK) from patient with mutation in myotilin (technical replication in triplicate). Regulation (accumulation) factors referred to relative abundances of a protein species within aggregates vs. aggregate-free areas. Myotilin concentration was highly increased considering attenuation effects associated with iTRAQ technology (highlighted in bold).

Anlage 5**Bach JE**, Oldenburg J, Müller CR, Rost S*Mutational spectrum and identification of deep intronic variants by next generation sequencing in the factor VIII gene of haemophilia A patients*

Hämostaseologie 2015 [Accepted for publication]. doi.org/10.5482/HAMO-15-05-0017

Mutational spectrum and deep intronic variants in the factor VIII gene of haemophilia A patients

Identification by next generation sequencing

J. E. Bach¹; J. Oldenburg²; C. R. Müller¹; S. Rost¹

¹Department of Human Genetics, University of Würzburg, Germany; ²Institute of Experimental Haematology and Transfusion Medicine, University of Bonn, Germany

Keywords

Factor VIII, haemophilia A, mutation spectrum, next generation sequencing

Summary

Haemophilia A (HA) is caused by a broad spectrum of different mutation types in the factor VIII gene (*F8*). In our patient cohort of more than 2600 HA patients as well as in other published studies, the most frequent cause are missense mutations in different *F8* exons or the recurrent intron 22 inversion. Some exons and several specific nucleotide positions represent hot spots for point mutations in the examined cohort. About 4 % of cases remain without mutation after routine HA diagnostic methods including inversion PCRs, Sanger sequencing and multiplex ligation-dependent probe amplification

(MLPA). Deep intronic mutations cannot be detected by current standard HA diagnostics but have been reported for several genetic disorders. However, next generation sequencing (NGS) of the whole genomic sequence of the *F8* gene allows to identify deep intronic variants. **Conclusion:** In general, NGS provides an effective approach to screen for different HA causing mutation types in the *F8* gene.

Schlüsselwörter

Faktor VIII, Hämophilie A, Mutationsspektrum, Next-Generation-Sequenzierung

Zusammenfassung

Hämophilie A (HA) wird durch ein breites Spektrum verschiedener Mutationstypen im

Faktor-VIII-Gen (*F8*) verursacht. In unserem Patientenkollektiv von über 2600 HA-Patienten und in anderen veröffentlichten Studien sind Missense-Mutationen in verschiedenen *F8*-Exons sowie die rekurrente Intron-22-Inversion die häufigsten Ursachen der HA. Einige Exons und bestimmte Nukleotid-Positionen stellen Hot-Spots für Punktmutationen in der untersuchten Kohorte dar. Etwa 4% der Fälle bleiben jedoch ungelöst nach Durchführung der Routine-Diagnostik, die die Inversions-PCRs, Sanger-Sequenzierung und MLPA (Multiplex Ligation-Dependent Probe Amplification) beinhaltet. Tiefe intronische Mutationen können im Rahmen der gegenwärtigen Routine-Diagnostik nicht entdeckt werden, wurden jedoch im Zusammenhang mit verschiedenen genetisch bedingten Erkrankungen berichtet. Im Rahmen der Analyse der kompletten genomischen Sequenz des *F8*-Gens mittels NGS (Next Generation Sequencing) können solche tiefen intronischen Varianten identifiziert werden. **Schlussfolgerung:** Generell stellt NGS einen effektiven Ansatz dar, das *F8*-Gen auf verschiedene HA verursachende Mutationstypen hin zu untersuchen.

Korrespondenzadresse

J. Elisa Bach
Department of Human Genetics, University of Würzburg, Biocenter, Am Hubland,
97074 Würzburg, Germany
Tel. +49(0)931/318-4095; Fax -4069
elisa.bach@uni-wuerzburg.de

Mutationsspektrum und tiefe intronische Varianten im Faktor-VIII-Gen von Hämophilie-A-Patienten

Identifikation mittels
Next-Generation-Sequenzierung
Haemostaseologie Supplement 2016;36:
<http://dx.doi.org/10.5482/HAMO-15-05-0017>
received: May 12, 2015
accepted in revised form: November 27, 2015

Haemophilia A (HA, MIM#306700), the most common X-linked bleeding disorder with an incidence of 1 in 5000 male newborns, is associated with a deficiency of coagulation factor VIII (FVIII) caused by mutations in the factor VIII gene (*F8*) (1). The clinical severity of the disorder is subdivided into mild, moderate and severe according to the plasma levels of residual

FVIII protein activities of 5–40%, 1–5% and < 1%, respectively (2).

In the past 18 years, a cohort of 2671 male index patients has been screened for multiple mutation types in the *F8* gene by different diagnostic methods in our laboratory. In a first step each patient was tested for the recurrent intron 22 inversion occurring by intrachromosomal recombination

events between homologous extragenic sequences (3).

- In the first five years (1997–2002) this was performed by Southern blot (4),
- between 2003 and 2010 by single-tube polymerase chain reaction (PCR) according to Liu et al. (5) and
- since 2009 an inverse PCR assay published by Rossetti et al. was used (6).

- For two years both PCR methods were applied in parallel.

If no abnormality was detected, a PCR-based assay to detect the recurrent intron 1 inversion was performed since 2002 (7). Until mid-2008, patients negative for inversions were screened for variations in each exon by denaturing gradient gel electrophoresis (DGGE) before analyzing single conspicuous exons by Sanger sequencing (8). Since then, the complete coding sequences of the *F8* gene including 20 bp of flanking intronic sequence were routinely sequenced to search for point mutations. Since mid-2007, the last analysis step for patients without abnormalities identified was multiplex ligation-dependent probe amplification (MLPA) in order to screen for large deletions and/or duplications (9). In case of some mutations with presumed regulatory consequence (splice, promoter etc.), mRNA analyses were performed to confirm the mutation's effect (10). Meanwhile, different next generation sequencing (NGS) methods have become an effective approach for high-throughput genetic analyses (11) and have already been used for the analysis of the complete *F8* gene especially in patients in whom no causative mutations in the *F8* gene could be detected by the diagnostic methods described above (12).

A large spectrum of HA causing mutation types in the *F8* gene is known and recorded in different databases, the most frequently used was HAMSTeRS/HADB, the Haemophilia A Mutation, Structure, Test and Resource Site of the University College London (www.hadb.org.uk) which has only recently shifted to "The Factor VIII Variant Database" (www.factorviii-db.org/).

Mutation spectrum

The cohort of HA patients tested in our laboratory showed the mutational spectrum depicted (► Tab. 1). In about 44% of the cases a point mutation of the missense type was identified leading to an amino acid change in the FVIII protein. Most of them were found in exon 23, exon 14 and exon 11 with 158, 122 and 117 findings

Tab. 1 Mutational spectrum in all male HA index patients examined within 18 years: The table shows the absolute (n) and relative (%) number of findings of different mutation types in the *F8* gene causing HA in a total of 2671 patients analyzed between January 1997 and December 2014.

mutation type	n	%
missense mutation	1169	43.77
intron 22 inversion	678	25.38
small deletion	205	7.68
nonsense mutation	189	7.08
splice mutation	96	3.59
small insertion/duplication	81	3.03
large deletion	75	2.81
intron 1 inversion	30	1.12
synonymous/silent mutation	18	0.67
large duplication	17	0.64
promoter mutation	2	0.07
no mutation identified	111	4.16
total	2671	100

(14%, 10% and 10% of all missense mutations), respectively. The most frequent missense mutations were c.1636C>T, p.Arg546Trp in exon 11 (54 cases), c.5096A>T, p.Tyr1699Phe in exon 14 (39x) and c.6506G>A, p.Arg2169His in exon 23 (35x; all mutations are named according to HGVS nomenclature). As described by others (1, 13) a high percentage of the missense mutations occurred at CpG dinucleotides leading to an amino acid change of usually arginine to histidine or cysteine. In our patient cohort, this was the case in 52% of the 21 most frequent missense mutations.

Only for 646 out of 1169 patients with missense mutations, information on the severity of the phenotype was available:

- 310 of them showed mild HA (= 48% of patients with known severity),
- 182 moderate (= 28%) and
- 154 severe HA (= 24%).

Hence, the ratio of patients with missense mutations showing a mild phenotype versus patients with missense mutations showing a moderate to severe phenotype is about fifty-fifty.

The recurrent intron 22 inversion was found in a quarter of all patients, an intron 1 inversion only in about 1%. Besides classical intron 1 and 22 inversions also unusual rearrangements with an inversion in intron 1 or 22 combined with copy number variations (CNVs) could be detected and studied in detail in some cases (14).

Small deletions accounted for 7.7% of cases and nonsense mutations leading to a premature stop codon for 7.1%. Similarly to the missense mutations, a high frequency of nonsense mutations occurred at CpG sites or rather arginine codons.

The rest of the patients could be diagnosed with less frequent (< 4%) mutation types like splice mutations, small insertions, large deletions, synonymous mutations, duplications of several exons and a few regulatory mutations in the promoter region of the *F8* gene (9, 10, 15, 16).

Six of the potential splice mutations could be verified as pathogenic by mRNA analyses in leukocytes from the affected patients (10). Of the small deletions and insertions/duplications, 33.2% occurred at one of three poly-A stretches (A_n) in exon 14 known to be hot spots for HA causing frameshift mutations (13, 17). An adenine deletion could be found in 37 cases and an insertion in 15 cases in the A_9 stretch at position c.3629_3637. In the A_8 stretch at position c.4372_4379 we detected an adenine insertion in 17 cases and a deletion in 8 cases as well as five cases of a thymine deletion separating the A_9 from another two adenines (A_9TA_2). Finally, we identified 12 cases of an adenine insertion and one case of an adenine deletion in the A_6 stretch at position c.4820_4825. Interestingly, in most cases these mutations in poly-A stretches were found in severe HA patients but in some cases they led to an unexpectedly moderate phenotype presumably due to DNA replication errors as described by Young et al. (18).

The well-known silent mutation c.1569G>T, p.Leu523Leu (= p.Leu504Leu according to traditional nomenclature) which was described before as leading to the deletion of 36 bp of exon 11 or rather 12 amino acids (19, 20) was found in nine cases in our cohort. A further silent mutation (c.6108C>T, p.Tyr2036Tyr) detected in our patient cohort could be confirmed

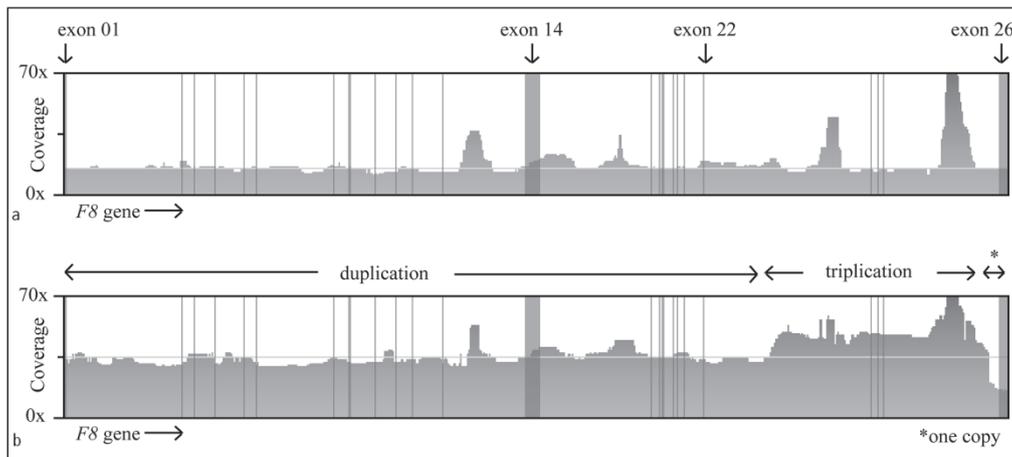


Fig. 1 Coverage of the whole *F8* gene in an individual with normal gene dosage (a) and in an individual with CNVs (b): Exons are marked by grey vertical bars. The light grey horizontal line indicates the mean coverage. Some regions considerably exceeded this average due to repeat regions in several *F8* introns that were overrepresented by the enrichment procedure. A typical coverage distribution over the complete *F8* gene with normal dose, i.e. one copy per exon, is given in (a) with a mean coverage of 21.

b) The coverage of patient 16 with severe HA and a known duplication of exons 1–22 and a triplication of exons 23–25 of the *F8* gene: The mean coverage is 36. The CNVs are apparent in comparison to the normal dose of (a), i.e. the shift from double to triple dose between exon 22 and 23 as well as from triple to normal dose between exon 25 and 26. The coverage of exon 26 is comparable to exon 26 of (a).

as causative mutation by mRNA analysis showing that skipping of exon 19 is the consequence of this synonymous variant (10). Large duplications comprising whole exons or almost the complete *F8* gene are very rare. Ten of the 17 large duplications detected in our cohort could be demonstrated by breakpoint analyses to be arranged in tandem (15). The least common mutation type in the *F8* gene are single base substitutions in the promoter region. We have found two promoter variants in our cohort which could both be confirmed as causative by luciferase assays (16).

In about 4% of our cases (111 patients) no mutation in the coding region of the *F8* gene could be detected by standard HA diagnostics. This group was composed of mainly mild and a few moderate phenotypes as severely affected HA patients could be diagnosed with a specific causative mutation in all cases.

Our mutational spectrum follows a similar pattern in comparison to the data of other reported HA patient cohorts (17) and databases like “The Factor VIII Variant Database” (<http://www.factorviii-db.org/>)

or the Human Gene Mutation Database (HGMD) available at <http://www.hgmd.cf.ac.uk/docs/login.html>.

Deep intronic mutations

In order to solve cases where no mutation in the coding regions of the *F8* gene could be found by routine HA diagnostics, several studies have searched for deep intronic mutations in the *F8* gene, mostly by mRNA analyses. As a result, several groups identified mutations in different *F8* introns creating *de novo* splice sites leading to the insertion or deletion of sequences in the normal *F8* mRNA and thus to alternative, pathogenic *F8* transcripts (21–24). Meanwhile, additionally or alternatively to mRNA analysis, next generation sequencing (NGS) has been used to screen the complete *F8* gene for disease causing variants. This has proven an effective way to analyze coding and non-coding *F8* regions for most mutation types including deep intronic mutations in a one-step sequencing approach (12). In our laboratory, we per-

formed a NGS study for the analysis of the entire *F8* gene in 16 male HA patients using a hybridization based enrichment strategy (SureSelect, Agilent) followed by sequence analysis on Illumina’s MiSeq desktop sequencer (25). Fifteen of them were unsolved cases with no alteration in the *F8* gene identified so far with mild and in one case moderate HA phenotypes. Patient 16 suffered from severe HA and was pre-diagnosed by MLPA with a duplication of exons 1–22 and a triplication of exons 23–25 of the *F8* gene. NGS data revealed a total of 23 deep intronic candidate variants in different *F8* introns, i.e. one to five variants in each of the patients. Three of the single nucleotide variants (SNVs) have been published before as pathogenic mutations causing HA: c.1537+325A>G and c.1538–18G>A in intron 10 and c.5587–93C>T in intron 16 (23, 24, 26), the others were unknown or listed in dbSNP but with minor allele frequencies (MAF) < 1 %. Different bioinformatic tools were used to score the potential deleterious effect of all candidate variants. Algorithms predicting potential *de novo* splice sites

suggested high splicing probabilities for several of the SNVs comparable to or even higher than predictions for the original splice sites. Another tool, the Combined Annotation Dependent Depletion (CADD) integrates various functional aspects and annotations into the so-called C-score indicating the potential deleteriousness of SNVs (27). The scores provided for published HA causing deep intronic mutations ranged between 0.006 and 8.0 and our deep intronic candidate variants were scored in a similar range of 0.006 to 8.7.

Additionally to five SNVs, the patient with moderate HA showed a deletion of 9.2 kb in intron 1. The known copy number variations (CNVs) of patient 16 with severe HA could be easily seen in the NGS data in a similar way as by MLPA (► Fig. 1). Yet, it was not possible to pinpoint the definite CNV breakpoints or to establish the order of the additional exons in the patient's NGS data.

Conclusion

Routine HA diagnostics with standard molecular methods succeeded in identifying the causative mutation in more than 95% of our HA patients. The mutation spectrum in our cohort of more than 2600 male HA patients is similar to other publications and gene-specific mutation databases. Missense mutations are the most common mutations in the F8 gene of our cohort as well as in other published patient cohorts and are distributed all over the complete factor VIII gene / protein with mutation hotspots mainly affecting arginine residues in particular exons. Several poly-A stretches constitute mutation hotspots for small insertions or deletions leading to a frameshift and in most but not all cases a severe HA phenotype.

In our cohort, 111 patients mainly suffering from mild HA showed no pathogenic variant in the F8 gene using the standard diagnostic methods described. We have analyzed 15 of the unsolved cases for deep intronic mutations by a NGS study targeting the complete F8 gene. We were able to detect deep intronic candidate variants in each of the unsolved cases. In an additional patient with known CNVs, the copy number variation was clearly detectable in

the NGS data. Candidate variants were *in silico* scored as potentially disease causing by way of generating *de novo* splice sites.

NGS represents an effective way to screen coding and non-coding regions of the F8 gene in a one-step sequencing approach with the potential to eventually replace the time-consuming Sanger sequencing and MLPA. Using NGS in routine HA diagnostics one could detect more than two thirds of HA mutations, as a broad spectrum of HA causing mutation types can be identified by this method. Point mutations and small insertions or deletions in coding and non-coding regions as well as hemizygous CNVs (large deletions and duplications) are easily detectable in NGS data. Still, complex rearrangements like the intron 22 and intron 1 inversions presumably have to be detected by traditional methods. Moreover, before reporting on deep intronic mutations in a diagnostic setting, functional studies like mRNA analyses would be reasonable in order to verify *in silico* predictions.

Conflict of interest

The authors declare that there is no conflict of interest.

References

- Graw J, Brackmann HH, Oldenburg J et al. Haemophilia A: from mutation analysis to new therapies. *Nat Rev Genet* 2005; 6: 488–501.
- Bolton-Maggs PH, Pasi KJ. Haemophilias A and B. *Lancet*. 2003; 361: 1801–1809.
- Naylor JA, Green PM, Rizza CR, Giannelli F. Analysis of factor VIII mRNA reveals defects in everyone of 28 haemophilia A patients. *Hum Mol Genet* 1993; 2: 11–17.
- Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet* 1993; 5: 236–241.
- Liu Q, Nozari G, Sommer SS. Single-tube polymerase chain reaction for rapid diagnosis of the inversion hotspot of mutation in hemophilia A. *Blood* 1998; 92: 1458–1459.
- Rossetti LC, Radic CP, Larripa IB, De Brasi CD. Genotyping the hemophilia inversion hotspot by use of inverse PCR. *Clin Chem* 2005; 51: 1154–1158.
- Bagnall RD, Waseem N, Green PM, Giannelli F. Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood* 2002; 99: 168–174.
- Higuchi M, Antonarakis SE, Kasch L et al. Molecular characterization of mild-to-moderate hemophilia A. *Proc Natl Acad Sci USA* 1991; 88: 8307–8311.
- Rost S, Löffler S, Pavlova A et al. Detection of large duplications within the factor VIII gene by MLPA. *J Thromb Haemost* 2008; 6: 1996–1999.
- Zimmermann MA, Gehrig A, Oldenburg J et al. Analysis of F8 mRNA in haemophilia A patients with silent mutations or presumptive splice site mutations. *Haemophilia* 2013; 19: 310–317.
- Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet* 2010; 11: 31–46.
- Pezeshkpoor B, Zimmer N, Marquardt N et al. Deep intronic 'mutations' cause hemophilia A. *J Thromb Haemost* 2013; 11: 1679–1687.
- Margaglione M, Castaman G, Morfini M et al. The Italian AICE-Genetics hemophilia A database. *Haematologica*. 2008; 93: 722–728.
- Zimmermann MA, Oldenburg J, Müller CR, Rost S. Unusual genomic rearrangements in introns 1 and 22 of the F8 gene. *Hämostaseologie* 2011; 31 (Suppl 1): S69–S73.
- Zimmermann MA, Oldenburg J, Müller CR, Rost S. Characterization of duplication breakpoints in the factor VIII gene. *J Thromb Haemost* 2010; 8: 2696–2704.
- Zimmermann MA, Meier D, Oldenburg J et al. Identification and characterization of mutations in the promoter region of the factor VIII gene. *J Thromb Haemost* 2012; 10: 314–317.
- Oldenburg J, El-Maarri O. New insight into the molecular basis of hemophilia A. *Int J Hematol* 2006; 83: 96–102.
- Young M, Inaba H, Hoyer LW et al. Partial correction of a severe molecular defect in hemophilia A, because of errors during expression of the factor VIII gene. *Am J Hum Genet* 1997; 60: 565–573.
- Tavassoli K, Eigel A, Wilke K et al. Molecular diagnostics of 15 hemophilia A patients. *Hum Mutat* 1998; 12: 301–303.
- El-Maarri O, Herbinaux U, Graw J et al. Analysis of mRNA in hemophilia A patients with undetectable mutations reveals normal splicing in the factor VIII gene. *J Thromb Haemost* 2005; 3: 332–339.
- Youssefian H, Kazazian HH Jr, Patel A et al. Mild hemophilia A associated with a cryptic donor splice site mutation in intron 4 of the factor VIII gene. *Genomics* 1988; 2: 32–36.
- Bagnall RD, Waseem NH, Green PM et al. Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two hemophilia A patients. *Br J Haematol* 1999; 107: 766–771.
- Castaman G, Giacomelli SH, Mancuso ME et al. Deep intronic variations may cause mild hemophilia A. *J Thromb Haemost* 2011; 9: 1541–1548.
- Inaba H, Koyama T, Shinozawa K et al. Identification and characterization of an adenine to guanine transition within intron 10 of the factor VIII gene as a causative mutation in a patient with mild hemophilia A. *Haemophilia* 2013; 19: 100–105.
- Bach JE, Wolf E, Oldenburg J et al. Identification of deep intronic variants in 15 hemophilia A patients by next generation sequencing of the whole factor VIII gene. *Thromb Haemost* 2015; 114.
- Santacroce R, Santoro R, Sessa F et al. Screening of mutations of hemophilia A in 40 Italian patients. *Blood Coagul Fibrinolysis* 2008; 19: 197–202.
- Kircher M, Witten DM, Jain P et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; 46: 310–315.

Anlage 6**Bach JE**, Wolf B, Oldenburg J, Müller CR, Rost S*Identification of deep intronic variants in 15 hemophilia A patients by Next Generation Sequencing of the whole factor VIII gene*

Thromb Haemost. 2015 Oct;114(4):757-67. doi: 10.1160/TH14-12-1011.

Identification of deep intronic variants in 15 haemophilia A patients by next generation sequencing of the whole factor VIII gene

J. Elisa Bach¹; Beat Wolf^{2,3}; Johannes Oldenburg⁴; Clemens R. Müller¹; Simone Rost¹

¹Department of Human Genetics, University of Würzburg, Würzburg, Germany; ²University of Applied Sciences Western Switzerland, Fribourg, Switzerland; ³Department of Bioinformatics, University of Würzburg, Würzburg, Germany; ⁴Institute of Experimental Haematology and Transfusion Medicine, University of Bonn, Bonn, Germany

Summary

Current screening methods for factor VIII gene (*F8*) mutations can reveal the causative alteration in the vast majority of haemophilia A patients. Yet, standard diagnostic methods fail in about 2% of cases. This study aimed at analysing the entire intronic sequences of the *F8* gene in 15 haemophilia A patients by next generation sequencing. All patients had a mild to moderate phenotype and no mutation in the coding sequence and splice sites of the *F8* gene could be diagnosed so far. Next generation sequencing data revealed 23 deep intronic candidate variants in several *F8* introns, including six recurrent variants and three variants that have been described before. One patient additionally showed a deletion of 9.2 kb in intron 1, mediated by Alu-type repeats. Several bioinformatic tools were used to score the variants in comparison to known pathogenic *F8* mutations in order to predict

their deleteriousness. Pedigree analyses showed a correct segregation pattern for three of the presumptive mutations. In each of the 15 patients analysed, at least one deep intronic variant in the *F8* gene was identified and predicted to alter *F8* mRNA splicing. Reduced *F8* mRNA levels and/or stability would be well compatible with the patients' mild to moderate haemophilia A phenotypes. The next generation sequencing approach used proved an efficient method to screen the complete *F8* gene and could be applied as a one-stop sequencing method for molecular diagnostics of haemophilia A.

Keywords

Alternative splice sites, deep intronic variant, factor VIII, haemophilia A, next generation sequencing

Correspondence to:

J. Elisa Bach
Department of Human Genetics
University of Würzburg, Biocenter
Am Hubland, 97074 Würzburg, Germany
Tel.: +49 931 3189102, Fax: +49 931 3184069
E-mail: elisa.bach@uni-wuerzburg.de

Received: December 5, 2014
Accepted after March 17, 2015
Epub ahead of print: May 7, 2015
<http://dx.doi.org/10.1160/TH14-12-1011>
Thromb Haemost 2015; 114: ■■■

Note: The work was carried out at the Department of Human Genetics, University of Würzburg, Biocenter, Am Hubland, 97074 Würzburg, Germany.

Introduction

Deficiency of coagulation factor VIII activity leads to haemophilia A (HA), the most common X-linked bleeding disorder with an incidence of 1 in 5,000 males (1). According to residual activities (FVIII:C), HA is classified into severe, moderate and mild phenotypes with <1%, 1–5% and 5–40% activity, respectively (2, 3). Almost 50% of HA cases are severe, 10% and 30–40% of HA patients are affected by the moderate and mild form, respectively (4).

The factor VIII gene (*F8*) is located on chromosome Xq28 and spans 187 kb with 26 coding exons (transcript variant 1, ENST00000360256/NM_000132.3). The 25 introns span sizes of 207 bp to 32.4 kb (5). Intron 1 and intron 22 each represent a mutation hot spot for recurrent inversions occurring by intrachromosomal recombination between homologous extragenic sequences (6–8).

With frequencies of almost 30% of all HA patients and 50% of severe cases the intron 22 inversion constitutes the most frequent HA mutation type (2). The intron 1 inversion is seen in about 1% of cases. Missense mutations distributed over all exons are found

in almost 40% of HA patients and nonsense mutations in 9%. Small deletions or insertions account for 10% and large deletions as well as splice site mutations for 3% of cases each. However, in approx. 2% of cases none of the mutations mentioned above could be found (6).

In order to solve such cases, several studies have searched for deep intronic mutations in the *F8* gene, mostly by mRNA analyses revealing the creation of new donor splice sites and thus leading to the insertion of intronic sequences into mRNA and consequently to premature stop codons (9–13).

Meanwhile, high-throughput sequencing methods ("next generation sequencing", NGS) have become an affordable new standard for genetic analysis. Hence, the aim of this study was to screen the whole *F8* gene by a NGS approach for deep intronic mutations causing HA and to test the method for routine diagnostic purposes. We analysed a cohort of 14 patients with mild HA plus one moderate case in whom the intron 22 and intron 1 inversions as well as mutations in the coding *F8* sequence had been excluded before.

Materials and methods

Patient samples

Fourteen male patients with mild HA with residual activities of FVIII protein (FVIII:C) between 5% and 38% plus one male patient with moderate HA (FVIII:C 2%) were included in this study. These patients were chosen from the 110 unsolved cases out of more than 2,600 HA patients tested in our laboratory. This group was largely composed of mild phenotypes and a few moderate cases. By our standard molecular HA diagnostic procedures none of the known mutation types in the *F8* gene could be determined in any of these patients. Whereas in all HA patients with severe phenotypes we were able to identify mutations by standard diagnostic methods and hence, severe HA patients were not included in this study. Von Willebrand disease type 2N (Normandy) had been excluded by Sanger sequencing of exons 18–24 including 20 bp of flanking intronic sequence of the *VWF* gene in all 15 selected patients under standard conditions.

Informed consent for genetic analyses was given by all patients and ascertained family members.

Next generation sequencing (NGS)

The whole *F8* gene (transcript variant 1, ENST00000360256/NM_000132.3) with the complete 5' and 3' UTR and additional 1000 bp of 5' upstream sequence (ChrX: 154,064,070–154,252,028; GRCh37) was analysed using NGS. Target enrichment was performed with the SureSelect^{XT} System (Agilent Technologies, Santa Clara, CA, USA), according to the manufacturer's instructions. Agilent SureDesign was used to design the probes with probe tiling parameters as follows: 5× tiling density, no masking, maximum boosting performance. This resulted in a total of 9,417 probes and 178,464 bp total target region size, i.e. approx. 95% of the *F8* gene. Alu-type repeats were excluded in the probe design in order to avoid unspecific sequence alignments, repeat regions were identified with RepeatMasker version open-4.0.3 (14). As the SureSelect probes comprise 120 bp, only Alu regions longer than 90 bp were excluded and Alus separated by less than 40 bp of specific sequence were defined as one excluded region. Nevertheless, flanking probes covered at least 50 bp of the 5'- and 3'-end of excluded regions. NGS was performed on a MiSeq sequencer with a MiSeq Reagent Kit V2 and 150 bp paired end reads (Illumina Inc, San Diego, CA, USA), according to the manufacturer's protocols.

Bioinformatic analyses

GenSearchNGS v1.5 (PhenoSystems, Lillois, Belgium) was used for analysis of the NGS data. Raw data were aligned to the complete *F8* gene and to the whole X chromosome of the human reference sequence hg19 (GRCh37), respectively, and variants detected with a GenSearchNGS-specific algorithm. For variant identification, filters were set for variant frequency at >80% (for hemizygous variants), for balance of forward and reverse reads (>0, i.e. at

least one read of each direction) and for quality (Phred quality score >20). Coverage statistic tools of GenSearchNGS were used to compare the coverage distribution over the whole introns. Potential splice effects of candidate variants were evaluated by bioinformatic prediction tools embedded in Alamut version 2.3 (Interactive Biosoftware, Rouen, France): SpliceSiteFinder-like (Alamut-specific tool), Max-EntScan (15), NNSPLICE (16), GeneSplicer (17) and Human Splicing Finder (18) provide probability scores for the use of potential donor and acceptor splice sites (dss and ass, respectively), ESEfinder (19) and RESCUE-ESE (20) predict and identify sequences with exonic splice enhancer (ESE) activity serving as binding sites for several serine/arginine-rich (SR) proteins. A further tool, the Combined Annotation Dependent Depletion (CADD) provided another method to predict the functional impact of the observed variants by integrating different functional aspects and diverse annotations into a single outcome, the C-score (21, 22).

PCR and Sanger sequencing

Potential mutations detected in the NGS data were confirmed in patients and if available also their relatives by PCR-based Sanger sequencing using standard protocols. Long-range PCR (LR-PCR) was used to detect deletion breakpoints of patient 8 in 25 µl reactions with the Expand Long Range dNTPack (Roche Diagnostics, Basel, Switzerland) according to the manufacturer's instructions. Several primers surrounding the assumed breakpoints were tested (primer sequences on request). Cycle sequencing was performed with the BigDye Terminator v1.1 mix (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. Sequence data were generated on an ABI 3130xl capillary sequencer under standard conditions and analysed by CodonCode Aligner v5.0 (CodonCode Corporation, Centerville, MA, USA).

Haplotype analysis

Six short tandem repeat (STR) markers located in the introns (Int) of the *F8* gene (F8Int1, F8Int13, F8Int22.1, F8Int22.2, F8Int25.1, F8Int25.2) (23) and four polymorphic STR loci adjacent to the *F8* gene locus (ST14/DXS52 on Xq28, DXS9929/*MTM1* on Xq28, DXS8377 on Xq28 and *HPRT1* on Xq26) were examined in patient 2–4, 7, 9, 11, 12, 14 and 15. Standard PCR was performed using fluorescence-labelled primers. An ABI 3130xl capillary sequencer was used for fragment analyses under standard conditions and data were examined with GeneMapper v4.0 (Applied Biosystems).

Results

Patient and NGS data

Fifteen male patients with moderate or mild HA and FVIII:C of 2–38% were included in this study. NGS of the whole *F8* gene generated a mean of 1.2 M sequence pairs (150 bp paired end reads) for each patient (► Table 1).

Table 1: Clinical data and bioinformatic raw data of 15 haemophilia A patients. Factor VIII activities (FVIII:C) are given as percentage of normal activity. The next column shows the number of paired end reads obtained by sequence analysis on Illumina's MiSeq, the next two columns illustrate the number of individual *F8* variants of each patient in comparison to the other patients' variants and the number of variants shared by two or three patients. The last column shows the number of remaining candidate variants after further filtering the patients' individual and shared variants for MAF and misaligned sequences due to highly repetitive regions. These variants were then further analysed with bioinformatic tools (see Table 2).

Patient	Phenotype	FVIII:C	# Paired end sequences	# individual variants	# variants shared with ≥ 1 patient(s)	# candidate variants
1	mild	38% *	1,235,616	3	–	1
2	mild	20%	1,283,768	–	3 (with pat. 9)	3
3	mild	24% *	1,119,785	–	5 (with pat. 7 and 14)	2
4	mild	approx. 18%	1,408,911	11	21 (with pat. 12); 1 (with pat. 14)	2
5	mild	8%	1,170,802	1	–	1
6	mild	18%	916,316	2	–	1
7	mild	13%	1,191,714	1	5 (with pat. 3 and 14)	2
8	moderate	2%	1,530,003	5	–	5
9	mild	25% *	1,070,680	–	3 (with pat. 2)	3
10	mild	>5%	1,135,966	3	–	3
11	mild	14%	1,239,352	–	1 (with pat. 15)	1
12	mild	30%	1,029,023	5	21 (with pat. 4)	2
13	mild	10%	1,216,397	2	–	2
14	mild	35% *	1,106,491	1	1 (with pat. 4); 5 (with pat. 3 and 7)	2
15	mild	10%	1,152,700	–	1 (with pat. 11)	1

* average value of two measurements.

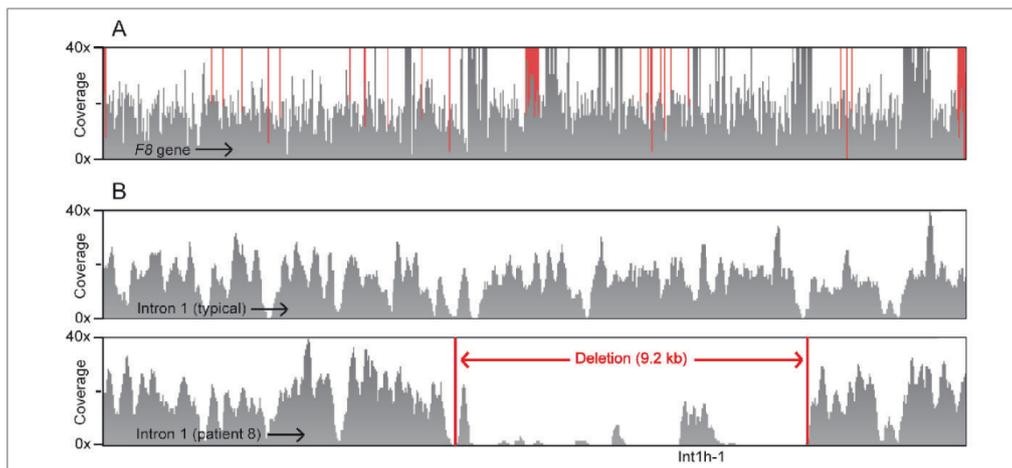


Figure 1: Coverage of the whole *F8* gene and intron 1 with deletion. Overview on the coverage distribution over the whole *F8* gene (A) and intron 1 (B) with a scale of 40x, pictures were obtained from an alignment of the NGS data to the complete X chromosome as reads of highly homologous repeat regions hereby were aligned more specifically onto X chromosomal-wide LINE/L1- and LTR-repeat sequences. A) A typical coverage pattern over the complete *F8* gene is given with exons marked in red. The average coverage was 23, several repeat regions were overrepresented with coverage maxima

beyond the scale of 40x. B) A typical coverage pattern of intron 1 is shown in comparison to the coverage of intron 1 of patient 8 showing a deletion of approx. 9.2 kb (breakpoint regions marked with red bars). In between the deleted part some reads were misaligned on the one hand owing to regions with highly homologous repeats and on the other hand owing to the Int1h-2 region which could be enriched and sequenced because of its high homology to the Int1h-1 region in the middle of the deleted part.

Alignment and coverage

At first, sequences were aligned to the *F8* gene reference sequence leading to a yield of approx. 48% aligned reads per patient. Coverage in the target region ranged broadly from 0 (for 0.3% of target regions) up to >32,000 with both extremes appearing in LINE/L1-, LTR- and other repeat-type sequences in the introns. Mean overall coverage was 921, the median coverage of the covered region was 14. On average, 85% of the target region was covered with >10 reads and 35% with >20 reads.

A second alignment to the complete X-chromosomal reference sequence provided a mean yield of 66% aligned sequences per patient as reads of highly homologous repeat regions could now be aligned more specifically onto LINE/L1- and LTR-repeat sequences distributed over the whole X chromosome. Coverage maxima now ranged from 793 in the *F8* gene up to approx. 3,000 reads in X-chromosomal LINE/L1- and LTR-type repeats. On average 99.5% of the target region was covered, among 79% with >10 and 25% with >20 reads, this constitutes a mean coverage of 23 and a median coverage of 13 of the covered region.

We compared the coverage distribution of each *F8* intron from each patient in order to identify potential copy number variations (CNVs) of intronic sequences. In most cases, coverage patterns in all introns of all patients were similar, but a large region of approx. 9–10 kb in intron 1 of patient 8 seemed to be deleted (► Figure 1). Based on the NGS data, both breakpoints were assumed to be located in Alu-type repeats. Within the deleted region several small sections were covered with a few unspecific reads (max. coverage 8) owing to inaccurately aligned sequences in SINE/Alu- or LTR-type repeats. The deleted region also encompassed the Int1h-1 sequence (7) which was covered by up to 17 reads. These were supposed to originate from the highly homologous Int1h-2 region upstream of the *F8* gene because the difference of 1 bp between Int1h-1 and Int1h-2 at position g.154,234,415T>G appeared in all reads.

Variants

On average, 561 variants were detected for each patient after an alignment against the *F8* gene reference sequence and filtering as described. An alignment to the complete X chromosome provided a mean of only 70 variants in the *F8* gene of each patient due to fewer reads aligned inaccurately onto repeat regions.

To look for variants unique to a patient and for variants shared by two or three patients, we compared the variant lists of the X-chromosomal alignment to each other leading to at least one and up to 21 variants per patient (► Table 1). After sorting out annotated polymorphisms with minor allele frequencies (MAFs) >1% and variants caused by unspecific alignments in LTR- and LINE/L1-repeats a total of 23 conspicuous variants remained (► Table 2). These variants were further analysed by splice prediction algorithms and CADD analysis providing a 'C-score' for the variants' deleteriousness (► Figure 2 and ► Figure 3, ► Table 2). Based on this, they were categorised in two groups: known pathogenic mutations (category 1; 3 variants) and variants suspected to

have a disease causing effect (category 2; 20 variants) if at least one of the bioinformatic tools predicted a certain degree of deleteriousness. The three mutations of category 1, c.1537+325A>G (patient 5), c.1538-18G>A (patient 13) and c.5587-93C>T (patient 6), have been described before as pathogenic mutations causing HA and verified by mRNA analyses (9, 11, 24). Six of the variants of category 2 are listed in dbSNP (25) with global MAFs <1%, with 0 to 5 hemizygotes observed in the data of the 1000 Genomes project. The other 14 variants have not been described before. According to the 1000 Genomes project (26) the *F8* gene regions flanking the unknown variants described here were sequenced in 1,654 alleles of presumed healthy control individuals. ► Figure 2 shows graphical displays of bioinformatic prediction analyses for six variants from category 2 (other variants see Suppl. Figure 1, available online at www.thrombosis-online.com). In most cases, several of the algorithms predicted an increase of the probability for the use of the mutated sequence as *de novo* donor (dss) or acceptor splice site (ass). Comparing the scores predicted for the variants to those of the nearest wildtype dss or ass, respectively, supports their use as alternative splice sites.

► Figure 3 shows the C-scores for the observed 23 variants of this project compared to 21 known causative mutations in the *F8* gene of different types. In accordance with the CADD analysis of all possible genome-wide substitutions (21), the known nonsense *F8* mutations showed the highest C-scores (36–48) followed by known missense (approx. 10–28), splice (approx. 5.4–10) and synonymous mutations (approx. 6.1–8.4). C-scores of published deep intronic variants ranged between 2.9 and 8.0. Known SNPs showed scores between 0.001 and 0.36. The 23 deep intronic variants found in this project were ranked with C-scores between 0.006 and 8.8, with the three published mutations of category 1 scored 0.006, 0.458 and 1.79, respectively.

Segregation analysis

Relatives were available for three patients: an affected brother, a sister (suspected to be carrier) and the mother (known carrier) of patient 4; the sister (carrier status unknown) and mother (carrier) of patient 9 and seven relatives of patient 12 (► Figure 4) – his daughter (carrier) and her affected son, the patient's mother, sister and niece (all carriers) as well as his niece's two daughters (one carrier, one healthy daughter). Within all three families the respective variants were shown to co-segregate correctly with the disease. The results would predict the sisters of patients 4 and 9 to be carriers.

Analysis of the intronic deletion of patient 8

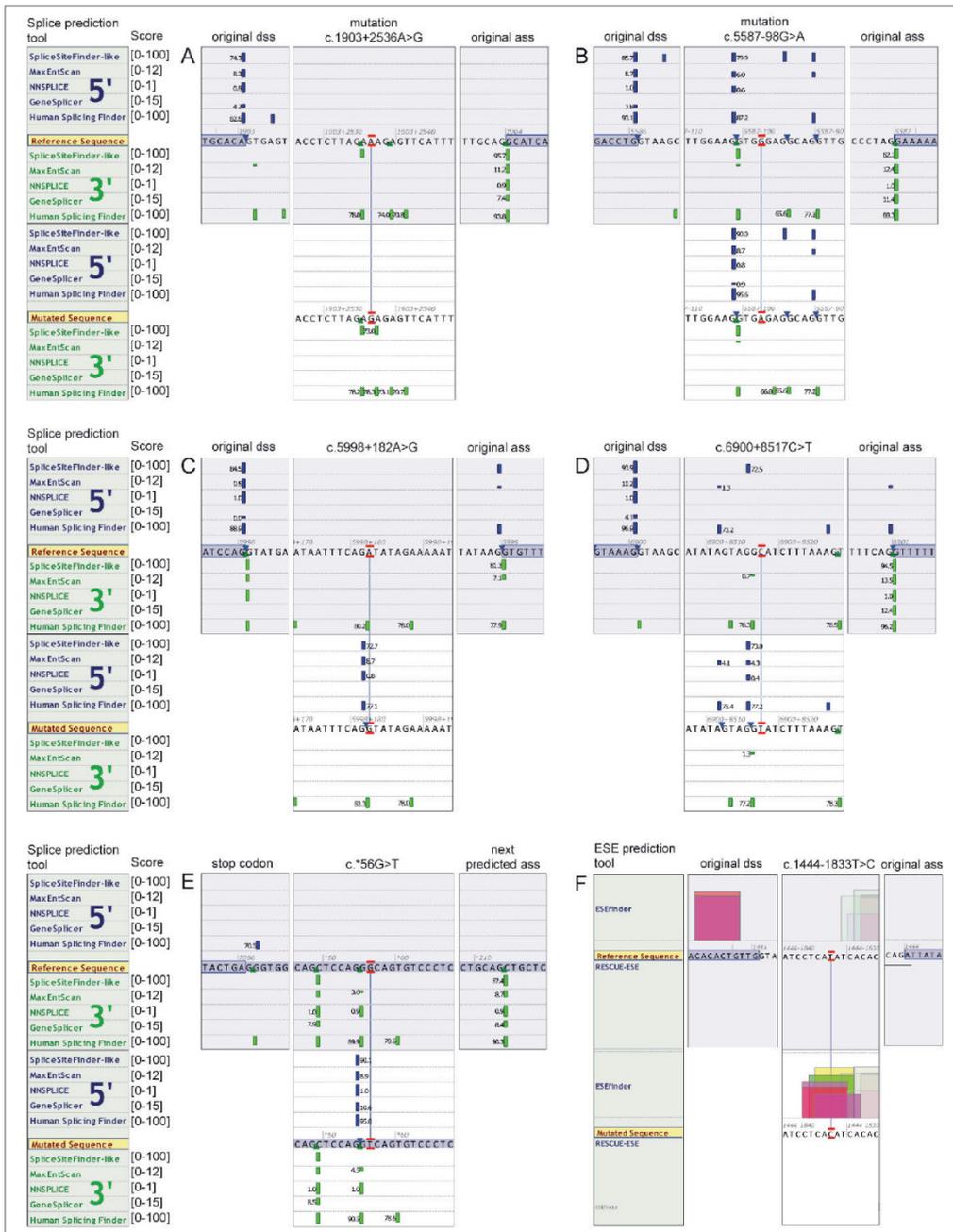
Both breakpoints of the deletion in intron 1 of patient 8 seemed to be located in repeat elements of the AluY-type with a homology of 90% as compared by BLAST search (27). As Alu repeats had been excluded from the targeted regions by our probe design, definite breakpoints were not covered in the NGS data. In order to characterise the breakpoints, we performed a series of LR-PCRs with several primer combinations. By sequencing the LR-PCR products,

Table 2: Intronic candidate variants found in the NGS data of the 15 HA patients after filtering out variants with MAF > 1 % and variants in misaligned repeat reads. Intron numbers, genomic and cDNA coordinates of the variants are given according to the official nomenclature of the Human Genome Variation Society (HGVS) using the *F8* gene transcript variant 1 (ENST00000360256/NM_000132.3) as reference sequence. Variants were divided in category 1 (known pathogenic mutations; bold) and category 2 (suspected to be causative; regular letters). If available, a reference number for known single nucleotide polymorphisms (rs SNP number) is given according to the Single Nucleotide Polymorphism Database (dbSNP) or an accession number of known pathogenic mutations according to the Human Gene

Mutation Database (HGMD) (39). Minor allele frequencies (MAFs) are indicated as given by the data of the 1000 Genomes project. The predicted effect of the variants on splice behaviour is shown as calculated by the bioinformatic tools embedded in Alamut software (a graphical display of variants marked with * is shown in Figure 2). A further prediction is given by the C-score based on the Combined Annotation Dependent Depletion (CADD) analysis predicting the deleteriousness of the observed variants by integrating different functional aspects and diverse annotations (21). The last column shows the results of a Sanger sequencing based segregation analysis of the variants of patients 4, 9 and 12 (n. a., not available). dss, donor splice site; ass, acceptor splice site.

Intron	Position	Variant	Patient	Reference	Alamut splice prediction	C-score	Segregation
1	g.154,238,632	c.144-10758_144-10757insTATA	12	–	de novo ass possible	8.095	YES
	g.154,233,589	c.144-5714C>T	10	–	de novo ass possible	0.476	n. a.
	g.154,231,575	c.144-3700C>T	8	rs138903905; MAF: ALL 0.5% (5 hemizygotes), EUR 0.2% (1 hemizygote)	de novo ass possible	1.858	n. a.
	g.154,229,134	c.144-1259C>T	4	rs184460929; MAF: ALL 0.2% (2 hemizygotes), EUR 0%	de novo ass possible	2.389	YES
6	g.154,209,741	c.787+3221A>G	2, 9	–	de novo ass possible	1.290	n. a.
7	g.154,195,804	c.1010-842G>A	3, 7, 14	rs148954517; MAF: ALL 0.7% (3 hemizygotes), EUR 2.1% (3 hemizygotes)	de novo ass possible	2.088	n. a.
9	*g.154,191,276	c.1444-1833T>C	10	–	de novo ESE motifs	8.749	n. a.
10	g.154,189,025	c.1537+325A>G	5	CS1211666; (11)	de novo dss possible	0.458	n. a.
	g.154,185,464	c.1538-18G>A	13	CS080683; (34)	de novo ass possible, weakens normal ass	1.790	n. a.
12	*g.154,179,631	c.1903+2536A>G	8	rs150944728; MAF: ALL 0.4% (3 hemizygotes), EUR 0%	de novo ass possible	0.380	n. a.
13	g.154,166,090	c.2114-6139C>T	10	–	de novo ass possible	0.130	n. a.
	g.154,164,332	c.2114-4382_2114-4381insG	3, 7, 14	rs200992472; (10); MAF: ALL 0.7% (3 hemizygotes), EUR 2.1% (3 hemizygotes)	de novo ass and dss possible	0.637	n. a.
	g.154,161,632	c.2114-1681T>A	13	–	de novo dss possible	0.236	n. a.
14	g.154,156,553	c.5219+293T>A	1	–	de novo dss possible	2.915	n. a.
16	*g.154,132,897	c.5587-98G>A	2, 9	–	de novo dss (and ass) possible	0.006	YES (pat.9)
	g.154,132,892	c.5587-93C>T	6	CS117569; (9, 10)	de novo dss possible	0.006	n. a.
18	*g.154,131,999	c.5998+182A>G	8	–	de novo dss (and ass) possible	4.658	n. a.
	g.154,131,182	c.5999-740T>C	4	–	no effect	3.328	YES
22	g.154,121,564	c.6429+2788T>C	12	rs28430439; MAF: ALL 0.1% (no hemizygotes), EUR 0%	de novo ass possible	6.540	YES
	g.154,096,327	c.6430-4825T>C	8	–	de novo ESE motifs	0.312	n. a.
25	*g.154,080,190	c.6900+8517C>T	8	–	de novo dss (and ass) possible	3.711	n. a.
	g.154,066,752	c.6901-725T>C	2, 9	–	de novo ass possible	2.285	YES (pat.9)
3'UTR	*g.154,065,816	c.*56G>T	11, 15	–	de novo dss (and ass) possible	5.270	n. a.

6 Bach et al. NGS of the whole F8 gene in 15 HA patients



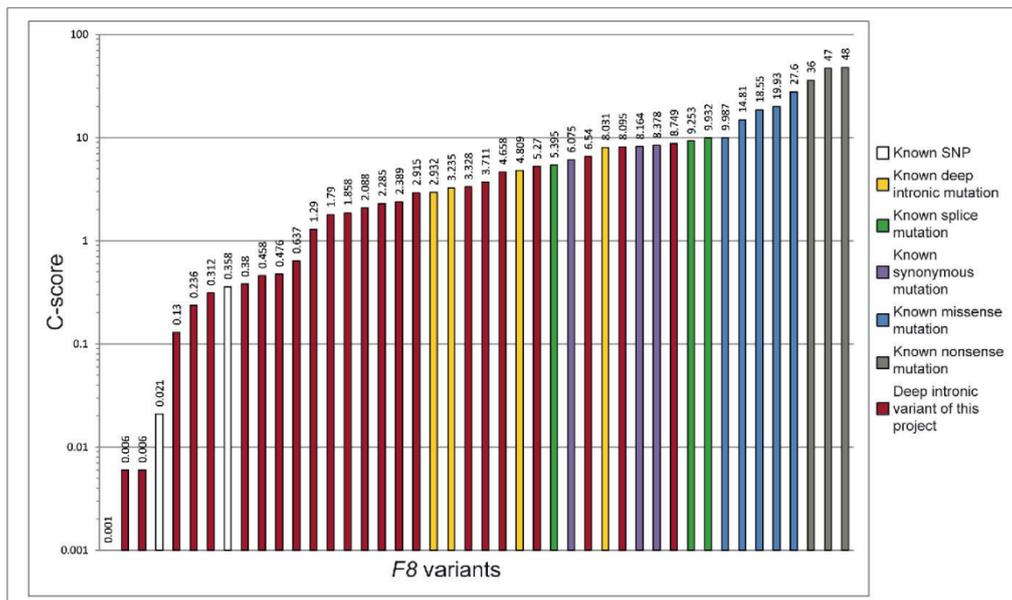


Figure 3: C-scores. C-scores with logarithmic scale of the 23 variants found in this study as well as three known SNPs, four known deep intronic mutations, three known splice mutations, three known synonymous mutations, five known missense and three known nonsense mutations causing haemophilia A. Scores were calculated based on the Combined Annotation Dependent Depletion (CADD) analysis predicting the deleteriousness of variants by combining different functional aspects and diverse annotations.

the breakpoints could be confined to a motif of 11 bp in the middle of AluY elements localising the deletion to g.154,232,081_154,241,265 or c.143+9420_144-4206del and thus a total of 9185 bp (► Figure 5).

Haplotype analysis

As some of the variants were recurrent in our patient cohort, we performed a haplotype analysis for patients sharing one or more variants and one healthy male control sample for 10 polymorphic STR loci in or adjacent to the *F8* gene in order to identify a poten-

tial genetic relationship (► Table 3). Based on this, a relationship of patient 4 and 14 and patient 3, 7 and 14 appears unlikely as different alleles were identified for at least one intragenic and several of the flanking *F8* markers. A relationship of patients 4 and 12 as well as 2 and 9 could not completely be excluded. Patient 4 and 12 differed in F8Int1, but shared the same alleles for the other tested intragenic STRs as well as 21 single nucleotide polymorphisms located between exon 6 and 26. Patient 2 and 9 shared the same alleles in all tested intragenic STRs, but there are differences in three of four adjacent markers. Patient 11 and 15 shared the same haplotype in all *F8* markers and for ST14/DXS52 (28). Even if there are differences in the remaining three flanking markers, a relationship was considered possible.

Discussion

We performed a NGS analysis of the complete *F8* gene (coding and non-coding sequences, with most of the Alu-type repeats, i.e. approx. 5% of the *F8* gene sequence excluded) in 15 HA patients in whom no causative mutation had been identified by routine molecular diagnostic workup.

As with all genomic sequences, data analysis was impeded by the ubiquitous occurrence of repeat elements in the human *F8*

Figure 2: Splice site prediction. Splice site predictions by Alamut software showing the simulated effect of six variants highly suspected to be deleterious for *F8* mRNA splicing. A-E) Probabilities for the use of five variants as de novo donor (dss/5') or acceptor splice site (ass/3') are indicated for five splice prediction tools with scores for the wild-type sequence on the top (highlighted grey) and mutated sequence beneath (white). Additionally, the scores for the original dss and ass lying upstream and downstream of the variant, respectively, are shown for A-D and the upstream lying stop codon and the next predicted ass in the 3'UTR region for E. F) Prediction of de novo binding motifs for exonic splice enhancers (ESE) is shown for two further tools (ESEfinder and RESCUE-ESE) for another variant as well as the corresponding original dss and ass.

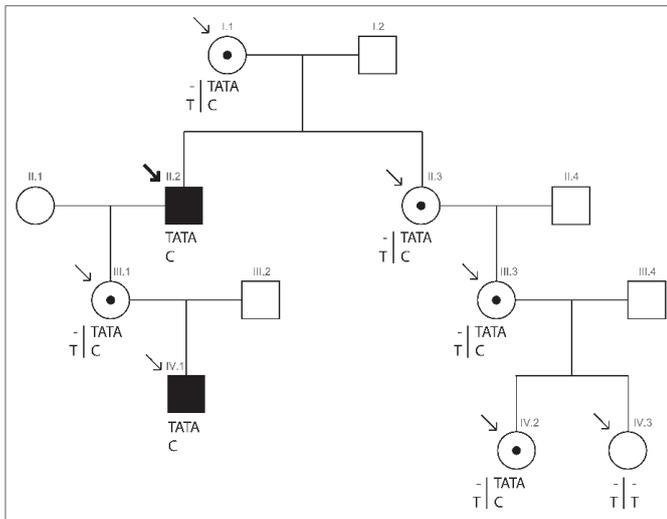


Figure 4: Pedigree of patient 12 with segregation of two variants. Segregation was analysed in patient 12 (II.2, bold arrow) and seven relatives (regular arrows). Affected males are indicated by a black square, carrier females by circles with black dots and healthy persons by blank squares or circles. The detected alleles at the position of the two variants c.144–10758_144–10757insTATA and c.6429+2788T>C are given for the tested family members: “-|-” indicates homozygous wild-type sequence, “TATA” hemizygotously affected persons and “-|TATA” heterozygotously affected persons for the variant at c.144–10785. “T|T” indicates homozygous wildtype sequence, “C” hemizygotously affected persons and “T|C” heterozygotously affected persons for the variant at c.6429+2788.

gene leading to a large number of falsely mapped reads at the expense of an even coverage. This problem could be partly overcome by an alignment of the raw data to the complete X chromosome instead of only the *F8* genomic sequence without influencing the coverage of non-repeat regions in the *F8* gene.

Our study identified 23 single nucleotide variants (SNVs) distributed over several introns and a deletion in intron 1 as potentially causative genomic alterations.

The 9.2 kb deletion of patient 8 was characterised with both breakpoints localised in AluY repeats (► Figure 5). Deletions in *F8* and other genes caused by Alu-mediated unequal homologous recombination have been described as disease causing before (29–32). In addition, the patient’s NGS data revealed five individual intronic SNVs (► Table 2). For c.5998+182A>G and c.6900+8517C>T bioinformatic tools consistently predicted a strong probability for a *de novo* dss (► Figure 2). Both variants in-

Table 3: Haplotype analysis with several X-chromosomal markers in or adjacent to the *F8* gene in four patient groups and a healthy male control. The table shows the results of the analysis of six short tandem repeat (STR) markers located in the introns (Int) of the *F8* gene and four polymorphic STR loci adjacent to the *F8* gene locus (markers are arranged in order of their appearance on the X chromosome from centromere to telo-

mere). Analysis was performed in order to look for a potential genetic relationship of patients 4 and 12, patients 3, 7 and 14, patients 2 and 9, patients 11 and 15 as well as patients 4 and 14 because each “group” shared one or more conspicuous variants. In each column the respective allele is given, alleles were enumerated each beginning with the shortest repeat.

	HPRT1	DXS8377	DXS9929/MTM1	ST14/DXS52	F8Int25.2	F8Int25.1	F8Int22.2	F8Int22.1	F8Int13	F8Int1
Pat. 4	3	8	2	1	1	1	1	2	3	2
Pat. 12	3	1	3	3	1	1	1	2	3	3
Pat. 3	3	6	1	2	1	2	1	2	2	2
Pat. 7	5	5	2	2	1	2	3	2	2	2
Pat. 14	7	4	3	2	1	2	1	1	2	2
Pat. 2	1	2	2	4	1	3	2	3	1	1
Pat. 9	6	3	2	5	1	3	2	3	1	1
Pat. 11	4	7	2	2	1	3	2	3	1	1
Pat. 15	2	8	1	2	1	3	2	3	1	1
Control	3	8	2	2	1	3	2	3	1	1

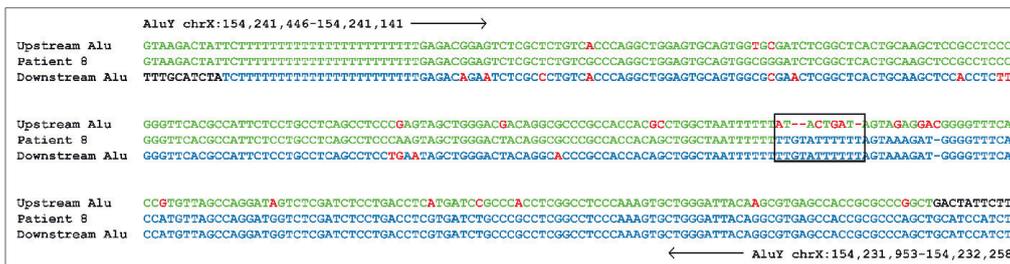


Figure 5: Sequences surrounding the breakpoints of the deletion in intron 1 of patient 8. The region of the breakpoint (motif of 11 bp) is framed. The above line shows the sequence of the Alu element 5' of the breakpoint, below is the patient's sequence surrounding the deletion breakpoints, the bottom line shows the sequence of the Alu element 3' of the deletion. Sequence identities with the upstream and downstream Alu are indicated with green and blue letters, respectively, discrepancies in red.

introduce new GT(GU) splice motifs, critical for the definition of the 5' end of a sequence that has to be removed during mRNA splicing (33). The predictions for c.144–3700C>T, c.1903+2536A>G and c.6430–4825T>C showed smaller effects on cryptic splicing (► Table 2, ► Figure 2) but for the first two variants at least two prediction scores increased at an AG splice motif in the mutated region enhancing the probability for a new ass. The third variant is scored less harmful, but nevertheless two new ESE motifs are predicted. Altogether, patient 8 carries a 9.2 kb intronic deletion containing the *Int1h-1* region known to be involved in the formation of intron 1 inversion and at least two intronic SNVs with strong hints for their use as *de novo* splice sites which may exhibit a modifying effect.

Patients 5, 6 and 13 could be diagnosed with a pathogenic intronic mutation each that had already been described before as leading to the inclusion of intronic sequences into the *F8* mRNA (9–11, 24, 34). This effect is also correctly predicted by the bioinformatic tools used in this project, i.e. two variants are strongly suggested to create a *de novo* dss (new GT-motif created) whereas the effect of the third one gave lower scores although a new AG-motif is created. The variant c.1538–18G>A was identified by Santacroce et al. in an Italian HA patient (34). Interestingly, patient 13 is also of Italian origin, but does not share the same haplotype with the patients of Santacroce et al. The patient carries a second individual variant, c.2114–1681T>A in intron 13 increasing the probability for a *de novo* dss which in combination with the variant in intron 10 may lead to alternatively spliced *F8* transcripts.

Three SNVs were shared by patient 2 and 9 who also share a *F8* haplotype (► Table 2 and ► Table 3). All variants co-segregated with the disease in the available family members of patient 9. Highest evidence for pathogenicity was scored for variant c.5587–98G>A which is strongly predicted to serve as dss (► Figure 2) and to create a new ESE-motif. It is located 5 bp upstream of the published causative mutation of patient 6 and both mutations are uniformly predicted as strong *de novo* dss and rated with the same C-score of 0.006. The other two variants, c.787+3221A>G and c.6901–725T>C, both may serve as *de novo* ass and induce an additional ESE-motif, though with lower scores.

The 3'UTR variant c.*56G>T was identified in patients 11 and 15 who also share a *F8* haplotype. The SNV introduces a new GT-motif predicted to serve as *de novo* dss with high scores (► Figure 2). In general, the 3'UTR is critical for diverse regulatory functions concerning gene expression (35). According to McGlinchy et al., an alteration in 3'UTR sequence may lead to NMD of mRNA and thus to reduced gene expression (36) which could explain the patients' reduced FVIII:C of 14% and 10%, respectively.

The variants c.1010–842G>A and c.2114–4382_2114–438 insG were recurrent in patients 3, 7 and 14 who do not share a *F8* haplotype. Both variants are annotated in the 1000 Genomes Project with overall MAFs of 0.7%. The G insertion in intron 13 was also found in two HA patients by Pezeshkpoor et al. Causality could not finally be proven, but the variant was not found in 288 healthy control alleles (10). A possible use as *de novo* dss or ass and an additional ESE-motif were predicted for the insertion and a *de novo* ass effect for the second variant which in combination may cause the mild HA phenotype of these patients.

In the data of patient 10, three individual intronic SNVs with possible effects on splicing were identified. Strongest consequence was predicted for c.1444–1833T>C with a high C-score of 8.75 and five *de novo* ESE-motifs (► Figure 2). Two bioinformatic scores indicated a possible *de novo* ass caused by c.2114–6139C>T whereas for the variant c.144–5714C>T only a slight increase of 1.1% and 1.5%, respectively, was predicted for two neighbouring *de novo* ass. Each of the three variants could slightly contribute to some amounts of alternative *F8* transcripts.

Patient 12 was diagnosed with two singular variants. The variant c.6429+2788T>C has already been found heterozygously in the 1000 Genomes Project with a global MAF of 0.1%, but not in the European population. The C-score of 6.54 was quite high, less conspicuous was the splice prediction with a slight probability increase for a *de novo* ass. The patient's second variant is the insertion c.144–10758_144–10757insTATA with an even higher C-score of 8.10. Splice prediction was again less conspicuous, only one of five scores predicted a *de novo* dss and ESEfinder identified a *de novo* ESE-motif. A further indication for the possible causality

of the variants was their perfect co-segregation with the disease status among seven tested family members, including his affected grandson and several obligatory female carriers (Figure 4). Interestingly, patient 4 and patient 12 shared 21 sequence variants which were not observed in any of the other 13 patients. All variants are known polymorphisms with MAFs of 9–38%. As the two patients also shared the same alleles for at least five STR markers within the *F8* gene this strongly indicates a shared haplotype.

Two SNVs were individual for patient 4. A slight increase for the probability of a *de novo* ass was predicted by one of five splice algorithms for c.144–1259C>T which has also been found in the 1000 Genomes Project with a global MAF of 0.2%, but not in the European population. No effect on correct splicing was predicted for the second variant c.5999–740T>C. Nevertheless, both variants showed correct co-segregation with the disease in the nuclear family.

Finally, the variant c.5219+293T>A in intron 14 was individual for patient 1. A *de novo* dss was predicted by one of five splice algorithms. ESEfinder and RESCUE-ESE both identified *de novo* motifs with enhancer activity. The C-score of almost 3 lies in the upper half of C-scores for all deep intronic variants tested.

In summary, this NGS approach proved to be an effective method to screen the complete *F8* gene for deep intronic variants in a selected cohort of 15 HA patients. In each of the patients analysed, at least one candidate variant was identified and *in silico* predicted to affect correct splicing. The sum of the detected variants in each patient may lead to the synthesis of several alternative transcripts which either undergo nonsense mediated decay (NMD) or result in truncated FVIII proteins and thus cause the patients' HA phenotypes. The prediction algorithms used proved their reliability when applied to the published causative variants identified in this paper and by other studies (37). Yet, it needs to be borne in mind that it is an *in silico* prediction. Functional assays to confirm the causality of the variants were not possible as this was a pseudonymised study. For three of the cases studied, segregation analyses in healthy and affected family members supported the

causality of the variants. Relatives of the other patients were not available.

Several patients showed an accumulation of presumably causative variants in their *F8* genes. This is further support to the observation that multiple mutations can occur simultaneously in the same gene in one individual and as a complex haplotype could contribute to a certain phenotype. It has been noted that the observed number of multiple mutations is higher than predicted for a random distribution of mutations (38). These authors also point out that multiple mutations could have either synergistic or opposing effects on the clinical phenotype. If so, monogenic inheritance could still be multi-allelic.

Our NGS approach resulted in a success rate of almost 100% for the identification of deep intronic variants in the *F8* gene of the examined HA patients. This may encourage other working groups to consider such a screening as worthwhile for patients lacking mutations in the *F8* coding regions. However, the identification and bioinformatic analysis of such single variants in non-coding regions of the *F8* gene is not a definite proof of their genetic cause of the disease unless functional experiments will show a modified expression of the factor VIII protein and/or its clotting activity. In a diagnostic setting, we would strongly suggest verification by functional assays like mRNA analyses before reporting such variants as causative. Yet it appears feasible in the near future to use NGS as the method of choice for molecular HA diagnostics as it allows for an affordable one-step sequence analysis of the entire *F8* gene. Before its application in routine diagnostics, however, we would recommend optimising the enrichment in terms of minimising probes for highly repetitive regions which were strongly overrepresented and thus hampering a straightforward data analysis.

Acknowledgements

We thank Rosi Bott, Simone Günther and Katja Schrooten for technical assistance.

Conflicts of interest

B. Wolf reports having received personal fees and non-financial support from Phenosystems <sa, outside the submitted work. None of the other authors declares any conflicts of interest.

References

- Graw J, Brackmann HH, Oldenburg J, et al. Haemophilia A: from mutation analysis to new therapies. *Nat Rev Genet* 2005; 6: 488–501.
- Oldenburg J, El-Maarri O. New insight into the molecular basis of hemophilia A. *Int J Hematol* 2006; 83: 96–102.
- Bolton-Maggs PH, Pasi KJ. Haemophilias A and B. *Lancet* 2003; 361: 1801–1809.
- Bogdanova N, Markoff A, Eisert R, et al. Spectrum of molecular defects and mutation detection rate in patients with mild and moderate hemophilia A. *Hum Mutat* 2007; 28: 54–60.
- Gitschier J, Wood WL, Goralka TM, et al. Characterization of the human factor VIII gene. *Nature* 1984; 312: 326–330.
- Oldenburg J, Ananyeva NM, Saenko EL. Molecular basis of haemophilia A. *Haemophilia* 2004; 10 (Suppl 4): 133–139.

What is known about this topic?

- In approx. 2 % of patients with haemophilia A (HA) no mutation in the coding region of the *F8* gene could be identified by means of standard molecular HA diagnostics.
- Deep intronic mutations in the *F8* gene are only rarely described so far.

What does this paper add?

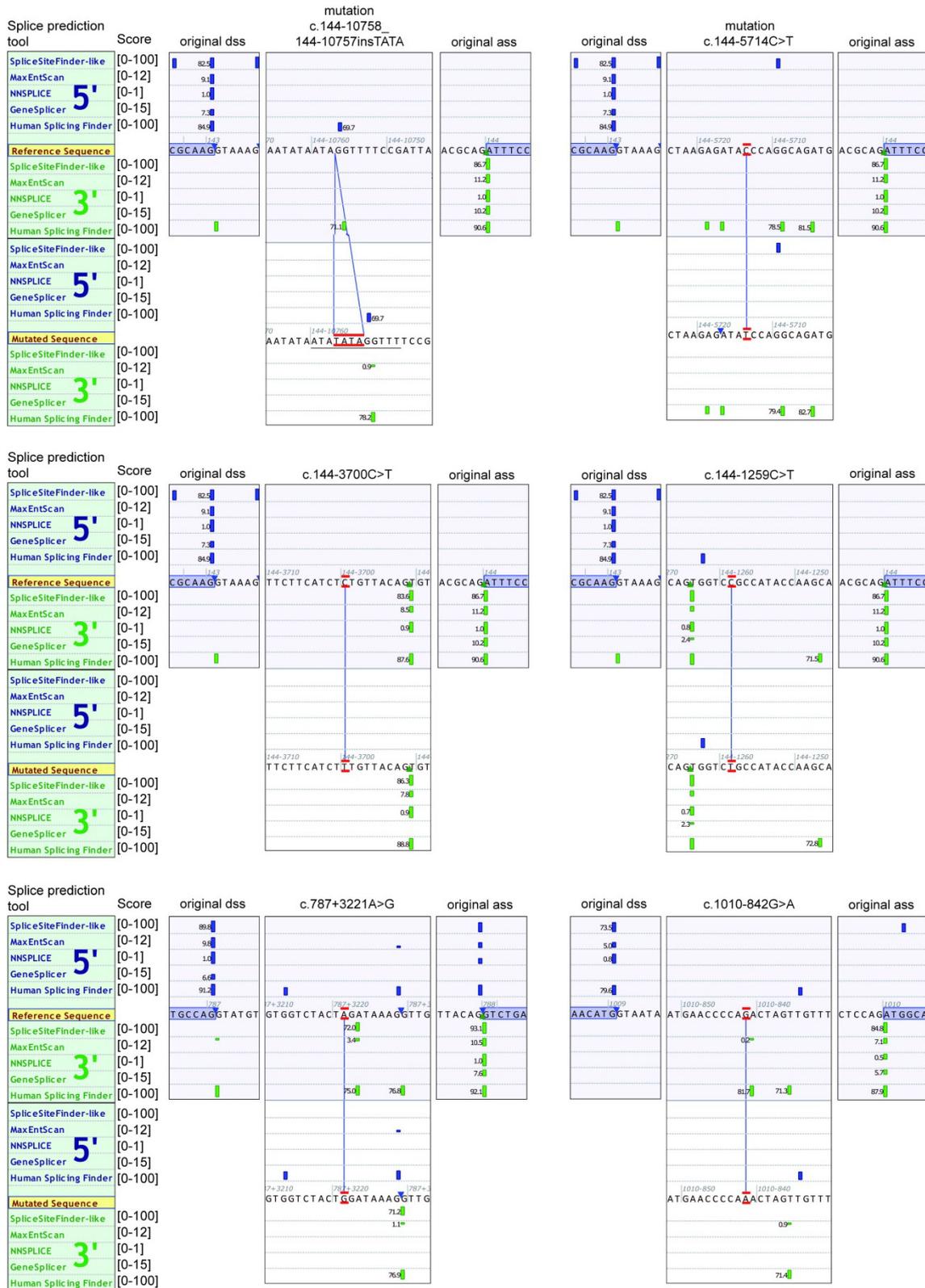
- In this study we screened the complete *F8* gene of HA patients without mutation in *F8* coding regions by next generation sequencing (NGS).
- By this we detected 23 conspicuous deep intronic variants and an intronic deletion of 9.2 kb.
- The NGS approach used in this study could be used after optimisation in routine diagnostics for analysis of the whole *F8* gene in one step.

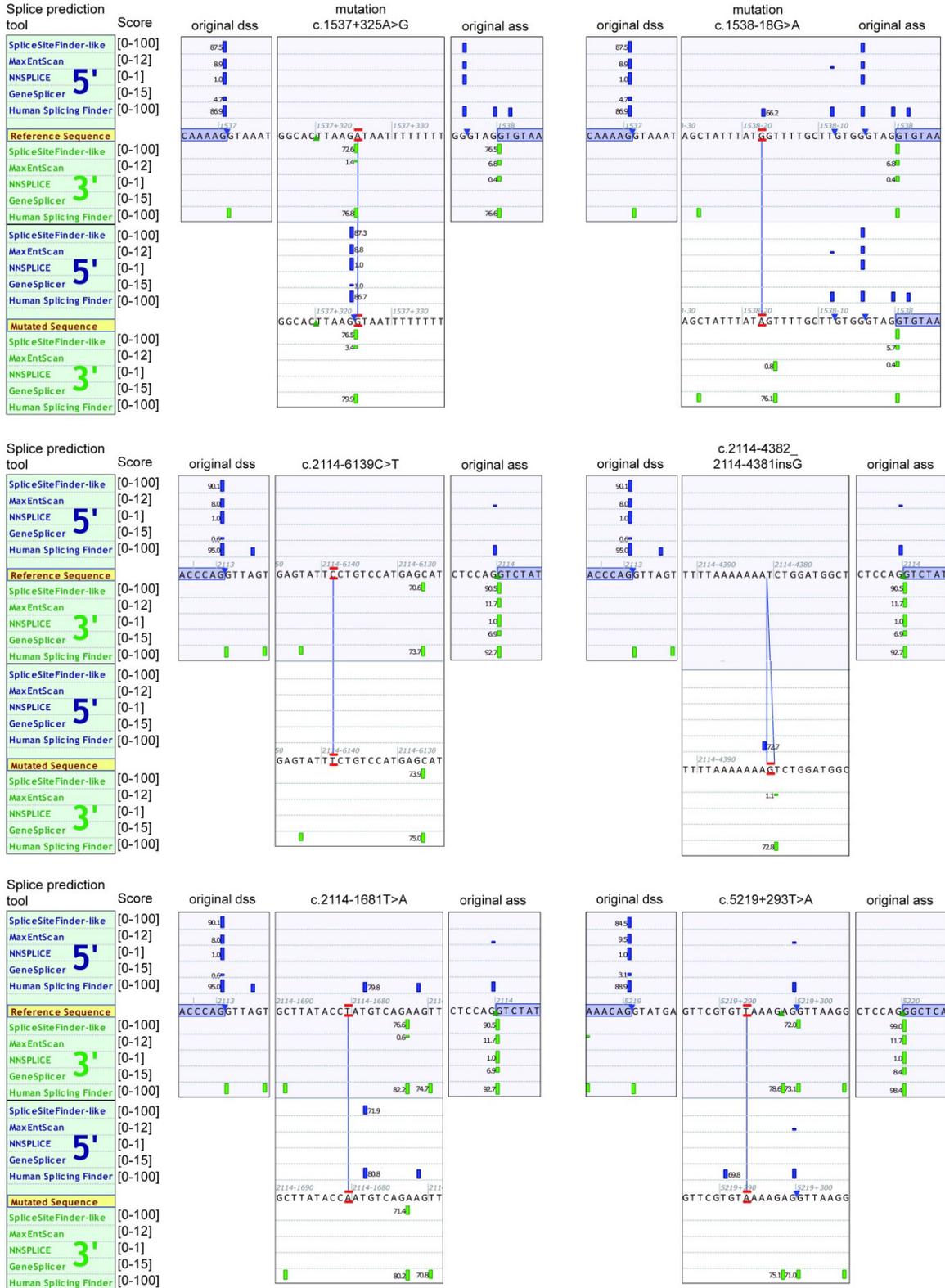
7. Bagnall RD, Waseem N, Green PM, et al. Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood* 2002; 99: 168–174.
8. Naylor JA, Buck D, Green P, et al. Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions. *Hum Mol Genet* 1995; 4: 1217–1224.
9. Castaman G, Giacomelli SH, Mancuso ME, et al. Deep intronic variations may cause mild hemophilia A. *J Thromb Haemost* 2011; 9: 1541–1548.
10. Pezeshkpoor B, Zimmer N, Marquardt N, et al. Deep intronic 'mutations' cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA. *J Thromb Haemost* 2013; 11: 1679–1687.
11. Inaba H, Koyama T, Shinozawa K, et al. Identification and characterization of an adenine to guanine transition within intron 10 of the factor VIII gene as a causative mutation in a patient with mild haemophilia A. *Haemophilia* 2013; 19: 100–105.
12. Bagnall RD, Waseem NH, Green PM, et al. Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients. *Br J Haematol* 1999; 107: 766–771.
13. Youssoufian H, Kazazian HH, Jr, Patel A, et al. Mild hemophilia A associated with a cryptic donor splice site mutation in intron 4 of the factor VIII gene. *Genomics* 1988; 2: 32–36.
14. Quinlivan RM, Muller CR, Davis M, et al. Central core disease: clinical, pathological, and genetic features. *Arch Dis Child* 2003; 88: 1051–1055.
15. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; 11: 377–394.
16. Reese MG, Eeckman FH, Knipf D, et al. Improved splice site detection in Genie. *J Comput Biol* 1997; 4: 311–323.
17. Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001; 29: 1185–1190.
18. Desmet FO, Hamroun D, Lalande M, et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; 37: e67.
19. Cartegni L, Wang J, Zhu Z, et al. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003; 31: 3568–3571.
20. Fairbrother WG, Yeh RF, Sharp PA, et al. Predictive identification of exonic splicing enhancers in human genes. *Science* 2002; 297: 1007–1013.
21. Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; 46: 310–315.
22. Nishio H, Sato T, Fukunishi S, et al. Identification of malignant hyperthermia-susceptible ryanodine receptor type 1 gene (RYR1) mutations in a child who died in a car after exposure to a high environmental temperature. *Leg Med* 2009; 11: 142–143.
23. Machado FB, Medina-Acosta E. High-resolution combined linkage physical map of short tandem repeat loci on human chromosome band Xq28 for indirect haemophilia A carrier detection. *Haemophilia* 2009; 15: 297–308.
24. Castaman G, Giacomelli SH, Mancuso ME, et al. F8 mRNA studies in haemophilia A patients with different splice site mutations. *Haemophilia* 2010; 16: 786–790.
25. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; 29: 308–311.
26. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491: 56–65.
27. Altschul SE, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990; 215: 403–410.
28. Yang YH, Song KS, Kim IK, et al. Rapid polymerase chain reaction analysis of SstI4 (DXS52) VNTR: carrier detection of hemophilia A. *J Obstet Gynaecol Res* 1997; 23: 399–406.
29. Vidal F, Farsac E, Tusell J, et al. First molecular characterization of an unequal homologous alu-mediated recombination event responsible for hemophilia. *Thromb Haemost* 2002; 88: 12–16.
30. Rossetti LC, Goodeve A, Larripa IB, et al. Homeologous recombination between AluX_s-sequences as a cause of hemophilia. *Hum Mutat* 2004; 24: 440.
31. Nakaya SM, Hsu TC, Geraghty SJ, et al. Severe hemophilia A due to a 1.3 kb factor VIII gene deletion including exon 24: homologous recombination between 41 bp within an Alu repeat sequence in introns 23 and 24. *J Thromb Haemost* 2004; 2: 1941–1945.
32. Deininger PL, Batzer MA. Alu repeats and human disease. *Mol Genet Metab* 1999; 67: 183–193.
33. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003; 72: 291–336.
34. Santacroce R, Santoro R, Sessa E, et al. Screening of mutations of hemophilia A in 40 Italian patients: a novel G-to-A mutation in intron 10 of the F8 gene as a putative cause of mild hemophilia A in southern Italy. *Blood Coagul Fibrinolysis* 2008; 19: 197–202.
35. Hughes TA. Regulation of gene expression by alternative untranslated regions. *Trends Genet* 2006; 22: 119–122.
36. McGlincy NJ, Smith CW. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem Sci* 2008; 33: 385–393.
37. Zimmermann MA, Gehrig A, Oldenburg J, et al. Analysis of F8 mRNA in haemophilia A patients with silent mutations or presumptive splice site mutations. *Haemophilia* 2013; 19: 310–317.
38. Chen JM, Ferec C, Cooper DN. Closely spaced multiple mutations as potential signatures of transient hypermutability in human genes. *Hum Mutat* 2009; 30: 1435–1448.
39. Cooper DN, Krawczak M. Human Gene Mutation Database. *Hum Genet* 1996; 98: 629.

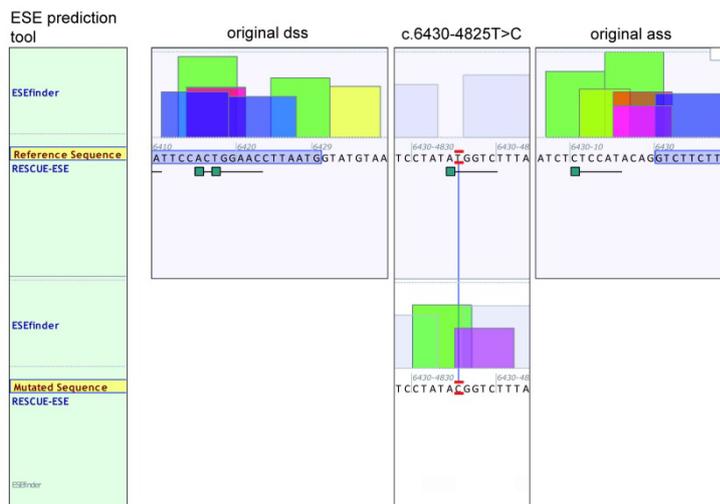
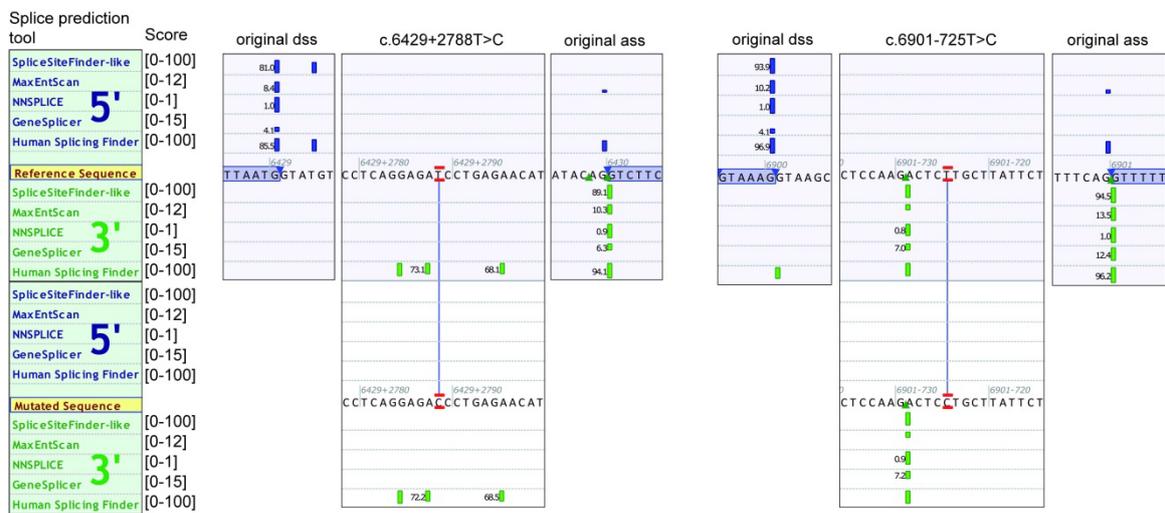
Supplementary Figure

Supplementary Material to Bach et al. “Identification of deep intronic variants in 15 haemophilia A patients by Next Generation Sequencing of the whole factor VIII gene” (Thromb Haemost 2015; 114.3)

Suppl. Figure 1: Bioinformatic predictions by Alamut software showing the simulated effect of the rest of the discussed variants: Splice site prediction with calculated probabilities for the use of the variants as de novo donor (dss/5') or acceptor splice site (ass/3') are indicated for five splice prediction tools with scores for the wild type sequence on the top (grey) and mutated sequence beneath (white). Additionally, the scores for the original dss and ass lying upstream and downstream of the variants, respectively, are shown. The last picture shows de novo binding motifs for exonic splice enhancers (ESE) calculated by two further tools (ESEfinder and RESCUE-ESE) for the last variant as well as the corresponding original dss and ass.







Anlage 7**Bach JE**, Müller CR, Rost S*Mini-gene assays confirm the splicing effect of deep intronic variants in the factor VIII gene*

Thromb Haemost. 2015 Dec 22;115(1):222-4. doi: 10.1160/TH15-05-0399

Mini-gene assays confirm the splicing effect of deep intronic variants in the factor VIII gene

J. Elisa Bach; Clemens R. Müller; Simone Rost

Department of Human Genetics, University of Würzburg, Würzburg, Germany

Dear Sirs,

Certain mutations in the factor VIII gene (*F8*) lead to reduced activity and deficiency of coagulation factor VIII (FVIII) and thus cause haemophilia A (HA) (1, 2). A broad spectrum of mutation types results in HA of variable clinical severity, including inversions, missense mutations, small and large deletions, insertions or duplications, mutations at splice sites and a few promoter mutations (3, 4). With current technology and approaches, routine molecular diagnostics fail to identify the putative HA-causing mutations in approx. 2% of cases (3). Analysis of mRNA has identified several deep intronic mutations that create novel splice sites leading to pathogenic transcript variants, most of which cause mild HA phenotypes (5–8). Using next-generation sequencing (NGS) of the complete *F8* genomic sequence, we recently have identified deep intronic variants in 15 out of 15 patients with mild to moderate HA whose disease causing mutation was not identified by conventional methods (9). At least one variant per patient was predicted to be pathogenic by several bioinformatic algorithms. As mRNA from these patients was not available, to confirm the causality we designed mini-gene splicing assays to functionally characterise these variants.

We focused on two candidate variants located in relatively small and “clonable”

introns of 286 bp and 1738 bp in size: c.5587–98G>A (chrX:154,132,897) in intron 16 (mini-gene 1; variant found in two male index patients with a FVIII:C of 20% and 25%, respectively) and c.5998+182A>G (chrX:154,131,999) in intron 18 (mini-gene 2; found in one male index patient with a FVIII:C of 2%). Each of the five *in silico* splice prediction algorithms included in the Alamut software (Interactive Biosoftware, Rouen, France) predicted the variant c.5587–98G>A to strengthen a cryptic donor splice site (dss) and four out of five algorithms predicted c.5998+182A>G to create a novel dss. Castaman et al. had demonstrated that another variant in intron 16 (chrX:154,132,892, c.5587–93C>T) located 5 bp upstream of our mini-gene 1 variant led to the inclusion of intronic sequences into *F8* mRNA (6). We used this variant as a positive control in our assay.

The wild-type sequence of the introns of interest (*F8* genomic reference sequence of transcript variant 1, ENST00000360256/NM_000132.3) plus the two 5' and 3' flanking exons and introns each, i.e. exons 15–18 for mini-gene 1 and exons 17–20 for mini-gene 2 (►Figure 1A) were cloned into pcDNA3.1^{+/myc}-His vector (Invitrogen, Life technologies, Carlsbad, CA, USA) using *Sac*II, *Xho*I and T4 DNA Ligase (New England Biolabs (NEB), Frankfurt, Germany). Mutations were introduced by *in vitro* mutagenesis with specific 30 bp mutagenesis primers (available on request) and *Pfu* DNA Polymerase (Promega, Mannheim, Germany) following the QuikChange protocol (Agilent Technologies, Santa Clara, CA, USA). Original DNA templates were digested with *Dpn*I (NEB) and products transformed into One Shot Top10 Chemically Competent cells (Invitrogen). Positive clones were identified by Sanger sequencing (ABI 3130xl, Life technologies) and plasmids prepared using the Nucleo Bond PC 100 kit (Macherey-

Nagel, Düren, Germany). HEK-293 cells were transfected with both, wild-type and mutated constructs using Lipofectamine 2000 Transfection Reagent (Invitrogen). After 24 hours, RNA was isolated with the High Pure RNA Isolation Kit (Roche Diagnostics, Mannheim, Germany) and cDNA was synthesised with SuperScript II Reverse Transcriptase (Invitrogen) and an oligo-dT18 primer. PCR products were obtained by suitable primer combinations and analysed by gel electrophoresis and Sanger sequencing. Sequence data were aligned to the *F8* genomic reference sequence (transcript variant 1, ENST00000360256/NM_000132.3) and analysed with CodonCode Aligner v5.0 (CodonCode Corporation, Centerville, MA, USA).

Gel electrophoresis and sequencing of the wild-type controls showed correct splicing in this artificial *in vitro* environment of mini-gene constructs 1 and 2. For the positive control (c.5587–93C>T) we obtained the same results as Castaman et al. (6): the insertion of 56 bp of intron 16

Figure 1: Mini-gene assay: constructs and results. A) Overview on the cloned sequences. Mini-gene 1 comprises exons/introns 15–18 and mini-gene 2 exons/introns 17–20. The size of introns and exons is given in bp. The position of the mutated bases in mini-gene 1, positive control and mini-gene 2, respectively, is marked with red arrows. B) Gel electrophoresis of PCR products of the cDNA of mini-gene 1 assay (wt1 = wild-type sequence of mini-gene 1, mut1 = mutated sequence with c.5587–98G>A, contr = positive control sequence with c.5587–93C>T) and mini-gene 2 assay (wt2 = wild-type sequence of mini-gene 2, mut2 = mutated sequence with c.5998+182A>G). C) The panel shows the predicted effect for c.5587–98G>A as displayed by Alamut software (Interactive Biosoftware, Rouen, France) with five prediction algorithms: SpliceSiteFinder-like (Alamut-specific tool), MaxEntScan (10), NNSPLICE (11), GeneSplicer (12) and Human Splicing Finder (13). Probability scores for the use of donor (dss/5') or acceptor (ass/3') splice sites are given for the wild-type sequence on the top (light grey background) and mutated sequence beneath (white). By strengthening a dss at c.5587–102 another ass at c.5586+137 is also used and the 49 bp fragment is spliced onto the 3' end of exon 16 leading to a frameshift with a premature termination codon in the mRNA. The position of the positive control variant c.5587–93C>T reported by Castaman et al. (6) is marked with a red asterisk.

Correspondence to:

J. Elisa Bach
Department of Human Genetics
University of Würzburg, Biocenter
Am Hubland, 97074 Würzburg, Germany
Tel.: +49 931 3189102, Fax: +49 931 3184069
E-mail: elisa.bach@uni-wuerzburg.de

Received: May 12, 2015

Accepted after minor revision: June 19, 2015

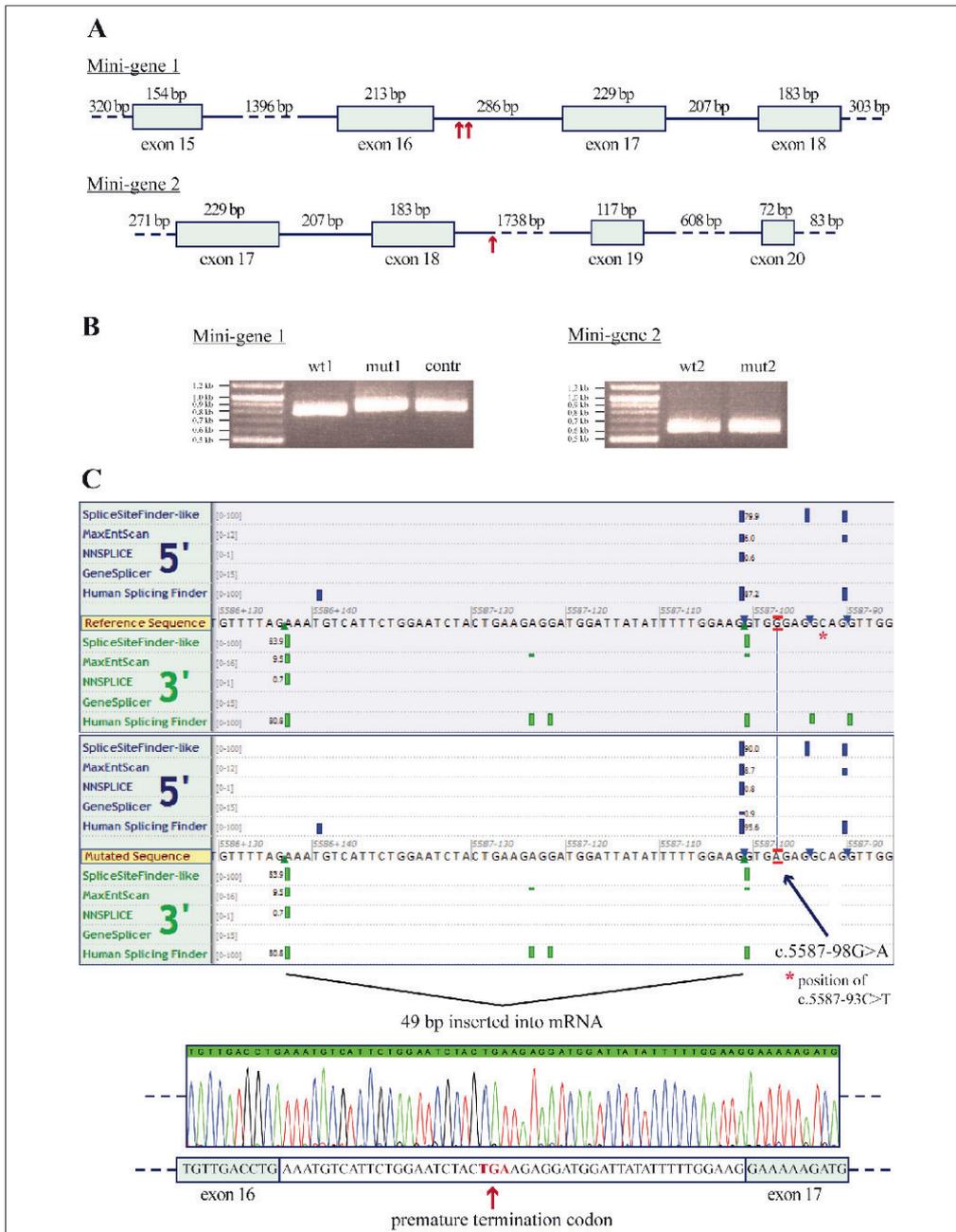
Epub ahead of print: August 6, 2015

<http://dx.doi.org/10.1160/TH15-05-0399>

Thromb Haemost 2016; 115: 222–224

© Schattauer 2016

Thrombosis and Haemostasis 115.1/2016



into the *F8* mRNA resulting in a frameshift and a new termination codon 24 bp downstream of exon 16. Similar results were obtained for our novel variant c.5587-98G>A detected in two HA patients of our recent NGS study (► Figure 1B) (9). As predicted by bioinformatic algorithms the variant seems to strengthen and activate a cryptic dss at c.5587-102. A cryptic acceptor splice site (ass) located a few bp upstream seems also to be used – the same site as shown for c.5587-93C>T. Hence, a fragment of 49 bp (c.5586+137_5587-102) is spliced onto the 3'-end of exon 16. It is identical to the beginning of the 56 bp fragment introduced by c.5587-93C>T in the positive control and causes the same frameshift with the same novel termination codon after 24 bp within the inserted fragment (► Figure 1C).

For mini-gene 2, we did not observe any difference between wild-type and mutated constructs containing c.5998+182A>G. Both were spliced correctly and identical results were seen in gel electrophoresis (► Figure 1B) and sequence data. Although a strong novel dss was predicted *in silico* (9), based on our *in vitro* analysis, pathogenicity of this variant seems very unlikely. As described in our previous study, the patient with c.5998+182A>G showed a complex haplotype with four additional deep intronic single nucleotide variants (SNVs) and a large deletion of approx. 9.2 kb located deep in intron 1 which may generate certain amounts of alternative *F8* transcripts and thus cause the patient's moderate HA phenotype (9).

In conclusion, the mini-gene assay is a convenient alternative to mRNA studies to analyse the effect of deep intronic variants in an artificial system especially for variants in smaller introns leading to "clonable" constructs. Our system was able to reproduce the results for the c.5587-93C>T mutation of Castaman et al. (6) and to show that the nearby variant c.5587-98G>A leads to the same premature termination codon by aberrant splicing. The loci c.5587-93C>T and c.5587-98G>A in intron 16 of the *F8* gene might represent an intronic mutation hotspot as seven unrelated HA patients with one of these two variants have been identified (6, 8 and this study). Coincidentally, these two variants appear to result in the same pathogenic mechanism. The identification and confirmation of pathogenicity of such mutation hotspots in the *F8* intronic sequence could improve molecular HA diagnostics especially for laboratories in which NGS analyses of the complete *F8* gene are not available.

Conflicts of interest

None declared.

References

- Gitschier J, Wood WI, Goralka TM, et al. Characterization of the human factor VIII gene. *Nature* 1984; 312: 326–330.
- White GC, 2nd, Rosendaal F, Aledort LM, et al. Definitions in hemophilia. Recommendation of the scientific subcommittee on factor VIII and factor IX of the scientific and standardization committee of the International Society on Thrombosis and Haemostasis. *Thromb Haemost* 2001; 85: 560.
- Oldenburg J, El-Maarri O. New insight into the molecular basis of hemophilia A. *Int J Hematol* 2006; 83: 96–102.
- Zimmermann MA, Meier D, Oldenburg J, et al. Identification and characterization of mutations in the promoter region of the factor VIII gene. *J Thromb Haemost* 2012; 10: 314–317.
- Bagnall RD, Waseem NH, Green PM, et al. Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients. *Br J Haematol* 1999; 107: 766–771.
- Castaman G, Giacomelli SH, Mancuso ME, et al. Deep intronic variations may cause mild hemophilia A. *J Thromb Haemost* 2011; 9: 1541–1548.
- Inaba H, Koyama T, Shinozawa K, et al. Identification and characterization of an adenine to guanine transition within intron 10 of the factor VIII gene as a causative mutation in a patient with mild hemophilia A. *Haemophilia* 2013; 19: 100–105.
- Pezeshkpoor B, Zimmer N, Marquardt N, et al. Deep intronic 'mutations' cause hemophilia A: application of next generation sequencing in patients without detectable mutation in *F8* cDNA. *J Thromb Haemost* 2013; 11: 1679–1687.
- Bach JE, Wolf B, Oldenburg J, et al. Identification of deep intronic variants in 15 haemophilia A patients by next generation sequencing of the whole factor VIII gene. *Thromb Haemost* 2015; 114: Epub ahead of print.
- Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; 11: 377–394.
- Reese MG, Eeckman FH, Kulp D, et al. Improved splice site detection in Genie. *J Comput Biol* 1997; 4: 311–323.
- Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001; 29: 1185–1190.
- Desmet FO, Hamroun D, Lalande M, et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; 37: e67.

Aus Gründen des Datenschutzes sind die Unterschriften der Koautoren, Lebenslauf und Danksagung in der Online-Version der Dissertation nicht enthalten.