

Summary

Neuromuscular diseases exist in many subtypes which are caused by variants in genes involved in the structure, function and stability of the musculoskeletal system, the nervous system, or the interface between both. Differentiated subtypes exist by the clinical phenotype and genetic cause. Muscular dystrophies (MD) are leading often to death, because no therapies were available until recently, as discovered for DMD or SMA. Dystroglycanopathies (DGpathies), a subtype of MDs, could lead to a wide spectrum of MDs including from a severe CMD named Walker-Warburg-Syndrome to milder Muscle-Eye-Brain disease, CMDs and to certain types of LGMDs [10]. The availability of the NGS has led to the discovery of novel DGpathy associated genes, but the genetic basis in many cases was still elusive at the start of my thesis and the biochemical function of these genes in the dystroglycan post-translation processing pathway were unknown.

To discover novel neuromuscular disease associated genes, we recruited neuromuscular patients over the last years from Germany and collaborators abroad with a diagnosis of CM or MD, with a total of 276 patients. Nearly 60 % (156) have been myopathy cases, and it is known that in general around 50% of myopathy cohorts could be solved by whole exome sequencing. Using NGS analysis, we solved 56% of the CM cases revealed novel genes, as the congenital myopathy gene *UNC45B* with biallelic variants, or novel variants extending the phenotypic spectrum in known CM genes as *MYH2* or *NEB*. Moreover, we discovered a founder mutation in the Turkish population in *PYROXDI* in a large homozygous region, detected by a Haplotype analysis. Two patients with variants in *TTN*, the largest protein in the human body, were contributed to a large cohort of *TTN* cases with many collaborators, to gain a better understanding of the missense variants in *TTN* by biochemical and structural analysis of the variants to be distinguished into pathogenic and benign ones. Further filtering and copy number variant analyses revealed a deletion of at least three genes including the known myopathy gene *SEPN1*, and two unknown genes in two affected brothers. Still, 89 cases were unsolved, which will be solved in the future by using Whole Genome Sequencing.

The 120 MD cases have been further divided into 99 DGpathy cases, because we wanted to unfold the genetic landscape of DGpathies based on NGS analyses. Back then, only 16 DGpathy associated genes were described, with mostly unknown functions in the glycosylation processing pathway of α -DG. Variants in each of the DGpathy genes causing a variety of phenotypes. No phenotype-genotype correlation exists because the mostly discovered missense variants could cause different phenotypes, as the most severe and lethal WWS or MEB, but also the mildest, LGMD, with a later age of onset and no mental retardation. The genetic analysis of the cohort with 99 patients showed that the most prevalent genes in our cohort were *POMT1*, *POMT2* and *POMGNT1*. In rare genes, as *B3GALNT2*, *B4GAT1* or *INPP5K* we only detected one case in each, additionally, 4 cases have been solved by CNV analysis. We also detected *AHCY* variants in two families, which could be a novel DGpathy gene, but further functional analysis is needed to understand the function in the glycosylation process. Still, 88% of the

cases solved by WES including CNV analysis, leaving 10 cases which we already submitted for Trio-WGS analysis, expecting intronic variants or another novel DGpathy gene.

In each cohort many missense variants were detected, which were the most complicated variants to define as pathogenic or non-pathogenic. Many in silico tools were designed and published over the last years to predict the impact of missense variants. To distinguish DGpathy missense variants by using most known prediction tools, I chose 6 DGpathy genes with several pathogenic missense variants to analyze them compared to benign variants. These results showed that pathogenicity scores could only be used for comparing variants within one gene, but not to other genes. So, each gene seemed to have his own cut-off values, which we further analyzed by developing informative 3D plots by including the CADD scores, and allele frequency both on the protein sequence of the analyzed protein. Results showed that some domains in the genes were including more pathogenic or benign variants, and also different cut-offs for the CADD score between pathogenic and non-pathogenic variants. We had a goal to develop personalized treatment approaches based on each variant's properties. For this purpose we investigated the impact of 19 CRPPA variants, by using in silico analysis with a designed CRPPA model and experimentally validated data of the same variants in terms of stability, structural impact on CRPPA, enzyme activity and ligand binding. For this purpose, we analyzed the effect of 19 *CRPPA* variants, in detail and prepared a systematic pipeline to understand the structural and biochemical alterations caused by each variant and how the physiological activity restored for personalized treatment approaches. The resulting in-silico and experimentally data have been compared and used for the classification of the *CRPPA* variants in terms of stability, structural changes in the secondary structure, enzymatic activity and ligand binding in 6 groups (I-VI).

Zusammenfassung

Neuromuskuläre Krankheiten sind eine heterogene Gruppe von vererbten Muskelerkrankungen, unterteilt in viele unterschiedliche Subtypen, darunter auch kongenitale Myopathien und muskuläre Dystrophien. Diese Krankheiten werden verursacht durch Mutationen in einem der vielen entdeckten Gene. Muskuläre Dystrophien führen in vielen Fällen zum Tod der Patienten. Für die meisten Muskelerkrankungen gibt es keine Therapien zur Heilung der betroffenen Patienten, nur die Symptome können behandelt und gelindert werden. In den letzten Jahren hat die Forschung dazu beigetragen, erfolgreich Ergebnisse bei DMD oder SMA zu erzielen. Dystroglykanopathien (DGpathies), ein Subtyp von muskulären Dystrophien, können zu unterschiedlichen Phänotypen im Patienten führen können, von sehr tödlichem WWS oder MEB, meistens ausgelöst bei der Geburt, zum milderen LGMD. Trotz einiger publizierter Fälle und entdeckter DGpathie Gene zu Beginn des Projektes, waren jedoch die Funktionen der meisten DGpathie Gene im Glykosylierungsprozess von α -DG noch nicht ausreichend erforscht.

Um neue Gene entdecken zu können, habe ich eine Kohorte von 275 neuromuskulären Patienten aus Deutschland und von Kollaboraturen aus dem Ausland rekrutiert. Die klinisch diagnostizierten Patienten wurden in zwei Gruppen unterteilt, in Patienten mit einer Myopathie und Patienten mit einer muskulären Dystrophie. Der Einsatz von hoch moderner NGS Analyse sollte helfen, die krankheitsverursachenden Mutationen zu finden und dabei auch neue Gene zu entdecken. In der Myopathiekohorte habe ich 156 Patienten gesammelt, welche durch die Anwendung einer NGS Analyse alle genetisch untersucht wurden. Dadurch war es mir möglich, über 56% der Fälle zu lösen, d.h. die krankheitsverursachende Mutation in einem bekannten oder neuen Myopathie-Gen zu entdecken und zu bestätigen. Die gelösten Fälle beinhalteten eine Varietät an Genen und Varianten, wie zum Beispiel die bi-allelischen Varianten in dem neuen Myopathie Gen benannt, *UNC45B*, dass wir erfolgreich als Gruppe publiziert haben. Aber auch Mutationen in publizierten Genen, die zu einer Erweiterung des Phänotyps geführt haben, wie für die Myopathie Gene *MYH2* und *NEB* dies der Fall war, welches zur Publikation von *NEB* führte. Ein sehr interessanter Fall war die Entdeckung von Varianten in zwei türkischen Familien in *PYROXDI*, welches als neues kongenitales Myopathie Gen publiziert wurde. Die gefundene Variante c.464A>G, p.N155S wurde schon in 90% der publizierten Fälle detektiert und eine Haplotypisierung der Variante bestätigte unsere Theorie einer Founder- bzw. Gründer-Mutation in der türkischen Population. Die von mir durchgeführte Altersberechnung der Founder-Mutation ergab ein Alter von über 2000 Jahren. Wie in den meisten Myopathiekohorten hatten wir auch Varianten in dem Gen *TTN*, das größte Gen im menschlichen Körper. In einer umfangreichen Publikation mit internationalen Kollaboraturen wurden viele *TTN* Fälle gesammelt und deren strukturelle Effekte auf das Protein analysiert, zu dem ich auch einen Beitrag zur genetischen Analyse von Missens-Mutationen geleistet habe. Eine CNV Analyse aller verbleibenden Fälle in meiner Kohorte führte zu der Entdeckung einer noch nicht publizierten großen Deletion von 3 Genen, darunter auch das bekannte Gen *SEPNI* und zwei weitere unbekannte Gene. Detektiert wurde die Deletion in zwei betroffenen Brüdern mit zwei weiteren gesunden Geschwistern.

Nach der genetischen Analyse aller Fälle waren viele davon noch ungelöst, welche wir durch weitere NSG Analysen, wie durch die Anwendung von Trio-WGS Analysen in der Zukunft lösen werden.

Die 119 Muskeldystrophie Patienten wurden weiter unterschieden in 20 muskuläre Dystrophien und 99 DGpathie. Mein Ziel war es, eine sehr große Kohorte, sowohl klinisch als auch genetisch, zu untersuchen, und die genetische Ursache zu finden, um die Krankheit und den Verlauf besser zu verstehen. Zu Beginn meiner Doktorarbeit waren nur 16 DGpathie Gene publiziert worden, mit meist unbekannter Funktion. Durch die geringen publizierten Fälle war es zudem nicht möglich, Genotyp- Phänotyp Korrelationen zu erstellen, verstärkt durch die große Anzahl an detektierten Missense- Mutationen mit sehr unterschiedlich verursachten Phänotypen. Durch die NGS Analyse konnten 88% der Fälle in dieser Kohorte gelöst werden, mit den meisten Varianten in den Genen *POMT1*, *POMT2* und *POMGNT1*. Seltener waren Varianten in den Dgpathie Genen *B3GALNT2*, *B4GAT1* oder auch *INPP5K*. Zudem konnten vier Fälle in dieser Kohorte durch eine CNV Analyse gelöst werden. Im Vergleich zu der Myopathie Kohorte, wurden 88% der DGpathie Fälle durch die Anwendung von NGS gelöst und auch bestätigt.

Üblicherweise werden in jedem Patienten oder auch jeder Kohorte viele Missense-Varianten entdeckt, deren Pathogenitäten schwierig sind zu bestimmen. In den letzten Jahren wurden viele neue Programme für Punktmutationen publiziert, die online angewendet, Aussagen zur Pathogenität von Varianten vorhersagen, durch bestimmen verschiedener Eigenschaften. Um die Effektivität der Programme zu analysieren, habe ich 15 DGpathie Missense-Mutationen ausgewählt und mit aktuellen Onlineprogrammen die Eigenschaften bestimmt. Die Resultate zeigten, dass die erhaltenen Werte durch eher für den Vergleich von Varianten innerhalb eines Gens sinnvoll waren, jedoch nicht zum Vergleich von Mutationen in unterschiedlichen Genen miteinander. Somit besitzen viele Gene einen eigenen Grenzwert, den wir durch die Entwicklung von 3D Diagrammen durch die Gegenüberstellung der CADD-Werte, der Allelfrequenz und der jeweiligen Proteinsequenz bestimmt haben. Die Ergebnisse verdeutlichten, dass einige Domänen in Genen anfälliger für pathogene Varianten waren als andere. Zudem konnte ich eine Grenze für den CADD-Wert zwischen den pathogenen und nicht pathogenen Varianten in einigen Genen feststellen. Als nächstes hatten wir uns als Ziel vorgenommen, die erhaltenen in Silico Daten von Varianten im tertiären DGpathie Gen *CRPPA* mit experimentell erhaltenen Resultaten für die gleichen Varianten in Bezug auf die Stabilität, strukturellen und enzymatischen Effekte durch den Aminosäureaustausch in den Varianten zu vergleichen. In der Zukunft sollte dies dazu beitragen, durch die erhaltenen Daten eine personalisierte Therapiemöglichkeit für Patienten zu entwickeln, basierend auf den Eigenschaften jeder Mutations. Für diesen Zweck habe ich 19 *CRPPA* Varianten untersucht und eine systematische Pipeline zur Analyse der Varianten entwickelt. Alle Varianten wurden in 6 Gruppen (I-VI) eingeteilt, mit einer speziell empfohlenen Therapie in jeder Gruppe.