

## **Abstract**

A rare disease is defined as any disease that affects less than 1 in 2,000 people. Rare diseases are individually uncommon but collectively common and encompass approximately 7000 different disorders. Mendelian disorders, which are typically rare diseases caused by a single gene variant, follow Mendelian patterns of inheritance. Neurogenetic and musculoskeletal disorders are among the main categories of Mendelian disorders and represent a significant burden on individuals, society, and healthcare systems especially in consanguineous populations.

Yemen is a less developed country with a high rate of consanguineous marriages and a poor health service system, including an extreme shortage of medical genetics facilities. Hence, only very few studies on genetic disorders in the Yemeni population have been performed.

The current study has been conducted on a large cohort of Yemeni patients with different neurogenetic and musculoskeletal disorders. It aimed at uncovering the molecular genetic cause(s) of the disorders and functional characterization of the consequences of novel variants.

For my PhD project, I recruited and clinically characterized 81 families with different neurogenetic and musculoskeletal disorders from all over Yemen. All of them were subjected to different genetic testing approaches, including next-generation sequencing. Additionally, various molecular genetic and biochemical methods, including RNA-sequencing, ChIP-sequencing, Sanger sequencing, linkage analysis, western blotting, immunofluorescence, and pull-down assays, were employed to reveal the consequences of the identified variants.

A molecular genetic diagnosis was established in 53 of the 81 families (65.4%). The highest diagnostic yield was achieved in the group of families with hereditary myopathies (85.7%), in contrast to only 62% in the group of families with central nervous system disorders. As to the most successful testing approach, whole-exome sequencing was superior to Mendeliome sequencing and gene panel sequencing, with 73.7%, 54.3%, and 42.9%, respectively. Among the solved cases, I found one family with a dual molecular diagnosis of spastic paraplegia 54 and congenital muscular dystrophy with integrin alpha-7 deficiency.

Moreover, in this study, I discovered a novel disease-associated gene (*LEF1*) and twenty-four novel variants in different known disease-associated genes. The phenotypic spectrum was extended for two genes, *CEP63* and *NDE1*, which are known to be involved in Seckel syndrome and microlissencephaly, respectively. Three possible founder mutations could be proposed, in *C12orf57*, *NDE1*, and *SMOC1*, either specific for the Yemeni population or shared with other inhabitants of the Arabian Peninsula.

Interestingly, the data generated from patient-derived fibroblasts with an *NDE1* nonsense variant (p.Arg220\*), which was found in a family with non-syndromic microcephaly, revealed an escape from nonsense mediated decay. Thus, the truncated mutant protein may retain some residual activity and account for the less severe phenotype and the absence of brain malformation.

An intriguing finding of this thesis is the identification and characterization of *LEF1* as a novel gene involved in autosomal recessive ectrodactyly ectodermal dysplasia without cleft lip/palate. LEF-1 (lymphoid enhancer-binding factor 1) is a transcription factor that directly interacts with  $\beta$ -catenin to regulate the expression of Wnt target genes. A homozygous variant (NM\_016269.4:c.69\_71dupGAT, p.Met23dup) was identified in the proband of a consanguineous family who manifested ectrodactyly (split hand/split foot), hair loss, oligodontia, nipple aplasia, and mild intellectual disability. The LEF-1 mutation (p.Met23dup) is located in the highly conserved  $\beta$ -catenin binding domain and could be shown *in vitro* to disrupt the LEF-1– $\beta$ -catenin interaction and to result in a significant dysregulation of the Wnt and HOX genes, which are important for the development of limbs, teeth, and hair.

The results of this study are expected to contribute to the improvement of genetic testing and counseling services in Yemen as well as a growing awareness of genetic disorders among healthcare personnel and in the general population.

## Zusammenfassung

Eine seltene Krankheit ist definiert als eine Krankheit, die weniger als 1 von 2.000 Menschen betrifft. Seltene Krankheiten sind individuell selten aber kollektiv häufig und umfassen etwa 7000 verschiedene Erkrankungen. Monogenetische Erkrankungen sind typische seltene Krankheiten, die durch einen Defekt in einem einzelnen Gen hervorgerufen werden und Mendelschen Vererbungsregeln folgen. Die neurogenetische und muskuloskelettale Erkrankungen gehören zu den Hauptkategorien der monogenetischen Erkrankungen und stellen eine hohe Belastung für den Einzelnen, die Gesellschaft und die Gesundheitssysteme dar, insbesondere in blutsverwandten (konsanguinen) Populationen.

Der Jemen ist eines der am wenigsten entwickelten Länder mit einer hohen Rate an konsanguinen Ehen und einem schlechten Gesundheitssystem, einschließlich eines extremen Mangels an medizinisch-genetischen Einrichtungen. Daher wurden bisher nur sehr wenige Studien über genetische Erkrankungen in der jemenitischen Bevölkerung durchgeführt.

Die vorliegende Studie wurde an einer großen Kohorte von jemenitischen Patienten mit verschiedenen neurogenetischen und muskuloskelettalen Erkrankungen durchgeführt. Sie zielte darauf ab, die molekulargenetische(n) Ursache(n) der Erkrankungen aufzudecken und eine funktionelle Charakterisierung der neu gefundenen Varianten vorzunehmen.

Für mein Promotionsprojekt habe ich 81 Familien mit verschiedenen neurogenetischen und muskuloskelettalen Erkrankungen aus dem gesamten Jemen rekrutiert und klinisch charakterisiert. Alle wurden unterschiedlichen genetischen Testansätzen unterzogen, einschließlich des Next Generation Sequencing (NGS). Zusätzlich wurden verschiedene molekulargenetische und biochemische Methoden eingesetzt, einschließlich RNA-Sequenzierung, ChIP-Sequenzierung, Sanger-Sequenzierung, genetische Genkopplungs-Analyse, Western Blot, Immunfluoreszenz und Pull-Down-Assays, um die Konsequenzen der identifizierten Varianten aufzudecken.

Eine molekulargenetische Diagnose wurde in 53 der 81 Familien (65,4%) gestellt. Die höchste diagnostische Ausbeute wurde in der Gruppe der Familien mit erblichen Myopathien (85,7%) erzielt, im Gegensatz zu nur 62% in der Gruppe der

Familien mit Erkrankungen des Zentralnervensystems. Die Gesamt-Exom-Sequenzierung war mit 73.7% der Mendeliom-Sequenzierung mit 54.3% und der Gen-Panel-Sequenzierung mit 42.9% überlegen. Unter den geklärten Fällen fand ich eine Familie mit einer dualen molekularen Diagnose von spastischer Paraplegie 54 und kongenitaler Muskeldystrophie mit Integrin-Alpha-7-Mangel.

Darüber hinaus entdeckte ich in dieser Studie ein neues krankheitsassoziiertes Gen (*LEF1*) und 24 neue Varianten in verschiedenen bereits bekannten krankheitsassoziierten Genen. Das phänotypische Spektrum wurde für zwei Gene, *CEP63* und *NDE1*, erweitert, von denen bekannt ist, dass sie am Seckel-Syndrom bzw. an der Mikrolissenzephalie beteiligt sind. In *C12orf57*, *NDE1* und *SMOC1* könnten drei mögliche Gründermutationen vorgeschlagen werden, die entweder spezifisch für die jemenitische Bevölkerung sind oder mit anderen Bewohnern der Arabischen Halbinsel geteilt werden.

Von Patienten einer Familie mit nicht-syndromaler Mikrozephalie wurden Fibroblasten mit einer *NDE1*-Nonsense-Variante (p.Arg220\*) generiert. Interessanterweise zeigten die Daten, dass ein „Nonsense-mediated decay“ umgangen werden konnte, da das verkürzte mutierte Protein eine gewisse Restaktivität behalten hat und daher für den weniger schweren Phänotyp und das Fehlen einer Fehlbildung des Gehirns verantwortlich sein könnte.

Ein faszinierendes Ergebnis dieser Arbeit ist die Identifizierung und Charakterisierung von *LEF1* als neuartigem Gen, das an der autosomal rezessiven Ektrodaktylie-ektodermalen Dysplasie ohne Lippen-/Gaumenspalte beteiligt ist. LEF-1 (Lymphoid Enhancer-Bindungsfaktor 1) ist ein Transkriptionsfaktor, der direkt mit  $\beta$ -Catenin interagiert, um die Expression von Wnt-Zielgenen zu regulieren. Eine homozygote Variante (NM\_016269.4: c.69\_71dupGAT, p.Met23dup) wurde im Probanden einer konsanguinen Familie identifiziert, die Ektrodaktylie (gespaltene Hand / gespaltener Fuß), Haarausfall, Oligodontie, Brustwarzenaplasie und leichte geistige Behinderung zeigte. Die LEF-1-Mutation (p.Met23dup) befindet sich in der hochkonservierten  $\beta$ -Catenin-Bindungsdomäne und es konnte in vitro gezeigt werden, dass sie die LEF-1- $\beta$ -Catenin-Wechselwirkung stört und zu einer signifikanten Dysregulation der Wnt- und *HOX*- Gene führt, die für die Entwicklung von Gliedmaßen, Zähnen und Haaren wichtig sind.

Die Ergebnisse dieser Studie sollen zur Verbesserung der genetischen Untersuchungen und Beratungsdienste im Jemen beitragen, sowie zu einem wachsenden Bewusstsein für genetische Erkrankungen beim Gesundheitspersonal und in der allgemeinen Bevölkerung führen.