WES analysis reveals a homozygous *SGCB* variant in a Pakhtun family with LGMDR4 phenotype

Muhammad Tariq[1,α], Muhammad Latif[2,α], Memona Inam[1], Amin Jan[3], Nousheen Bibi[4], Hussein Sheikh Ali Mohamoud[5], Isse Ali[6], Habib Ahmad[2], Aziz Khan[7], Jamal Nasir[8], Abdul Wadood[9,*], Musharraf Jelani[1,*]

[1]Centre for Omic Sciences, Islamia College Peshawar, Khyber-Pakhtunkhwa, Pakistan

[2]Centre for Genetics and Inherited Diseases, Taibah University, Al-Madinah Al-Munawwarah, Saudi Arabia

[3]Department of Physiology, North-West School of Medicine, Peshawar, Khyber-Pakhtunkhwa, Pakistan

[4]Department of Bioinformatics, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan

[5]Viapath at Guy's Hospital Great Maze Pond SE1 9RT, London, United Kingdom

[6]School of Science & Technology, Nottingham Trent University, NG11 8NS, United Kingdom

[7]Government Higher Secondary School Dilasa Khan Mandew, Bannu Khyber-Pakhtunkhwa, Pakistan

[8]Division of Life Sciences, University of Northampton, NN1 5PH, United Kingdom

[9]Department of Biochemistry, Abdul Wali Khan University, Mardan, Khyber-Pakhtunkhwa, Pakistan

[α]Authors contributed equally to this work

***Corresponding authors**

1. Dr. Musharraf Jelani, Associate Professor, Centre for Omic Sciences, Islamia College Peshawar, Khyber-Pakhtunkhwa, Pakistan. Email: mjelani@icp.edu.pk; Phone: 0092 333 5207053

2. Dr. Abdul Wadood Associate Professor Department of Biochemistry, Abdul Wali Khan University Mardan, Khyber-Pakhtunkhwa, Pakistan. Email: awadood@awkum.edu.pk; Phone: 0092 345 2683291

Abstract

Limb-girdle muscular dystrophy (LGMD) is a term used for the proximal muscles' weakness mainly affecting arms, shoulders, legs and thighs. These patients have altered body posture due to weak muscles, and have difficulty in holding, standing or walking. Genetic causes of both autosomal dominant (LGMDD) and recessive (LGMDR) forms have been identified. In this study, we analyzed a consanguineous Pakistani family of Pakhtun ethnicity. The disease had started at early childhood (7-8 years). Slowly and gradually the phenotype had worsened, and the patients had become totally wheelchair bound. Whole exome sequencing (WES) at 100X coverage on Illumina NovaSeq6000 platform followed by Sanger sequencing revealed a homozygous variant (c.610T>C; p.Ser204Pro) in *SGCB* gene known for LGMDR4 phenotype. Structural protein prediction tools and molecular docking analyses showed critical structural changed in the binding interface and SGCB protein tunnel. To the best of our knowledge this is the first report of *SGCB* variant identified in a Pakhtun family. Structural analysis of p.Ser204Pro substitution widens SGCB pathogenicity causing LGMDR4 phenotype. WES analysis can be used as a first line molecular diagnostic test for rare diseases like LGMD in Pakhtun ethnic groups of Pakistani population. Encouraging premarital testing in the closest relatives of such patients may have positive impact on reduction of the recurrence risk in their subsequent generations.

Keywords

Sarcoglycan beta gene; LGMDR4; WES; Pakhtun family; protein modeling

1. Introduction

Limb-girdle muscular dystrophies (LGMDs) are postnatal onset progressive muscles disorders which mainly affect shoulders, pelvic and girdle muscles (Bushby, 1995). The nomenclature has been refined and patients with proximal muscles weakness, high serum CK levels, and have dystrophic changes on muscles histology are considered LGMDs (Straub et al., 2018; Wicklund, 2019). The disease may appear in early childhood in which patients exhibit loss of independent walking and sudden falls. Respiratory and cardiac muscles abnormalities have also been found associated with LGMDs (Alonso-Pérez et al., 2020; Guimarães-Costa et al., 2020).

First gene of muscular dystrophy was mapped on chromosome X in 1987 (Hoffman et al., 1987), and first locus of LGMD on autosome was mapped to chromosome 5q31.2 in 1992 (Speer et al., 1992). Since then several genes have been reported for various types of LGMD segregating in X-linked, autosomal dominant and recessive modes of inheritance. Broadly LGMD can be classified into autosomal dominant (LGMDD) and recessive (LGMDR) forms. Based on their causative genes the two types are further classified into five (LGMDD1-5) and twenty-four (LGMDR1-24) sub-types, respectively (Wicklund, 2019). Numerical numbers for sub-clinical types have been assigned in the order of their causative gene's discovery (Straub et al., 2018).

A meta-analysis study has provided 1.63 per one million individuals as an estimated prevalence of LGMDs (Mah et al., 2016). LGMDR are more common especially in populations with high rate of consanguineous marriages (Nigro and Savarese, 2014; Mojbafan et al., 2020). Though some subtypes have been considered more exclusive to specific populations for example α-SG (LGMDR3) in Europe (Trabelsi et al., 2008), γ-SG (LGMDR5) in the North Africa (Dalichaouche et al., 2017) and δ-SG (LGMDR6) in Brazil (Passos-Bueno et al., 1999). However, relative

frequencies among LGMDRs subtypes vary in different populations. Just like β-SG (LGMDR4) variations are considered more frequent in middle east, though they have been reported world-wide (Alavi et al., 2017; Ozyilmaz et al., 2019; Winckler et al., 2019; Mojbafan et al., 2020).

The Khyber-Pakhtunkhwa province, located in the north-western region of Pakistan, is mostly occupied by Pakhtun tribes. This population, based on ethnic customs and traditions, prefer to marry within their families. According to one estimate cousin marriages in Pakistan are approximately 58% and may be higher in rural areas (Pervaiz et al., 2018). Frequent cousin marriages increase the chance of homozygosity in preceding generations which could be a high risk for autosomal recessive disorders (Bhinder et al., 2019). As genetic disorders are routed to zygote level they are hardly cured and most of them are non-treatable or have irreversible abnormalities which is an economic, psychological and social burden on community.

In this study, we aimed to perform genetic diagnosis of a Pakhtun family residing in district Nowshera, Khyber Pakhtunkhwa province, Pakistan.

2. Materials and methods

2.1 Ethical approval

Prior to start this study, Institutional Bioethics Committee (IBC), Islamia College Peshawar provided ethical approval for samples' collection and research (ref. # 529/oric/icp). An informed consent was signed by the study participants, and all biosafety measures were undertaken according to guidelines of "Helsinki's declaration 2013".

2.2 Samples' collection

Pedigree (Figure 1) was prepared interviewing the family elder (mother, II-2, of three affected individuals). Peripheral blood samples about 5 to 6 ml were collected in EDTA tubes from each participant. Genomic DNA was extracted using commercially available kits (GeneJET, ThermoFisher Scientific, USA). The purified genomic DNA was stored at 4ºC, which was later on used in WES and SS analyses.

2.3 Genetic analyses

Based on the LGMD phenotype and autosomal recessive mode of inheritance, we expected the causative variant in a homozygous or compound heterozygous forms in affected individuals of the family. As the LGMD candidate genes list was longer and could not be screened through SS. Thus WES analysis of one affected individual (III-1) was performed. Candidate variants in WES data were prioritized focusing on the genes previously known in LGMD or overlapping phenotypes. A list of 44 genes was retrieved under the keywords "autosomal recessive limb girdle muscular dystrophy" from NCBI genes list (https://www.ncbi.nlm.nih.gov/gene) for this purpose (Supplementary Materials Table S1).

2.3.1 WES analysis

WES data generation and basic bioinformatics was performed at Macrogen Inc. Republic of Korea. It included the following steps as per our requirement.

a. Genomic DNA (2-5µg) of III-1 was used to prepare WES libraries using SureSelectV-6 (Agilent Technologies, USA) and sequencing was performed on NovaSeq6000 NGS platform (Illumina, USA).
b. The raw data FASTQ sequence files were aligned using BWA (Burrows-Wheeler Alignment Tool) software (Jo and Koh, 2015). Duplicate sequences removed using PICARD software (http://broadinstitute.github.io/picard/), and variants annotation was done with GATK (Genome Analysis Toolkit) software (McKenna et al., 2010; Van der Auwera et al., 2013).
c. Potential candidate variants were short-listed using, autosomal recessive model (variants in homozygous or compound heterozygous form), variant with equal or less than 0.001 minor allele frequency, variants with at least depth quality of >20 and variants in protein coding regions and intron-exon boundaries. All other variants in deep intronic regions, coding synonymous amino acids were filtered out.
d. Shortlisted variants were further assessed for its potential "disease causing or pathogenic" effect through online prediction tools.
e. The final variant was subjected to SS to confirm its segregation with the disease phenotype.

2.3.2 Sanger validation

SS in all family members was performed to confirm autosomal recessive segregation of the selected variant *SGCB*. Wild-type genomic sequences downloaded from Ensembl Genome Browser (https://asia.ensembl.org/Homo_sapiens/Info/Index accession number ENSG00000163069). Primers (forward SGCB_4F: 5'- AGCAAGGGACAACAAAGCTC-3'and reverse SGCB_4R:5'-AGCTTGCAGTGAGTCGAGAT-3') were selected via Primer3Plus online

software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Standard PCR amplification program was followed to amplify the target region in all family members.

The SS reactions were prepared using BigDye™ terminator v3.1 cycle sequencing kits and after purification the products were run on ABI 3500 Genetic Analyzer (Life Technologies, USA). To verify variations, SS files of each individual were compared with wild type sequences using BioEdit software (www.mbio.ncsu.edu/BioEdit/bioedit.html).

2.3.3 Secondary and tertiary structure prediction

The crystal structure of SGCB was not available; therefore, 3D model ($SGCB^{WT}$) was built through homology modeling on I-Tasser server (Zhang, 2008), and validated through MolProbity server (Chen et al., 2010). 5nxk structure was adopted as a template with template modeling score (TM score of 0.853) and sequence identity of 22%. The altered $SGCB^{Ser204Pro}$ was developed using $SGCB^{WT}$ protein structure as a template. The predicted 3D structures were visualized and analyzed with UCSF Chimera tool (Pettersen et al., 2004). Secondary structure of the two SGCB proteins were predicted through a sequence based secondary structure prediction server, J-Pred (http://www.compbio.dundee.ac.uk/jpred/). COFACTOR (https://zhanglab.ccmb.med.umich.edu/COFACTOR/) and COACH (https://zhanglab.ccmb.med.umich.edu/COACH/) server were used to predict the critical binding residues of SGCB protein. Using protein structure and network comparison COFACTOR infers the critical binding residues while COACH combines the results of COFACTOR, S-SITE and TM-SITE programs to deduce functionally important residues.

2.3.3.1 Hydropathy and topology analysis

The transmembrane (TM) helices of $SGCB^{WT}$ and $SGCB^{Ser204Pro}$, were evaluate through membrane protein explorer (MPEx) (Snider et al., 2009), and hydrophobicity of the respective amino acid was analyzed. The experimental hydrophobicity values were used by MPEx tool to predict the TM region of query molecule and it was cross validated by Protter tool (http://wlab.ethz.ch/protter/start/). Pores and channels meant for aid-in passages of ions via the secondary structures of SGCB were predicted and analyzed by CAVER analyst 2.0 tool (Jurcik et al., 2018).

2.3.3.2 Effect of alteration on protein structure and function

The structural and functional impact of identified non-synonymous SGCB variant (p.Ser204Pro) was determined through HOPE server (Venselaar et al., 2010), and the effects of amino-acid variants in physiochemical properties, structure and spatial function of the proteins were predicted. Functional motifs in the SGCB protein scanned via in the PhosphoELM database (http://elm.eu.org/) and evolutionary conservation analysis was performed by using ClustalW server (https://www.genome.jp/tools-bin/clustalw). *In-silico* site directed mutagenesis approach was used to build the $SGCB^{Ser204Pro}$ model using $SGCB^{WT}$ as a template. Structure analysis and verification servers PROCHECK and ERRAT were used to assess the stereo chemical quality and environmental profile (Colovos and Yeates, 1993; Laskowski et al., 1996).

2.3.3.3 Molecular docking

Molecular docking of $SGCB^{WT}$ with Sarcoglycan Delta (SGCD) and Sarcoglycan Zeta (SGCZ) were carried out using AutoDock4.0 (Morris et al., 2009). Kollamen united charges were assigned and Polar hydrogen atom were added with the annealing parameters of 4.0 A° and 2.5 A° for

hydrogen bonding and Van der Waals interactions, respectively. The grid map was generated on the whole protein structure with 80_80_80 points and spacing of 0.875 A°. For each docking experiment number of runs was set to 100. AUTODOCK automatically group the resulting complex conformation into cluster based on RMSD value with default threshold of 2.0 Å R.M.S.D. The best docked complex for SGCB with SGCD and SGCZ were selected on the basis binding free energy value and molecular interactions were monitored using ligplot (Wallace et al., 1995), Discovery Studio visualizer (http://accelrys.com/products/collaborative-science/biovia-discovery-studio) and UCSF chimera (Pettersen et al., 2004).

3 Results

3.3 Clinical details of our enrolled family

At the time of our visit, the family had six affected members (three males III-1, III-2, III-8 and three females III-3, III-7, IV-1). The disease had autosomal recessive segregation and there were multiple cousin marriages in the pedigree (Figure 1). One affected girl (III-3) had died at the age of 20 without any reported medical records. There was no record of previous clinical investigations; for this reason, we could not specify the exact sub-clinical type of LGMD in this family. However, an affected individual (III-1 aged 26-years at the time of our visit) narrated his disease progression himself. They had normal life till the ages of seven to eight, after that slowly and gradually they faced difficulty in walking, balancing, and running. They were not able to carry their school bags and would fall down if walk continuously up to 100 meters. From the early teenage period, they were completely dependent for their daily life being unable to stand, walk or hold anything. Due to their latest conditions laying on wheel-chairs, we could not investigate radiological examinations. Apart from muscular weakness, they otherwise had neither hearing

loss, nor speech articulation abnormalities. Their father (II-1) had died and their unaffected mother (II-2) narrated that he had no complications like their children.

3.4 Exome sequencing and candidate gene prioritization

Exome sequencing results, of patient III-1, revealed 114,624 variants in total in which 604 variants were found in the 44 candidate genes (Supplementary Table S1). There were 290 variants in homozygous and 314 in heterozygous form. Applying the filtering criteria described earlier we found a homozygous variant in the sarcoglycan beta gene (NM_000232.4: c.610T>C: p.Ser204Pro). SS confirmed it homozygous segregation in all available affected individuals and was heterozygous in obligate carriers (Figure 1 and 2). This variant was not present in ethically matched unaffected or healthy individuals (n=200). Computational analysis of p.Ser204Pro substitution in SGCB gene with six pathogenicity prediction tools predicted the damaging effect of identified variant (Supplementary Table S1).

3.3 Structural characterization and mapping of Ser204Pro in SGCB protein

The SGCB secondary structure comprised of 17 β-sheets 3 alpha helices, 12 β-turns and 10 ϒ-turns. Ser204 mapped in the β-12 in sarcoglycan domain of SGCB protein. Comparative analysis of $SGCB^{WT}$ $SGCB^{Ser204Pro}$ secondary structures revealed minor changes in the irregular secondary structure elements without significant change in the regular secondary structure elements (Figure 3a). A high quality model was predicted $SGCB^{WT}$ with the C-score of 0.98. TM score of 0.8 for the $SGCB^{WT}$ signified the accurate topology for further analysis (Figure 3b).

Moreover, parameters such as peptide bond planarity, non-bonding interactions, Cα tetrahedral distortion, main chain H-bond energy, and G-factor for all 3 structures were in favorable ranges.

Root-mean-square deviation (RMSD) of $SGCB^{WT}$ and $SGCB^{Ser204Pro}$ was 0.345Å. SGCB residues 117,192,193,200, 203, 204,229 were predicted as binding site residues with c score of 0.04 and cluster size of 5 (Figure 3c).

The size of wild type residue Ser204 is different from the altered Pro204 residue. Proline is bigger in size and shape and favors the tight turns so it might create bump in the $SGCB^{Ser204Pro}$ protein structure. Furthermore, the increased hydrophobicity of altered residue at position 204 can disturb the correct folding and interaction of the SGCB protein (Figure 4a and b). Results indicated the conservation of serine at position 204 in the sarcoglycan domain of SGCB protein across seven homologues species (Figure 4c). Similarly, screening functional motif in the SGCB protein revealed that Ser204 is located in the GSK3 phosphorylation recognition motif (KASTERIT) with the conservation score of1.00 and probability of $2.679x10^{-2}$.

3.3.1 Hydropathy, topology and tunnel analysis

Hydropathy analysis showed that N-terminal part (1-62 aa) of the SGCB protein lied towards the intracellular part of the membrane with one transmembrane domain encompassing residue 63-93 and C-terminal part of protein lied extracellular. Similarly, no significant change was observed between the wild type and altered SCGB protein hydropathy plot (Supplementary Figure S1a-c). Protein tunnels aid in developing the connectivity of functional buried cavities with bulk solvent and protein channels, in addition to movement of ions across the membrane. Remarkably, the p.Ser204Pro was observed in tunnel formation (Supplementary Figure S1d) so the substitution of serine at position 204 with Pro might change the shape of tunnel resulting in decrease activity and stability of the protein. Overall, tunnel length was 73.729 with score of 0.833. Bottle neck was 1.176 and straightness 0.789.

3.3.2 Binding interaction analysis

The sarcoglycan complex (SG) comprises of α-, β-, γ-, and δ-sarcoglycan, and is essential to the dystrophin-associated glycoprotein complex, which provide structural stability to the sarcolemma and protect the muscle fibers from mechanical stress during muscle contraction. β-sarcoglycan (SGCB) binds to δ-sarcoglycan (SGCD) and newly identified member zeta (ζ)-SGCZ of SG complex. Network construction in STRING (https://string-db.org/) revealed strong binding with experimental evidence for interaction of SGCB with SGCD and SGCZ. The number of nodes were 11 with 55 edges while the average local clustering coefficient was 1 and p-value of < 1.0e-16 for the protein-protein interaction network of SGCB protein (Figure 4d,e).

Detailed molecular docking interaction analysis was performed to further assess the conformational switches of SGCB upon binding with SGCD and SGCZ (Supplementary Figure S2b). Docking analysis indicated the Gly117, Pro192, Ser193, Val200, Ala203, Ser204 and Asn229 of SGCB were involved in number of Vander Waal and electrostatic interactions with the SGCD and SGCZ protein with binding free energy of -12.34kcal/mol and -11.09kcal/mol respectively. Ser204 of SGCB made H-bond with the Pro186 of SGCD and second H-bond between Ala203 of SGCB with Phe187 of SGCB. Apart from these number of hydrophobic interactions were observed between Gly117, Ser193 and Asn229 of SGCB with important interface residues of SGCD (Supplementary Figure S2a and S2b). Similarly docking analysis of SGCB and SGCD also revealed different type of interactions between important binding residues of both SG members (Supplementary Figure S2a and S2c). So due to serine to proline substitution at position 204 will disturb the binding motif and might contribute in weakening the interaction between members of SG family and eventually will decrease the stability of SG complex.

4 Discussion

Pathogenic variations in more than 30 genes have been reported to leading to various types of LGMDs, in which sarcoglycanopathies are considered the most frequent form (Straub et al., 2018; Wicklund, 2019; Winckler et al., 2019). There are four types of sarcoglycans (SG) proteins α, β, γ and δ have been reported in LGMDR3, 4, 5, and 6 respectively. SG are transmembrane glycoproteins which form tetrameric complex across the cell membranes of striated (cardiac and skeletal) muscles fibers (Tarakci and Berger, 2016; Belhasan and Akaaboune, 2020). These complexes along with dystrophin have important role in maintain muscles' membrane integrity and protects sarcolemma from injuries during muscles contraction and relaxation (Tarakci and Berger, 2016; Gawor and Prószyński, 2018). Pathogenic variations in any member of these SG-complexes lead to disease phenotypes.

LGMDR4 (previously known as LGMD2E) is caused by variations in beta-sarcoglycan sub-unit encoding (*SGCB*, OMIM 600900) gene, located at 4q12. These variations are known for biochemical deficiency in the entire sarcoglycan protein complex (Bonnemann et al., 1995; Duggan et al., 1997). *SGCB* is considered the most common cause of LGMDR4 in Iranian population (Alavi et al., 2017). The variant, c.610T>C; p.Ser204Pro, identified in the patients under investigation lied in extracellular domain of SGCB protein. So far 79 pathogenic variants have been identified throughout the SGCB protein from various populations of the world (Supplementary Table 2). Reviewing all the 79 variations and the clinical features described in the literature; we could not draw a clear genotype-phenotype correlation among LGMDR4 patients. Though three missense variants have been recently reported in monogenic dilated cardiomyopathies (Mazzarotto et al., 2020), and some others previously reported in Duchene-like muscular dystrophy patients (Duggan et al., 1997), give an impression that some variant could cause severe forms of LGMDR4. However, cardiac involvement or phenotype severity could not be considered as unique entity of some specific variations, instead they could be found frequently in patients with *SGCB* alterations (Bonnemann et al., 1995; Barresi et al., 2000; Sveen et al., 2008; Alonso-Pérez et al., 2020).

In-silico analysis of this study revealed that Pro204 residue create a bump and disturb the correct folding of SGCB. Remarkably the altered residue was mapped on GSK3 phosphorylation

recognition motif. So, p.Ser204Pro substitution will result in the loss of phosphorylation and may disturb the normal functioning of SGCB protein. SGCB is crucial member of sarcoglycan complex and any loss of interaction with other component of complex can lead to a disease phenotype (Gao and McNally, 2015). Protein-protein interaction revealed that Ser204 is an important residue in SGCB biding interface and its substitution to Pro204 demolishes normal function of β-SG and may result in reduction of SG complex stability.

Linkage analysis followed by SS has remained a successful tool and still can be considered for mapping and screening of causative genes in LGMD families (Speer et al., 1992; Mojbafan et al., 2020); however, NGS has almost replaced the older tools introducing WES or targeted genes' panels in sporadic and familial cases of LGMD (Ghaoui et al., 2015; Ozyilmaz et al., 2019). NGS tools enables more efficient and cost-effective genetic screening. WES analysis has confirmed causative gene identification in 45% of LGMD patients which were declared "difficult-to-diagnose" clinically (Ghaoui et al., 2015). Similarly, LGMD specialized laboratories in US and UK have improved WES sensitivity for the molecular diagnostics. Turkish and Brazilian patients also been successfully diagnosed through customized NGS panels (Ozyilmaz et al., 2019; Winckler et al., 2019). To the best of our knowledge, only two other families with subclinical types LGMDR6 and LGMDR10 have been reported so far from Khyber-Pakhtunkhwa province, investigated through WES analysis (Khan et al., 2019; Younus et al., 2019). This study describes the first ever LGMDR4 report from Pakhtun ethnicity of Pakistan.

Genetic screening and counselling could be challenging, especially of LGMD patients from remote areas of Pakistan. Lack of clinical history makes it difficult to comprehend the disease progression. In such conditions, only research laboratories make their efforts to visit these remote areas and make it possible to perform genetic testing through sequencing for these families. Taking muscles biopsy is a painful invasive procedure for patient (Joyce et al., 2012). Furthermore, diminished or absent immunostaining in muscle biopsies can though identify SG abnormalities and early loss of ambulation (Semplicini et al., 2015); but, it does not detect causative genes' alterations. Thus, customized NGS genes panels, WES or genome sequencing could be better options to successfully and efficiently identify the LGMD clinical subtypes.

5 Conclusion

LGMDs have been assigned to more than 30 genes so far, with an equal number of clinical and sub-clinical types. WES analysis can be used as a first line molecular diagnostic tool in single gene disorders like LGMDs and it has the potential to uncover the underlying genetic causes. In-time and cost-effective genetic testing through NGS technologies will have a positive impact on genetic counselling, premarital and embryonic screening of various ethnic groups of Pakistan, where cousin marriages are more frequent. It is suggested that WES analysis be adopted as premarital testing for LGMD and other related rare genetic disorders of this population. The identified variant (p.Ser204Pro) increased the variation spectra of the *SGCB* gene and detailed structural and functional analysis uncovered underlying molecular mechanism of its pathogenesis.

Acknowledgements

The authors would like to acknowledge Higher Education Commission (HEC), Pakistan for the financial support (Project No. 20.3479), under the National Research Program for Universities. Dr. Abdul Wadood is PI and Dr. Musharraf Jelani is Co-I of this project.

Authors' contributions

All authors contributed to the study conception and design. Material preparation, data collection and analyses and first draft writing were performed by Memona Inam, Amin Jan, Muhammad Tariq, Habib Ahmad, Isse Ali, and Aziz Khan. Computational experiments design and execution by Nousheen Bibi. The manuscript was reviewed, edited and finalized by Muhammad Latif, Hussein Sheikh Ali Mohamoud, Jamal Nasir, Abdul Wadood and Musharraf Jelani. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript

Conflict of interest

Authors declare no conflict of interest

Data availability

Exome and Sanger sequencing raw data of the subjects is available upon special request.

References

References

Alavi, A., Esmaeili, S., Nilipour, Y., Nafissi, S., Tonekaboni, S.H., Zamani, G., Ashrafi, M.R., Kahrizi, K., Najmabadi, H. and Jazayeri, F., 2017. LGMD2E is the most common type of sarcoglycanopathies in the Iranian population. J Neurogenet 31, 161-169.

Alonso-Pérez, J., González-Quereda, L., Bello, L., Guglieri, M., Straub, V., Gallano, P., Semplicini, C., Pegoraro, E., Zangaro, V., Nascimento, A., Ortez, C., Comi, G.P., Dam, L.T., De Visser, M., van der Kooi, A.J., Garrido, C., Santos, M., Schara, U., Gangfuß, A., Løkken, N., Storgaard, J.H., Vissing, J., Schoser, B., Dekomien, G., Udd, B., Palmio, J., D'Amico, A., Politano, L., Nigro, V., Bruno, C., Panicucci, C., Sarkozy, A., Abdel-Mannan, O., Alonso-Jimenez, A., Claeys, K.G., Gomez-Andrés, D., Munell, F., Costa-Comellas, L., Haberlová, J., Rohlenová, M., Elke, V., De Bleecker, J.L., Dominguez-González, C., Tasca, G., Weiss, C., Deconinck, N., Fernández-Torrón, R., López de Munain, A., Camacho-Salas, A., Melegh, B., Hadzsiev, K., Leonardis, L., Koritnik, B., Garibaldi, M., de Leon-Hernández, J.C., Malfatti, E., Fraga-Bau, A., Richard, I., Illa, I. and Díaz-Manera, J., 2020. New genotype-phenotype correlations in a large European cohort of patients with sarcoglycanopathy. Brain 143, 2696-2708.

Barresi, R., Di Blasi, C., Negri, T., Brugnoni, R., Vitali, A., Felisari, G., Salandi, A., Daniel, S., Cornelio, F., Morandi, L. and Mora, M., 2000. Disruption of heart sarcoglycan complex and severe cardiomyopathy caused by beta sarcoglycan mutations. J Med Genet 37, 102-7.

Belhasan, D.C. and Akaaboune, M., 2020. The role of the dystrophin glycoprotein complex on the neuromuscular system. Neurosci Lett 722, 134833.

Bonnemann, C.G., Modi, R., Noguchi, S., Mizuno, Y., Yoshida, M., Gussoni, E., McNally, E.M., Duggan, D.J., Angelini, C. and Hoffman, E.P., 1995. Beta-sarcoglycan (A3b) mutations

cause autosomal recessive muscular dystrophy with loss of the sarcoglycan complex. Nat Genet 11, 266-73.

Duggan, D.J., Gorospe, J.R., Fanin, M., Hoffman, E.P. and Angelini, C., 1997. Mutations in the sarcoglycan genes in patients with myopathy. N Engl J Med 336, 618-24.

Gao, Q.Q. and McNally, E.M., 2015. The Dystrophin Complex: Structure, Function, and Implications for Therapy. Compr Physiol 5, 1223-39.

Gawor, M. and Prószyński, T.J., 2018. The molecular cross talk of the dystrophin-glycoprotein complex. Ann N Y Acad Sci 1412, 62-72.

Ghaoui, R., Cooper, S.T., Lek, M., Jones, K., Corbett, A., Reddel, S.W., Needham, M., Liang, C., Waddell, L.B., Nicholson, G., O'Grady, G., Kaur, S., Ong, R., Davis, M., Sue, C.M., Laing, N.G., North, K.N., MacArthur, D.G. and Clarke, N.F., 2015. Use of Whole-Exome Sequencing for Diagnosis of Limb-Girdle Muscular Dystrophy: Outcomes and Lessons Learned. JAMA Neurol 72, 1424-32.

Joyce, N.C., Oskarsson, B. and Jin, L.W., 2012. Muscle biopsy evaluation in neuromuscular disorders. Phys Med Rehabil Clin N Am 23, 609-31.

Khan, A., Wang, R., Han, S., Umair, M., Abbas, S., Khan, M.I., Alshabeeb, M.A., Alfadhel, M. and Zhang, X., 2019. Homozygous missense variant in the TTN gene causing autosomal recessive limb-girdle muscular dystrophy type 10. BMC Med Genet 20, 166.

Mazzarotto, F., Tayal, U., Buchan, R.J., Midwinter, W., Wilk, A., Whiffin, N., Govind, R., Mazaika, E., de Marvao, A., Dawes, T.J.W., Felkin, L.E., Ahmad, M., Theotokis, P.I., Edwards, E., Ing, A.Y., Thomson, K.L., Chan, L.L.H., Sim, D., Baksi, A.J., Pantazis, A., Roberts, A.M., Watkins, H., Funke, B., O'Regan, D.P., Olivotto, I., Barton, P.J.R., Prasad, S.K., Cook, S.A., Ware, J.S. and Walsh, R., 2020. Reevaluating the Genetic Contribution of Monogenic Dilated Cardiomyopathy. Circulation 141, 387-398.

Mojbafan, M., Bahmani, R., Bagheri, S.D., Sharifi, Z. and Zeinali, S., 2020. Mutational spectrum of autosomal recessive limb-girdle muscular dystrophies in a cohort of 112 Iranian patients and reporting of a possible founder effect. Orphanet J Rare Dis 15, 14.

Ozyilmaz, B., Kirbiyik, O., Ozdemir, T.R., Kaya Ozer, O., Kutbay, Y.B., Erdogan, K.M., Guvenc, M.S., Kale, M.Y., Gazeteci, H., Kilic, B., Sertpoyraz, F., Diniz, G., Baydan, F., Gencpinar, P., Dundar, N.O. and Yis, U., 2019. Impact of next-generation sequencing panels in the evaluation of limb-girdle muscular dystrophies. Ann Hum Genet.

Semplicini, C., Vissing, J., Dahlqvist, J.R., Stojkovic, T., Bello, L., Witting, N., Duno, M., Leturcq, F., Bertolin, C., D'Ambrosio, P., Eymard, B., Angelini, C., Politano, L., Laforêt, P. and Pegoraro, E., 2015. Clinical and genetic spectrum in limb-girdle muscular dystrophy type 2E. Neurology 84, 1772-81.

Speer, M.C., Yamaoka, L.H., Gilchrist, J.H., Gaskell, C.P., Stajich, J.M., Vance, J.M., Kazantsev, A., Lastra, A.A., Haynes, C.S., Beckmann, J.S. and et al., 1992. Confirmation of genetic heterogeneity in limb-girdle muscular dystrophy: linkage of an autosomal dominant form to chromosome 5q. Am J Hum Genet 50, 1211-7.

Straub, V., Murphy, A. and Udd, B., 2018. 229th ENMC international workshop: Limb girdle muscular dystrophies - Nomenclature and reformed classification Naarden, the Netherlands, 17-19 March 2017. Neuromuscul Disord 28, 702-710.

Sveen, M.L., Thune, J.J., Køber, L. and Vissing, J., 2008. Cardiac involvement in patients with limb-girdle muscular dystrophy type 2 and Becker muscular dystrophy. Arch Neurol 65, 1196-201.

Tarakci, H. and Berger, J., 2016. The sarcoglycan complex in skeletal muscle. Front Biosci (Landmark Ed) 21, 744-56.

Wicklund, M.P., 2019. The Limb-Girdle Muscular Dystrophies. Continuum (Minneap Minn) 25, 1599-1618.

Winckler, P.B., da Silva, A.M.S., Coimbra-Neto, A.R., Carvalho, E., Cavalcanti, E.B.U., Sobreira, C.F.R., Marrone, C.D., Machado-Costa, M.C., Carvalho, A.A.S., Feio, R.H.F., Rodrigues, C.L., Gonçalves, M.V.M., Tenório, R.B., Mendonça, R.H., Cotta, A., Paim, J.F.O., Costa, E.S.C., de Aquino Cruz, C., Benà, M.I., Betancur, D.F.A., El Husny, A.S., de Souza, I.C.N., Duarte, R.C.B., Reed, U.C., Chaves, M.L.F., Zanoteli, E., França, M.C., Jr. and Saute, J.A., 2019. Clinicogenetic lessons from 370 patients with autosomal recessive limb-girdle muscular dystrophy. Clin Genet 96, 341-353.

Younus, M., Ahmad, F., Malik, E., Bilal, M., Kausar, M., Abbas, S., Shaheen, S., Kakar, M.U., Alfadhel, M. and Umair, M., 2019. SGCD Homozygous Nonsense Mutation (p.Arg97*) Causing Limb-Girdle Muscular Dystrophy Type 2F (LGMD2F) in a Consanguineous Family, a Case Report. Front Genet 9, 727.

Alavi, A., Esmaeili, S., Nilipour, Y., Nafissi, S., Tonekaboni, S.H., Zamani, G., Ashrafi, M.R., Kahrizi, K., Najmabadi, H. and Jazayeri, F., 2017. LGMD2E is the most common type of sarcoglycanopathies in the Iranian population. J Neurogenet 31, 161-169.

Alonso-Pérez, J., González-Quereda, L., Bello, L., Guglieri, M., Straub, V., Gallano, P., Semplicini, C., Pegoraro, E., Zangaro, V., Nascimento, A., Ortez, C., Comi, G.P., Dam, L.T., De Visser, M., van der Kooi, A.J., Garrido, C., Santos, M., Schara, U., Gangfuß, A., Løkken, N., Storgaard, J.H., Vissing, J., Schoser, B., Dekomien, G., Udd, B., Palmio, J., D'Amico, A., Politano, L., Nigro, V., Bruno, C., Panicucci, C., Sarkozy, A., Abdel-Mannan, O., Alonso-Jimenez, A., Claeys, K.G., Gomez-Andrés, D., Munell, F., Costa-Comellas, L., Haberlová, J., Rohlenová, M., Elke, V., De Bleecker, J.L., Dominguez-González, C., Tasca, G., Weiss, C., Deconinck, N., Fernández-Torrón, R., López de Munain, A., Camacho-Salas, A., Melegh, B., Hadzsiev, K., Leonardis, L., Koritnik, B., Garibaldi, M., de Leon-Hernández, J.C., Malfatti, E., Fraga-Bau, A., Richard, I., Illa, I. and Díaz-Manera, J., 2020. New genotype-phenotype correlations in a large European cohort of patients with sarcoglycanopathy. Brain 143, 2696-2708.

Bhinder, M.A., Sadia, H., Mahmood, N., Qasim, M., Hussain, Z., Rashid, M.M., Zahoor, M.Y., Bhatti, R., Shehzad, W., Waryah, A.M. and Jahan, S., 2019. Consanguinity: A blessing or menace at population level? Ann Hum Genet 83, 214-219.

Bonnemann, C.G., Modi, R., Noguchi, S., Mizuno, Y., Yoshida, M., Gussoni, E., McNally, E.M., Duggan, D.J., Angelini, C. and Hoffman, E.P., 1995. Beta-sarcoglycan (A3b) mutations cause autosomal recessive muscular dystrophy with loss of the sarcoglycan complex. Nat Genet 11, 266-73.

Bushby, K.M., 1995. Diagnostic criteria for the limb-girdle muscular dystrophies: report of the ENMC Consortium on Limb-Girdle Dystrophies. Neuromuscul Disord 5, 71-4.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C.J.A.C.S.D.B.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. 66, 12-21.

Colovos, C. and Yeates, T.O., 1993. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2, 1511-9.

Dalichaouche, I., Sifi, Y., Roudaut, C., Sifi, K., Hamri, A., Rouabah, L., Abadi, N. and Richard, I., 2017. γ-sarcoglycan and dystrophin mutation spectrum in an Algerian cohort. Muscle Nerve 56, 129-135.

Duggan, D.J., Gorospe, J.R., Fanin, M., Hoffman, E.P. and Angelini, C., 1997. Mutations in the sarcoglycan genes in patients with myopathy. N Engl J Med 336, 618-24.

Gao, Q.Q. and McNally, E.M., 2015. The Dystrophin Complex: Structure, Function, and Implications for Therapy. Compr Physiol 5, 1223-39.

Ghaoui, R., Cooper, S.T., Lek, M., Jones, K., Corbett, A., Reddel, S.W., Needham, M., Liang, C., Waddell, L.B., Nicholson, G., O'Grady, G., Kaur, S., Ong, R., Davis, M., Sue, C.M., Laing, N.G., North, K.N., MacArthur, D.G. and Clarke, N.F., 2015. Use of Whole-Exome

Sequencing for Diagnosis of Limb-Girdle Muscular Dystrophy: Outcomes and Lessons Learned. JAMA Neurol 72, 1424-32.

Guimarães-Costa, R., Fernández-Eulate, G., Wahbi, K., Leturcq, F., Malfatti, E., Behin, A., Leonard-Louis, S., Desguerre, I., Barnerias, C., Nougues, M.C., Isapof, A., Estournet-Mathiaud, B., Quijano-Roy, S., Fayssoil, A., Orlikowski, D., Fauroux, B., Richard, I., Semplicini, C., Romero, N.B., Querin, G., Eymard, B., Laforêt, P. and Stojkovic, T., 2020. Clinical correlations and long-term follow-up in 100 patients with sarcoglycanopathies. Eur J Neurol.

Harris, E., Topf, A., Barresi, R., Hudson, J., Powell, H., Tellez, J., Hicks, D., Porter, A., Bertoli, M., Evangelista, T., Marini-Betollo, C., Magnusson, O., Lek, M., MacArthur, D., Bushby, K., Lochmuller, H. and Straub, V., 2017. Exome sequences versus sequential gene testing in the UK highly specialised Service for Limb Girdle Muscular Dystrophy. Orphanet J Rare Dis 12, 151.

Hoffman, E.P., Brown, R.H., Jr. and Kunkel, L.M., 1987. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 51, 919-28.

Jo, H. and Koh, G., 2015. Faster single-end alignment generation utilizing multi-thread for BWA. Biomed Mater Eng 26 Suppl 1, S1791-6.

Jurcik, A., Bednar, D., Byska, J., Marques, S.M., Furmanova, K., Daniel, L., Kokkonen, P., Brezovsky, J., Strnad, O., Stourac, J., Pavelka, A., Manak, M., Damborsky, J. and Kozlikova, B., 2018. CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. Bioinformatics 34, 3586-3588.

Khan, A., Wang, R., Han, S., Umair, M., Abbas, S., Khan, M.I., Alshabeeb, M.A., Alfadhel, M. and Zhang, X., 2019. Homozygous missense variant in the TTN gene causing autosomal recessive limb-girdle muscular dystrophy type 10. BMC Med Genet 20, 166.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8, 477-86.

Mah, J.K., Korngut, L., Fiest, K.M., Dykeman, J., Day, L.J., Pringsheim, T. and Jette, N., 2016. A Systematic Review and Meta-analysis on the Epidemiology of the Muscular Dystrophies. Can J Neurol Sci 43, 163-77.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-303.

Mojbafan, M., Bahmani, R., Bagheri, S.D., Sharifi, Z. and Zeinali, S., 2020. Mutational spectrum of autosomal recessive limb-girdle muscular dystrophies in a cohort of 112 Iranian patients and reporting of a possible founder effect. Orphanet J Rare Dis 15, 14.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30, 2785-91.

Nigro, V. and Savarese, M., 2014. Genetic basis of limb-girdle muscular dystrophies: the 2014 update. Acta Myol 33, 1-12.

Ozyilmaz, B., Kirbiyik, O., Ozdemir, T.R., Kaya Ozer, O., Kutbay, Y.B., Erdogan, K.M., Guvenc, M.S., Kale, M.Y., Gazeteci, H., Kilic, B., Sertpoyraz, F., Diniz, G., Baydan, F., Gencpinar, P., Dundar, N.O. and Yis, U., 2019. Impact of next-generation sequencing panels in the evaluation of limb-girdle muscular dystrophies. Ann Hum Genet.

Passos-Bueno, M.R., Vainzof, M., Moreira, E.S. and Zatz, M., 1999. Seven autosomal recessive limb-girdle muscular dystrophies in the Brazilian population: from LGMD2A to LGMD2G. Am J Med Genet 82, 392-8.

Pervaiz, R., Faisal, F. and Serakinci, N., 2018. Practive of consanguinity and attitudes towards risk in the Pashtun population of Khyber Pakhtunkhwa, Pakistan. J Biosoc Sci 50, 414-420.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-12.

Reddy, H.M., Cho, K.A., Lek, M., Estrella, E., Valkanas, E., Jones, M.D., Mitsuhashi, S., Darras, B.T., Amato, A.A., Lidov, H.G., Brownstein, C.A., Margulies, D.M., Yu, T.W., Salih, M.A., Kunkel, L.M., MacArthur, D.G. and Kang, P.B., 2017. The sensitivity of exome sequencing in identifying pathogenic mutations for LGMD in the United States. J Hum Genet 62, 243-252.

Snider, C., Jayasinghe, S., Hristova, K. and White, S.H., 2009. MPEx: a tool for exploring membrane proteins. Protein Sci 18, 2624-8.

Speer, M.C., Yamaoka, L.H., Gilchrist, J.H., Gaskell, C.P., Stajich, J.M., Vance, J.M., Kazantsev, A., Lastra, A.A., Haynes, C.S., Beckmann, J.S. and et al., 1992. Confirmation of genetic heterogeneity in limb-girdle muscular dystrophy: linkage of an autosomal dominant form to chromosome 5q. Am J Hum Genet 50, 1211-7.

Straub, V., Murphy, A. and Udd, B., 2018. 229th ENMC international workshop: Limb girdle muscular dystrophies - Nomenclature and reformed classification Naarden, the Netherlands, 17-19 March 2017. Neuromuscul Disord 28, 702-710.

Trabelsi, M., Kavian, N., Daoud, F., Commere, V., Deburgrave, N., Beugnet, C., Llense, S., Barbot, J.C., Vasson, A., Kaplan, J.C., Leturcq, F. and Chelly, J., 2008. Revised spectrum of mutations in sarcoglycanopathies. Eur J Hum Genet 16, 793-803.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S. and DePristo, M.A., 2013. From FastQ data to high confidence variant calls:

the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43, 11.10.1-33.

Venselaar, H., te Beek, T.A., Kuipers, R.K., Hekkelman, M.L. and Vriend, G.J.B.b., 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. 11, 548.

Wallace, A.C., Laskowski, R.A. and Thornton, J.M., 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8, 127-34.

Wicklund, M.P., 2019. The Limb-Girdle Muscular Dystrophies. Continuum (Minneap Minn) 25, 1599-1618.

Winckler, P.B., da Silva, A.M.S., Coimbra-Neto, A.R., Carvalho, E., Cavalcanti, E.B.U., Sobreira, C.F.R., Marrone, C.D., Machado-Costa, M.C., Carvalho, A.A.S., Feio, R.H.F., Rodrigues, C.L., Gonçalves, M.V.M., Tenório, R.B., Mendonça, R.H., Cotta, A., Paim, J.F.O., Costa, E.S.C., de Aquino Cruz, C., Benà, M.I., Betancur, D.F.A., El Husny, A.S., de Souza, I.C.N., Duarte, R.C.B., Reed, U.C., Chaves, M.L.F., Zanoteli, E., França, M.C., Jr. and Saute, J.A., 2019. Clinicogenetic lessons from 370 patients with autosomal recessive limb-girdle muscular dystrophy. Clin Genet 96, 341-353.

Younus, M., Ahmad, F., Malik, E., Bilal, M., Kausar, M., Abbas, S., Shaheen, S., Kakar, M.U., Alfadhel, M. and Umair, M., 2019. SGCD Homozygous Nonsense Mutation (p.Arg97*) Causing Limb-Girdle Muscular Dystrophy Type 2F (LGMD2F) in a Consanguineous Family, a Case Report. Front Genet 9, 727.

Zhang, Y.J.B.b., 2008. I-TASSER server for protein 3D structure prediction. 9, 40.