Sequence Variants in Three Genes Underlying Leukodystrophy in Pakistani Families

Authors: Farah Bibi\textsuperscript{1,4*}, Nighat Haider\textsuperscript{2*}, Shahab Ud Din\textsuperscript{3}, Muqadar Shah\textsuperscript{2}, Jai Krishin\textsuperscript{2}, Naila Qayyum\textsuperscript{5}, Ghazala Kaukab Raja\textsuperscript{1}, Henry Houlden\textsuperscript{4}, Wasim Ahmad\textsuperscript{5}, Tanwir Khaliq\textsuperscript{3}, Asmat Ullah\textsuperscript{3,5}

* Equal contribution

1 University Institute of Biochemistry & Biotechnology, PMAS – Arid Agriculture University, Rawalpindi, 43600, Pakistan

2 Department of Paediatric Medicine, Shaheed Zulfiqar Ali Bhutto Medical University, Children Hospital, Pakistan Institute of Medical Sciences, Islamabad

3 Department of Molecular Biology, Shaheed Zulfiqar Ali Bhutto Medical University, Islamabad

4 Department of Neuromuscular disorders, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK

5 Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan

Corresponding author

Dr. Asmat Ullah, Department of Molecular Biology, Shaheed Zulfiqar Ali Bhutto Medical University, Islamabad; Tel: +92-51-9107511, E-mail: asmatullah@bs.qau.edu.pk,

ORCID: https://orcid.org/0000-0002-5715-3425

Co-author ORCIDS:

Farah Bibi, https://orcid.org/0000-0002-3468-2586

Nighat Haider: https://orcid.org/0000-0001-5753-3359

Shahab UD Din: https://orcid.org/0000-0001-7299-3388

Muqadar Shah: muqadar75@gmail.com

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/JDN.10036

This article is protected by copyright. All rights reserved
Funding statement
This work was funded by the Higher Education Commission (HEC), Islamabad, Pakistan. F.B. was supported by International Research Support Initiative Program (IRSIP) from HEC, Islamabad, Pakistan.

Conflict of interest disclosure
The authors declare no conflict of interest.

Ethics approval statement
The present study was reviewed and approved by the Ethics Committee for the use of human subjects at University Institute of Biochemistry and Biotechnology, Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, and Shaheed Zulfiqar Ali Bhutto Medical University Islamabad Pakistan.

Patient Consent Statement
Signed informed consent to perform the study and publication of radiographs was obtained from the guardians/parents of affected individuals.

Permission to reproduce material from other sources
All the sources from which material is adapted are properly cited in the manuscript.
Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.
Abstract

Leukodystrophies (LDs) are a heterogeneous group of rare and progressive genetic diseases that affect brain, spinal cord and often the peripheral nerves. They are characterized by abnormal development or destruction of the myelin sheath of the brain. This study was aimed to search for the causative variants in three unrelated consanguineous families presented with LD. Detailed clinical investigations were carried out on probands in three unrelated consanguineous families of Pakistani origin. Targeted gene sequencing and Whole Exome Sequencing (WES) were performed for variant identification. Candidate variants were checked for co-segregation with the phenotype using Sanger sequencing. Public databases including ExAC, gnomAD, dbSNP and the 1,000 Genome Project were searched to determine frequencies of the alleles. Conservation of the missense variants was ensured by aligning orthologous protein sequences from diverse vertebrate species.

Targeted gene sequencing identified a novel homozygous missense mutation [c.2135G>A, p.(Arg712His)] in the ATP Binding Cassette Subfamily D Member 1 (ABCD1; OMIM# 300371) in three affected siblings in family A. WES followed by validation by Sanger sequencing revealed previously reported homozygous missense variants [c.162C>A; p.(Asn54Lys)] in ASPA (OMIM# 608034) in family B and [c.361G>C;p.(Gly121Arg)] in ARSA (OMIM# 607574) in family C.
Investigation of three families underlies importance of WES as an amazing diagnostic tool for conclusive determination of a specific type of LD. Further, the study would assist in carrier testing and prenatal diagnosis of the affected families. In addition, searching for common variants in the genes causing LD would help in designing low-cost targeted variation testing in patients.

**Key Words:** Leukodystrophy, Targeted Gene Sequencing, WES, Variants in *ABCD1*, *ASPA*, and *ARSA*
Myelin is an important part of the basic makeup of the central nervous system (CNS). Defects in myelin biogenesis and process of myelination can lead to a group of disorders known as leukodystrophies (Bonkowsky et al., 2010). Leukodystrophies (LDs) are heterogeneous group of inherited abnormalities of white matter of the CNS with or without peripheral nervous system involvement (Vanderver et al., 2015). Up till now, thirty types of leukodystrophies are described with specific clinical characteristics and genetic causes (Froukh, 2019). All LDs types have genetic basis and do not include acquired myelin disorders. While the related term Leukoencephalopathy (LE) refers to white matter abnormalities in general and may either be acquired or inherited in nature (Vanderver et al., 2015). Since each type of LD affects a different part of the myelin sheath, therefore they lead to a range of neurological problems. However, the specific-characteristic symptoms of LDs include decreased motor function, muscle rigidity, and eventual degeneration of sight and hearing. Some of the LDs are progressive and often fatal and its age of onset is presented across all age groups including infantile (first year), late infantile (1–5 years), juvenile (5–12 years) or adolescence and adulthood.

The inheritance patterns of the described LD types are autosomal recessive, de novo dominant, X-linked recessive and X-linked dominant suggesting LDs to be monogenic/Mendelian disorders (Froukh, 2019). The most common types of LDs are Metachromatic leukodystrophy, Canavan disease, Krabbe disease, Alexander disease, and X-linked adrenoleukodystrophy (OMIM# 300100) (Vanderver et al., 2015). Definitive diagnosis of the LD type is very challenging as less is known about etiology of each type and the curative treatment of LDs is currently limited.

The basic and key source of investigation in patient with suspected LD is brain MRI. Neurologic features such as loss of cerebellar volume, autonomic dysfunction, alterations in head circumference, seizures or neurobehavioral abnormalities and extra neurologic features such as adrenal insufficiency, endocrine disturbances, visual, skeletal and hearing impairment, and gastrointestinal symptoms can further help refine diagnostic considerations (Parikh et al., 2015). However there are certain limitations to these diagnostics recommendations since a single brain MRI is not sufficient to distinguish between delayed myelination, hypomyelination and the early stages of a LD and serial MRI scans are often required, usually with a minimum of 6–12 months interval between studies and also the degree to which certain neurologic and extra neurologic features are present may vary with the age of presentation and the specific LD. Following MRI
pattern analysis and clinical examination, the serial biochemical testing is another primary diagnostic approach to LDs. However, the next-generation sequencing (NGS) can act as a powerful diagnostic tool for conclusive determination of specific type of LD.

Recently, NGS has been proved to be very useful and powerful definitive diagnostic tool in identifying many genes that cause myelin defects. Since the number of genes associated with LD is increasing, it is arguable to use NGS genetic testing using gene panels, whole exome sequencing (WES), or whole genome sequencing (WGS) as the best genetic testing option (Parikh et al., 2015). Therefore, if the initial evaluation of the patient indicates the possibility of LD, then molecular genetic testing is required.

In Pakistan, the epidemiological data on LD is very scarce and genetic causes are not known in several of these cases. Therefore, this study aims to identify the genetic causes of LD in five affected patients, including four males and one female, from three unrelated consanguineous families of Pakistani origin.

Materials and Methods

Ethical Compliance and Family Recruitment

The study, conducted and presented here, was approved by the Ethics Committee for the use of human subjects at University Institute of Biochemistry and Biotechnology, Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, and Shaheed Zulfiqar Ali Bhutto Medical University Islamabad Pakistan.

Three families (A-C), resident of Punjab province in Pakistan, were recruited for the present study (Figure.1a-c). Pedigree analysis showed inheritance pattern of X-linked in family A and autosomal recessive in two other families (B, C). Based on brain MRI, neurological and extra neurological features, patients in each family were diagnosed with LD (Figure 2). Phenotypes recorded in the patients in the three families were distinct. Written informed consent was obtained from all those who participated in the study.

DNA Extraction and Targeted Gene Sequencing

This article is protected by copyright. All rights reserved
DNA was extracted from blood samples, collected from the participants, by a protocol involving sucrose lysis, proteinase K digestion, and salting out. Based upon the inheritance pattern of disease and association of $ABCD1$ gene, located on X-chromosome (Xq28), with adrenoleukodystrophy, primers were designed to sequence the same gene in DNA samples of participants in family A. Gene specific primers were designed using Primer3 software (http://bioinfo.ut.ee/primer3-0.4.0/). Standard thermocycler conditions were used for gene amplification.

**Whole Exome Sequencing**

DNA sample from one patient in each of the two families, B and C, was subjected to WES. Nextera RapidCapture Enrichment kit (Illumina) was used according to the manufacturer instructions, which provides a fast simple method for isolating the human exome. Libraries were sequenced in an Illumina HiSeq3000 using a 100-bp paired-end reads protocol (Macrogen, Seoul, Korea).

**Data Annotation**

Exome data were annotated using the wANNOVAR application (http://wannovar.wglab.org/). Only homozygous variants were considered given that the mode of inheritance was autosomal recessive and the families were consanguineous. Variants were filtered so that those with an allele frequency ≥0.01 in any public database (ExAC, gnomAD, dbSNP, and the 1,000 Genome Project) were excluded. Variants located in the exons and those predicted to affect splicing were considered. Regions of autozygosity in the whole-exome data were determined using the Homozygosity mapper software (https://www.homozygositymapper.org). Synonymous variants in these autozygous regions were also considered. The conservation of variants in respective orthologues was checked using the HomoloGene site (https://www.ncbi.nlm.nih.gov/homologene), the UCSC multiz alignment tool (http://genome.ucsc.edu/), and by aligning the gene sequences from different species using the Clustal Omega program (https://www.ebi.ac.uk/Tools/msa/clustalo/).

**Segregation Analysis**
Specific primers were designed for PCR amplification of regions containing the variants, and the products were subjected to Sanger sequencing using BigDye™ Terminator V.3.1 cycle sequencing kit and ABI PRISM 3730XL sequencer (Applied Biosystems Inc., USA). PCR was carried out using the conditions previously applied by Ullah et al., (2017). Segregation analysis was performed on all available family samples, and sequencing results were analyzed with the BioEdit version 6.0.7 (Ibis Biosciences, Carlsbad, CA, USA).

Results

Clinical Findings

Family A: This family had three males (IV-3, IV-4, IV-5) which were available for clinical and genetic analysis. Affected individual (IV-3) was 12 years old male presented with complaints of difficulty in speech and inability to read and write. The child was born at full term through spontaneous vertex delivery with immediate cry and unremarkable antenatal history. He was developmentally delayed with neck holding at 2 years, sitting at 3 years, walking at 4 years and still unable to speak proper words but can understand the language. He is the fourth product of the consanguineous marriage. All the base line laboratory studies were within the normal range. Adrenocorticotropic hormone and cortisol levels were also normal. MRI brain showed abnormal signals in the periventricular deep white matter predominantly along bilateral atria and occipital horns appearing hyperintense on T2W1 FLAIR is to hypointense on T1W1 not showing diffusion restriction. Subcortical U fibers were not affected. Findings were suggestive of adrenoleukodystrophy.

Two other male siblings IV-4 and IV-5 aged 10 and 7 years respectively had expressive dysphasia although they were developmentally normal. History of similar complaint was recorded in a maternal uncle (III-6). On examination he was conscious and alert with normal vitals, weight and height falling on 50th centile. General physical examination was unremarkable. Tone, power and reflexes were normal with intact cranial nerves and no cerebellar dysfunction. No abnormality was detected on rest of systemic review.

Family B: Affected individual (V-2) was 2 years-old female born to a consanguineous couple. Clinically, the patient was found to have macrocephaly with delayed milestone. She was noted to
be floppy at 3 month of age, and had not attained head control even at 6 months. Brain ultrasound, performed at an age of 6 months, was normal. CT scan of brain noted symmetrical diffuse hypodensity in white matter suggesting likely presentation of normal myelination pattern for age. Later at age of 10 months, patient suffered high fever and drowsiness for almost 3-4 days. MRI brain findings suggested leukodystrophy with extensive signal abnormality in bilateral periventricular and subcortical white matter, bilateral basal ganglia and thalami extending into brainstem and cerebellar hemispheres. Her general condition deteriorated over the next 12–18 months. She was unable to sit, stand or turning over in bed and had progressively enlarging head size. At age of 2 years, she was bed-ridden with no eye to eye contact, and poor response to vocal stimuli and surroundings.

Family C: Patient (IV-1) in the family was 4 years old boy conceived from a consanguineous marriage. The proband was presented with regression of milestones. An impaired motor development became apparent by the age of 8 to 15 months, when developmental milestones such as unaided sitting and/or walking were delayed or could not be achieved. He was unable to hold his neck and sit independently. Patient was conscious but non-responsive to his surroundings. He had rigid body tone along with microcephaly and unable to speak. He had a history of multiple generalized seizures which first appeared at third year of life. The MRI brain detected widespread symmetrical demyelination of the central and periventricular white matter of both cerebral hemispheres. The abdominal ultrasound was unremarkable. No sign of cardiac and respiratory issues, and visual loss were seen at the time of study. None of the other family members had any of the above-described complaints.

**Genetic Findings**

Considering X-linked inheritance pattern of ALD in family A, the previously reported causative gene ABCD1 (transcript NM_000033) was sequenced in all available affected ad unaffected members. Sequence analysis revealed a novel missense variant [c.2135G>A, p.(Arg712His)] segregating with the disease phenotype in the family (Figure 1d). The variant was not found in ExAC, 1000Genome, gnomAD, dbSNP and was predicted to be disease causing by MutationTaster (www.mutationtatser.org), Polyphene-2 (http://genetics.bwh.harvard.edu/pph2/).
and CADD (https://cadd.gs.washington.edu/). By conservation analysis, arginine (R) at amino acid position 712 was found conserved among different species (Figure 1e).

In family B, analysis of the filtered exome data revealed two-years old female proband (V-1) was homozygous for the variant [c.162C>A, p.(Asn54Lys)] in the ASPA causing Canavan disease (OMIM# 271900). Parents of the affected individual showed heterozygosity for the mutant allele.

In family C, analysis of the filtered exome data revealed four-year old male proband (IV-1) was homozygous for the variant [c.361G>C, p.(Gly121Arg)] in the ARSA causing Metachromatic leukodystrophy (OMIM# 250100). Parents of the affected individual were found to be heterozygous carriers of the variant (Figure 1d,f,g).

**Discussion**

The present study described investigation of three unrelated families segregating LD either in X-linked or autosomal recessive manner. Sanger sequencing revealed a novel variant p.(Arg712His) in the ABCD1 in family A with X-linked inheritance of ALD. In other two families, with autosomal recessive inheritance pattern of LD, two previously reported variants were identified. This included p.(Asn54Lys) in the ASPA in family B and p.(Gly121Arg) in the ARSA in family C.

The X-linked ALD is a rare type of LD caused by an impairment of peroxisomal beta-oxidation of a very long chain fatty acids (VLCFAs) and subsequently their accumulation in the CNS, adrenal cortex and testes due to absent or abnormal functioning of adrenoleukodystrophy protein (ALDP). 745 amino acids ALDP is encoded by ABCD1, which is located on chromosome X and composed of 10 exons. Its N-terminal contains a transmembrane domain (TMD) which consists of six membrane-passing segments (amino acids 1-506), hydrophilic cytosolic nucleotide binding fold (NBF) containing the highly conserved motifs Walker A, Walker B and 19-mer that are important for substrate specificity and nucleotide binding (amino acids 507-630), and the C-terminal region (amino acids631-745) (Hillebrand et al., 2007). All known peroxisomal ABC transporters are half transporters and require a partner half transporter to form a functional homodimer or heterodimer transporter which is likely involved in the peroxisomal transport or

This article is protected by copyright. All rights reserved
catabolism of very long chain fatty acid (VLCFA). The last 87-C terminal amino acids of ALDP ranging from 658-745 harbors the most significant protein area intervening the ALDP homo- or heterodimerization. Along these lines the change identified in our family A p.(Arg712His) is highly likely to interfere with these interactions and affect the function of the ALDP.

Until now, more than 500 different mutations in the \( ABCD1 \) gene have been identified which are distributed along the whole gene. The exact prevalence of X-ALD in Pakistani population is unknown. Sultan et al. (2014) reported that out of total 366 Pakistani cases of degenerative brain disease, 19.7% cases were presented with X-ALD. In another study, a single Pakistani child was reported with X-ALD. However, in both these studies, diagnosis was based on abnormal findings on brain MRI complemented by selected electrophysiological tests and clinical clues. To the best of our knowledge, present case from Pakistan is the first report of X-linked ALD validated by molecular analysis.

Pathogenic mutations in the \( ASPA \) lead to deficiency of aspartoacylase (ASPA, aminoacylase 2). Accumulation of its substrate N-acetylaspartic acid in brain cause Canavan disease, a fatal progressive LD affecting young children. The ASPA is composed of N-terminal domain (amino acids 1-212) and C-terminal domain (amino acids 213–313). Few amino acid residues are highly conserved in the ASPA and cluster within the area that likely represents an active site of the protein. Mutations involving any of these amino acids delineate the active site and can cause deficiency of the aspartoacylase leading to Canavan disease. The variant p.(Asn54Lys) found in family B was previously reported in affected children from Australia and Pakistan (Howell et al., 2004; Bijarnia et al., 2013) (Table 1). Since the same mutation has been found from two far off places therefore it is highly likely that the variant p.(Asn54Lys) can be considered as mutational hotspot.

Metachromatic leukodystrophy is an inherited disorder caused by the pathogenic mutations in \( ARSA \) leading to deficiency of lysosomal enzyme arylsulfatase A (ASA) which hydrolyzes cerebrosidesulfate to cerebroside and sulfate. Storage of sulfatides in myelin producing cells is associated with progressive demyelination and deterioration of intellectual functions and motor skills such as the ability to walk (Kolodny and Fluharty, 1995). The variant p.(Gly121Arg) in the \( ARSA \), identified in family C, was previously reported by Draghia et al. (1997) in a patient of
French origin (Table 1). Amino acid glycine at position 121 is located in a highly conserved region consisting of a box of six conserved residues containing a histidine that may participate in the active site of arylsulfatase A suggesting that it is essential for normal function of the enzyme (Draghia et al., 1997).

Conclusion

In summary, the three variants identified in the LD causative genes lead to an abnormal development or destruction of the white matter in the brain. The study not only further expanded spectrum of the mutations in the genes responsible for causing LD but also emphasizes importance of the use of WES as a powerful diagnostic tool for conclusive determination of specific type of LD. In addition, the study will also help in carrier testing and prenatal diagnosis in the affected families.

Acknowledgements

We are highly indebted to the family members for their invaluable cooperation and participation in this study. This work was funded by the Higher Education Commission (HEC), Islamabad, Pakistan. F.B. was supported by International Research Support Initiative Program (IRSIP) from HEC, Islamabad, Pakistan.

Conflict of Interest:
Declared None

References


Legends to Figure

**Figure 1:** Pedigrees of families showing X-linked (A) and autosomal recessive (B, C) inheritance pattern of the disease (a-c). Partial nucleotide sequence of the *ABCD1* generated through Sanger sequencing showing variation (c.2135G>A) in affected and carrier individuals (d). Comparison of partial amino acid sequence of ABCD1 protein with other orthologous showed conservation of arginine (R) at position 712 (highlighted in red) among different species (e). Sequencing chromatogram of the *ASPA* obtained from whole exome data showing variant (c.162C>A) in affected and carrier individuals of family B (f). Partial nucleotide sequence of the *ARSA* showing pathogenic mutation (c.361G>C) in affected and carrier individuals in family C (g).

**Figure 2:** Annotated figure containing MRIs of the brain as well as concisely written clinical details of each family.

The upper panel shows MRI findings including abnormal signals in the periventricular deep white matter predominantly along bilateral atria and occipital horns appearing hyperintense on T2W1 FLAIR is to hypointense on T1W1 not showing diffusion restriction of affected individual in family A.

Middle panel shows Coronal Flair MRI sequence showing periventricular and subcortical white matter hyperintensities, Axial T2 MRI sequence showing periventricular and subcortical white matter hyperintensities extending to basal ganglia and brainstem, Sagittal T1 MRI sequence showing periventricular and subcortical white matter signal abnormalities, and DWI B1000 sequence showing white matter hyperintense signal, with low signal on ADC suggesting restricted diffusion in affected individual of family B.

Lower panel shows MRI findings of family C including Axial FLAIR MRI sequence showing confluent symmetrical hyperintensities on the sub-cortical and deep white matter, and Axial T2 MRI sequence showing periventricular and subcortical white matter signal change.
Table 1: Comparison of clinical details of published cases with same mutations and present cases.

<table>
<thead>
<tr>
<th>Clinical details</th>
<th>ASPA; [c.162C&gt;A; p.(Asn54Lys)]</th>
<th>ARSA; [c.361G&gt;C; p.(Gly121Arg)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td>Bijarnia et al., 2013</td>
<td>Howell et al., 2004</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Ethnic origin</td>
<td>Pakistani</td>
<td>Pakistani</td>
</tr>
<tr>
<td>Consanguinity</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Age of presentation</td>
<td>2 year</td>
<td>2 year</td>
</tr>
<tr>
<td>Family history</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Neurological status at presentation</td>
<td>Macrocephaly with delayed milestone</td>
<td>Spasticity, enlarged head size, history of floppiness in early infancy</td>
</tr>
<tr>
<td>MRI findings</td>
<td>Leukodystrophy (Extensive signal abnormality in bilateral)</td>
<td>Leukodystrophy (diffuse leukodystrophy with almost cystic)</td>
</tr>
<tr>
<td>Periventricular and subcortical white matter, bilateral basal ganglia and thalami extending into brainstem and cerebellar hemispheres)</td>
<td>Degeneration seen in frontal and temporal lobes. Generalized cortical atrophy with ventricular dilatation)</td>
<td>The central and periventricular white matter of both cerebral hemispheres)</td>
</tr>
</tbody>
</table>
jdn_10036_f1.jpg