ORIGINAL RESEARCH

Genotype–phenotype associations in Alström syndrome: a systematic review and meta-analysis

Brais Bea-Mascato 1,2, Diana Valverde 1,2

ABSTRACT

Background Alström syndrome (ALMS; #203800) is an ultrarare monogenic recessive disease. This syndrome is associated with variants in the ALMS1 gene, which encodes a centrosome-associated protein involved in the regulation of several ciliary and extraciliary processes, such as centrosome cohesion, apoptosis, cell cycle control and receptor trafficking. The type of variant associated with ALMS is mostly complete loss-of-function variants (97%) and they are mainly located in exons 8, 10 and 16 of the gene. Other studies in the literature have tried to establish a genotype–phenotype correlation in this syndrome with limited success. The difficulty in recruiting a large cohort in rare diseases is the main barrier to conducting this type of study.

Methods In this study we collected all cases of ALMS published to date. We created a database of patients who had a genetic diagnosis and an individualised clinical history. Lastly, we attempted to establish a genotype–phenotype correlation using the truncation site of the patient’s longest allele as a grouping criteria.

Results We collected a total of 357 patients, of whom 227 had complete clinical information, complete genetic diagnosis and meta-information on sex and age. We have seen that there are five variants with high frequency, with p.(Arg2722Ter) being the most common variant, with 28 alleles. No gender differences in disease progression were detected. Finally, truncating variants in exon 10 seem to be correlated with a higher prevalence of liver disorders in patients with ALMS.

Conclusion Pathogenic variants in exon 10 of the ALMS1 gene were associated with a higher prevalence of liver disease. However, the location of the variant in the ALMS1 gene does not have a major impact on the phenotype developed by the patient.

INTRODUCTION

Alström syndrome (ALMS; #203800) is an ultrarare monogenic disease caused by variants in the ALMS1 gene. It is an autosomal recessive disorder with an estimated incidence of 1–9 cases per 1 000 000 inhabitants. Currently there are approximately 1000 cases described worldwide (Orphanet; 3 May 2022).

Most of the variants associated with this disease generate a stop codon, either at the variant site (nonsense variant) or as a result of a frameshift alteration, leading to complete loss-of-function (cLOF) variants. 3 Currently, there are 388 cLOF variants reported in ClinVar and 253 in gnomAD. These pathogenic variants have a uniform distribution along the gene and are mainly located in the coding regions. Whole exome sequencing is a standard practice for genetic testing of rare diseases, which means that intronic regions are poorly studied, explaining why most pathogenic variants are detected in coding regions. 4 Exons 8, 10 and 16 are considered variant hotspots, but this seems to be due to their large size rather than a specific regulatory correlation. For example, exon 8 (6108 bp) covers 50% of the total gene sequence (12 844 bp). ALMS presents a very heterogeneous phenotype in which symptoms can be aggregated into two main groups. 4 5 The first group includes the presence of retinal dystrophy from early age, several metabolic disorders (obesity and hypertriglyceridaemia and/or type 2 diabetes mellitus (T2DM)), hearing loss, liver and kidney dysfunctions, and cardiac disorders such as dilated cardiomyopathy (DCM). 5 6 The second group of symptoms would include short stature, recurrent pulmonary infections, mental and cognitive impairments, and several endocrine disorders, affecting the thyroid and reproductive systems. 5 6 This second group would also include
other symptoms of uncertain frequency, such as alterations in the fingers, alopecia and spinal abnormalities.3

ALMS is characterised by high intrafamilial and interfamilial phenotypic heterogeneity.1–6 This means that siblings with the same genotype may develop different phenotypes, which complicates the establishment of a genotype–phenotype correlation.7 However, in recent years, great efforts have been made to define the clinical criteria for the management of patients with ALMS.8

Several studies have attempted to establish a genotype–phenotype correlation, with limited success, using cohorts with 12–18 patients.9,10 Studies in larger cohorts (58 patients) have shown that there is a correlation between disease-causing variants in exon 16 and the presence of retinal dystrophy before 1 year of age and the occurrence of urological dysfunction, DCM and T2DM.11 Moreover, this study found a significant correlation between disease-causing variants in exon 8 and absent, mild or delayed kidney disease.

Premature termination codon (PTC) variants are often classified as cLOF and are associated with the activation of nonsense-mediated mRNA decay (NMD), which results in the elimination of the expression of the mutated gene.11,12 However, it has been shown that NMD is not 100% efficient when PTC variants are located in the last exons of genes.13–15 In the case of ALMS, some patients continue to express the ALMS1 protein even when carrying two cLOF variants.14 In that study, the clinical manifestations of 23 patients were related to the expression or non-expression of the ALMS1 protein, and it was observed that those patients with residual ALMS1 protein expression had milder phenotypes.14

Although several databases list the variants described in the ALMS1 gene (gnomAD, ClinVar), only one, LOVD, offers a register with 140 patients; however, clinical data are absent or incomplete in many cases.16–18 There is still no open patient registry that comprehensively collects clinical and genetic information. In this study, we have reviewed 105 publications related to ALMS compiling all available clinical and genetic information. We have collected a total of 357 patients, manually curated, in whom various analyses have been performed to look for a genotype–phenotype correlation. Lastly, we have made this data set available to the scientific community in a single public access database (http://genesandfunctions.uvigo.es/alms-db/).

METHODS

PRISMA guidelines

This meta-analysis was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, as described in Niederlova et al18 for Bardet-Biedl syndrome (BBS). In this case, the main PICO question our study sought to address was: do patients with ALMS have different phenotypes depending on where in the gene the stop variant occurs? Similar studies have been conducted in smaller cohorts grouping patients by major variant hotspots in the ALMS1 gene.2,6,19

Search strategy

The PubMed and Google Scholar databases were searched in January 2022 using the following keyword combination: (“Alström syndrome” OR ALMS) AND (“genotype-phenotype” OR “cohort” OR “case report”). The screening of the search results was carried out by BB-M by initially examining the title and abstract to determine whether the article corresponded to the subject of the analysis. Some of the articles analysed in this study were not initially found by the database search but were detected by recommendations and references of other more relevant works. The search was conducted only in English, covering a period between the origin of the databases and the date of the search (January 2022).

Study selection

All articles selected in the first search were carefully reviewed to meet the criteria for inclusion in the meta-analysis. After a thorough reading of the article, the study was included if the following characteristics were met: the cohort of the article included a patient with a diagnosis of ALMS, and the diagnosis had a genetic and clinical characterisation. Articles whose diagnosis of ALMS was based solely on phenotype were discarded,20–35 as well as those that simply presented or reported the patient’s causal variants without giving a complete36 or individualised37–41 clinical history. From an initial selection of 105 studies, 31 studies were discarded and 74 were selected for information extraction and subsequent analyses.

Data extraction and curation

For extraction of information, three main groups were defined: meta-data, genetic information and clinical information.

Meta-data refer to the identification (ID) of each patient within the data set, the ID of the family to which they belong (patients with the same family ID are siblings) and the reference from where the information of the patient was extracted.

For genetic information, the following were extracted, whenever possible: the original allele reference, the nomenclature of the variant for the cDNA and protein sequence, the exon/intron where the variant is located, and the genotype of the patient (homozygote or compound heterozygote). All reported variants were named following the Human Genome Variation Society (HGVS) guidelines with the transcript NM_015120.4 and validated with the name checker of the Muta layswer software.19 The back translator tool of the Muta layswer software was also used to determine the variant nomenclature at the cDNA level when only the protein annotation of the variant was provided in the article. In these cases, only nonsense variants could be determined, as the algorithm does not work with frameshift variants.

Regarding clinical information, sex, age and ethnicity were extracted (whenever possible) and the diagnostic criteria described by Marshall et al were used to try to homogenise them: history of nystagmus in infancy/childhood, legal blindness, and cone and rod dystrophy by electroretinogram (ERG), obesity and/or insulin resistance and/or T2DM, history of DCM/congestive heart failure (CHF), hearing loss, hepatic dysfunction, renal failure, pulmonary dysfunction, short stature, hypogonadism in males, irregular menses and/or hyperandrogenism in females, thyroid disorders, intellectual disability, abnormal appearance of a finger, intestinal problems, sclerosis/flat wide feet, epilepsy, and alopecia (table 1). Initially, the full information reported in the article was noted for each individual symptom and the age of onset if this was reported. This was then converted into a binary matrix to simplify and homogenise the downstream analysis.

Statistical frequency analysis

To determine the existence of a genotype–phenotype correlation in our patient registry, we selected a subcohort of 227 patients who met the following criteria: (1) complete genetic information (detection of two mutated alleles with a complete notation at the cDNA and protein level) and (2) complete clinical information (sex, age and presence or absence of the five most prevalent clinical manifestations reported in these patients: vision...
impairments, metabolic anomalies, hearing anomalies, heart anomalies and liver anomalies). The heterogeneity of the clinical notation associated with these five categories was summarised in a binary matrix (1=presence, 0=absence), which was used for frequency calculations to determine the prevalence of each symptom in the cohort and in the different subgroups.

Following the methodology described by Niederlova et al., a syndromic score between 0 and 1 was calculated for each patient using the following formula:

\[
\text{Syndromic score} = \sum_i x_i
\]

This syndromic score gives an estimate of how many of the five most prevalent clinical manifestations described are present in each patient, where \(x_i\) is the presence (1) or absence (0) of each of the five most prevalent symptoms in each patient of the ALMS cohort.

In this study the lack of race was not considered an exclusion criteria. Due to the high heterogeneity of this characteristic and the small cohort size, correlations with enough statistical power could not be established.

Patient ages were initially aggregated into 10-year age intervals: 0–9, 10–19, 20–29, 30–39, 40–49 and 50–59. After studying the evolution of the syndromic score in the different age intervals, it was found that the number of patients above 39 years was very small. This is consistent with the life expectancy of patients with ALMS, which is usually not more than 50 years for this reason, a regrouping of the last three intervals in patients over 30 years of age was made for the subgroup analysis.

Following the methodology used by Marshall et al., patients were genotypically grouped according to the different variant hotspots: exon 8 (E8), exon 10 (E10) and exon 16 (E16). For compound heterozygous patients, this aggregation was performed based on the patient's longest allele. Since ALMS is an autosomal recessive disease, a single functional copy of the gene is sufficient to prevent the onset of disease symptoms. Diseases with a similar mechanism, such as recessive titinopathies, have already been described in the literature.

Approximately 90% of the variants were found in the hotspots (E8, E10 and E16); however, to include the remaining percentage in the study, three intervals were defined: longest allele truncated before exon 9 (group 1, G1), longest allele truncated between exons 9 and 14 (group 2, G2), and longest allele truncated after exon 14 (group 3, G3).

In cases where only two groups were compared, Wilcoxon test was used. In trials where comparisons were made between several groups (more than two), an overall p value calculation was performed using non-parametric Kruskal-Wallis test, followed by a comparison of means by peer groups using the Wilcoxon test with a false discovery rate (FDR) correction. Results were considered significant when corrected p values were less than 0.05.

For the analysis of the prevalence of symptoms in the different genetic groups, contingency tables were created showing the number of positive/negative cases within each patient group. Differences in the prevalence of phenotypes in the different patient groups were assessed using pairwise Fisher’s exact test. The statistical significance of the differences between individual groups was determined with the FDR correction for multiple comparisons.

RESULTS

Cohort description

A total of 357 patients were collected from 74 scientific publications that passed the initial screening. Only patients with information on sex, age and complete genetic characterisation (description of two non-functional alleles; see the Methods section) were used for further analysis. Only variants in the coding regions were considered in the study. This reduced the

---

**Table 1** Abbreviations used in the analysis, full name of the phenotypic categories, phenotypes aggregated in each category, prevalence of each phenotypic category in the study cohort (n=227) and representative HPO terms for each category

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Complete name</th>
<th>Phenotype</th>
<th>Prevalence (%)</th>
<th>HPO terms added in each category</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI</td>
<td>Vision impairments</td>
<td>History of nystagmus in infancy/childhood, legal blindness, and cone and rod dystrophy by electroretinogram (ERG)</td>
<td>97.80</td>
<td>HP:0000556 (retinal dystrophy), HP:0000639 (nystagmus), HP:0000618 (blindness)</td>
</tr>
<tr>
<td>MT</td>
<td>Metabolic anomalies</td>
<td>Obesity and/or insulin resistance and/or T2DM</td>
<td>85.02</td>
<td>HP:0001513 (obesity), HP:0005978 (T2DM)</td>
</tr>
<tr>
<td>HL</td>
<td>Hearing anomalies</td>
<td>Hearing loss</td>
<td>59.91</td>
<td>HP:0000365 (hearing loss)</td>
</tr>
<tr>
<td>HRT</td>
<td>Heart anomalies</td>
<td>History of DCM/CHF</td>
<td>49.34</td>
<td>HP:0001644 (DCM), HP:0001635 (CHF)</td>
</tr>
<tr>
<td>LIV</td>
<td>Liver anomalies</td>
<td>Hepatic dysfunction</td>
<td>36.56</td>
<td>HP:0001410 (decreased liver function)</td>
</tr>
<tr>
<td>REN</td>
<td>Renal anomalies</td>
<td>Renal failure</td>
<td>29.52</td>
<td>HP:0000077 (renal anomaly)</td>
</tr>
<tr>
<td>MEND</td>
<td>Mental anomalies</td>
<td>Mental disability</td>
<td>24.67</td>
<td>HP:0001249 (intellectual disability)</td>
</tr>
<tr>
<td>PUL</td>
<td>Pulmonary anomalies</td>
<td>Pulmonary dysfunction</td>
<td>19.38</td>
<td>HP:0002795 (impaired pulmonary function)</td>
</tr>
<tr>
<td>REP</td>
<td>Reproductive system anomalies</td>
<td>Males: hypogonadism; females: irregular menses and/or hyperandrogenism</td>
<td>17.62</td>
<td>HP:0000026 (male hypogonadism), HP:000858 (irregular menstruation)</td>
</tr>
<tr>
<td>TYD</td>
<td>Thyroid metabolism anomalies</td>
<td>Thyroid disorders</td>
<td>14.54</td>
<td>HP:0000820 (abnormality of the thyroid gland)</td>
</tr>
<tr>
<td>SCO</td>
<td>Spine/feet anomalies</td>
<td>Scoliosis/flat wide feet</td>
<td>13.22</td>
<td>HP:0002650 (scoliosis), HP:0001763 (pes planus)</td>
</tr>
<tr>
<td>SHS</td>
<td>Stature anomalies</td>
<td>Short stature</td>
<td>9.25</td>
<td>HP:0004322 (short stature)</td>
</tr>
<tr>
<td>ALO</td>
<td>Alopecia</td>
<td>Alopecia</td>
<td>6.17</td>
<td>HP:0001596 (alopecia)</td>
</tr>
<tr>
<td>NER</td>
<td>Nervous system anomalies</td>
<td>Epilepsy</td>
<td>3.08</td>
<td>HP:0001250 (seizure)</td>
</tr>
<tr>
<td>ABFING</td>
<td>Finger anomalies</td>
<td>Abnormal appearance of a finger</td>
<td>1.32</td>
<td>HP:0001155 (abnormality of the hand)</td>
</tr>
<tr>
<td>INT</td>
<td>Intestinal anomalies</td>
<td>Intestinal problems</td>
<td>1.32</td>
<td>HP:0011024 (abnormality of the gastrointestinal tract)</td>
</tr>
</tbody>
</table>

CHF: congestive heart failure; DCM, dilated cardiomyopathy; HPO, Human Phenotype Ontology; T2DM, type 2 diabetes mellitus.
Genotype-phenotype correlations

Figure 1 Cohort description of 227 patients with Alström syndrome (ALMS). (A) Counting of the different alleles in the cohort. Alleles with more than two copies in the cohort are represented. (B) Number of alleles per exon in the study cohort. (C) Patients grouped by their allele of the ALMS1 gene with the furthest truncation variant. (D) Age composition in the subgroups with the longest allele of the ALMS1 gene truncated before exon 9 (group 1, G1), between exon 9 and exon 14 (group 2, G2), and after exon 14 (group 3, G3).

initial cohort to 227 patients, 128 (56.38%) males and 99 females (43.61%). This study cohort contained 176 variants, where 168 were cLOF and 8 missense variants (online supplemental figure S1). Most of them were private variants, but five pathogenic variants were shown to have a high frequency: p.(Arg2722Ter) (28 alleles), p.(Gln3495Ter) (22 alleles), p.(Thr3592LysfsTer6) (14 alleles), p.(Glu3773TrpfsTer18) (11 alleles) and p.(Pro3911GlnfsTer16) (10 alleles) (figure 1A). The number of alleles described for these variants shows certain inconsistencies with the gnomAD database, and in some cases the variant does not even appear in this database (table 2).

Patients were initially grouped into age ranges by decade (from 0 to 59 years), but due to the low number of patients in the 40–49 and 50–59 age ranges they were added to the 30–39 age group and reclassified as over 30 years for further analysis.

Out of a total of 454 alleles, 207 (45.59%) were in E8, 107 (23.57%) in E16 and 81 (17.84%) in E10 (figure 1B). Thus, these exons were the main variant hotspots in our cohort, consistent with the literature. In other cohorts, the pathogenic variants in these hotspots comprise 21%–57% in E8, 19%–40% in E16 and 12%–32% in E10 of the total.1 5 14 19 30 66 89 104

Due to the recessive nature of ALMS, we decided to define the characteristic allele of each patient according to the pathogenic variant furthest from the transcription start site of the gene (figure 1C). Many of the patients carrying variants in E8 were compound heterozygotes with variants in E16. After this regrouping, we detected that 80 (35.34%) patients had the largest allele truncated in E8, 42 (18.50%) in E10 and 75 (33.03%) in E16.

In accordance with other similar studies,5 we used the variant hotspots (exons 8, 10 and 16) to subdivide the patients into genetic groups 1, 2 and 3, respectively. Due to the low frequency of pathogenic variants in other exons of the ALMS1 gene, we decided to include patients with truncated alleles in the adjacent exons to these variant hotspots. G1 includes all patients with the longest truncated allele before exon 9. G2 includes patients with the longest truncated allele between exon 9 and exon 13. Lastly, G3 included patients with the longest truncated allele from exon 14 to exon 20. As a result, G1, G2 and G3 contained 85, 45 and 97 patients, respectively (figure 1D).

 Patients with the longest allele truncated around E10 have a higher syndromic score than the other subgroups

For the phenotypic manifestations collection strategy, 16 syndromic groups (see the Methods section) were initially defined. In the genotype-phenotype correlation analysis, a minimum prevalence threshold of 15% (n=33) was established in the cohort (online supplemental figure S2). This reduced the initial phenotypic manifestations to nine groups: vision impairments (VI; 97.80%), metabolic anomalies (MT; 85.02%), hearing anomalies (HL; 59.91%), heart anomalies (HR; 49.34%), liver anomalies (LI; 36.56%), renal anomalies (29.52%), mental anomalies (24.67%), pulmonary anomalies (19.38%) and reproductive system anomalies (17.62%).

Following the methodology developed by Niederlova et al.,18 the five main syndromic groups (VI, MT, HL, HR and LI; see the Methods section) were used to create a discrete syndromic score ranging from 0 to 1 (figure 2A). The mean of this syndromic score in the total cohort is around 0.7, with the most common values being 0.6 (three of the five symptoms) and 0.8 (four of the five symptoms) (figure 2B). No significant differences in the syndromic score between sexes were observed (figure 2C and online supplemental figure S3). However, the syndromic score increased significantly in the different age ranges of the patients (p=3.1 e-10; figure 2D and online supplemental figure S4). This is consistent with the progressive worsening that these patients undergo throughout their lives.14 Finally, we observed how the syndromic score was distributed in the different groups defined according to their longest allele. Patients whose longest allele was truncated around E10 (G2) had a higher syndromic score compared with the G1 (p=0.017) and G3 (p=0.016) groups (figure 2E).

 Patients with truncation variants around E10 have a higher prevalence of liver disease

To determine whether the differences between the G1, G2 and G3 groups were due to an unequal composition between the different age ranges in each group, the syndromic score was studied by combining both parameters (age and genetic group) (figure 3A). In the first decade of life, no significant differences in the syndromic score were observed between the groups. Between the age of 10 and 19 years, a higher syndromic score was observed in the G2 group compared with the G3 group (FDR=0.061). From 20 to 29 years of age, a higher syndromic score was observed in the E10 group compared with the G1 and G3 groups (FDR=0.009, in both cases) (figure 3A). Lastly, in patients over 30 years of age, no significant differences were observed between the genetic groups (figure 3A).

After concluding that the differences observed between the groups were not due to an unequal age composition between

<table>
<thead>
<tr>
<th>Pathogenic variant</th>
<th>ALMSDB</th>
<th>gnomAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.(Arg2722Ter)/p.(Arg2722Ter)</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>p.(Gln3495Ter)/p.(Gln3493Ter)</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>p.(Thr3592LysfsTer6)/p.(Thr3592LysfsTer6)</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>p.(Glu3773TrpfsTer18)/p.(Glu3771TrpfsTer18)</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>p.(Pro3911GlnfsTer16)</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

The differences in nomenclature are due to the Mutalyzer software using RefSeq (NM_015120.4) as a reference, while gnomAD uses Ensembl (ENST00000264448.6). ALMSDB: Alström syndrome database.

Genotype-phenotype correlations

them, the prevalence of the nine syndromic manifestations in each of these groups was studied (Figure 3B). Patients within the G2 group were found to have a higher prevalence of liver disorders compared with patients in the G3 group (FDR=0.00792). For the remaining eight syndromic groups, no significant differences were found between the genetic groups (Figure 3B).

The G2 group consisted of 23 heterozygous patients and 22 homozygous patients. The prevalence of the presence/absence of liver problems was, respectively, 9/14 in the heterozygous group and 15/7 in the homozygous group.

DISCUSSION

Meta-analyses are a useful tool to address one of the main limitations when attempting to establish a genotype-phenotype correlation in a rare disease: the inability to recruit a large cohort to draw robust and statistically significant conclusions. Although meta-analyses mainly focus on the aggregation of several studies with well-defined and well-studied cohorts, this is often not possible in rare diseases. Alternatively, one can try to aggregate information obtained from case reports in the literature. However, the lack of uniform criteria in the way the authors describe their patients is one of the main biases of this approach. In this paper we adapt and apply the methodology described by Niederlova et al from a polygenic disease such as BBS to a monogenic disease such as ALMS.

Even though ALMS is a monogenic disease,123 to date several studies have already discussed the possible existence of several tissue-specific isoforms for the ALMS1 gene, which may have different regulatory roles that explain the high symptomatic heterogeneity.120 123–125 Although this is an interesting approach, the isolated nature of the data published in the case reports does not allow for validation. To validate this hypothesis, it is necessary to recruit a large cohort where the level and functionality of the protein product in different tissues should be investigated to correlate the clinical phenotype of the patient with the different regulatory mechanisms of ALMS1. Different cohorts have been described in the literature, the largest being that of Marshall et al,5 followed by Ozantürk et al,86 the National Institutes of Health (NIH) clinical centre cohort,50 104 Chen et al14 and Rethanavelu et al,89 with 58, 44, 38, 23 and 21 patients with ALMS, respectively. Significant genotype–phenotype correlations were only detected in Marshall et al,5 highlighting the importance of the sample size used in this type of study.

In our analysis we found that there were discrepancies between the absolute frequency in the most reported alleles concerning databases such as gnomAD (Table 2). Most of these variants were reported in three or more studies.1, 5 14 50 86 104 However, the pathogenic variant p.(Pro3911GlnfsTer16) described by Khan et al in a Saudi Arabian population is not reported in either

Figure 2  Phenotypic impact of different subgroups of patients with ALMS. (A) Prevalence of the five main syndromic groups in the cohort: vision impairments (VI), metabolic anomalies (MT), hearing anomalies (HL), heart anomalies (HRT) and liver anomalies (LIV). (B) Distribution of syndromic scores calculated from the presence or absence of the five most relevant syndromic groups. (C) Gender comparison of syndromic scores. (D) Comparison of the syndromic scores between age groups. (E) Comparison of the syndromic scores between subgroups created from the longest allele of the ALMS1 gene. ALMS, Alström syndrome; F, female; G1, group 1; G2, group 2; G3, group 3; M, male.

Figure 3  Correlation study between the prevalence of symptoms and the truncation site of the ALMS1 gene. (A) Evolution of the syndromic score among the subgroups by the longest allele of the ALMS1 gene in the different age groups. (B) Prevalence of the different symptom clusters in the subgroups by the longest allele of the ALMS1 gene. G1, group 1; G2, group 2; G3, group 3; HL, hearing anomalies; HRT, heart anomalies; LIV, liver anomalies; MEND, mental anomalies; MT, metabolic anomalies; PUL, pulmonary anomalies; REN, renal anomalies; REP, reproductive system anomalies; VI, vision impairments.
gnomAD or ClinVar. This highlights the importance of requiring researchers to notify public repositories of variants they have detected. These variants must be manually reviewed and curated before publication.

The causal variants of ALMS are mainly of the cLOF type (generating a PTC). Such variants activate NMD, preventing the translation of the sequence from mRNA to protein. However, certain situations have been described, such as PTC variants in the last exons of a gene, in which NMD can be prevented or incompletely produced. Chen et al have detected ALMS1 gene expression in patients carrying two cLOF variants leading to PTC. Their study has also correlated residual ALMS1 gene expression with the development of milder phenotypes. Taking this into account, it could be that cLOF variants in the terminal exons of the ALMS1 gene do not activate NMD, resulting in a partially functional protein or a non-functional misfolded protein. Under this hypothesis, we grouped the patients in our study according to the longest transcript they could have. Three different groups related to the main variant hotspots of the ALMS1 gene, exons 8, 10 and 16, were defined. A similar approach has already been used on a cohort of 58 patients with ALMS by Marshall et al. Subsequently, a syndromic score was created by adding the five clinical manifestations more prevalent in ALMS. This methodology was adapted from Niederlova et al.

The results showed that the syndromic score increases with age (figure 2D), which was consistent with the gradual worsening that these patients suffer throughout their lives. This helped us to validate the effectiveness of the syndromic score in measuring the severity of the patient’s symptoms. In addition, we have also determined the Pearson linear correlation between the syndromic score and the age of the patient (online supplemental figure S4). Although the correlation is significant, it is not very strong (R = 0.41). The correlation between these two variables appears to be exponential rather than linear. Another consideration is that we are comparing a continuous versus a discrete variable.

On the other hand, another possible explanation for the weak correlation (R = 0.41) between syndromic score and age, could be that the progression of the disease from childhood (>9 years) to adulthood/adolescence is independent of age after the first decade of life. In fact, three of the four age groups (10–19, 20–29, >30) share a similar median value. It is possible that many symptoms of the syndrome worsen by the genetic disease per se more than by ageing, as it occurs in other diseases such as diabetes in the general population and in patients with obesity. For this reason, ALMS could be regarded as a disease model of accelerating ageing.

When comparing the syndromic scores of the different genetic groups by age ranges, it was observed that patients with truncation variants in E10 evolve more unfavourably than the rest (figure 3A). The fact that these differences are not appreciable after the age of 30 years could be due to a higher mortality in this group in the second and third decades of life. This supposed higher mortality after 30 years could be correlated with the greater prevalence of liver problems that patients with truncation variants at E10 appear to have (figure 3B). This correlation was not detected in previous genotype–phenotype analyses. Marshall et al described disease-causing variants in E16 as leading to urological dysfunction, DCM/CHF and T2DM. In our study, we did not look for correlations in urological dysfunction due to the lack of data reported in the literature, leading to a low prevalence of these types of symptoms. In the case of DCM/CHF and T2DM, the prevalence was the same among the established genetic groups. Marshall et al also found a significant association between alterations in E8 and absent, mild or delayed kidney disease, findings that were not observed in our analysis. Mortality in childhood, especially due to cardiomyopathy, increases the difficulty of establishing genotype–phenotype correlations and could explain the correlation found by Marshall et al which did not appear in our study. On the other hand, metabolic symptoms such as T2DM or obesity are influenced by lifestyle and environment in ALMS. Although genetics plays a role in predisposition to these metabolic symptoms, it is possible to partially manage them with appropriate lifestyle habits, which makes it difficult to establish a genotype–phenotype relationship in these cases. Recently, it has been described that metabolic disorders in ALMS seem to act as a comorbidity of liver diseases, starting with hepatic steatosis and progressing to hepatic fibrosis. Interesting points to discuss in this study are the lack of patients carrying variants in or adjacent to E10 and the small size of the cohort (n=18). In our cohort, the prevalence of metabolic disorders appears to be similar between the different groups, but as mentioned above the prevalence of liver disorders is not. Thus, although metabolic disorders are potential comorbidities for the development of liver problems, patients’ genotype seems to weigh more heavily.

A possible explanation for this event could be related to the residual expression of the ALMS1 protein. Interestingly, the phenotype in patients with variants in E8 appears to be like those with variants in E16, so the differential activation of the NMD seems to be related to other upregulatory mechanisms beyond the length of the allele. Some authors have described that patients with pathogenic variants in early exons, such as exon 5, can also develop mild phenotypes. However, we did not detect this in our cohort, where three of the four patients carrying homoygous mutations in exon 5 had a syndromic score between 0.8 and 1. Tissue-dependent alternative splicing alterations (intronic or splice site variants) could be the main upstream mechanism for these events. Pathogenic variants in E10 could prevent NMD activation and give rise to a misfolded protein that causes greater hepatotoxicity. In the case of variants in E16 or adjacent, perhaps the NMD is not activated either, but the generated protein, despite not being functional, could have a seminormal conformation that prevents the formation of aggregates. This hypothesis could be easily tested if these protein isoforms can be simulated by homology, but unfortunately the protein structure of the ALMS1 gene remains unknown and cannot be simulated using artificial intelligence models such as AlphaFold. This limitation could currently only be overcome by doing tissue-dependent expression studies in patients with ALMS. In any case, given the findings of this study, it would be recommended that patients with causal variants between exons 9 and 14 have a more exhaustive monitoring of liver function, compared with patients carrying causal variants in other exons.

Lastly, the ALMS1 sequence has already described the presence of a long non-coding RNA (lncRNA), ALMS1-IT1, with a role in regulating proliferation in various types of cancers and neuroinflammation in rats, and a pseudogene, ALMS1P1, whose function is still unknown. However, no cases of ALMS with these symptoms have been described to date. Due to the young age of the ALMS1 gene, the presence of more regulatory elements would not be uncommon and could explain why the localisation of the cLOF variant can lead to different tissue-specific phenotypes. Intronic or splicing acceptor variants could affect the regulatory elements (lncRNA or miRNA) or the exonic composition of the ALMS1 gene by altering gene regulatory
networks, which are tissue-dependent and cell type-dependent. These events could be of help in understanding the regulation of ALMS1 in the different tissues involved in the disease and the great heterogeneity observed in the clinical symptoms.

Some of the main limitations of this study are the lack of homogeneous criteria when collecting patients’ clinical data and the biased study of certain genes, preventing the assessment of common mutational load and polygenic epistasis events between different causal genes. Furthermore, the effect of the described causal variants on the stability, expression and functionality of the ALMS1 protein has not been evaluated. Finally, the ethnicity of the patients was not considered as a variable to establish genotype–phenotype correlations.

CONCLUSION

In conclusion, five highly prevalent pathogenic variants were detected in our cohort, but not all of them are present in public databases. There are no gender differences in the prevalence of ALMS symptoms. The syndromic score used increases with age. Patients whose longest allele of the ALMS1 gene is truncated around E10 display higher prevalence of liver dysfunction and a worse disease progression. No differences in the prevalence of DCM/HCM and T2DM are observed among patients grouped by their longest truncated allele.

Contributors

BB-M and DV designed the study. BB-M selected, reviewed, collected and curated data from the scientific articles; designed and executed the analysis pipeline; and created the publicly accessible online database. BB-M and DV drafted the article. Both authors reviewed the manuscript, provided approval for publication and are the guarantors of the article.

Funding

This work was funded by Instituto de Salud Carlos III de Madrid FIS Project PI15/00049 and PI19/00332, Xunta de Galicia (Centro de Investigación de Galicia CINBIO 2019-2022) Ref ED431G-2019/06, and Consolidacion e estructuración de unidades de investigación competitivas e outras accións de fomento (ED431C-2018/54). BB-M (FPU17/01567) was supported by a graduate studentship award (FPU predoctoral fellowship) from the Spanish Ministry of Education, Culture and Sport.

Competing interests

None declared.

Patient consent for publication

Not required.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

All data relevant to the study are included in the article or uploaded as supplementary information. The complete data can also be viewed at http://genesandfunctions.uvigo.es/alms/db.

Supplemental material

This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

ORCID ID

Brais Beà-Mascato http://orcid.org/0000-0003-1588-2897

REFERENCES

by two independent yet linked mutations in ALMS1 and DYSF. *Hum Mutat* 2015;36:836–41.
Genotype-phenotype correlations


