
7

Exome sequencing identifies a branch point variant in Aarskog-Scott syndrome

Emmelien Aten, * Y. Sun, * R. Almomani, G.W.E Santen, T. Messemaker, S.M. Maas, M.H Breuning, J.T. den Dunnen

**These authors contributed equally to this work*

Human Mutation. 2012. In Press.

Abstract

Aarskog-Scott syndrome (ASS) is a rare disorder with characteristic facial, skeletal, and genital abnormalities. Mutations in the *FGD1* gene (Xp11.21) are responsible for ASS. However, mutation detection rates are low. Here, we report a family with ASS where conventional Sanger sequencing failed to detect a pathogenic change in *FGD1*. To identify the causative gene we performed whole-exome sequencing in two patients. An initial analysis did not reveal a likely candidate gene. After relaxing our filtering criteria, accepting larger intronic segments, we unexpectedly identified a branch point (BP) variant in *FGD1*. Analysis of patient-derived RNA showed complete skipping of exon 13, leading to premature translation termination. The BP variant detected is one of very few reported so far proven to affect splicing. Our results show that, besides digging deeper to reveal non-obvious variants, isolation and analysis of RNA provides a valuable but under-appreciated tool to resolve cases with unknown genetic defects.

Key words
FGD1 protein, Exome sequencing, Aarskog syndrome, RNA splice sites, Branch point mutations

Introduction

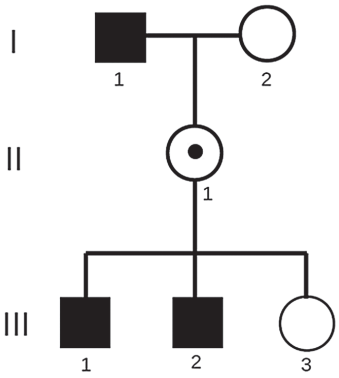
The large scale sequencing of thousands of exomes and genomes from disease cohorts using next generation sequencing (NGS) is expected to uncover causal gene variants in many thus far unresolved cases. Currently, high-throughput clinical sequencing of the entire genome is still too expensive but selective enrichment of genomic regions of interest, such as the exome, provides a cost-effective and scalable approach. The normal strategy is to compare variants with clear functional consequences (nonsense, frame shift, splice site, conserved missense) from several patients/families with identical phenotypes and searching for a shared defective gene or from parent-child trios searching for *de novo* variants in the child. The approach has revealed the cause of a range of rare Mendelian disorders. However, exome sequencing comes with several disadvantages. It is labor intensive (costly), it misses variants outside the targeted (protein coding exons) regions and it generates sequence data with a large variability in coverage across the regions sequenced, missing 5-15% of the targeted area. Furthermore, to cope with the sheer number of variants encountered, choices have to be made regarding the variants that deserve follow-up work (truncating/missense variants, intronic variants, splice site variants, etc). When the pathogenic consequence of a variant is unclear, the effect has to be studied in detail at other levels (e.g. RNA, protein, in vitro functional assay), in other tissues or in additional patients/families making the final proof of causality a tremendous effort. In many cases this may not be possible at all, e.g. since the function of the gene identified is not known or because no other patients are available.
Besides the first nucleotides, intronic regions are usually not studied. This has two main reasons: first introns contain many variants yielding too many candidates to follow-up. Second, RNA is often not available making it impossible to prove predicted effects on RNA splicing.

Aarskog-Scott syndrome (ASS or faciogenital dysplasia, OMIM #305400) is a clinically and genetically heterogeneous disorder characterized by proportionate short stature, short limbs (usually rhizomelic), broad hands and feet, genital hypoplasia (shawl scrotum) and facial dysmorphisms. Mental retardation has been described but is not a common finding (Orrico et al., 2005; Kaname et al., 2006; Shalev et al., 2006). Both autosomal dominant and recessive inheritance patterns have been demonstrated (Teebi et al., 1988; van de Vooren et al., 1983) but currently only one gene has been directly linked to ASS, *FGD1* at Xp11.21. Both substitutions and deletions have been described in *FGD1* in X-linked cases (Shalev et al., 2006; Bedoyan et al., 2009).

The *FGD1* gene (faciogenital dysplasia 1) is ubiquitously expressed and composed of 18 exons encoding a 961 amino acid guanine nucleotide exchange factor (GEF), binding

specifically to the Rho protein Cdc42. The Cdc42 pathway is involved in regulating cell growth and differentiation and plays a role in skeletal development (Pasteris et al., 1997; Hou et al., 2003). The protein consists of several important segments such as a proline-rich SH3-binding motif, a RhoGEF homology domain (DH), two pleckstrin-homology domains (PH) and a FYVE-zinc finger domain (ZF). In ASS, variants have been reported throughout the gene, both missense and protein truncating but there is no clear correlation between the nature and location of variants and disease severity (Orrico et al., 2007). Most variants are private, with the exception of three (c.529dupC, c.1966C>T and several affecting amino acid Arg443; www.LOVD.nl/FGD1) (Orrico et al., 2010; Orrico et al., 2004). *FGD1* variants are found in only ~ 20% of Aarskog-Scott syndrome patients (Orrico et al., 2004). This can be attributed to inaccurate clinical diagnosis, the existence of overlapping syndromes (notably Noonan, LEOPARD, Teebi hypertelorism and Robinow syndrome) or genetic heterogeneity (Bottani et al., 2007). In diagnostics, *FGD1* analysis of the entire coding region is performed using DHPLC and/or Sanger sequencing and deletion/duplication analysis using MLPA.
Two siblings (III-1 and III-2) from a family of Dutch origin were referred with a clinical diagnosis of ASS. The phenotype is summarized in Figure 1 and Table 1 and is illustrated in Figure S1. Conventional diagnostic Sanger sequencing revealed no changes in the *FGD1* gene (Greenwood Genetic Center, USA).

Figure 1: Pedigree of the Dutch Aarskog family.



SNP-array analysis of II-2, III-1 and III-2 excluded the presence of large causative deletions/duplications in the genome, including the region containing *FGD1*. To resolve the genetic basis of ASS in this family, we performed exome sequencing of the two siblings (III-1 and III-2). Illumina GAIIX generated 51 nt, paired end reads. A coverage plot can be found in

Table 1: Phenotype overview of the Dutch ASS family.
na = not applicable

	III-1	III-2	II-1	I-1
Craniofacial				
Widow's peak	-	+	+	+
Cow's lick	+	+	+	+
Hypertelorisme	+	+	+	+
Ptosis	-	-	-	-
Downward slant palpebral fissures	-	-	-	-
Short nose	+	+	+	-
Wide philtrum	+	+	-	+
Cleft lip and palate	+	-	-	-
Underdeveloped maxillae	-	-	-	-
Crease below lower lip	-	-	-	-
Abnormal auricles	+	+	-	-
Skeletal				
Short stature	+	+	+	+
Short broad hands	-	-	-	-
Clinodactyly fifth finger	-	-	-	-
Midl interdigital webbing	-	-	-	-
Joint laxity	+	+	+	-
Contractures of fingers	+	-	-	+
Broad feet	+	-	-	-
Genital				
Shawl scrotum	+	+	na	-
cryptorchidism	+	+	na	-

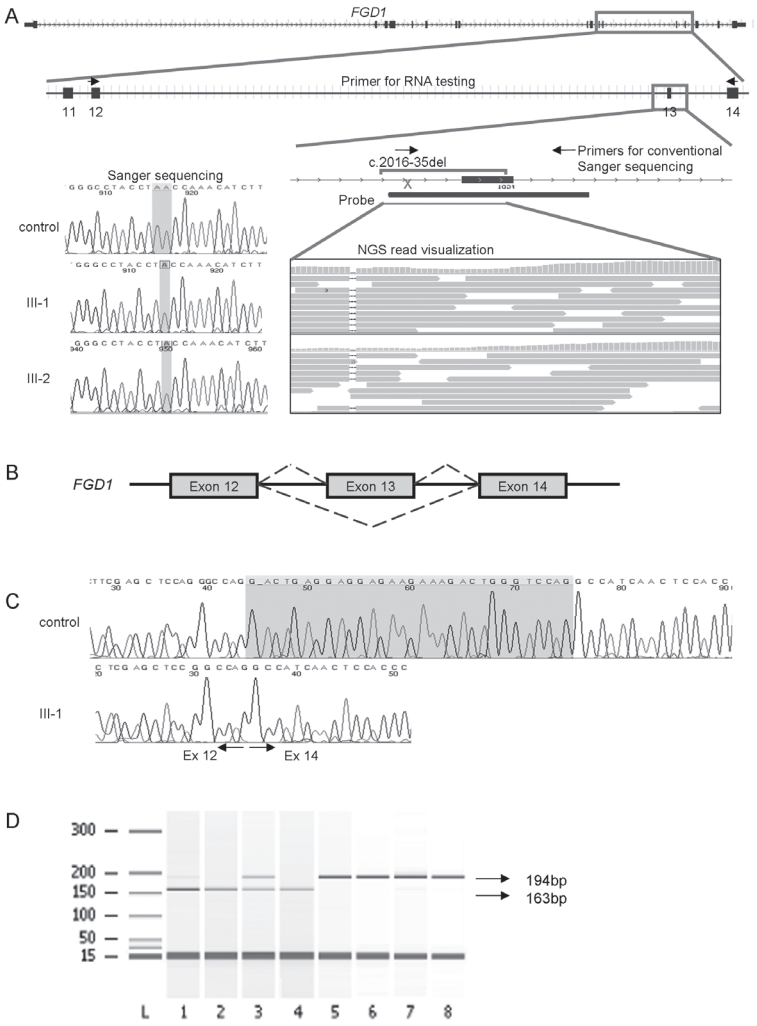
Figure S2. Standard filtering parameters were set for variants with a minimal depth of 8 and initially selecting nonsense, frame shift, splice-site and conserved missense variants not present in variant databases (dbSNP, 1000Genomes, in-house exome variants). Prioritization of variants with functional consequences in targeted exonic regions (Table S1) listed a frameshift insertion in *OXGR1* and a missense variant in *C9orf72*. Additional Sanger Sequencing showed that the *OXGR1* variant did not cosegregate with the disorder in the family. A (GGGGCC)n expansion between exon 1a and 1b in *C9orf72* has recently been reported to cause frontotemporal dementia and/or amyotrophic lateral sclerosis (FTLD/ALS). However, the phenotype of FTLD/ALS is quite different from ASS. Therefore, both variants were excluded as candidate genes for AAS. Criteria for variant analysis were then relieved to include intronic variants up to 50 bp from the splice sites, yielding three candidate variants (Table S2). Unexpectedly, a single nucleotide deletion (c.2016-35del) in the *FGD1* gene emerged that potentially affects the branch point of the splice acceptor site of exon 13. The variant was not present in dbSNP131 or the 1000 Genomes project (May 2011) and has not been reported in the *FGD1* gene variant database (www.LOVD.nl/FGD1), which we established in the course of this study. Using Sanger sequencing we confirmed the variant in the two siblings. In addition we detected perfect co-segregation with ASS in the family (Figure2A).

In Silico analysis predicted the variant to disrupt normal splicing, skipping exon 13, disrupting the reading frame and to lead to a premature stop codon (Desmet et al., 2009) (Figure 2B). This would result in a truncated protein of 677 amino acids.
The effect on splicing was analyzed using RNA derived from cultured lymphocytes. RT-PCR analysis of individuals I-1 (not shown), III-1 (Figure 2C) and III-2 (not shown) showed near complete skipping of exon 13. Semi-quantification of skipped products (Agilent 2100 Bioanalyzer) showed complete skipping in I-1 and III-2, 94% in III-1 and 56% in carrier female II-1 (Figure 2D, Figure S3). We did not observe nonsense-mediated mRNA decay (NMD). RT-qPCR using primers covering exon-exon junctions confirmed these results. Compared to three controls, all affected and carrier individuals had a significantly reduced *FGD1* expression (data not shown).

To explore the general application of exome sequencing for intronic variants, we investigated our datasets to give an estimate of the chance to find an intronic variant by exome sequencing. Although exome sequencing is not designed to detect intronic variants, one can detect them. In our experiment, coverage was sufficient to call variants for ~ 80% of the intronic sequences up to position -35 bp (Figure S4A).We projected the number of variants detected when sequences up to 100 bp intronic would have been targeted by capture (read depth >=8). If probe extension is combined with filtering on conservation scores and information from variant databases (dbSNP, 1000 genomes project, inhouse exome variants), this would increase the number of candidate variants only marginally, making the effort sensible. Figure S4B illustrates this by showing the additional number of candidate variants remaining after standard filtering upon inclusion of an increasing part of the intron (various distances to the splice site) in our dataset.

Exome sequencing has solved several Mendelian diseases successfully. The obvious focus is to look for nonsense, frameshift, splice-site and missense variants and analyze the pathogenic consequence of candidate variants. If the variant cannot be classified, the effect has to be studied in detail at other levels (e.g. RNA, protein, in vitro functional assays). Since RNA is often not available for analysis and since only short intronic segments are analyzed, the true fraction of variants disrupting splicing remains largely unknown but is estimated between 15-50% (Ward and Cooper, 2010). Every intron contains core sequence elements that are essential for splicing: a 5' splice donor site (SD), 3' splice acceptor site (SA) and a branch point site (BP). In experimentally verified distant BPs the region between the BP and 3' SD is devoid of AG dinucleotides, the so called 'AG exclusion zone' (AGEZ) (Gooding et al., 2006) . Since BPs reside deeper into the intron, BP variants proven to affect splicing are quite rare. Observed consequences of variants in BP sites include exon skipping, intron retention and partial intron retention. Several variants have been published that result in genetic diseases. Substitutions resulting in disease have so far only been described in

Figure 2: Genomic analysis of the mutation. A) Analysis result of GAII-X sequencing, displayed in IGV browser. III-1 and III-2 show a single nucleotide deletion, c.2016-35del in intron 12 of *FGD1*. On the left side, DNA Sanger sequencing results confirming the variant in the Dutch Aarskog family (III-1 and III-2 versus a WT control). B) A diagram of the normal and abnormal splicing in exon 12, 13 and 14. C) Sanger sequencing of the RT-PCR shows a skip of exon 13 in individuals III-1 versus a WT control, confirming the branch point variant. The skip of exon 13 leads to a 31 bp shorter mRNA. D) The 2100 Bioanalyzer results show semi-quantification of skipped products. III-1(lane 1) shows a 94% skip, containing two products, a large skipped (163 bp) and a small WT band (194 bp). III-2(lane 2) and I-1(lane4) show a 100% skip. II-1 (mother, lane 3) has a 50-50 distribution of WT and skipped products. Controls(lane 5~8) only carry the WT band.



the two most conserved BPs positions, the BP adenosine itself and the -2 position (relative to the BP)(Kralovicova et al., 2006). To our knowledge, a deletion of the BP has not been described. Useful approaches for follow-up studies of BP variants have been communicated (reviewed by Pagani and Baralle (Pagani and Baralle, 2004)). *In silico prediction* of BP variants, combined with segregation studies and RNA analysis is very useful, especially for clear disease-associated genes. As an alternative, or in conjunction with RNA analysis, *in vitro* splicing reporter assays using minigene reporter constructs and lariat PCR have been used to prove pathogenicity of variants affecting splicing (Gao et al., 2008; Kishore et al., 2008; Kralovicova et al., 2006).

The deletion of the adenosine at position -35 of *FGD1* intron 12 is the first report of an insertion or deletion affecting the BP. Our data make it likely that the A represents the signal for lariat structure formation; 1) its location matches the usual branch point distance from the 3' SA, 2) its sequence "cTaAc" matches the human branch point consensus sequence yUnAy (Gao et al., 2008) , 3) it is conserved (Phastcons score 1), 4) the region between BP and SA is devoid of a AG dinucleotide (an AGEZ , Figure S5) (Gooding et al., 2006) and, most convincingly, 5) we showed abnormal *FGD1* mRNA splicing. Interestingly, since we see a nearly 100% skip of exon 13 in males, the flanking 'A' residue seems not to act as a BP site.

As mentioned, diagnostic *FGD1* Sanger sequencing was performed in the two siblings but failed to identify a causative variant. The 5' primer used to amplify exon 13 covers the deleted A (Figure 2A), but it obviously did not prevent amplification in males and therefore the variant went undetected. Based on the guidelines for diagnostic molecular testing(American college of medical genetics, 2006), there is no clear consensus on the regions that must be included in a standard variant screen. Although the direct splice sites (+1 to +6 and -10 bp to -1) are usually included, the branch point is not. An important reason for this is that BP is difficult to define (weak consensus sequence). Furthermore, RNA is not usually isolated making it difficult to study the effect of any variant detected. It is interesting to note that when blood-derived RNA of the patients would have been analyzed, exon 13 skipping had been detected, triggering an analysis of introns 12 and 13 and the detection of the -35delA variant. The fact that the mutation detection rate for *FGD1* variants in clinically diagnosed ASS patients is less than 20%, may be partly due to the fact that variants affecting splicing are not routinely investigated. We tested four other clinically diagnosed ASS patients with unknown genetic background (3 Dutch, 1 Belgian) for c.2016-35del, but no abnormality was found. Since RNA was not available we could not perform a RNA analysis.

As we demonstrate here, although exome sequencing is not designed to detect intronic variants, they can be partially detected. The intron/exons borders are targeted by the whole-exome capture kits but it does not include the branch point. Even if it is captured, those

variants would be filtered out by the standard strategy to prioritize variants (nonsense, splice site, frameshift). Moreover, sufficient coverage for the branch points is required to reliably call variants while sequence coverage drops significantly near these sites because they only flank the targeted regions. The BP variant in *FGD1* was detected and subjected to detailed analysis mainly because it had a high coverage and because the gene was known to be involved in Aarskog syndrome. Depending on the current capture design, sequences can go up to 200 bp into the introns. To ensure 'standard' capture of the majority of branch points, the exome-kit probe design should be extended with sequences deeper into the intron region taking AGEZ regions into account. Sequencing deeper into the intron will uncover many more variants most of them probably not affecting splicing at all. However, validation of these variants will not be too difficult if they have sufficient coverage.

In Summary, our findings show that traditional clinical gene sequencing is not perfect and, due to practical and cost considerations, may miss causative variants. Application of exome sequencing in a diagnostic setting should give a significant increase in the detection of pathogenic variants by partially covering the intron-exon flanking regions. Ultimately, only full genome sequencing will guarantee that all possible regions are taken into account. Our data indicate that variant analysis in exome sequencing should include analysis of intronic variants, with special attention to branch point sites. In addition, RNA should never be neglected as a valuable source to confirm variants that may affect splicing or to detect such variants directly. In fact, sequencing RNA isolated from the same blood sample as currently used to extract the DNA for exome/genome sequencing may provide an attractive, cost-effective approach when no obvious pathogenic variants are identified.

Acknowledgements

The authors have no conflict of interest to declare.

We are grateful to the family, clinicians, and researchers who contributed to this study. Professor R.C.M. Hennekam is acknowledged for establishing the clinical diagnosis of Aarskog-Scott syndrome in this family. We thank Dr. E.K.Bijlsma, Dr N.S. den Hollander, Dr. A. van Haeringen and Dr. R.McGowan for contributing patient samples not directly used in this study and the technicians in the laboratory of diagnostic genome analysis for their help in lymphocyte culturing. Dr. P.A.C. 't Hoen is acknowledged for his advice and critical reading of the manuscript and J. Laros for the bioinformatic analysis. Y. Sun is supported by China Scholarship Council. This work was supported by the EC's 7th Framework Programme under grant agreement n°s 223026 (NMD-chip), 223143 (TechGene) and 200754 (GEN2PHEN).

1. American college of medical genetics. 2006. Practice guidelines for Sanger Sequencing Analysis and Interpretation .

2. Bedoyan JK, Friez MJ, DuPont B, Ahmad A. 2009. First case of deletion of the faciogenital dysplasia 1 (FGD1) gene in a patient with Aarskog-Scott syndrome. Eur J Med Genet 52:262-264.

3. Bottani A, Orrico A, Galli L, Karam O, Haenggeli CA, Ferey S, Conrad B. 2007. Unilateral focal polymicrogyria in a patient with classical Aarskog-Scott syndrome due to a novel missense mutation in an evolutionary conserved RhoGEF domain of the faciogenital dysplasia gene FGD1. Am J Med Genet A 143A:2334-2338.

4. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37:e67.

5. Gao K, Masuda A, Matsuura T, Ohno K. 2008. Human branch point consensus sequence is yUnAy. Nucleic Acids Res 36:2257-2267.

6. Gooding C, Clark F, Wollerton MC, Grellscheid SN, Groom H, Smith CW. 2006. A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. Genome Biol 7:R1.

7. Hou P, Estrada L, Kinley AW, Parsons JT, Vojtek AB, Gorski JL. 2003. Fgd1, the Cdc42 GEF responsible for Faciogenital Dysplasia, directly interacts with cortactin and mAbp1 to modulate cell shape. Hum Mol Genet 12:1981-1993.

8. Kaname T, Yanagi K, Okamoto N, Naritomi K. 2006. Neurobehavioral disorders in patients with Aarskog-Scott syndrome affected by novel FGD1 mutations. Am J Med Genet A 140:1331-1332.

9. Kishore S, Khanna A, Stamm S. 2008. Rapid generation of splicing reporters with pSpliceExpress. Gene 427:104-110.

10. Kralovicova J, Lei H, Vorechovsky I. 2006. Phenotypic consequences of branch point substitutions. Hum Mutat 27:803-813.

11. Orrico A, Galli L, Buoni S, Hayek G, Luchetti A, Lorenzini S, Zappella M, Pomponi MG, Sorrentino V. 2005. Attention-deficit/hyperactivity disorder (ADHD) and variable clinical expression of Aarskog-Scott syndrome due to a novel FGD1 gene mutation (R408Q). Am J Med Genet A 135:99-102.

12. Orrico A, Galli L, Cavaliere ML, Garavelli L, Fryns JP, Crushell E, Rinaldi MM, Medeira A, Sorrentino V. 2004. Phenotypic and molecular characterisation of the Aarskog-Scott syndrome: a survey of the clinical variability in light of FGD1 mutation analysis in 46 patients. Eur J Hum Genet 12:16-23.

13. Orrico A, Galli L, Faivre L, Clayton-Smith J, Azzarello-Burri SM, Hertz JM, Jacquemont S, Taurisano R, Arroyo C, I, Tarantino E, Devriendt K, Melis D, Thelle T, Meinhardt U, Sorrentino V. 2010. Aarskog-Scott syndrome: clinical update and report of nine novel mutations of the FGD1 gene. Am J Med Genet A 152A:313-318.

14. Orrico A, Galli L, Obregon MG, de Castro Perez MF, Falciani M, Sorrentino V. 2007. Unusually severe expression of craniofacial features in Aarskog-Scott syndrome due to a novel truncating mutation of the FGD1 gene. Am J Med Genet A 143:58-63.

15. Pagani F, Baralle FE. 2004. Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5:389-396.

16. Pasteris NG, Buckler J, Cadle AB, Gorski JL. 1997. Genomic organization of the faciogenital dysplasia (FGD1; Aarskog syndrome) gene. Genomics 43:390-394.

17. Shalev SA, Chervinski E, Weiner E, Mazor G, Friez MJ, Schwartz CE. 2006. Clinical variation of Aarskog syndrome in a large family with 2189delA in the FGD1 gene. Am J Med Genet A 140:162-165.

18. Teebi AS, Naguib KK, Al-Awadi S, Al-Saleh QA. 1988. New autosomal recessive faciodigitogenital syndrome. J Med Genet 25:400-406.

19. van de Vooren MJ, Niermeijer MF, Hoogeboom AJ. 1983. The Aarskog syndrome in a large family, suggestive for autosomal dominant inheritance. Clin Genet 24:439-445.

20. Ward AJ, Cooper TA. 2010. The pathobiology of splicing. J Pathol 220:152-163.

Supplementary Material

Patient DNA

Blood samples of patients and relatives were collected and used for segregation analysis. Consent to genetic testing was obtained from adult probands or parents in the case of minors.

DNA and RNA Analysis

DNA was extracted from whole blood following standard protocols. Leucocyte RNA with or without cycloheximide (CHX) treatment was isolated from blood using Nucleospin RNA II kit (MACHEREY-NAGEL) following the official protocol. cDNA was synthesized from 500 ng of total RNA by RevertAid RNaseH-M-MuLV reverse transcriptase in a total volume of 20 µl according to the protocol provided by the supplier (MBI-Fermentas). Target regions were amplified by RT-PCR (primer sequences are available upon request). The products were evaluated with the Bioanalyzer 2100 DNA chip 1000 (Agilent), according to the manufacturer's instructions.

SNP arrays

SNP arrays (1M-Duov3-0 Illumina Inc., San Diego, CA) were performed and a total of 750 ng DNA was processed according to the manufacturer's instructions. SNP copy number (log R ratio) and B-allele frequency were assessed for II-2, III-1 and III-2 to exclude large copy number variations (CNVs) as a possible genetic cause.

Mutation detection by exome sequencing

To resolve the genetic basis of ASS in this family, we performed exome sequencing of the two siblings (III-1 and III-2) using the SureSelect human all exon kit v2 (Agilent), following the manufacturer's instructions. In brief, 5 ug genomic DNA was fragmented to 300-800 bp using a Covaris S-series and Illumina pair-end adapters were ligated to the fragments. 500 ng adapter ligated DNA was hybridized with 500 ng SureSelect probe mix. Sequencing (one lane per patient) was performed on the Illumina GAIIX. All reads were aligned with BWA-0.5.8 (Li and Durbin, 2010) to the human reference genome hg19. Variant calling, including both single nucleotide substitutions (SNVs) and small insertions, deletions and indels, were done by Samtools-0.1.12 (Li et al., 2009) followed by annotation using SeattleSeq Annot.131.

Sanger sequencing

PCR was performed by using Phire Hot Start II DNA polymerase (Finnzyme) following the official protocol. Primers used in PCR reactions are available upon request. PCR products were first purified by QIAquick PCR purification kit (QIAGEN), then mixed with 25 pmol of the forward or reverse primers respectively and sequenced by the Applied Biosystems 96-capillary 3730XL system.

In Silico Analysis

Prediction of a splice effect was tested using the Human Splicing Finder (HSF) tool which gives a position Weight matrices-derived scores (HSF) for potential splice sites (Desmet et al., 2009), and also integrates the results from the MaxEnt program.

Accession Numbers

GenBank reference sequences: *FGD1*: NM_004463.2
UNIPROT: FYVE, RhoGEF and PH domain-containing protein 1: P98174

Table S1: Standard filtering for exonic regions.

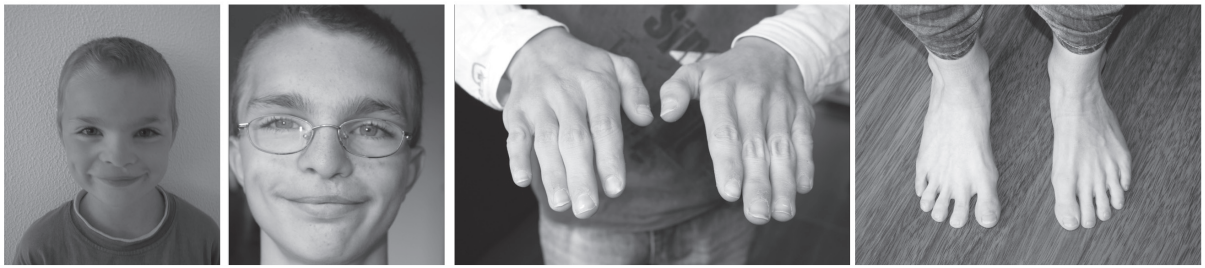
	III-1	III-2
Unique genomic variants	74019	43209
Exonic + 5' and 3'-SS	25642	21491
In sureselect 38M probeset	20080	20187
Depth >= 8	18378	18990
Not coding synonymous	8394	8556
Not validated DbSNP/1K	570	672
Not present In house database	53	169
Conserved variants[a]	30	62
Variants present in two sibs	2	

[a] phylop >3 for SNVs, Phastcons >0.8 for indels

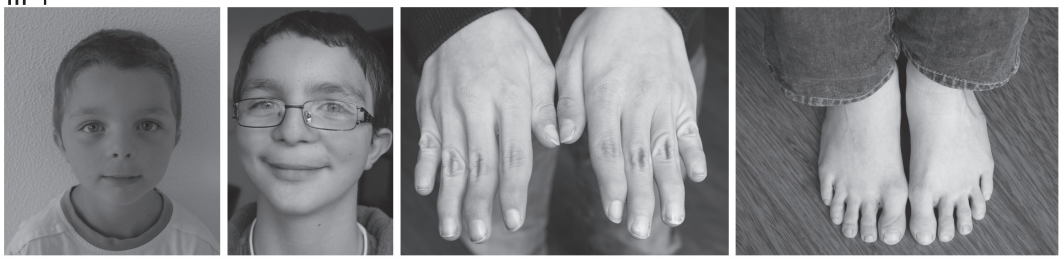
Table S2: Reported variants after standard analysis for intronic positions up to -50 bp.

Sample	Variant	Gene	Function	Depth	dbSNP	OMIM
III-1	NM_018325.2:c.607G>C	C9orf72	missense	9	none	No
III-2	NM_018325.2:c.607G>C	C9orf72	missense	9	none	No
III-1	NM_004463.2:c.2016-35delA	FGD1	intron	11	none	Yes
III-2	NM_004463.2:c.2016-35delA	FGD1	intron	11	none	Yes
III-1	NM_080818.3:c.512_513insA	OXGR1	frameshift	43	none	No
III-2	NM_080818.3:c.512_513insA	OXGR1	frameshift	67	none	No

Figure S1: Phenotype of the two Aarskog siblings (III-1, III-2) and their maternal grandfather (I-1). Facial features are illustrated at age 8 years and 13 years (III-1) and at age 6 years and 11 years (III-2). Craniofacial features, skeletal and genital features are described in Table 1.



III-1



III-2

I-1

Figure S2: Coverage plots of GAIIX sequencing for III-1 and III-2.

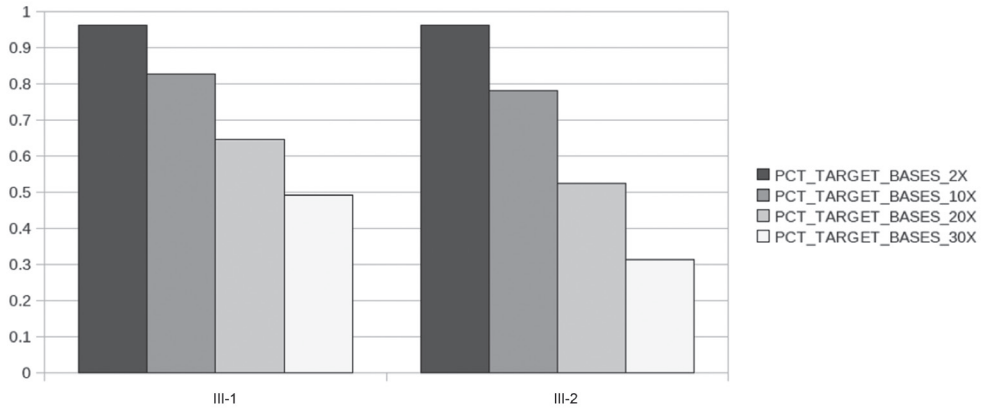


Figure S3: The 2100 Bioanalyzer Traces of RT-PCR show semi-quantification of skipped products. The peaks represented are the lower ladder (15 bp) and upper ladder (1500 bp), the skipped FGD1 product (163 bp) and the WT product.

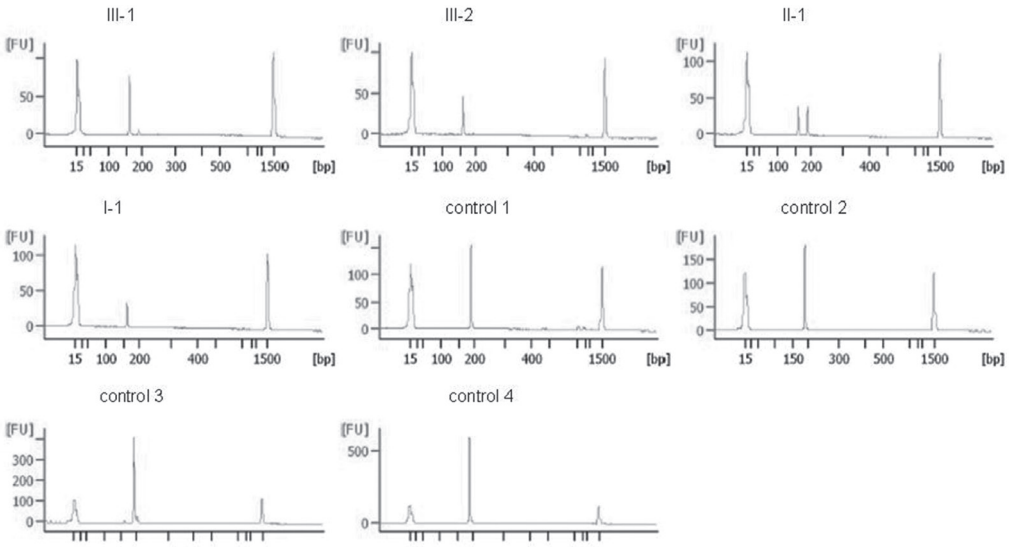


Figure S4: Application of exome sequencing for detection of intronic variants.
A) Intronic coverage plot for a range of distances to the splice site in our dataset. Percentage of the introns covered (minimal depth 8) decreases with an increasing distance from splicesite. B) Calculation of the number of filtered variants upon inclusion of intronic variants by a range of distances to the splice site (depth >8).

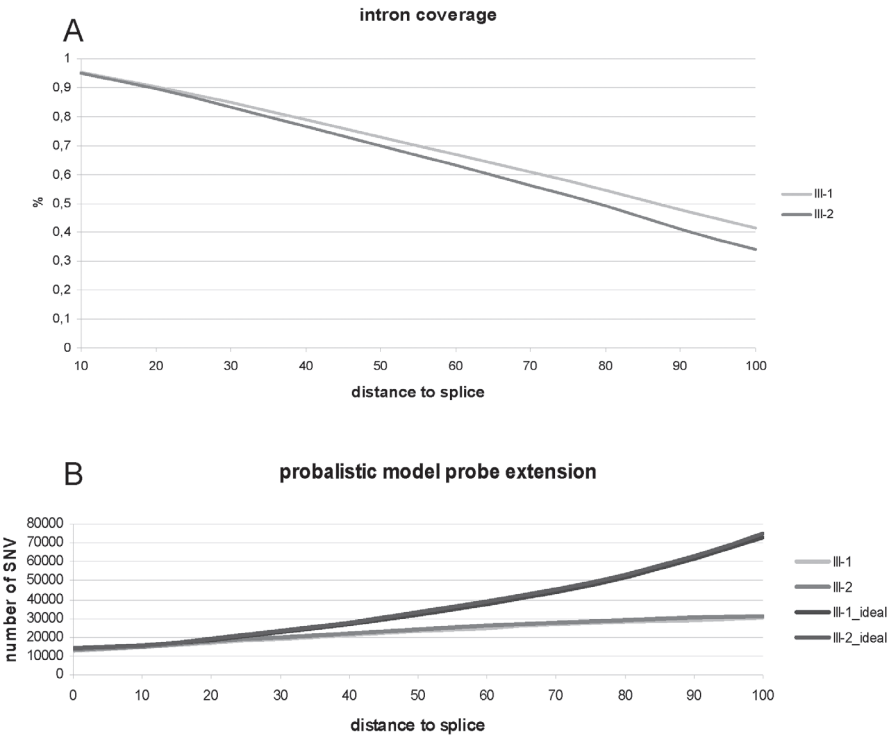


Figure S5: Overview of splice elements of FGD1 exon 12-13. 5'splice site (GU dinucleotide), branch point sequence YUNAY with adenine for lariat formation, AG exlusion zone (AGEZ), polypyrimidine tract (ppt) and 3'splice site (AG dinucleotide).