Identification of a novel mutation in the SGSH gene in an Indian child with Sanfilippo syndrome type IIIA

Krupa Torne a, Parag M Tamhankar b, Vasundhara Tamhankar b, Lakshmi Vasudevan b, Shilpa Mithbawkar b, Rakhi Kesarkar b

aDr. Balabhai Nanavati Hospital, S.V. Road, Vile Parle (West), Mumbai 400 056, India.
bCentre for Medical Genetics, Ecstasy Business Park, JSD Road, Mulund West, Mumbai 400081

*Corresponding author e-mail: paragtmd2@gmail.com

ABSTRACT
We describe an Indian child with Sanfilippo syndrome type IIIA with a novel mutation. The child is six years old, fifth in birth order, born of a non-consanguineous marriage. The child presented with global developmental delay, autistic features, hyperactivity, wide smile, thick bushy eyebrows and umbilical hernia. Clinical exome analysis (next generation sequencing analysis of 8527 genes) identified homozygosity for Chr17:78187615:G>A or c.C733T or p.R245C mutation in the SGSH gene. This is a novel substitution at the common hotspot Arg245, which has the most frequently reported mutation, namely p.R245H. Bioinformatics analysis by Hope software revealed that its pathogenicity could be due to disturbed interactions with other residues, such as aspartic acid 179 and glutamic acid 195. This is the third mutation proven case report of Sanfilippo syndrome type IIIA from India.

KEYWORDS: SGSH, novel, Sanfilippo syndrome, R245C mutation

Citation: Torne et al. Identification of a novel mutation in the SGSH gene in an Indian child with Sanfilippo syndrome type IIIA. Polymorphism 2020;5:7-16.
INTRODUCTION

Sanfilippo syndrome or mucopolysaccharidoses (MPS) type III is an inherited error of metabolism of heparan sulphate and is autosomal recessive in inheritance. Initially, they were described as mild cases of MPS with isolated heparin sulphate excretion in urine, as reported first by Lorincz, Meyer (1958) and Harris (1961). Later, Sylvester Sanfilippo et al. (1963), a pediatrician described eight cases with varying degrees of mental retardation with mild effects on skeleton, viscera and facial features. Later on, four biochemical phenotypes with indistinguishable clinical findings were described: type A: deficiency of heparan N sulphatase (gene SGSH) (MIM#252900); type B: alpha-N-acetylglicosaminidase (gene NAGLU) (MIM#252920); type C: acetyl CoA:alpha-glicosaminde N-acetyltransferase (gene HGSNAT) (MIM#252930) and type D: N-acetylglicosamine-6-sulphatase (gene GNS) (MIM#252940) All four types are autosomal recessive genetic diseases. MPS type IIIA is the commonest type (Andrade et al., 2015). Very few cases of MPS type IIIA have been identified from India. We describe clinical and molecular findings in a child with Sanfilippo syndrome having a novel mutation at the hotspot residue Arg245 in the SGSH gene.

Table 1: List of known pathogenic missense mutations, truncating mutations and small indels in the SGSH gene

<table>
<thead>
<tr>
<th>Exon wise</th>
<th>Number</th>
<th>Likely pathogenic or pathogenic mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter and 5’UTR</td>
<td>1</td>
<td>c.-687_664-23del,</td>
</tr>
<tr>
<td>Intron 1</td>
<td>1</td>
<td>c.89-2A&gt;G</td>
</tr>
<tr>
<td>Intron 2</td>
<td>1</td>
<td>IVS2-2A&gt;G</td>
</tr>
<tr>
<td>Intron 3</td>
<td>1</td>
<td>c.356-2A&gt;G</td>
</tr>
<tr>
<td>Intron 4</td>
<td>Nil</td>
<td></td>
</tr>
<tr>
<td>Intron 5</td>
<td>2</td>
<td>IVS5+17C&gt;T, IVS5-36insTG</td>
</tr>
<tr>
<td>Intron 6</td>
<td>Nil</td>
<td></td>
</tr>
<tr>
<td>Intron 7</td>
<td>Nil</td>
<td></td>
</tr>
</tbody>
</table>
**MATERIALS AND METHODS**

**Clinical findings**
The child was born of a non-consanguineous marriage and was fifth in birth order. He was born preterm, required neonatal intensive care for 54 days for low birth weight and respiratory distress. He required ventilation and surfactant. His milestones were as follows: head holding after 6 months, turning over after 9 months, sitting at 1 year, standing after 1.5 years, walking after 2 years and running at 3 years. He had speech delay, behavioural issues (hyperactivity, attention deficit), autistic features, and social regression. He had umbilical hernia operated at the 1st year. There is a history of recurrent respiratory tract infections. For several years, he was considered by his pediatrician as an autistic individual. Complete blood count and thyroid studies were normal. Fragile X triplet primer polymerase chain reaction showed a normal CGG repeat size of 24 repeats. Karyotype was normal 46, XY,inv(Y)(p11.2q11.2). Parental karyotypes were normal as well. Chromosomal microarray test showed the absence of pathogenic deletion/duplications. MRI of brain done at six months of age showed paucity of white matter in the cerebral hemispheres with diffuse thinning and non-myelination of the corpus callosum. The patient had umbilical hernia, which was operated at one year of age. A follow-up MRI scan at 3.5 years showed mild diffuse atrophy with marginally reduced periventricular white matter with faint hyperintensity on FLAIR sequence in peririgonal white matter (poor myelination). Marginally reduced volume of the corpus callosum was seen. The family history showed recurrent fetal losses. The first pregnancy resulted in medical termination of pregnancy at 6.5 months for antenatal scan showing hydrocephalus and neural tube defect. The second pregnancy resulted in missed abortion at 1.5 months. The third pregnancy resulted in still born male baby at 28 weeks of gestation. The fourth pregnancy resulted in missed abortion at 1.5 months.

On examination at six years, he has mild coarse facial features, tall forehead, spiky hair, thick bushy eyebrows, long face, prominent jaw, large ears, and excessive hair on body. Head circumference was normal (50.5 cm). The patient’s height was normal. The cornea was clear. There was no pallor or icterus. The liver was firm and palpable. Cardiac sounds were normal and there was no murmur. Feet showed pes planus. He was ambulant. He did not respond to name being called and was hyperactive with behavioural issues, such as aggression and inappropriate laughter. The patient spoke very few words and communication was primarily non-verbal. His hearing test was normal. Fundus examination was normal. The clinical features were characteristic of lysosomal storage diseases, particularly Sanfilippo syndrome. Since there are four different enzymes for Sanfilippo

syndrome; a DNA test was preferred over enzyme test to be done first. Clinical exome test (next-generation sequencing analysis of 8527 genes with known phenotypes) was advised to identify the single gene cause for the patients' clinical syndrome.

**Clinical exome sequencing and bioinformatics**

The Exome (8527 genes) was captured using custom-designed (Agilent Sureselect) target specific probes and these targeted regions were then sequenced using the Illumina sequencing system at a mean coverage of more than 80-100X and read quality more than Q20. The target region included the exon and 10bp of flanking intronic sequence. We followed the GATK best practices framework for identification of variants in the sample using Sentieon (v201808.01). The sequences obtained were aligned to the human reference genome (GRCh37/hg19) using Sentieon aligner and analyzed using Sentieon for removing duplicates, recalibration and re-alignment of indels (Freed et al, 2017; Li et al., 2010). Sentieon haplotypcaller was used to identify variants relevant to the clinical indication. Gene annotation of the variants was performed using the VEP program against the Ensemble release 91 human gene model (McLaren et al., 2010; Zerbino et al., 2018). In addition to SNVs and small Indels, copy number variants (CNVs) were detected from targeted sequence data using the ExomeDepth (v1.1.10) method. This algorithm detects rare CNVs based on the comparison of the read-depths of the test data with the matched aggregate reference dataset (Plagnol et al., 2012). Bioinformatics analysis from VCF file and interpretation of sequence variants was done. The bioinformatics pipeline used for analyzing/annotation of VCF files was Annovar (http://wannovar.wglab.org/). Variants were considered pathogenic if they were previously reported in the OMIM indexed disorders or if they are pathogenic as per the ACMG criteria (https://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/). Variants were also considered pathogenic if previously reported in ClinVar database and thought to have important clinical implications. Low somatic mosaic novel variants for dominant diseases with variable penetrance were not considered significant. Low quality and low coverage data were filtered out. The variants were reported in genes which are causally linked to mucopolysaccharidoses group of diseases. Also, the type of mutation identified should have a known biological mechanism for the linked disease as per scientific publications or guidelines. Also, the variant should not have been reported in the homozygous state in ExAC databases. In silico analysis for novel variants was done using software Sorting Tolerant from Intolerant (SIFT) (http://provean.jcvi.org/index.php), Polyphen2 (http://genetics.bwh.harvard.edu/pph2/), MutationTaster (http://www.mutationtaster.org/).

Novel variants were also analyzed using HOPE automatic mutant analysis server available at http://www.cmbi.umcn.nl/hope which provides insights into the structural effects of a mutation. The program Yasara is used to build a homology model and is offered by a number of web services. These services can, for example, calculate residue accessibility, secondary structure, ligand contacts, metal contacts, ionic interactions, disulfide bonds, hydrogen bonds, etc. HOPE uses Uniprot to know about the active sites, domain, motifs, region, glycosylation sites, metal-contacts, DNA contacts, transmembrane domains, variations and mutagenesis sites. It also uses Reprof, a secondary structure prediction program to acquire information about both secondary structure and solvent accessibility. To verify for the founder or common ancestry effect, the locus of loss of heterozygosity (LOH) surrounding the SGSH gene was verified in the VCF file.
RESULTS

The VCF file contained 104506 variants, of which 13531 were exonic, of which 12571 qualified as pass, of which 11704 variants were discarded as having frequency of more than 1 percent in 1000Genomes or ExAC database. Of the selected 867 variants in 604 genes, the variant Chr17:78187615: G>A was identified in the SGSH gene (NM_000199: exon6:c.C733T: p.R245C) in the homozygous state with 391X coverage. The missense mutation R245C, arginine to cysteine seen in our patient is a novel variant predicted to be pathogenic using in silico analysis SIFT, Polyphen2 and MutationTaster software. HOPE web interface analysis identified PDB 4MIV (crystal Structure of Sulphamidase, crystal form L) as a possible modelling template. The mutant residue is smaller than the wild-type residue. The wild-type residue charge was positive; the mutant residue charge is neutral. The mutant residue is more hydrophobic than the wild-type residue. The wild-type residue forms a hydrogen bond with glutamic Acid at position 195. The size difference between wild-type and mutant residue means that the new residue is not in the correct position to make the same hydrogen bond as the original wild-type residue did. The difference in hydrophobicity would affect hydrogen bond formation. The wild-type residue forms a salt bridge with aspartic acid at position 179 and glutamic acid at position 195. The difference in charge will disturb the ionic interaction made by the original, wild-type residue.

Fig 1. 1A: Pedigree of the family 1B to 1D: Analysis of the mutation by HOPE software. 1B: The figure below shows the schematic structures of the original (left) and the mutant (right) amino acid. The backbone, which is the same for each amino acid, is coloured red. The side chain, unique for each amino acid, is coloured black. 1C: Close-up of the mutation at different angles. The protein is coloured grey, the side chains of both the wild-type and the mutant residue are shown and coloured green and red respectively,
1D: Close-up of the mutation. Both the wild-type and mutant side chain are shown in green and red, respectively. The rest of the protein is shown in grey. 1E: Close-up of the mutation, same colours as animation 1. The animation shows alternating the wild-type side chain and the mutant side chain.

No likely pathogenic variant was seen in the following genes: NAGLU (MPS type IIIB), HGSNAT (MPS type IIIC) and GNS (MPS type IIID). The variant R245C satisfies the following criteria: PM1 - located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation; PM2 - absent from controls (or at extremely low frequency if recessive) in exome sequencing 1000 genomes project, or exome aggregation consortium, PM5 - novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before, PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.). The mutation is classified likely pathogenic by criteria 4: three or more than 3 moderate criteria satisfied.

The locus of loss of heterozygosity (LOH) as observed in the VCF file was present between chromosome coordinates chr17: 78087428 and chr17: 79502475, a region of 1415047 base pairs (1.41 MB) with the following genes included: GAA (partial), EIF4A3, CARD14 in upstream; SLC26A11, RNF213, ENDOV, RPTOR, BAIAP2, AATK, ACTG1, FSCN2 (partial) in downstream. One end (upstream) of the LOH was 100187 base pair away from the mutation site and the other end (downstream) of the LOH mutation was 1314860 base pairs away from the mutation site. The following intragenic SGSH gene polymorphic markers was observed in the patient: rs7503034:TT, rs2269376:GG, rs2269375:AA, rs2269374:AA, rs35087113:GG, rs67188486:delAG/delAG, rs6565647:GG, rs6565648:TT (SGSH gene haplotype).

DISCUSSION

In this particular case, mucopolysaccharidosis with primary neurological involvement without major skeletal and visceral involvement, particularly Sanfilippo syndrome, mild types of Schie syndrome, Hunter syndrome, mucolipidosis type III/IV and sialidosis were considered as clinical differential diagnosis. Sanfilippo syndrome has an incidence of 0.68 per 1,00,000 live births in France and 1.21 per 1,00,000 live births in the United Kingdom. MPS type IIIA predominates in the northen Europe, whereas IIIB predominates in the southern countries. The incidence of this disease in our population is unknown and indeed ours is the third mutation confirmed case being reported. It results from failure to degrade glycosaminoglycan (GAG) heparan sulphate. Other mucopolysaccharidoses such as Hurler Schie phenotype and Hunter phenotype present with classical “gargoyleism” depicting coarse facies, radiological manifestation called as “dysostosis multiplex” due to increased accumulation of GAGs in the bones and large liver and spleen enlargement. In contrast, MPS type III presents with subtlety, mild hyperactivity, behavioural abnormality, aggressiveness in the age group three to six years. The neuroregression leads to severe to profound mental retardation typically over a period of 5 to 10 years. Urine demonstrates the typical excretion of heparan sulphate, which may be highly variable and the respective enzymes can be analyzed in leucocytes or skin fibroblasts which are cumbersome and not easily available. X rays scans were not available and the patient was from a distant village and unable to stay in the city for detailed investigations. Thus, the typical facial gestalt of Sanfilippo syndrome with bushy eyebrows in the patient prompted us to take a rapid genotyping approach.
Genotyping first approach by clinical exome analysis (8527 genes with known clinical phenotypes) is fast becoming the test of choice for such patients due to availability, comprehensiveness and definitiveness. Identification of DNA mutation not only provides definitive diagnosis by avoiding other tests, but also provides the possibility of prevention by prenatal diagnosis or pre-implantation genetic diagnosis and carrier detection by targeted mutation analysis for familial mutations. Sanfilippo syndrome was diagnosed in our case by a genotyping approach as per the ACMG classification of variants, thereby, avoiding the need for biochemical testing. A review of mutation 1000 Genome, Human Gene Mutation Database identified 150 unique mutations with predominantly missense/nonsense type mutations (107), 2 splice site variants and the rest being insertion/deletion types. The mutation p.R245H is a common mutation with reported frequency varying from 57.8% (Netherlands), 35% (Germany), 31% (Australasia), 20% (United Kingdom) to 3% (Poland), making it one of the commonly observed mutation (Weber et al., 1997). Other founder mutations include p.S66W (Sardinia, Italy) (DiNatale et al, 1998), c.1080delC (Spain) (Montfort et al, 1998; Chabas et al, 2001) and p.R74C (Poland) (Bunge et al, 1997). In Indian patients, SGSH mutation, such as p.G205R, was reported in a 16 years old girl from Kerala (Singh et al., 2018). In another patient from Andhra Pradesh, p.R377C mutation was identified (Kadali et al., 2019). As per the genotype-phenotype correlation, the mutations p.R245H, p.Q380R, p.S66W, c.1080delC are associated with severe classic (severe) phenotype, whereas other mutations, such as p.G122R, p.R206P, p.S298P, p.I322S, p.D444G, p.S347F, p.E369K are associated with an attenuated phenotype, delayed presentation and slow progression of the disease (Yogalingam et al., 2001; Miyazaki et al., 2002; Gabrielli et al., 2005). In particular, patients with p.S298P on one allele may lead to the preservation of cognitive functions with longer survival even in the presence of severe alleles (Meyer et al., 2008). Other mutations, such as p.L12Q and p.T421R lead to more subtle/mild phenotype (Valstar et al., 2010).

The prognosis of Sanfilippo syndrome patients is supportive treatment since patients do not have curative treatments. The disease has variable progression and the treatment in Sanfilippo syndrome remains supportive in the form of therapy. However, recently Anja John et al. (2020) showed the benefit of hematopoietic stem cell transplant in a girl with Sanfilippo syndrome (p.R74C/R245H genotype in the SGSH gene) when performed early at the age of 2.5 years. When followed up for eight years, there was no regression, she could talk in small sentences, had good motor abilities, performed daily activities, but had profound behavioural problems.

The cause for recurrent fetal losses in the present family could not be ascertained retrospectively. However, it is important to note that stillbirths due to fetal Sanfillipo syndrome or hydrops fetalis have been reported (Muenzer, 2011). Since the recurrence risk of this autosomal recessive syndrome in further pregnancies of the parents is 25% every time, prenatal diagnosis for Sanfillipo syndrome by fetal DNA testing after chorionic villus sampling (CVS) or amniocentesis around 12 to 16 weeks was advised. The option of in vitro fertilization and pre-implantation genetic diagnosis for Sanfillipo syndrome as a reproductive option was also discussed with the parents.

Acknowledgments
We thank the patient’s family for their support. No specific funding was received for this work.

Conflict of interest statement
The authors have declared to have no conflict of interest.

Authors’ contributions
KT, PMT, VT performed clinical evaluation; PMT,
VT, LV, SM, RK performed genetic analysis; PMT performed bioinformatics analysis; KT, PMT wrote the first draft and all authors approved the final draft.

Declaration of originality
The authors declare that they have not copied text, figure or data from a particular source without appropriately citing it.

REFERENCES


Rush N, Hawley D, Gener M, Hattab E. Phenotypic Manifestations in a Patient With Mucopolysaccharidosis Type IIIA due to Two Novel Mutations in the Gene
Encoding N-Sulfoglucosamine Sulfohydrolase: An Autopsy Case Study. AJCP 2015; 144 (2): A022