

# The GALNS p.P77R variant is a probable Gujarati-Indian founder mutation causing Mucopolysaccharidosis IVA syndrome

**Harsh Sheth** (✉ [harsh.sheth@frige.co.in](mailto:harsh.sheth@frige.co.in))

FRIGE's Institute of Human Genetics, Ahmedabad, India

**Premal Naik**

Rainbow Pediatric Orthopedic Hospital

**Maulin Shah**

Orthokids Clinic

**Riddhi Bhavsar**

FRIGE's Institute of Human Genetics, Ahmedabad, India

**Aadhira Nair**

FRIGE's Institute of Human Genetics, Ahmedabad, India

**Frenny Sheth**

FRIGE's Institute of Human Genetics, Ahmedabad, India

**Jayesh Sheth**

FRIGE's Institute of Human Genetics, Ahmedabad, India

---

## Research Article

**Keywords:** Mucopolysaccharidosis IVA (MPS IVA), Morquio A syndrome, GALNS, founder variant, p.P77R

**Posted Date:** March 25th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1449904/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background:** Mucopolysaccharidosis IVA (Morquio syndrome A, MPS IVA) is an autosomal recessive lysosomal storage disorder caused due to biallelic variants in the N-acetylgalactosamine-6-sulfatase (*GALNS*) gene. The mutation spectrum in this condition is determined amongst sub-populations belonging to the north, south and east India geography, however, sub-populations of west Indian origin, especially Gujarati-Indians, are yet to be studied. We aimed to analyse the variants present in the *GALNS* gene amongst the population of Gujarat by sequencing all exons and exon-intron boundaries of the *GALNS* gene in patients from 23 unrelated families.

**Results:** We report 11 variants that include eight missense variants: (p.L36R, p.D39G, p.P77R, p.C79R, p.P125L, p.P151L, p.G255A and p.L350P), one splice site variant: (c.121-7C>G), one small insertion: (c.1241\_1242insA, p.I416HfsTer2) and one small deletion: (c.839\_41delACA). Of these, three missense variants (p.D39G, p.G255A and p.L350P), one splice site and the two indels mentioned above are novel. Interestingly, we observed a higher than anticipated prevalence of p.P77R variant in our cohort (n=14/25, 56%). Haplotype analysis in cases with p.P77R variant and 63 ethnicity matched healthy population controls suggested a 4 SNP haplotype block present in cases compared to controls (likelihood ratio test p-value=  $1.16 \times 10^{-13}$ ), thereby suggesting p.P77R variant as a founder variant in the Gujarati Indian population. Furthermore, age of mutation analysis suggested the variant to have arisen approximately 450 years ago in the population.

**Conclusion:** p.P77R variant in the *GALNS* gene is likely to be a founder variant in MPS IVA patients of Gujarati-Indian ancestry and appeared approximately 450 years ago in the population. To our knowledge, this the first variant to be posited as a founder variant in the *GALNS* gene in patients with MPS IVA syndrome.

## Background

Mucopolysaccharidosis IV A (MPS IVA, Morquio-A syndrome, OMIM#253000) is an autosomal recessive disorder which is caused by the deficiency in enzyme activity of N-acetylgalactosamine-6-sulfatase (*GALNS*) due to biallelic variants in the *GALNS* gene (OMIM#612222) [1, 2]. *GALNS* enzyme plays an essential role in the degradation of glycosaminoglycans (GAGs), keratan sulfate (KS) and chondroitin-6-sulfate (C6S) [3]. Therefore, deficiency in the *GALNS* enzyme activity leads to the accumulation of these substrates in the lysosomes, particularly in cartilage and cornea, leading to a wide gamut of clinical manifestations such as shorth trunk dwarfism, genu valgum, pectus carinatum, kyphoscoliosis, joint laxity, dysmorphic face, hepatomegaly and abnormal gait [4].

MPS IVA is a rare lysosomal storage disorder, with an estimated incidence ranging from one in 76,000 in Northern Ireland to one in 640,000 births in Western Australia [4–6]. Although the study in the Asian population is scarce, the available data estimates the birth prevalence of MPS IVA to be 1 in 500,000 live births in Japan, 1 in 304,000 in Taiwan [7], and 1 in 701,000 live births in Malaysia [8].

The *GALNS* gene is situated on chromosome 16q24.3 and comprises of 14 exons spanning over 50kb of genomic length and codes for a 522 amino acid *GALNS* protein with a signal peptide of 26 amino acids [9]. To date, a total of 333 variants in the *GALNS* gene have been reported which comprises of 248 missense/ nonsense variants, 32 small deletions, 5 small insertions, 2 small indels, 32 splice site variants and 3 complex rearrangements in the HGMD database [<http://www.hgmd.org>, accessed on 21st October 2021]. Of these, the 10 most commonly reported variants in MPS IVA patients from across the globe are: c.120 + 1G > A, c.337A > T, c.757C > T, c.860C > T, c.871G > A, c.901G > T, c.935C > G, c.953T > G, c.1156C > T and c.1171A > G [8]. Some of these variants are also commonly observed in particular sub-populations such as c.120 + 1G > A in 91% of Tunisian patients, c.1171A > G and c.337A > T in 64% and 52% of the Irish patients respectively, c.757C > T in 89% of Pakistani patients and c.953T > G in 58% of Chinese patients [8].

Interestingly for the Indian population, a large study of 68 unrelated MPS IVA patients from mostly northern and southern geographical regions identified 22 novel variants [10]. Of these, c.860C > T (8.82%), c.647T > C (7.35%), c.95A > C (6.61%) and c.871G > T (5.88%) were the most frequent variant in the observed population compared to other populations [10]. However, considering the diverse genetic architecture of the sub-populations residing in the Indian sub-continent, the study is likely to miss variants predominant in patients from western and eastern parts of India.

Here, we present the mutation spectrum of the *GALNS* gene in 23 patients from unrelated families from the western part of India, predominantly Gujarat. Furthermore, we present evidence for p.P77R variant as a founder variant in the Gujarati-Indian sub-population through haplotype analysis and estimate its age of emergence in the population.

## Results

### Clinical and *GALNS* enzyme activity spectrum

A total of 23 patients affected with MPS IVA were included in the present study. Out of the 23 families, seven had consanguineous marriages and the remaining 16 families practised endogamy. The age at diagnosis ranged from 11 months to 21 years with a mean age of 4.32 years. Detailed clinical and anthropometric information is presented in Table 1. Common clinical features included: short stature, short fingers, short neck-trunk, and frontal bossing. Kyphosis was seen in 43% of the patients (n = 10/23) and platyspondyly in 21% of the patients (n = 5/23). Of note, other prominent features of MPS IVA such as corneal clouding and knock knee were seen in 1/23 (5%) and 6/23 (24%) patients, respectively. On the radiological assessment of the patients, the key findings were anterior beaking of the vertebral bodies, acetabular dysplasia, platyspondyly, and multiple dysplastic epiphyses.

Table 1  
Clinical and anthropometric details of MPS IVA patients in the present study.

Patient ID	Age (in years)	Sex	Religion (Region)	Consanguinity	Phenotype							Additi
					Short stature	Pectus carinatum	Kyphosis	Scoliosis	Genu valgum	Platyspondyly	Anterior beaking of bone	
P1	3	Female	Hindu (Gujarat)	No	-	-	□	-	-	-	□	Short
P2	1.9	Male	Hindu (Gujarat)	No	-	-	□	-	-	-	□	Lumb
P3	0.9	Female	Muslim (Gujarat)	Yes	-	-	□	□	-	-	-	Gibbu
P4	1.5	Male	Muslim (Gujarat)	Yes	-	-	-	-	-	-	-	Skelet
P5	8	Female	Hindu (Gujarat)	No	□	-	□	-	□	-	-	Skelet extenc delaye develc fronta cornea short i trunk, wrist & mild h
P6	1	Female	Hindu (Gujarat)	No	-	-	□	□	-	□	-	Large bossir
P7	2	Male	Hindu (Gujarat)	No	□	-	□	-	-	-	-	Hydro
P8	6	Male	Darbar (Gujarat)	Yes	-	□	-	□	□	-	-	Skelet rickets
P9	2	Male	Hindu (Gujarat)	No	-	-	□	-	-	□	-	Bullet-metac bilater dyspl
P10	1	Male	Hindu (Gujarat)	No	-	-	-	-	-	□	-	Abnor
P11	6	Male	Hindu (Gujarat)	No	□	-	-	-	-	-	-	Short short t
P12	1.7	Female	Hindu (Gujarat)	No	-	-	-	-	-	-	-	Coars skelet abnor disturl
P13	2.5	Male	Hindu (Gujarat)	No	-	-	-	-	□	□	□	Bullet-phalla distort acetab
P14	3.7	Female	Hindu (Gujarat)	No	□	-	-	-	□	-	-	Sponc dysplac acetab dysplac epiphy dysplac
P15	4	Female	Hindu (Gujarat)	No	-	-	-	-	-	-	□	Short metac with p Madel in wris
P16	12	Female	Hindu (Gujarat)	No	-	-	-	-	-	-	-	Dysos multiplic acetab deform lip
P17	21	Male	Hindu (Gujarat)	No	-	-	-	-	-	-	-	Skelet

Patient ID	Age (in years)	Sex	Religion (Region)	Consanguinity	Phenotype							Additional
					Short stature	Pectus carinatum	Kyphosis	Scoliosis	Genu valgum	Platyspondyly	Anterior beaking of bone	
P18	5	Male	Javia (Gujarat)	No	-	-	-	-	□	-	-	Ankle bilateral valgus; spondyl epiphyseal dysplasia
P19	3.5	Male	Muslim (Gujarat)	Yes	□	-	-	-	-	-	-	Dysostosis multiplex
P20	4	Female	Muslim (Gujarat)	Yes	□	-	-	-	-	-	-	Lumbar wrist v
P21	1.6	Male	Hindu (Gujarat)	No	-	-	□	□	-	-	□	Short with p proximal
P22	6	Female	Muslim (Gujarat)	Yes	-	□	-	-	□	-	-	Coarsening angle
P23	2	Female	Muslim (Gujarat)	Yes	-	-	-	-	-	-	-	Skeletal

Elevated levels of urinary GAG were observed in all the patients with excess keratan sulfate (KS) and chondroitin sulfate (CS). The enzyme activity of N-acetylgalactosamine-6-sulfate sulfatase in all the patients was in the range of 0.02–11.1 nmol/hour/mg of protein and the mean enzyme activity was 2.30 nmol/hour/mg of protein which was less than 10% of the mean normal enzyme activity (Table 1).

#### GALNS gene mutation spectrum

The 23 patients harbored 11 different variants in the *GALNS* gene of which, 8 were missense variants (84%): c.107T > G (p.L36R), c.116A > G (p.D39G), c.230C > G (p.P77R), c.235T > C (p.C79R), c.374C > T (p.P125L), c.452C > T (p.P151L), c.764G > C (p.G255A) and c.1049T > C (p.L350P), 1 was a splice site variant (4%): c.121-7C > G, one small insertion (4%): c.1241\_1242insA (p.I416HfsTer2) and one small deletion (4%): c.839\_41delACA (Table 2; Fig. 1). Of these, 3 missense, one splice site and two indel variants were novel (p.D39G, p.G255A, p.L350P, c.121-7C > G, c.839\_41delACA and c.1241\_1242insA/p.I416HfsTer2) to our population (Table 2; Fig. 1). All 23 patients were observed to have a variant in the homozygous state except in two patients where the variants were likely to be compound heterozygous (p.P77R/p.G255A and c.839\_41delACA /p.P125L; Table 2; Fig. 1). All known and novel variants were classed as pathogenic/likely pathogenic according to the ACMG-AMP classification system [11].

Table 2  
*GALNS* gene mutation identified in Gujarati-Indian MPS IVA patients using Sanger sequencing.

Patient ID	cDNA change <sup>a</sup>	Amino acid change <sup>b</sup>	Location <sup>c</sup>	Mutation type	Allele frequency (gnomAD <sup>d</sup> )	Ethnicity	Reference	PolyPhen-2 analysis	DANN score	ACMG-AMP classification
P20	c.121-7C > G	-	Intron 1	Splice site	NA	Indian	Present study	-	-	Uncertain significance/likely pathogenic
P4, P19	c.107T > G	p.L36R	Exon 1	Missense	0.00003	Japanese, European, Italian, Indian	Morrone et al. 2014, Caciotti et al. 2015	1 Probably damaging	0.9889	Pathogenic
P7	c.116A > G	p.D39G	Exon 1	Missense	NA	Indian	Present study	0.997 Probably damaging	0.9844	Likely pathogenic
P1-P3, P5, P9, P10, P11*, P12, P14, P17, P18, P21-P23	c.230C > G	p.P77R	Exon 2	Missense	NA	Indian, Turkish	Tomatsu et al. 1995, Tuyuz et al. 2019, Present study	1 Probably damaging	0.9984	Pathogenic
P16	c.235T > C	p.C79R	Exon 2	Missense	0.00000807	Indian, Malaysian	Bidchol et al. 2014, Leong et al. 2019	1 Probably damaging	0.9974	Pathogenic
P15*	c.374C > T	p.P125L	Exon 4	Missense	0.00000399	Japanese, Chinese	Tomatsu et al. 1997, Xie et al. 2018	1 Probably damaging	0.9247	Pathogenic
P8	c.452C > T	p.P151L	Exon 5	Missense	0.0000279	Pakistani, Indian	Tomatsu et al. 1995, Bidchol et al. 2014	1 Probably damaging	0.9986	Pathogenic
P11*	c.764G > C	p.G255A	Exon 8	Missense	NA	Indian	Present study	1 Probably damaging	0.9979	Likely pathogenic
P15*	c.839_841delACA	p.N280del	Exon 8	Deletion	NA	Indian	Present study	-	-	Likely pathogenic
P13	c.1049T > C	p.L350P	Exon 10	Missense	NA	Indian	Present study	1 Probably damaging	0.9989	Likely pathogenic
P6	c.1241_1242insA	p.I416Hfs*2	Exon 11	Frame shift	NA	Indian	Present study	-	-	Pathogenic
NA: Not available. DANN: Deep learning approach to annotating variants.										
<sup>a</sup> cDNA numbering is based on RefSeq transcript NM_000512.5										
<sup>b</sup> Amino acid change is based on NP_00503.1										
<sup>c</sup> Genomic position is based on hg19/GCRh37 genome build										
<sup>d</sup> gnomAD allele frequency is based on version 2.1										

Interestingly, whilst the p.P77R variant had been reported previously in MPS IVA patients from India [10], the proportion of patients with this variant in our cohort was significantly high (60%; Table 2). Since our cohort was primarily derived from the same geo-ethnicity, a substantially higher than expected prevalence of the p.P77R variant in the cohort suggested it to be a potential founder variant.

#### GALNS p.P77R haplotype and age of variant estimation

A total of 31 SNPs were found between exons and exon-intron boundaries of the *GALNS* gene using the smMIP based sequencing approach within 13 cases and 63 controls. The haplotype associated with the p.P77R variant were formed by the SNP markers rs11076715, rs11076716 and rs377453859 with a total genomic length of 28.3kb (Fig. 2A and 2B). We observed a significant enrichment for this haplotype in cases compared to controls (likelihood ratio test  $p$ -value =  $1.16 \times 10^{-13}$ ), suggesting the p.P77R to be a founder variant in the MPS IVA patients of Gujarati-Indian ethnicity. Analysis performed by the DMLE + 2.3 software estimated that the age of the *GALNS*p.P77R variant might be approximately 450 years (95% CI: 306–647 years) (Fig. 2C) in the Gujarati-Indian population.

## Discussion

Mutational spectrum analysis for a given disease in diverse populations have provided critical evidence for understanding disease pathophysiology, gene domains that are intolerant to variation and development of therapeutic strategies. Indeed, communities that are geographically isolated or practice consanguinity/ endogamy are likely to have founder variants which are critical for developing and deploying targeted molecular diagnostics, as part of an initial screening, thereby reducing cost and turnaround time [12]. Founder variants exhibit linkage disequilibrium with nearby genetic markers. The size of the linkage disequilibrium interval is inversely correlated with the time since the variant emerged in a given population and studying the prevalence of a given founder variant amongst different sub-populations could provide evidence for their shared ancestry [13].

The current study is the first to report the mutation spectrum within *GALNS* gene in patients diagnosed with MPS IVA disease from the western part of India, especially Gujarat. Whilst ours is not the first study from India [10], the difference in the mutational spectrum between the two studies suggests either lack of or absence of patient cohort from the western part of India in the previous study. Furthermore, subtle difference in the phenotypic manifestation such as the mean age of diagnosis (4.32 years in the current study versus 6.7 years in Bidchol *et al.*) could be due to the differences in the awareness amongst the clinical fraternity, availability of screening/ diagnostic facility and the utility of the urinary GAG screening technique in the present study as a first line test.

However, there are several grounds of congruence between the two studies, such as, similar prevalence of missense variants between the two studies (~80%). Indeed, the mutation data for the *GALNS* gene in the HGMD database also shows missense variants to be the most common. Out of the eight missense mutations identified in our study, five have been previously reported in MPS IVA patients from different populations (p.L36R, p.P77R, p.C79R, p.P125L and p.P151L). The mutation p.C79R was first reported in an Indian MPS IVA patient by Bidchol *et al.* followed by Leong *et al.* who reported the same mutation in an Indian patient from Malaysia [7, 10]. Likewise, p.P125L and p.P151L have been previously reported in Mexican, Japanese and Italian populations. Structural studies have shown the amino acid Proline at the 125th position to be located in the N-cap domain of the alpha-helix. Also, the amino acid Proline at the 151st position is a conserved amino acid. Hence, any substitution at these positions is likely to result in modification of the packing and disturbing the tertiary structure of the *GALNS* protein [14]. The two most common variants identified in the present study cohort were p.L36R and p.P77R. The p.L36R variant is associated with the attenuated phenotype as reported by Tomatsu *et al.* [9]. In congruence with this, the patients in our cohort with this variant presented with milder phenotype. The variant p.P77R was first reported by Tomatsu *et al.* in a MPS IVA patient of Indian ethnicity [15]. This mutation has been reported in patients with severe phenotype [16]. Also, the amino acid Proline at position 77 has a key role in preventing the internalization in the hydrophobic patch of the *GALNS* protein. The positively charged side chain of arginine instead of proline is expected to have a destabilizing effect in this process [14]. Interestingly, none of the ten most common mutations in the *GALNS* gene mentioned in the earlier reports were found in our study. This observation is similar to that made by Bidchol *et al.* in their cohort of Indian patients [10]. This gives strong evidence about the molecular heterogeneity in India as well as the presence of a distinct mutation spectrum for the Indian sub-populations.

Haplotype analysis have previously been conducted lysosomal storage disorders to identify founder variants, for example, *GBA* p.N370S in Ashkenazi Jewish population [17] and *GBA* p.G85E in Korean population [18]. Indeed, our group has previously identified a common variant p.E462V in the *HEXA* gene in Tay Sach disease patients of Gujarati Indian ethnicity [19]. However, to our knowledge, no founder variant has been identified to date for MPS IVA patients throughout the world. To demonstrate that the variant p.P77R emerged due to a founder effect, a total of 30 SNPs were analysed of which 3 SNPs formed a strong linkage disequilibrium with the variant, which was observed across all 13 unrelated families in the current study. This variant may have emerged within the population itself or may have been brought through immigration and subsequent population bottleneck. The emergence of this variant coincides with establishment of the Mughal empire in the now state of Gujarat. However, due to the lack of mutational data availability from sub-populations residing in geographical units around Gujarat, it is difficult to assess the emergence of this variant and estimate number of carriers for this variant outside Gujarat. The Indian population is highly heterogeneous with each sub-population likely practicing endogamy/ consanguinity, hence, an accurate method for determination of the ethnicity of the studied individuals would be with the use of high-density SNP arrays, an analysis which is beyond the scope of the present study.

Technological advances coupled with population specific genetic architecture details provides a valuable tool for an efficient diagnosis and screening strategies for rare diseases like MPS IVA. In our study, we used urinary GAG as a screening test which is a relatively inexpensive approach for detection of MPS IVA patients. This screening test coupled with founder mutation is likely to help towards rapid and low-cost diagnosis of MPS IVA patients in Gujarat. With the Gujarati-Indians forming large diasporas in several western nations including the USA and UK, the impact of the current study on MPS IVA diagnostics is likely to be profound beyond India.

We hereby provide the first report showcasing the mutation spectrum in the *GALNS* gene amongst MPS IVA patients of the Gujarati-Indian origin. Furthermore, we provide evidence of a founder effect for the p.P77R variant in the Gujarati-Indian population with an estimated age of the variant to be approximately 450 years, which can be used as first-line marker for rapid genetic diagnosis in this population.

## Materials And Methods

### Patient recruitment

The present study comprises of 23 unrelated families with at least one member with clinical suspicion of MPS IVA (N = 23). These patients were recruited on the study after obtaining written informed consent from their parents/ guardians. The patients were clinically suspected of MPS IVA and their diagnosis was confirmed by the presence of kerton sulfate in urinary GAG one dimensional electrophoresis study followed by beta-galactosidase sulfate assay. Clinical history, family history and consanguinity details were recorded in pre-designed clinical *pro forma*. Sixty-three unrelated individuals from the same ethnogeographic group were recruited as controls. The study protocol was approved by the Institutional Ethics Committee of the Foundation for Research in Genetics and Endocrinology, Ahmedabad (Registration no: E/13237) as per the Helsinki Declaration. 5ml of peripheral whole blood was collected from cases and controls for GALNS enzyme and molecular diagnostic assays. Furthermore, 10-15ml of urine sample was collected from cases for urinary GAG testing.

## Urinary GAG testing and GALNS enzyme assay

Urinary GAG quantitative study was performed dimethylmethylene blue dye based spectroscopic method [20]. The concentration of urine creatinine of individual patients was measured using Liquixx Creatinine Kit (Erba Mannheim, Germany) as per the manufacturer's instructions. The pre-treatment of urine prior to the qualitative analysis of GAGs by electrophoresis was performed according to the protocol by Hopwood and Harrison 1982 [21]. Lysosomal enzyme  $\beta$ -galactose-6-sulfate-sulfatase activity was carried out from leukocytes using fluorogenic synthetic substrate 4-methylumbelliferyl- $\beta$ -galactose-6-sulfate-triethyl ammonium as described by Van Diggelen *et al*, 1990 [22]. The fluorescence of free 4-methylumbelliferone (4 MU) was measured by LS55 spectrofluorometer (Perkin Elmer, USA) to determine the  $\beta$ -galactose-6-sulfate-sulfatase activity. Protein concentration was determined by the Lowry method. The enzyme activity was expressed as the amount of substrate in nmol cleaved per hour per mg of protein in the cell lysates.

### GANLS gene mutation identification by Sanger sequencing

For all cases and controls, high molecular weight genomic DNA was isolated from peripheral blood by the salting-out method [23]. Exon and exon-intron boundaries of the *GALNS* gene were amplified by PCR in 14 fragments using 14 primer pairs (Supplementary Table 1). A 10 $\mu$ l reaction mixture for the DNA amplification of each fragment was made up of 100ng genomic DNA, 1mM dNTPs, and 10X Cetus buffer. 30 cycles of amplification were performed, each consisting of denaturation at 94°C for 1 minute, annealing at 60–65°C suitable for each exon for 45 seconds, and extension at 72°C for 45 seconds in a thermal cycler. The final extension time was at 72°C for 10 minutes. PCR products along with a 100 base-pair DNA ladder were then subjected to electrophoresis in 2% agarose gel for validation of amplification and the amplified products were purified using Exo-SAP-IT™ (USB Corporation, OH, USA). The purified products were sequenced using BigDye Terminator v3.1 and capillary electrophoresis was performed using an automated sequencer ABI-3500 (Applied Biosystems, CA, USA) for mutation analysis of *GALNS*. Bi-directional sequencing data was analysed by comparing the sequence read with the reference sequence of the *GALNS* gene (RefSeq cDNA NM\_000512.5). Identified variants were annotated with data from 1000genomes, gnomAD, dbSNP and the Human Gene Mutation Databases. *In silico* assessment of variant pathogenicity was carried out using SIFT (Sorting Intolerant From Tolerant) (<http://sift.jcvi.org>), Polyphen2 (Polymorphism Phenotyping v2) (<http://genetics.bwh.harvard.edu/pph2/>), MutPred (<http://mutpred.mutdb.org/>), PROVEAN (<http://provean.jcvi.org>) and MutationTaster (<http://www.mutationtaster.org/>). Finally, variants were annotated for pathogenicity using the ACMG-AMP classification system for single nucleotide variants [11].

### SNP genotyping in the GANLS gene for haplotype analysis of p.P77R variant

A total of 13 cases harboring p.P77R variant and 63 controls were genotyped for common single nucleotide polymorphisms (SNPs) in the *GALNS* gene for haplotype analysis using targeted capture by single molecule molecular inversion probe (smMIP) based sequencing assay. Briefly, smMIPs targeting exons and exon-intron boundaries of the *GALNS* gene were designed using the MIPgen tool [24]. smMIP probes were designed against GCRh37/hg19 human reference genome build with following set of parameters: a target capture size of 110bp, a combined length of 40bp for the targeting arms, 5bp unique molecular barcode sequence and no common single nucleotide polymorphisms (SNPs; dbSNP151 database) in the smMIP extension or ligation arm. A total of 40 smMIPs were designed (Supplementary Table 2). The smMIP capture was performed in accordance with the protocol described, with minor modifications [25]. In short, the regions of interest were captured in a reaction containing a molecular ratio between genomic DNA and smMIP of 1:1000. The smMIP capturing conditions were 10 minutes at 95°C for denaturation, followed by an incubation period of 16–18 hours during which hybridization of the phosphorylated smMIPs to the single stranded target DNA occurs together with gap-fill and probe circularization via ligation. All non-circular targets were digested by exonuclease treatment and the circular targets were amplified with primers containing sample specific barcoded reverse primers with following PCR conditions: 30 seconds at 98°C followed by 20 cycles of 10 seconds at 98°C, 30 seconds at 60°C and 30 seconds at 72°C and 2 minutes at 72°C. Primers used for amplification contained adapters compatible with Illumina sequencing [26]. After PCR, libraries were pooled each from 76 barcoded individual libraries. The pooled libraries were purified using Agencourt AMPure XP beads according to the manufacturer's protocol (Beckman Coulter, USA) and the final library was diluted to a concentration of 9pM and subsequently sequenced on a MiSeq platform (Illumina, USA) according to the manufacturer's protocol (300 cycles; V3 kit) resulting in 2x156bp paired end reads.

Data was analysed using an in-house smMIP pipeline which involved: trimming of 5bp unique molecular barcode (UMB) from fastq files and stored for later use, read alignment against GCRh37/hg19 human reference genome using BWA mem (v0.7.12) [27] with output presented as a sample specific BAM file amalgamated with UMB data, base quality score recalibration using GATK v4.1.12 and single nucleotide variants and indel calling using GATK's HaplotypeCaller v4.1.12 to create a VCF file. Data from VCF file was annotated with 1000 genomes and gnomAD databases to identify common SNPs (> 5% minor allele frequency) in the general population and the filtered SNPs were then transposed to a multi-sample excel sheet (Supplementary File 1).

## Haplotype reconstruction and age of p.P77R variant

All analyses were carried out using GCRh37/hg19 chromosomal positions. SNPs along the coding regions of the *GALNS* gene were used to obtain the haplotype that flanks the p.P77R variant in 13 cases and 63 controls. SNPs were assessed for Hardy-Weinberg equilibrium within the control cohort and any SNP not in equilibrium was not assessed in the downstream analyses. Pairwise linkage disequilibrium using `-pwlid-` command was used to estimate SNPs in linkage disequilibrium ( $R^2$  score). Association of the linkage disequilibrium block between cases and controls was assessed using the likelihood ratio test and

-hapipf- command. A two-sided p-value of  $< 0.05$  was considered statistically significant. All haplotype analysis was carried out using Stata v12. A 4 SNP haplotype block was selected for variant dating. The DMLE + 2.3 software [28] was used to estimate the age of p.P77R variant. The algorithm uses an intra-allelic coalescent model to assess the linkage disequilibrium across the marker set coupled to marker locations, population growth rate (0.01 for India), proportion of population samples (0.00248 for Gujarat) and a proportion of disease bearing chromosomes. Full input details are available in the Supplementary File 2.

## Declarations

### Ethics approval and consent to participate

Present study under submission has been approved by the institutional ethics committee [FRIGE's Institute of Human Genetics] with approval number FRIGE/IEC/5/2010 dated 7<sup>th</sup> March, 2010. This process is in accordance with the Helsinki declaration. An informed consent was obtained from the parents before enrolling the patients for the investigations.

### Consent for publication

An informed consent for publication of non-identifiable data was also obtained from parents of patients or individuals included in the submission. [This was in accordance with the requirement of the institutional ethics committee].

### Availability of data and materials

Raw data used for haplotype analysis and mutational age estimation is provided as supplementary files. Furthermore, raw sequence files generated from smMIP based targeted sequencing can be accessed from the European Nucleotide Archive (Project Accession: PRJEB51874).

### Competing interests

The authors declare that they have no conflict of interests.

### Funding

This work was partially supported by the Indian Council of Medical Research grant (BMS:54/2010) and the Gujarat State Biotechnology Mission grant (GSBTM/JDR&D/608/2020/459-461). The funding agencies were not directly or indirectly involved in the study design, specimen collection, analysis, interpretation and preparation of the manuscript.

### Author's contributions

HS, JS and FS conceived and planned experiments. PN, MS and JS carried out recruitment of cases and controls. RB and AN carried-out experiments and data analysis. HS, AN and JS were involved in statistical analysis and interpretation. HS, AN and FS wrote the first draft of the manuscript. All authors proofread the manuscript.

### Acknowledgements

We acknowledge Sudha Srinivasan for her valuable inputs in haplotype analyses. We are grateful to all the patients and their families for the support, as without their consent this study would not have been possible.

## References

1. Matalon R, Arbogast B, Justice P, Brandt IK, Dorfman A. Morquio's syndrome: deficiency of a chondroitin sulfate N-acetylhexosamine sulfate sulfatase. *Biochem Biophys Res Commun.* 1974;61:759–65.
2. Nakashima Y, Tomatsu S, Hori T, Fukuda S, Sukegawa K, Kondo N, et al. Mucopolysaccharidosis IV A: molecular cloning of the human N-acetylgalactosamine-6-sulfatase gene (GALNS) and analysis of the 5'-flanking region. *Genomics.* 1994;20:99–104.
3. Peracha H, Sawamoto K, Averill L, Kecskemethy H, Theroux M, Thacker M, et al. Molecular genetics and metabolism, special edition: Diagnosis, diagnosis and prognosis of Mucopolysaccharidosis IVA. *Mol Genet Metab.* 2018;125:18–37.
4. Hendriksz CJ, Harmatz P, Beck M, Jones S, Wood T, Lachman R, et al. Review of clinical presentation and diagnosis of mucopolysaccharidosis IVA. *Mol Genet Metab.* 2013;110:54–64.
5. Nelson J. Incidence of the mucopolysaccharidoses in Northern Ireland. *Hum Genet.* 1997;101:355–8.
6. Nelson J, Crowhurst J, Carey B, Greed L. Incidence of the mucopolysaccharidoses in Western Australia. *Am J Med Genet A.* 2003;123A:310–3.
7. Leong HY, Abdul Azize NA, Chew HB, Keng WT, Thong MK, Mohd Khalid MKN, et al. Clinical, biochemical and genetic profiles of patients with mucopolysaccharidosis type IVA (Morquio A syndrome) in Malaysia: the first national natural history cohort study. *Orphanet Journal of Rare Diseases.* 2019;14:143.
8. Morrone A, Tylee KL, Al-Sayed M, Brusius-Facchin AC, Caciotti A, Church HJ, et al. Molecular testing of 163 patients with Morquio A (Mucopolysaccharidosis IVA) identifies 39 novel GALNS mutations. *Mol Genet Metab.* 2014;112:160–70.
9. Tomatsu S, Fukuda S, Masue M, Sukegawa K, Fukao T, Yamagishi A, et al. Morquio disease: isolation, characterization and expression of full-length cDNA for human N-acetylgalactosamine-6-sulfate sulfatase. *Biochem Biophys Res Commun.* 1991;181:677–83.

10. Bidchol AM, Dalal A, Shah H, S S, Nampoothiri S, Kabra M, et al. GALNS mutations in Indian patients with mucopolysaccharidosis IVA. *Am J Med Genet A*. 2014;164A:2793–801.
11. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–24.
12. Ossa CA, Torres D. Founder and Recurrent Mutations in BRCA1 and BRCA2 Genes in Latin American Countries: State of the Art and Literature Review. *Oncologist*. 2016;21:832–9.
13. Marques AM, Ananina G, Costa VP, Vasconcellos JPC de, de Melo MB. Estimating the age of the p.Cys433Arg variant in the MYOC gene in patients with primary open-angle glaucoma. *PLOS ONE*. 2018;13:e0207409.
14. Sukegawa K, Nakamura H, Kato Z, Tomatsu S, Montañó AM, Fukao T, et al. Biochemical and structural analysis of missense mutations in N-acetylgalactosamine-6-sulfate sulfatase causing mucopolysaccharidosis IVA phenotypes. *Hum Mol Genet*. 2000;9:1283–90.
15. Tomatsu S, Fukuda S, Cooper A, Wraith JE, Rezvi GM, Yamagishi A, et al. Mucopolysaccharidosis IVA: identification of a common missense mutation I113F in the N-Acetylgalactosamine-6-sulfate sulfatase gene. *Am J Hum Genet*. 1995;57:556–63.
16. Tüysüz B, Alkaya DU, Toksoy G, Güneş N, Yıldırım T, Bayhan İA, et al. Mutation spectrum and pivotal features for differential diagnosis of Mucopolysaccharidosis IVA patients with severe and attenuated phenotype. *Gene*. 2019;704:59–67.
17. Diaz GA, Gelb BD, Risch N, Nygaard TG, Frisch A, Cohen IJ, et al. Gaucher Disease: The Origins of the Ashkenazi Jewish N370S and 84GG Acid  $\beta$ -Glucosidase Mutations. *The American Journal of Human Genetics*. 2000;66:1821–32.
18. Kim Y-M, Choi J-H, Kim G-H, Sohn YB, Ko JM, Lee BH, et al. The GBA p.G85E mutation in Korean patients with non-neuronopathic Gaucher disease: founder and neuroprotective effects. *Orphanet Journal of Rare Diseases*. 2020;15:318.
19. Mistri M, Tamhankar PM, Sheth F, Sanghavi D, Kondurkar P, Patil S, et al. Identification of novel mutations in HEXA gene in children affected with Tay Sachs disease from India. *PLoS One*. 2012;7:e39122.
20. de Jong JG, Wevers RA, Laarakkers C, Poorthuis BJ. Dimethylmethylene blue-based spectrophotometry of glycosaminoglycans in untreated urine: a rapid screening procedure for mucopolysaccharidoses. *Clin Chem*. 1989;35:1472–7.
21. Hopwood JJ, Harrison JR. High-resolution electrophoresis of urinary glycosaminoglycans: an improved screening test for the mucopolysaccharidoses. *Anal Biochem*. 1982;119:120–7.
22. van Diggelen OP, Zhao H, Kleijer WJ, Janse HC, Poorthuis BJ, van Pelt J, et al. A fluorimetric enzyme assay for the diagnosis of Morquio disease type A (MPS IV A). *Clin Chim Acta*. 1990;187:131–9.
23. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16:1215–1215.
24. Boyle EA, O’Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics*. 2014;30:2670–2.
25. Hiatt JB, Pritchard CC, Salipante SJ, O’Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res*. 2013;23:843–54.
26. O’Roak BJ, Vives L, Fu W, Egerton JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science*. 2012;338:1619–22.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
28. Reeve JP, Rannala B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics*. 2002;18:894–5.

## Figures

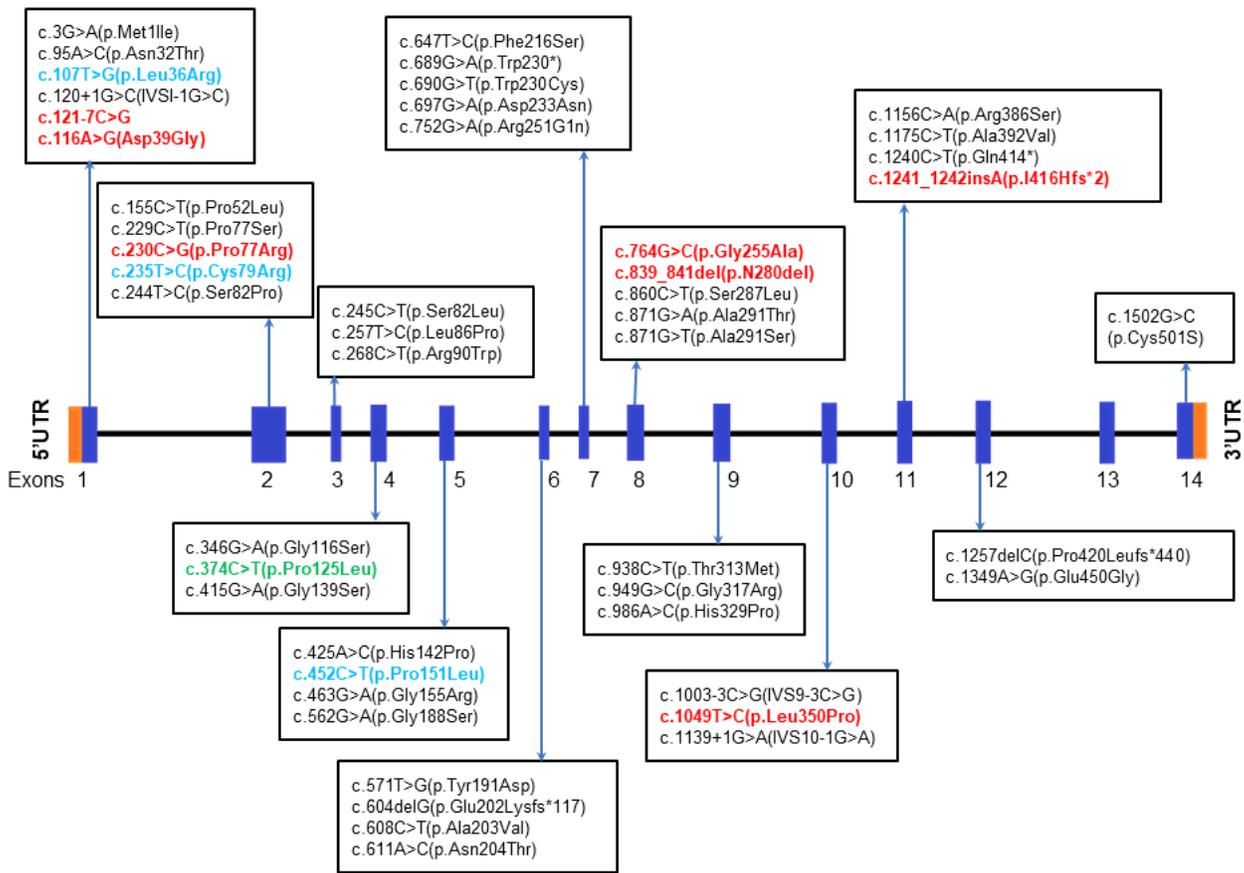
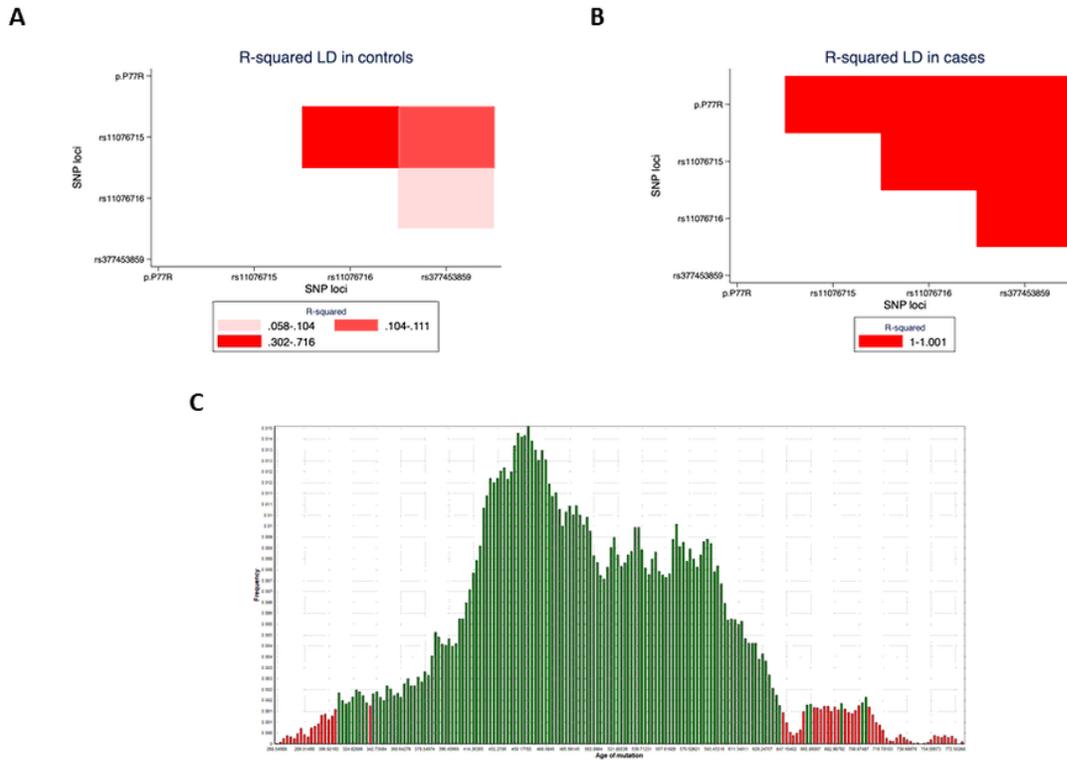


Figure 1

Schematic representation of the variants identified in the *GALNS* gene in patients with MPS IVA from India to date. Variants highlighted in red are observed in the present study whereas variants highlighted in blue are reported by Bidchol et al. 2014 and the present study. Variants highlighted in green has reported in the Chinese population and the present study.



**Figure 2**  
 Linkage disequilibrium, haplotype and age of variant analysis for p.P77R variant in the *GALNS* gene in Gujarati-Indian population. Linkage disequilibrium plot in (A) 63 controls (B) 13 cases with p.P77R variant in the *GALNS* gene. Red colour intensity represents strength of linkage disequilibrium (measured in  $R^2$ ) between 2 SNPs. (C) Predicted age of variant plot generated by the DMLE software. Green indicates values within 95% confidence intervals.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.doc](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryFile1.xlsx](#)
- [SupplementaryFile2.txt](#)