

# NGS based expanded carrier screening for genetic disorders in North Indian population reveals unexpected results – a pilot study.

**Kanika Singh**

Sir Ganga Ram Hospital

**Sunita Bijamia-Mahay** (✉ [bijamia@gmail.com](mailto:bijamia@gmail.com))

Sir Ganga Ram Hospital <https://orcid.org/0000-0002-5255-8212>

**Vedam Lakshmi Ramprasad**

MedGenome labs ltd.

**Ratna Dua Puri**

Sir Ganga Ram Hospital

**Sandhya Nair**

MedGenome labs ltd.

**Sheetal Sharda**

MedGenome Labs Ltd.

**Renu Saxena**

Sir Ganga Ram Hospital

**Sudha Kohli**

Sir Ganga Ram Hospital

**Samarth Kulshreshtha**

Sir Ganga Ram Hospital

**Kanwal Gujral**

Sir Ganga Ram Hospital

**Indrani Ganguli**

Sir Ganga Ram Hospital

**Ishwar Chander Verma**

Sir Ganga Ram Hospital

---

## Research article

**Keywords:** Carrier screening, Cystic fibrosis, Hearing loss, Pompe disease, Asian Indians.

**Posted Date:** October 23rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-23712/v7>

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

---

**Version of Record:** A version of this preprint was published on November 2nd, 2020. See the published version at <https://doi.org/10.1186/s12881-020-01153-4>.

## Abstract

**Background:** To determine the carrier frequency and pathogenic variants of common genetic disorders in the north Indian population by using next generation sequencing (NGS).

**Methods:** After pre-test counselling, 200 unrelated individuals (including 88 couples) were screened for pathogenic variants in 88 genes by NGS technology. The variants were classified as per American College of Medical Genetics criteria. Pathogenic and likely pathogenic variants were subjected to thorough literature-based curation in addition to the regular filters. Variants of unknown significance were not reported. Individuals were counselled explaining the implications of the results, and cascade screening was advised when necessary.

**Results:** Of the 200 participants, 52 (26 %) were found to be carrier of one or more disorders. Twelve individuals were identified to be carriers for congenital deafness, giving a carrier frequency of one in 17 for one of the four genes tested (*SLC26A4*, *GJB2*, *TMPRSS3* and *TMC1* in decreasing order). Nine individuals were observed to be carriers for cystic fibrosis, with a frequency of one in 22. Three individuals were detected to be carriers for Pompe disease (frequency one in 67). None of the 88 couples screened were found to be carriers for the same disorder. The pathogenic variants observed in many disorders (such as deafness, cystic fibrosis, Pompe disease, Canavan disease, primary hyperoxaluria, junctional epidermolysis bullosa, galactosemia, medium chain acyl CoA deficiency etc.) were different from those commonly observed in the West.

**Conclusion:** A higher carrier frequency for genetic deafness, cystic fibrosis and Pompe disease was unexpected, and contrary to the generally held view about their prevalence in Asian Indians. In spite of the small sample size, this study would suggest that population-based carrier screening panels for India would differ from those in the West, and need to be selected with due care. Testing should comprise the study of all the coding exons with its boundaries in the genes through NGS, as all the variants are not well characterized. Only study of entire coding regions in the genes will detect carriers with adequate efficiency, in order to reduce the burden of genetic disorders in India and other resource poor countries.

## Background

Birth defects, defined as abnormalities of structure and function present from birth, are progressively contributing to a greater proportion of fetal, neonatal, infant and childhood mortality in developing countries. This is due to the decline in infectious and nutritional causes due to extensive use of immunizations, control of diarrheal and respiratory infections, and improvements in health care [1]. In the West, 29.8 % of early mortality and 29.2 % of chronic problems, are due to birth defects [2]. The major difference in the type of disorders observed in developing and developed countries is a higher incidence of autosomal recessive single gene disorders due to consanguinity and endogamous marriages in the former [2,3]. WHO estimated that globally 206,000 deaths and about 7 % of all neonatal deaths are caused by birth defects [4]. In India, the Sample Registration System Survey during 2010-2013 showed the contribution of congenital malformations to childhood mortality, infant mortality and neonatal mortality as 4.4 %, 4.6 % and 4.0 % respectively [5]. In the tertiary care hospitals birth defects contributed from 4.2% to 13.4 % of perinatal mortality, making congenital malformations and genetic disorders as the third leading cause of neonatal mortality [6]. The burden of genetic disorders in India has been presented in a number of publications and their prevention through screening has been emphasized [3, 7].

Birth defects, including genetic disorders, can cause significant mortality, diminish productivity and quality of life and cause social stigmatization and economic burden especially in resource poor countries. Their prevention is therefore a priority in all countries.

Most countries carry out prevention by screening for infections during pregnancy by serology, chromosomal disorders by biochemical test and structural abnormalities by ultrasonography. However, population prevalence studies have shown that the number of single gene disorders is almost equal to or exceeds chromosomal disorders and congenital malformations combined [3, 7, 8, 9]. The cost of prevention through screening for single gene disorders is much less than the cost of treatment. For example, in Cyprus where thalassemia is common, it was shown that the cost of 8 weeks of prevention was equivalent to the cost of one week of treatment of the thalassemia population [10]. The Ministry of Health of Israel reported that the life time health care cost for persons with thalassemia vs the cost of national screening program was in a ratio of 4.22 to 1 [11]. The need to reduce the prevalence of genetic disorders in developing countries is greater now as the new treatments of genetic disorders are exorbitantly expensive and out of reach for these families [12,13]. Moreover most of this expenditure has to be covered by out of pocket expenses by the patients/parents themselves.

Screening for carriers of single gene disorders such as cystic fibrosis and Tay Sach disease has also been shown to be cost effective [14,15]. Beauchamp et al examined the clinical impact of a 176-condition expanded carrier screening and demonstrated its cost-effectiveness to reduce the burden of Mendelian disease as compared with minimal screening [16]. Zhanga et al considered the impact and cost-effectiveness of offering preventive population genomic screening for BRCA1/2, MLH1/MSH2 genes, cystic fibrosis, spinal muscular

atrophy and fragile X syndrome to all young adults (18-25 years) in a single-payer health-care system in Australia, and reported that it would be highly cost-effective, but ethical issues need to be considered [17].

The basic objective of carrier screening is to identify carriers and offer them reproductive options from choosing to marry someone who is not a carrier of the same disease (premarital screening) or prenatal diagnosis. In the event that both the husband and wife are carriers of the same disorder, preimplantation genetic diagnosis (after in vitro fertilisation) or prenatal diagnosis (during early stages of a naturally conceived pregnancy) can be carried out [18]. Screening only those families who have a previously affected child is very inefficient, as majority of affected children are born to couples with no previous family history. Similarly screening only those who have an a-priori increased risk of being a carrier based on their personal and family history or who are consanguineously married, or in couples who are opting for sperm or egg donation (Assisted Reproduction Technologies) would still be an inadequate strategy to identify the carriers of genetic disorders. It is best to screen all couples for the genetic disorders common in that population.

World-wide, carrier screening has evolved from an ancestry-based (e.g. in Jewish populations) to pan-ethnic testing, and from single gene disorders, such as cystic fibrosis or  $\alpha/\beta$ -thalassemia, by Sanger sequencing or hematologic techniques, to multiple disorders through Next Generation Sequencing (NGS) [19]. In the West, carrier screening by NGS was initially limited to targeted genotyping because most of the pathogenic variants in the Caucasian population had been characterized and the results were easier to interpret as the subjects were screened for known variants [19]. This approach is not suitable in resource poor countries as most of the pathogenic variants in different genes have not been characterized. However, screening later shifted to NGS of all coding exons of genes to identify carriers more efficiently. This is more suited in India and other resource-poor countries, identifying only the variants that are pathogenic or likely pathogenic and ignoring variants of uncertain significance.

Carrier screening studies for single gene disorders in India, as a service, have chiefly been carried out for  $\beta$ -thalassemia, based on hematologic technique [20]. Isolated studies for p.Phe508del in cystic fibrosis [21] and p.Trp24Ter in *GJB2* related hearing loss [22] and more recently *SMN1* deletion in SMA (Spinal muscular atrophy) [23] have been performed as research studies. The objectives of the present study were to determine the carrier frequency of variants in 88 genes expected to be common in Asian Indians and to identify the pathogenic or likely pathogenic variants.

## Methods

### Subjects

This study was carried out at Sir Ganga Ram Hospital, a tertiary care multispecialty facility, over a period of 22 months from October 2016 through June 2018. Institutional ethical clearance was obtained prior to commencing the study (Ethical clearance number EC/08/16/1066). The molecular analysis was performed at Medgenome Laboratories Ltd, Bangalore. Two hundred unrelated individual (n=101 male, n=99 female) between the age of 20-60 years, visiting the Medical Genetics and Obstetrics and/or Gynaecology out-patient clinic for various reasons unrelated to genetic disorders were enrolled, after pre-test counselling. Individuals known to be carriers of any genetic disease, or with history of a chronic medical disorder or familial genetic disorder were excluded from the study. The relevant history and clinical data of each individual was recorded on standard case record proformas (Supplementary file 1).

### Sample size

A sample size of 200 unrelated individuals was planned for enrolment in this pilot study.

### Statistical analysis

Descriptive analysis was done, and outcome reported as proportion of carriers upon total individuals tested (n/200). Confidence interval was calculated by Wilson method [24].

### Selection of gene panel

The selection of genes followed the Wilson and Jungner criteria [25]. Genes selected were those which cause high impact disorders that have significant effect on lifespan or reduce quality of life; or genes with moderate impact that do not reduce lifespan but impact quality of life; or disorders with significant socioeconomic burden for which couples would consider prenatal diagnosis. Limited literature available on the prevalence of various monogenic disorders in India was reviewed. The study by Ankala et al [26], summarised the prevalence of galactosemia in India 1:10,300, congenital adrenal hyperplasia (CAH) 1:2,600, phenylketonuria 1:18,300, and amino acid disorders 1:3,600. The prevalence of childhood hearing loss in India was estimated as 1:500 in one study in 2009 [27]. The true prevalence of cystic fibrosis in India is unknown but suspected to be high in a recent review done by Mandal et al [28], based on the increased citations in recent years.

Lazarin et al, [29] also observed a carrier frequency of 1:40 for cystic fibrosis in South Asian population, much higher than expected from data in India. The genetic register maintained about patients evaluated at our centre was analysed. Eighty eight genes [72 autosomal recessive (AR), 7 X-Linked (XL), 9 autosomal dominant (AD/AR) were selected for testing (Supplementary file 2). A smaller number of disorders were aimed at to develop a short but efficient panel that could be offered at a low cost. Two recessive disorders (cystic megalencephaly and calpainopathy) were included as they are common in an ethnic group (Agrawals) in North India [26]. Familial Hypercholesterolemia, though an autosomal dominant disorder was studied as it is life threatening and early treatment can save lives. The study was planned in coherence with American College of Medical Genetics (ACMG) position statement on prenatal/preconception expanded carrier screening [30]. Some common disorders were excluded either because these are not detectable by NGS technology with accuracy or the disorder can be screened easily by haematological tests. These included  $\beta$ -thalassemia, deletions in *SMN1* (survival motor neuron 1) causing SMA (spinal muscular atrophy), FXS (Fragile X syndrome), and DMD (Duchenne muscular dystrophy). Deletion study of CAH was excluded, although sequencing of the gene was performed. Large copy number variations in any of the 88 genes were also not included, in this sequence – based study.

### Pre and Post-test counselling

Prior to the testing, all individuals were counselled about the type of disorders being tested, the implications of being a carrier, the benefits of enrolment of the partner and voluntary nature of testing. Relevant personal, family and ethnic data were recorded. Subjects were clinically examined to rule out any chronic disorder. In post-test counselling the individuals were explained about carrier status and its implications, cascade screening of family members and residual risks remaining after the results (unscreened disorders, chromosomal disorders and indels). The study methodology is depicted in figure 1.

### Molecular and Bioinformatic analysis

DNA was extracted from blood using Qiagen kit, and targeted genes were captured by a custom kit. The libraries were sequenced to mean coverage of >80-100X on Illumina sequencing platform. The sequences obtained were aligned to human reference genome (GRCh37/hg19) using BWA program [31, 32] and analysed using Picard and GATK version 3.6 [33, 34]. Gene annotation of the variants was performed using VEP (Variant effect predictor) program against the Ensembl release 87 human gene model [35]. Clinically relevant pathogenic variants were annotated using published variants in the literature and a set of diseases databases – ClinVar [36], OMIM (Online Mendelian inheritance in man) [37], GWAS catalogue (Genome wide association study in man) [38], HGMD (Human gene pathogenic variant database) [39] and SwissVar [40].

Common variants were filtered based on allele frequency in 1000Genome Phase 3 [41], GnomAD [42], dbSNP147 [43], and an in house database of 100,000 exomes in Indian subjects (Medgenome). Non-synonymous variants effect was calculated using multiple algorithms such as PolyPhen-2 (polymorphism phenotyping v2) [44], SIFT (Sorts intolerant from tolerant) [45], Mutationtaster2 [46], Mutation Assessor [47] and LRT (Likelihood ratio test) [48]. Splicing prediction tools used were Mutationtaster2 [46], BDGP (Berkeley drosophila genome project) [49] and HSF (human splicing finder) [50]. Various filters applied to variants included Variant quality (pass), 1000 genomes MAF (<0.05), exonic and canonical splice site and read depth (>10x). Variants remaining after applying the listed filters were subjected to ACMG classification. Only those variants fulfilling the ACMG criteria for pathogenic and likely pathogenic were shortlisted [51]. The literature was reviewed for the filtered variants before assigning carrier status.

**Validation of NGS results:** All disease associated variants were manually inspected using IGV (integrative genomics viewer). It was observed that all variants had sequencing depth > 30. No strand biasness was observed. All variants were of good mapping quality. None of the variants were in highly repetitive regions. These were further validated using Sanger sequencing.

## Results

### Population demographics:

Of the 200 individuals enrolled, 61.5% belonged to the 31- 40 years age group (Table 1). Eighty eight percent had enrolled with their partner and none of them were consanguineously married. Maximum number of persons belonged to the northern states of India (Delhi (n=74), Punjab (n=50), Haryana (n=44), Uttar Pradesh (n=18), Himachal Pradesh (n=5), Jammu and Kashmir(n=5) and Rajasthan (n=4). Majority of the individuals identified themselves as Hindu Punjabi (20.5%). Detailed religious and ethnic characteristics of the subjects is listed in table 1. Some individuals (5%) could not be classified as they were either unsure of their caste and origin or were born of an inter-caste marriage.

### Table 1: Demographic characteristics of the individuals enrolled in the study

Parameter		No of individuals (n=200)	Percentage	
Age	20-30 yrs	61	30.5%	
	31-40 yrs	123	61.5%	
	41-50 yrs	13	6.5%	
	51-60 yrs	3	1.5%	
Sex	Male	101	50.5%	
	Female	99	49.5%	
Religion & Caste	Hindu	Punjabi	41	20.5 %
		Brahmin	30	15 %
		Agarwal	28	14 %
		Jat	18	9 %
		Punjabi	14	7 %
		Rajput	6	3 %
		Pahadi	6	3 %
		BrahminBengali	4	2 %
		Marwari	2	1 %
		Kashmiri Pandit	2	1 %
		Other	10	5 %
		Jain	18	9 %
	Sikh	12	6 %	
Muslim	9	4.5 %		

#### Carrier frequency:

Of the 200 participants, 52 (26 %) were found to be carrier of one or more disorders (Table 2). Congenital deafness as the most common disorder identified, with a carrier frequency of 1 in 17, for one of the four genes *SLC26A4* (solute carrier family 26, member 4), *GJB2* (gap junction beta 2 protein), *TMPRSS3* (transmembrane protease, serine 3) and *TMC1* (transmembrane channel like protein 1), in decreasing order. Cystic fibrosis was the second most commonly observed disorder with a carrier frequency of 1 in 22. Three subjects were detected to be carriers for Pompe disease (frequency 1 in 67) (Table 2).

**Table 2: Carrier frequency of the disorders screened**

S.no	Disease name (OMIM no.)	N (no. of carriers)/ 200 individuals	%	1 in _	Wilson 95% Confidence Interval	
					Lower %	Upper %
	Total no. of carrier individuals	52	26	3.84	19.9	31.9
1	Cystic fibrosis - <i>CFTR</i> (219700)	9	4.5	22.22	2.4	8.3
2	Deafness - <i>SCL26A4</i> (274600)	5	2.5	40.0	0.78	5
3	Deafness - <i>GJB2</i> (220290)	3	1.5	66.67	0.5	4.3
4	Deafness - <i>TMPRSS3</i> (601072)	3	1.5	66.67	0.5	4.3
5	GSD type II - <i>GAA</i> (232300)	3	1.5	66.67	0.5	4.3
6	Methyl malonicaciduria mut A - <i>MMAA</i> (251100)	2	1	100	0.27	3.6
7	AR polycystic kidney - <i>PKHD1</i> (263200)	2	1	100	0.27	3.6
8	Galactosemia - <i>GALT</i> (230400)	2	1	100	0.27	3.6
9	Smith Lemli Opitz syndrome - <i>DHCR7</i> (270400)	2	1	100	0.27	3.6
10	Albinism type II - <i>OCA2</i> (203200)	2	1	100	0.27	3.6
11	Megalencephalic leukoencephalopathy with cysts - <i>MLC1</i> (604004)	2	1	100	0.27	3.6
12	Gaucher disease - <i>GBA</i> (230800)	2	1	100	0.27	3.6
13	Phenylketonuria - <i>PAH</i> (261600)	2	1	100	0.27	3.6
14	Epidermolysis bullosa (Junctional) - <i>LAMC2</i> (226700, 226650)	2	1	100	0.27	3.6
15	Niemann Pick disease type C1 - <i>NPC1</i> (257220)	1	0.5	200	0.088	2.77
16	Deafness - <i>TMC1</i> (600974)	1	0.5	200	0.088	2.77
17	Biotinidase deficiency - <i>BTB</i> (253260)	1	0.5	200	0.088	2.77
18	Medium chain acyl CoA deficiency - <i>ACADM</i> (201450)	1	0.5	200	0.088	2.77
19	Limb girdle muscle dystrophy type 2A - <i>CAPN3</i> (253600)	1	0.5	200	0.088	2.77
20	Congenital adrenal hyperplasia - <i>CYP21A2</i> (201910)	1	0.5	200	0.088	2.77
21	Primary hyperoxaluria type 1 - <i>AGXT</i> (259900)	1	0.5	200	0.088	2.77
22	Argininosuccinic aciduria - <i>ASL</i> (207900)	1	0.5	200	0.088	2.77
23	Canavan disease - <i>ASPA</i> (271900)	1	0.5	200	0.088	2.77
24	Glutaric aciduria type 1 - <i>GCDH</i> (231670)	1	0.5	200	0.088	2.77
25	Krabbe disease - <i>GALC</i> (245200)	1	0.5	200	0.088	2.77
26	Congenital ichthyosis - <i>TGM1</i> (242300)	1	0.5	200	0.088	2.77
27	Metachromatic leukodystrophy - <i>ARSA</i> (250100)	1	0.5	200	0.088	2.77
28	Zellweger syndrome - <i>PEX1</i> (214100)	1	0.5	200	0.088	2.77
29	Epidermolysis bullosa dystrophica - <i>COL7A1</i> (226600)	1	0.5	200	0.088	2.77
30	Very long chain acyl CoA dehydrogenase deficiency - <i>ACADVL</i> (201475)	1	0.5	200	0.088	2.77

P - Pathogenic, LP - Likely pathogenic

There was no couple where both husband and wife were carriers for the same disorder. No woman was found to be a carrier for the seven X-linked disorders included in the panel (Fabry disease, Ornithine transcarbamylase deficiency, hemophilia A and B, Hunter syndrome, severe combined immunodeficiency (SCID) and adrenoleukodystrophy).

Of the 52 (26%) subjects found to be carriers, majority were carriers for one disorder (n=47/200=23.5%) and five for two disorders (n=5/200=2.5%). No individual was found to be a carrier for three or more disorders.

#### Disease causing variants

The disease-causing variants were identified 57 times in 52 of 200 subjects (Table 3). Number of variants were 47, as some variants were identified more than once. Majority were of missense type (72.34%). Among the already reported variants, 29.5% (n=13/44) have been described in patients belonging to the Indian subcontinent (India, Pakistan, Bangladesh). The individual variants are listed in Table 3 & 4

and discussed in more detail later. Three splice site variants were novel (not reported in the literature or locus specific databases) and fulfilled ACMG criteria for pathogenicity (Table 4).

**Table 3: Pathogenic and likely pathogenic variants observed in the three commonest disorders.**

S. no	Disorder and Gene	Transcript no.	Number of Variants	Variant description: cDNA position Protein change	ACMG criteria	No of individuals
1	Deafness, AR 4, with enlarged vestibular aqueduct, <i>SLC26A4</i>	ENST00000265715	4	c.1001G>T, p.Gly334Val	PS3+PM2+PP2+PP3+PP4+PP5	2
				c.1226G>C, p.Arg409Pro <sup>#</sup>	PM2+PM5+PP2+PP3+PP5	1
				c.1468A>C, p.Ile490Leu <sup>#</sup>	PS1+PM1+PP2+PP3+PP5	1
				c.1003T>C, p.Phe335Leu	PS1+PP2+PP3+PP5	1
2	Deafness, AR,1A, <i>GJB2</i>	ENST00000382844	2	c.231G>A, p.Trp77Ter <sup>#</sup>	PVS1+PS3+PM1+PM4+PP2+PP3	2
				c.71G>A, p.Trp24Ter <sup>#</sup>	PVS1+PS3+PM1+PM4+PP2+PP3	1
3	Deafness, AR, 8, <i>TMPRSS3</i>	ENST00000291532	2	c.413C>A, p.Ala138Glu	PM1+PM2+PP2+PP3	1
				c.323-6G>A <sup>#</sup>	PS3+PS4+PM2+PP3	2
4	Deafness AR,7, <i>TMC1</i>	ENST00000297784	1	c.1165C>T, p.Arg389Ter	PVS1+PS3+PM2+PM4+PP3	1
5	Cystic fibrosis, <i>CFTR</i>	ENST00000003084	9	c.223C>T, p.Arg75Ter	PVS1+PS3+PM2+PM4+PP2+PP3	1
				c.1646G>A, p.Ser549Asn <sup>#</sup>	PS3+PM1+PM2+PM5+PP2+PP3+PP5	1
				c.595C>T, p.His199Tyr	PS3+PM1+PM2+PM5+PP2+PP3+PP5	1
				c.3209G>A, p.Arg1070Gln	PS3+PM1+PM5+PP2+PP3+PP5	1
				p.Phe508del	PS3+PM1+PM4+PP2+PP3	1
				c.4096A>T, p.Ile1366Phe	PM1+PM2+PP2+PP3	1
				c.1472G>T, p.Cys491Phe	PM1+PM2+PP2+PP3	1
				c.4009T>G, p.Phe1337Val	PM1+PM2+PP2+PP3	1
				c.1859A>T, p.His620Leu	PM1+PM2+PM5+PP2+PP3	1
6	Glycogen storage disease II, <i>GAA</i>	ENST00000302262	1	c.1933G>A, p.Asp645Asn	PS4+PM1+PM2+PM5+PP2+PP3+PP5	3

PVS - pathogenic very strong, PS - pathogenic strong, PM- pathogenic moderate, PP - pathogenic supporting [51] (Supplementary file 3), # - described from the Indian subcontinent.

**Table 4: Pathogenic and likely pathogenic variants in other genes**

PVS - pathogenic very strong, PS - pathogenic strong, PM- pathogenic moderate, PP - pathogenic supporting [51] (Supplementary file 3), \* - Novel variant, # - described from the Indian subcontinent.

## Discussion

The study was designed to determine the carrier frequency of single gene disorders other than  $\beta$ -thalassemia for which has a carrier frequency of about 3- 4 % has been shown in many studies in India [52]. Disorders such as spinal muscular atrophy (SMA), fragile X

S. no	Disorder and Gene	Transcript	Number of variants	Variant description: cDNA position Protein change	ACMG criteria	No of individuals
1	Methylmalonic aciduria, cbla type, <i>MMAA</i>	ENST00000281317	1	c.433C>T p.Arg145Ter	PVS1+PS3+PM4+PP2+PP3	2
2	Epidermolysis bullosa dystrophica, AR, <i>COL7A1</i>	ENST00000328333	1	c.5287C>T p.Arg1763Ter	PVS1+PM1+PM4+PP2	1
3	Galactosemia, <i>GALT</i>	ENST00000378842	1	c.563A>G p.Gln188Arg	PS3+PM1+PP2+PP3+PP4+PP5	2
4	Smith-Lemli-Opitz syndrome, <i>DHCR7</i>	ENST00000355527	2	c.730G>A p.Gly244Arg	PS1+PM2+PP2+PP3+PP4+PP5	1
5				c.862G>A p.Glu288Lys	PM2+PP1+PP2+PP3+PP4+PP5	1
6	Biotinidase deficiency, <i>BTD</i>	ENST00000303498	1	c.469C>T p.Arg157Cys	PS3+PM1+PM2+PM5+PP2+PP3	1
7	Medium chain ACYL-CoA dehydrogenase deficiency, <i>ACADM</i>	ENST00000420607	1	c.811G>G/A p.Gly271Arg	PS1+PS3+PP2+PP3+PP5	1
8	Junctional Epidermolysis bullosa, <i>LAMB3</i>	ENST00000391911	1	c.2138-2A>G*	PVS1+PM2+PP3	2
9	Metachromatic leukodystrophy, <i>ARSA</i>	ENST00000216124	1	c.1210+1G>T*	PVS1+PM2+PP3+PP5	1
10	Peroxisome biogenesis disorder 1A (Zellweger), <i>PEX1</i>	ENST00000248633	1	c.2926+2T>C	PVS1+PM2+PP3+PP5	1
11	Polycystic kidney disease, AR, <i>PKHD1</i>	ENST00000371117	2	c.8441-1G>C*	PVS1+PM2+PP3	1
12				c.1480C>T, p.Arg494Ter	PVS1+PS3+PM2+PM4+PP2	1
13	Albinism type 2, <i>OCA2</i>	ENST00000354638	1	c.1580T>G p.Leu527Arg <sup>#</sup>	PM2+PP2+PP3+PP4+PP5	2
14	Phenylketonuria, <i>PAH</i>	ENST00000553106	1	c.688G>A p.Val230Ile	PS4+PM1+PP2+PP3+PP5	2
15	Limb Girdle muscle dystrophy type 2A, <i>CAPN3</i>	ENST00000397163	1	c.1504A>G p.Ile502Val	PM1+PM2+PM5+PP2+PP3	1
16	Congenital adrenal hyperplasia, <i>CYP21A2</i>	ENST00000418967	1	c.373C>T p.Arg125Cys	PM2+PP2+PP3+PP4+PP5	1
17	Megalencephalic leukoencephalopathy with subcortical cysts, <i>MLC1</i>	ENST00000311597	2	c.65G>A p.Arg22Gln	PS3+PP2+PP3+PP4	1
18				c.959C>A p.Thr320Lys	PS3+PM2+PP2+PP3+PP5	1
19	Primary hyperoxaluria type 1, <i>AGXT</i>	ENST00000307503	1	c.302T>C p.Leu101Pro <sup>#</sup>	PS3+PM2+PP2+PP3	1

20	Gaucher disease, <i>GBA</i>	ENST00000327247	2	c.866G>C p.Gly289Ala <sup>#</sup>	PM2+PP2+PP3+PP4+PP5	1
21				c.1448T>C p.Leu483Pro <sup>#</sup>	PS3+PM5+PP2+PP3+PP5	1
22	Argininosuccinic aciduria, <i>ASL</i>	ENST00000304874	1	c.857A>G p.Gln286Arg	PS4+PP2+PP3+PP5	1
23	Canavan disease, <i>ASPA</i>	ENST00000263080	1	c.902T>C p.Leu301Pro <sup>#</sup>	PM2+PP2+PP3+PP4+PP5	1
24	Glutaric acidemia type 1, <i>GCDH</i>	ENST00000222214	1	c.281G>A p.Arg94Gln <sup>#</sup>	PM2+PM5+PP2+PP3+PP4+PP5	1
25	Krabbe disease, <i>GALC</i>	ENST00000261304	1	c.956A>G p.Tyr319Cys <sup>#</sup>	PM1+PM2+PM5+PP2+PP3	1
26	Congenital ichthyosis, AR 1, <i>TGM1</i>	ENST00000206765	1	c.550C>T p.Pro184Ser	PM2+PM3+PP2+PP3+PP5	1
27	Very Long chain fatty acid acyl-CoA dehydrogenase deficiency, <i>ACADVL</i>	ENST00000543245	1	c.1480T>C p.Phe494Leu	PM1+PM2+PP2+PP3	1
28	Niemann Pick disease type C1, <i>NPCI</i>	ENST00000269228	1	c.3560C>T p.Ala1187Val	PM5+PP2+PP3+PP4+PP5	1

syndrome (FXS) and Duchenne muscular dystrophy (DMD) are common in all populations including South Asians were also excluded as these are difficult to detect with NGS [53]. Recently, we showed the carrier frequency of SMA in North India to be 2.25 % [23]. However, carrier frequency for other single gene recessive disorders is not known and significant differences in prevalence and pathogenic variants have been seen in different populations [54].

#### ***CFTR* (Cystic fibrosis transmembrane conductance regulator) pathogenic and likely pathogenic variants**

There were nine disease-causing variants identified in the *CFTR* gene in this cohort. Of these, only one case had the common p.Phe508del pathogenic variant i.e. 11% (n=1/9). Two pathogenic variants detected in *CFTR* gene in this study have been observed before in our laboratory (p.Arg75Ter and p.Ser549Asn). The remaining six pathogenic variants have not been reported in Indians earlier (Table 3). The variants p.Ser549Asn, p.His199Tyr, p.Arg1070Gln have been described by multiple authors and functional studies have been carried out classifying them as pathogenic as per ACMG criteria. The other four variants p.Ile1366Phe, p.Cys491Phe, p.Phe1337Val, p.His620Leu have been documented to be associated with disease, however lack adequate functional studies. They meet the ACMG criteria for likely pathogenicity (Table 3).

*CFTR* c.3854C>T, p.Ala1285Val variant was identified in three individuals, which though has been reported in the literature [55] associated with congenital bilateral absence of vas deferens (CBAVD), is more likely to represent a common polymorphism due to its observance in high frequency in the NGS data in the Indian population (0.5% minor allele frequency in South Asians in GnomAD exomes). This variant was not included in the list of disease associated variants.

Studies on the genetic profile of cystic fibrosis patients in India shows high variability, and many rare and new variants have been observed, while only few pathogenic variants (p.Arg1162Ter, p.Met1Thr, c.1161delC, p.Ser549Asp and c.1525-1G>A) are reported more than once [56, 57, 58]. This suggests the lack of founder or common mutations in *CFTR* gene and thus emphasises the need for sequencing of all coding regions of the *CFTR* gene in suspected cases in the Indian population. In the present study except for p.Phe508del no other pathogenic variant was present in the ACMG panel of cystic fibrosis [59]. In view of the heterogeneity in pathogenic variants, Mandal et al. also suggested that a single panel of pathogenic variants cannot be used for diagnosis or carrier testing of CF in India [28]. Archibald et al. also observed that the pathogenic variants in cystic fibrosis vary according to ethnic origin [53]. Lim et al. reported in ExAC database that the pathogenic variants in the *CFTR* in non-Europeans are different from those in people of European descent. They noted that none of the

current genetic screening panels or existing *CFTR* pathogenic variant databases cover a majority of deleterious variants in any geographical region outside of Europe [60].

Among the nine disease causing variants identified in the *CFTR* gene in the present cohort, only one case had the common p.Phe508del pathogenic variant i.e. 11% (n=1/9). Kapoor and Kabra et al. studied cord blood samples of 955 newborns and reported a p.Phe508del carrier frequency of one in 238 (0.42%) [21]. They estimated the frequency of homozygous p.Phe508del as 1/228,006. However, this cannot be considered to represent the true prevalence of cystic fibrosis in India as it took into account only one pathogenic variant. Comparison of p.Phe508del allele frequency with that reported from the West shows that Indians have a low frequency (19-44%) of the p.Phe508del pathogenic variant [61, 62, 63]. Cystic fibrosis was thought to be extremely rare in India. However, a growing number of publications in the last two decades have suggested a higher prevalence [28, 63, 64]. This indicates that CF may be much more common in the Indian population with majority of cases being missed or undiagnosed. *CFTR* related pathogenic variants may be rarely recognized in Indians in view of the different phenotypes (including cystic fibrosis and congenital absence of vas deferens), variable clinical severity and lack of availability of sweat testing, and absence of newborn screening.

#### ***GJB2* c.231G>A, p.Trp77Ter and c.71G>A, p.Trp24Ter**

Biallelic variants in the *GJB2* gene or deletion in the gene cause congenital nonprogressive mild to profound sensorineural hearing impairment. The pathogenic variants identified in *GJB2* represent have been previously reported in Indian subjects. Ram Shankar et al studied the pathogenic variants in *GJB2* gene in Indian patients with deafness and found p.Trp24Ter to be the most common pathogenic variant India [22]. In addition, they documented two other common pathogenic variants p.Trp77Ter and IVS1+1G>A. These differ from the common pathogenic variants identified in the Western (c.35delG) [65] and Japanese (c.235delC) and Korean (p.Val37Ile) populations [66, 67].

#### ***SLC26A4* related hearing loss**

Hearing loss due to *SLC26A4* has been reported as third most common cause of hearing loss in a study in a pan-ethnic population [68]. This occurs due to an enlarged vestibular aqueduct and temporal bone abnormalities which can be appreciated on imaging. In addition to hearing loss, these individuals may have euthyroid goitre (Pendred syndrome). In this study, two out of the four disease - causing variants reported have been previously described in individuals of Indian ethnic origin: p.Arg409Pro [69,70] and p.Ile490Leu [71]. Other variants found in our study include p.Gly334Val, that has been described chiefly in people of Mediterranean origin [72] and p.Phe335Leu which is a common variant reported worldwide [73].

Carrier screening and prenatal diagnosis for a disorder like hearing loss which impairs quality of life can have differing perceptions among families in different countries. The parental perceptions in Indian culture where resources are scarce towards congenital hearing loss have been pointed out by Nahar et al. previously [74]. While some families are interested in using the information to help in the management, planning and emotional adjustment to the birth of a child with deafness others opt for discontinuing an affected fetus especially if financial resources are scarce.

#### ***GBA* c.1448T>C, and c.866G>C, p.Gly289Ala**

Biallelic variants in the *GBA* gene causing a deficiency of acid  $\beta$ -glucosidase and cause Gaucher disease, the most common lysosomal storage disorder in the world [75]. The variant p.Gly289Ala and p.Leu483Pro were observed in one individual in the present cohort. Ankleshwari et al. studied 33 Indian patients with Gaucher disease, and identified p.Leu483Pro as the most common pathogenic variant 60.6 % (n=20/33). In addition, they reported p.Gly289Ala as a novel pathogenic variant in a patient with type I disease [76]. Homozygosity for the p.Leu483Pro variant is associated with neuronopathic involvement (type III) ranging from mild oculomotor apraxia to more severe involvement as well as lethal cases of collodion skin baby phenotype [77,78]. The variant most commonly observed in Western population (p.Asn370Ser) and associated with type I Gaucher disease is observed less commonly in India [77, 79].

#### ***GAA* c.1933G>A, p.Asp645Asn variant**

Biallelic pathogenic variants in the *GAA* gene cause deficiency of acid  $\alpha$ -glucosidase resulting in Pompe disease. We observed three individuals to be carriers for p.Asp645Asn variant in the *GAA* gene. This variant was reported for the first time in 1998 by Huie et al. and demonstrated low enzyme activity with this pathogenic variant in vitro and in vivo [80]. Subsequently this pathogenic variant has been reported in patients affected with infantile onset Pompe disease in several studies [81]. This variant lies in exon 14 of the gene, reported to be a hot spot for this gene [81]. However a study done on Indian ethnic patients reported no hot spots for this gene [82].

#### ***OCA2* c.1580T>G, p.Leu527Arg variant**

Oculocutaneous albinism type II (tyrosinase positive) is caused by biallelic pathogenic variants in the *OCA2* gene. These individuals acquire small amounts of pigment with age and tend to have less severe visual abnormalities. The p.Leu527Arg variant was observed in heterozygous in two individuals in our cohort. It was reported for the first time by Jowerek et al. in a Pakistani family with some pigmentation of hair [83]. They reported that this pathogenic variant lies in highly conserved residue of amino acids in the transmembrane 8 domain of the protein and segregated with affected member.

#### **AGXT c.302T>C, p.Leu101Pro variant**

Primary hyperoxaluria occurs due to deficiency of the liver peroxisomal enzyme alanine:glyoxylate-aminotransferase encoded by the *AGXT* gene. We observed one carrier (belonging to Punjabi community) for p.Leu101Pro variant in our cohort. This variant was reported for the first time by Williams et al [84], who demonstrated that the mutant gene protein had less than 1% of normal activity in vitro. Subsequently, Chanchlani et al. documented three patients with primary hyperoxaluria type 1 to have the p.Leu101Pro variant in homozygous state [85]. All the three patients belonged to north India or Pakistan. They suggested a possibility of this being a founder pathogenic variant in India although larger studies and haplotype analysis are required.

#### **ASPA c.902T>C, p.Leu301Pro**

The *ASPA* gene encodes for aspartoacylase enzyme, deficiency of which results in Canavan disease. One individual was found to be carrier for the p.Leu301Pro variant. This variant has been reported by our group previously in a patient of Indian ethnicity with classical Canavan disease and raised urine N-acetyl aspartate [86]. On the basis of the reported literature this variant has classified using ACMG criteria as likely pathogenic.

#### **ACADM c.811G>A, p.Gly271Arg**

Biallelic pathogenic variants in *ACADM* affect mitochondrial fatty acid  $\beta$ -oxidation due to deficiency of the enzyme medium-chain acyl-coenzyme A dehydrogenase. The p.Gly271Arg is a well reported pathogenic variant in the *ACADM* gene worldwide. It was observed in one individual in this study. The c.985A>G pathogenic variant commonly seen in the West, believed to be a founder pathogenic variant in Caucasians originating from an ancient Germanic tribe was not observed in the present cohort [87].

Disorders like AR polycystic kidney disease, methyl malonic acidemia, galactosemia, Smith-Lemli Opitz syndrome, oculocutaneous albinism type II, cystic megalencephalic leukoencephalopathy, phenylketonuria and junctional epidermolysis bullosa can be expected to be common in the Indian population as at least two cases were detected among the 200 individuals screened.

Other investigators and our group have identified a number of disorders with founder mutations among the Agarwal community [88, 89]. Carriers for only two of these were identified in the current panel of genes - calpainopathy and megalencephalic leukodystrophy with cysts. The mutations detected are not the common ones noted in the Agarwal community. However there were only 28 individuals in the cohort belonging to the Agarwal community and larger studies are indicated to determine their carrier frequency.

## **Conclusions**

Carrier screening has been widely offered in the preconception period in the West. However, in India it has mostly been offered to those who have a family history or consanguinity. With decreasing costs of NGS panels, carrier screening is being increasingly utilised in recent times in India. It is likely to become the method of choice to decrease the burden of genetic diseases in India, as treatment is not funded by state agencies and family's financial resources are scarce. This study also brings out the differences in common pathogenic variants between the West and in Asian Indians, an ethnically distinct population. The variant filtration and interpretation strategies in a healthy population are challenging and literature review is essential before assigning pathogenic status to a variant. With the availability of NGS based testing in India and growing amount of literature on Indian pathogenic variants and their representation in databases, the sensitivity of carrier screening is likely to improve. Targeted genotyping panels like the '23 pathogenic variant panel' developed by ACMG for cystic fibrosis are not suitable and will miss many carriers.

This study highlights the importance of an Indian database in improving the classification of variants. It is creditable that many genetic centres are pooling their data to develop such a database. The high carrier frequency of cystic fibrosis, if substantiated in larger population studies, would be sufficient ground to initiate new-born screening in the Indian population. One major limitation of this study is the small sample size, and a larger studies would be justified to serve as a valuable tool for reducing the burden of genetic disorders.

## **List Of Abbreviations**

ACMG – American College of Medical Genetics

AR – Autosomal recessive

AD – Autosomal Dominant

BDGP -Berkeley drosophila genome project

BWA – Burrow wheeler aligner

CAH – Congenital adrenal hyperplasia

CBAVD - congenital bilateral absence of vas deferens

DMD – Duchenne muscular dystrophy

ECS – Expanded carrier screening

FXS – Fragile X syndrome

GWAS catalogue - Genome wide association study in man

HGMD - Human gene pathogenic variant database

HSF - human splicing finder

IGV- integrative genomics viewer

LRT - Likelihood ratio test

NGS – Next generation sequencing

OMIM - Online mendelian inheritance in man

SCID - Severe combined immunodeficiency

SIFT - Sorts intolerant from tolerant

SMA – Spinal muscular atrophy

VEP - Variant effect predictor

VUS – Variants of uncertain significance

XL – X linked

## Declarations

Ethics Approval and Consent to participate - Ethical clearance was obtained by the Ethics committee, Sir Ganga Ram Hospital, New Delhi prior to commencing the study (Ethical clearance number EC/08/ 16/1066). All individuals were given pre-test counselling, informed about the disorders being tested and voluntary nature of participation. Written informed consent was obtained from all individuals prior to testing.

Consent for publication - Not applicable.

Availability of data and materials - The datasets generated and analysed during the current study are available in the European Variation Archive (EVA) repository, under the study browser at <https://www.ebi.ac.uk/eva>. The project number is PRJEB40310. The direct weblink to the data files is <https://www.ebi.ac.uk/ena/browser/view/PRJEB40310>

The direct weblink to the human reference genome (GRCh37/hg19) dataset used in this study is as follows:

[ftp://ftp.ensembl.org/pub/grch37/current/fasta/homo\\_sapiens/dna/](ftp://ftp.ensembl.org/pub/grch37/current/fasta/homo_sapiens/dna/)

Accession numbers are also listed in Table 3 and Table 4 and their respective datasets can be found in the Ensemble database.

The supplementary tables 2 and 3 provide the evidence of pathogenicity, and classification of pathogenic and likely pathogenic variants according to ACMG guidelines, 2015.

Competing interests – The authors declare that they have no competing interests.

Funding – The expense of NGS was funded by Medgenome Laboratories Pvt Ltd., Bangalore, India. Medgenome Laboratories Pvt Ltd., carried out the next generation sequencing and molecular analysis. The data was re-analysed at Sir Ganga Ram Hospital, New Delhi.

Author's contributions – ICV, SBM, RDP and VLR conceptualized and designed the details of the study. KS enrolled the subjects and counselled them as a part of a thesis dissertation under the guidance of ICV, SBM, RDP, SK1, KG and IG. VLR, SN and SS carried out the next generation sequencing and analysis of molecular data. SK2 re-analysed all the molecular data, formatted the files and uploaded them on public database. RS and SK1 interpreted the variants and evaluated their occurrence in house database. KS and SBM re-checked the individual annotated variant excel sheets, compiled all the results, carried out review of literature, and analysed the individual variants before post-test counselling of the individuals. KS, SBM and ICV drafted the manuscript and carried out the subsequent revisions. All authors read and approved the manuscript and agreed to be personally accountable for all the contributions in the manuscript.

Acknowledgements – To the patients for their kind participation.

## References

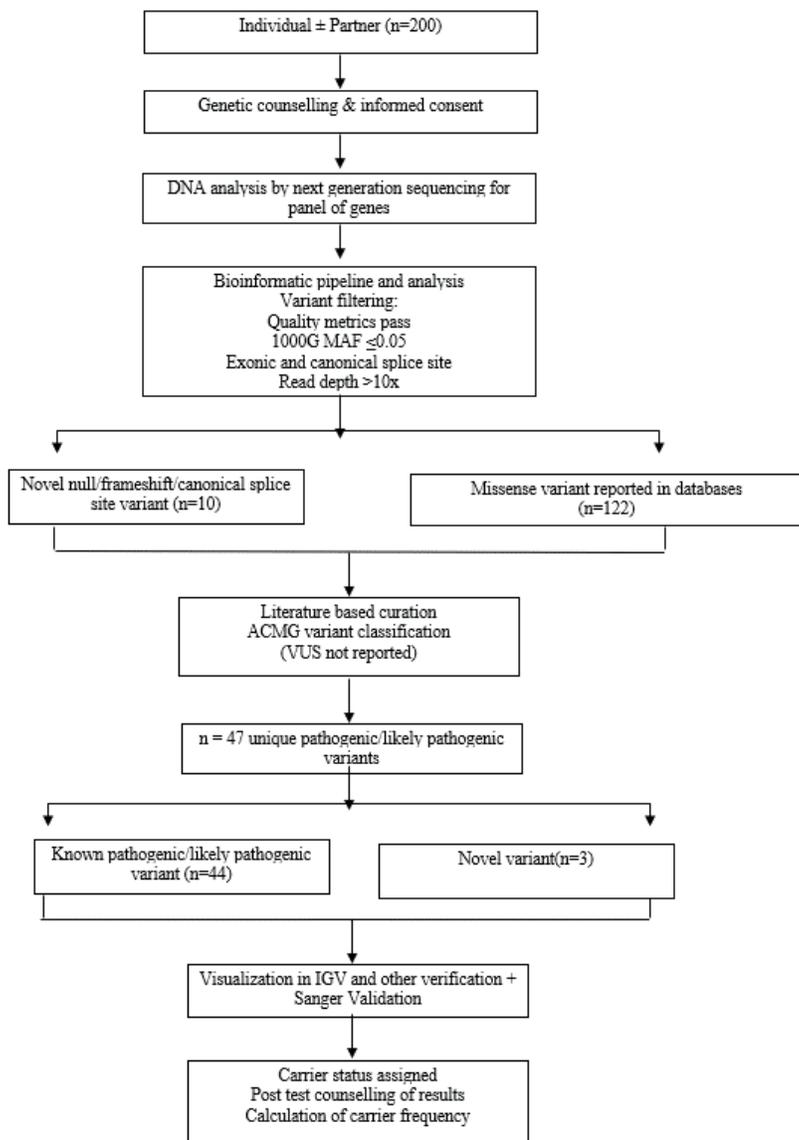
1. World Health Organization, Regional Office for South-East Asia. (2012) In: Capacity building for birth defects surveillance in South-East Asia: Report of the regional workshop on birth defects surveillance, Colombo, Sri Lanka. 2012. <https://apps.who.int/iris/handle/10665/204844>. Accessed 28 July, 2020.
2. Christianson A, Modell B. Medical genetics in developing countries. *Annu Rev Genomics Hum Genet.* 2004;5:219-65.
3. Verma IC, Puri RD. Global burden of genetic disease and the role of genetic screening. *Semin Fetal Neonatal Med.* 2015;20:354-63.
4. World Health Organization, Regional Office for South-East Asia. (2013). In: Prevention and control of birth defects in South-East Asia Region: strategic framework (2013-2017). <https://apps.who.int/iris/handle/10665/205644>. Accessed 28 July, 2020.
5. National Health Mission: Child Health. <https://nhm.gov.in/index1.php?lang=1&level=2&sublinkid=819&lid=219>. Accessed 28 July, 2020.
6. Swain S, Agrawal A, Bhatia BD. Congenital malformations at birth. *Indian Pediatrics.* 1994; 31:1187-1191.
7. Verma IC, Bijarnia S. The burden of genetic disorders in India and a framework for community control. *Community Genet.* 2002;5:192-196.
8. Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. *Am J Hum Genet* 1988;42:677-93.
9. Czeizel A, Sankaranarayanan K. The load of genetic and partially genetic disorders in man. I. Congenital anomalies: estimates of detriment in terms of years of life lost and years of impaired life. *Mutat Res.* 1984;128:73-103.
10. Angastiniotis M, Kyriakidou S, Hadjiminias M. How thalassemia was controlled in Cyprus. *World Health Forum.* 1986;7:291-7.
11. Ginsberg G, Tulchinsky T, Filon D, Goldfarb A, Abramov L, Rachmilevitz EA. Cost-benefit analysis of a national thalassaemia prevention programme in Israel. *J Med Screen.* 1998;5:120-6.
12. Sukumaran, A. Am I going to live? Genetic diseases leave 70 million Indians at God's mercy. In: Outlook Magazine. 2019. <https://www.outlookindia.com/magazine/story/india-news-am-i-going-to-live-genetic-diseases-leave-70-million-indians-at-gods-mercy/301809>. Accessed 26 July, 2020.
13. Adarsh A, Bhat D, Nataraj R. India Rare Disease Initiatives—A Review. *International Journal of Innovative Medicine and Health Science.* 2020,12:82-8.
14. Andrade E, Díaz J. Cost-effectiveness of the CFTR gene-sequencing test for asymptomatic carriers in the Colombian population. *CF Biomedica.* 2020;40:2table 183-95.
15. Warren E, Anderson R, Proos AL, Burnett LB, Barlow-Stewart K, Hall J. Cost-effectiveness of a school-based Tay-Sachs and cystic fibrosis genetic carrier screening program. *Genet Med.* 2005;7:484-94.
16. Beauchamp KA, Johansen Taber KA, Muzzey D. Clinical impact and cost-effectiveness of a 176-condition expanded carrier screen. *Genet Med.* 2019;21:1948-57.
17. Zhang L, Bao Y, Riaz M, Tiller J, Liew D, Zhuang X, et al. Population genomic screening of all young adults in a health-care system: a cost-effectiveness analysis. *Genet Med.* 2019;21:1958-68.

18. Bijarnia-Mahay S, Arora V, Saxena R, Verma IC. The changing scenario in prenatal diagnosis of genetic disorders: genetics to genomics. *Curr Med Res Pract* 2018;8:203–8.
19. Rose NC, Wick M. Carrier screening for single gene disorders. *Semin Fetal Neonatal Med*. 2018;23:78-84.
20. Madan N, Sharma S, Sood SK, Colah R, Bhatia LH. Frequency of  $\beta$ -thalassemia trait and other hemoglobinopathies in northern and western India. *Indian J Hum Genet*. 2010;16:16-25.
21. Kapoor V, Shastri SS, Kabra M, Kabra SK, Ramachandran V, Arora S et al. Carrier frequency of F508del mutation of cystic fibrosis in Indian population. *J Cyst Fibros*. 2006;5:43-6.
22. RamShankar M, Girirajan S, Dagan O, Ravi Shankar HM, Jalvi R, Rangasayee R, et al. Contribution of connexin26 (GJB2) mutations and founder effect to non-syndromic hearing loss in India. *J Med Genet*. 2003;40:e68.
23. Verma IC, Kohli S, Shenbagam S, Bijarnia-Mahay S, Puri RD, Saxena R. Carrier screening of spinal muscular atrophy in North Indian population and its public health implications. *Clin Genet*. 2020;98:198-9.
24. Wilson EB. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*. 1927;22:209-12.
25. Andermann A, Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the Genomic Age: A Review of Screening Criteria over the Past 40 Years. <http://www.who.int/bulletin/volumes/86/4/07-050112/en/>. Accessed 29 July,2020.
26. Ankala A, Tamhankar PM, Valencia CA, Rayam KK, Kumar MM, Hegde MR. Clinical applications and implications of common and founder mutations in Indian subpopulations. *Hum Mutat*. 2015;36:1-10.
27. Garg S, Chadha S, Malhotra S, Agarwal AK. Deafness: burden, prevention and control in India. *Natl Med J India*. 2009;22:79-81.
28. Mandal A, Kabra SK, Lodha R. Cystic fibrosis in India: Past, Present and Future. *J Pulm Med Respir Res* 2015;1:002.
29. Lazarin GA, Haque IS, Nazareth S, Lori K, Patterson AS, Jacobson JL, et al. An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals. *Genet Med*. 2013;15:178-86.
30. Grody WW, Thompson BH, Gregg AR, Bean LH, Monaghan KG, Schneider A, et al. ACMG position statement on prenatal/preconception expanded carrier screening. *Genet Med*. 2013;15(6):482-3.
31. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589-95.
32. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019;47(D1):D853-D858.
33. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297-303.
34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078-9. Accessed 5 May, 2018.
35. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;46(D1):D754-61. <https://asia.ensembl.org/index.html> Accessed 5 May, 2018.
36. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-7. <https://www.ncbi.nlm.nih.gov/clinvar/>. Accessed 1 August, 2020.
37. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 2005;33(Database issue):D514-7. <https://www.omim.org/>. Accessed 1 August, 2020.
38. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(Database issue):D1001-6. <https://www.ebi.ac.uk/gwas/>. Accessed 10 May, 2018.
39. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*. 2003;21(6):577-81. [hgmd.cf.ac.uk/](http://hgmd.cf.ac.uk/). Accessed 10 May, 2018.
40. Mottaz A, David FP, Veuthey AL, Yip YL. Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics*. 2010;26(6):851-2. <https://swissvar.expasy.org/>. Accessed 10 May, 2018.
41. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. [1000genomes.org/](http://1000genomes.org/). Accessed 10 May, 2018.
42. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-91. [broadinstitute.org/](http://broadinstitute.org/). Accessed 10 May, 2018.
43. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res*. 1999;9:677-9. <https://www.ncbi.nlm.nih.gov/projects/SNP/>. Accessed 13 May, 2018.

44. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248-9. <http://genetics.bwh.harvard.edu/pph2/>. Accessed 13 May, 2018.
45. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073-81. [http://sift.jcvi.org/www/SIFT\\_enst\\_submit.html](http://sift.jcvi.org/www/SIFT_enst_submit.html). Accessed 13 May, 2018.
46. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014;11:361-2. <http://www.mutationtaster.org/>. Accessed 15 May, 2018.
47. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011;39:e118. <http://www.org/>. Accessed 15 May, 2018.
48. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res*. 2009;19:1553-6.
49. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997;4:311-23. [http://fruitfly.org/seq\\_tools/splice.html](http://fruitfly.org/seq_tools/splice.html). Accessed 20 May, 2018.
50. Desmet FO, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009 May;37:e67. <http://umd.be/HSF3>. Accessed 20 May, 2018.
51. 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.
52. Colah R, Italia K, Gorakshakar A. Burden of thalassemia in India: the road map for control. *Pediatr Hemat Oncol J*. 2017; 2(4):79-84.
53. Archibald AD, Smith MJ, Burgess T, Scarff KL, Elliott J, Hunt CE et al. Reproductive genetic carrier screening for cystic fibrosis, fragile X syndrome, and spinal muscular atrophy in Australia: outcomes of 12,000 tests. *Genet Med*. 2018;20:513-23.
54. Woodcock IR, Fraser L, Norman P, Pysden K, Manning S, Childs AM. The prevalence of neuromuscular disease in the paediatric population in Yorkshire, UK; variation by ethnicity and deprivation status. *Dev Med Child Neurol*. 2016;58:877-83
55. Sachdeva K, Saxena R, Majumdar A, Chadha S, Verma IC. Mutation studies in the CFTR gene in Asian Indian Subjects with congenital bilateral absence of vas deferens: Report of two novel mutations and four novel variants. *Genetic Testing and Molecular Biomarkers*. 201;15:307-12.
56. Shastri SS, Kabra M, Kabra SK, Pandey RM, Menon PS. Characterisation of mutations and genotype-phenotype correlation in cystic fibrosis: experience from India. *J Cyst Fibros*. 2008;7:110-5.
57. Sharma N, Singh M, Kaur G, Thapa BR, Prasad R. Identification and characterization of CFTR gene mutations in Indian CF patients. *Ann Hum Genet*. 2009;73:26-33.
58. Watson MS, Cutting GR, Desnick RJ, Driscoll DA, Klinger K, Mennuti M, et al. Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet Med*. 2004;6:387-91.
59. Sachdeva K, Saxena R, Puri R, Bijarnia S, Kohli S, Verma IC. Mutation analysis of the CFTR gene in 225 children: identification of five novel severe and seven reported severe mutations. *Genet Test Mol Biomarkers*. 2012;16:798-801.
60. Lim RM, Silver AJ, Silver MJ, Borroto C, Spurrier B, Petrossian TC, et al. Targeted mutation screening panels expose systematic population bias in detection of cystic fibrosis risk. *Genet Med*. 2016;18:174-9.
61. Bowler IM, Estlin EJ, Littlewood JM. Cystic fibrosis in Asians. *Arch Dis Child*. 1993;68:120-2.
62. Powers CA, Potter EM, Wessel HU, Lloyd-Still JD. Cystic fibrosis in Asian Indians. *Arch Pediatr Adolesc Med*. 1996;150:554-5.
63. Kabra M, Kabra SK, Ghosh M, Khanna A, Arora S, Menon PS, et al. Is the spectrum of mutations in Indian patients with cystic fibrosis different? *Am J Med Genet*. 2000;93:161-3.
64. Mei-Zahav M, Durie P, Zielenski J, Solomon M, Tullis E, TSui LC, et al. The prevalence and clinical characteristics of cystic fibrosis in South Asian Canadian immigrants. *Arch Dis Child*. 2005;90:675-9.
65. Green GE, Scott DA, McDonald JM, Woodworth GG, Sheffield VC, Smith RJ. Carrier rates in the midwestern United States for GJB2 mutations causing inherited deafness. *JAMA*. 1999;281:2211-6.
66. Taniguchi M, Matsuo H, Shimizu S, Nakayama A, Suzuki K, Hamajima N, et al. Carrier frequency of the GJB2 mutations that cause hereditary hearing loss in the Japanese population. *J Hum Genet*. 2015;60:613-7.
67. Han SH, Park HJ, Kang EJ, Ryu JS, Lee A, Yang YH, et al. Carrier frequency of GJB2 (connexin-26) mutations causing inherited deafness in the Korean population. *J Hum Genet*. 2008;53:1022-8.
68. Sloan-Heggen CM, Bierer AO, Shearer AE, Kolbe DL, Nishimura CJ, Frees KL, et al. Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum Genet*. 2016;135:441–50.

69. Van Hauwe P, Everett LA, Coucke P, Scott DA, Kraft ML, Ris-Stalpers C, et al. Two frequent missense mutations in Pendred syndrome. *Hum Mol Genet.* 1998;7:1099-104.
70. Madden C, Halsted M, Meinzen-Derr J, Bardo D, Boston M, Arjmand E, et al. The influence of mutations in the SLC26A4 gene on the temporal bone in a population with enlarged vestibular aqueduct. *Arch Otolaryngol Head Neck Surg.* 2007;133:162-8.
71. Kühnen P, Turan S, Fröhler S, Güran T, Abali S, Biebermann H, et al. Identification of PENDRIN (SLC26A4) mutations in patients with congenital hypothyroidism and "apparent" thyroid dysgenesis. *J Clin Endocrinol Metab.* 2014;99:E169-76.
72. Dossena S, Nofziger C, Brownstein Z, Kanaan M, Avraham KB, Paulmichl M. Functional characterization of pendrin mutations found in the Israeli and Palestinian populations. *Cell Physiol Biochem.* 2011;28:477-84.
73. Pera A, Villamar M, Viñuela A, Gandía M, Medà C, Moreno F, et al. A mutational analysis of the SLC26A4 gene in Spanish hearing-impaired families provides new insights into the genetic causes of Pendred syndrome and DFNB4 hearing loss. *Eur J Hum Genet.* 2008;16:888-96.
74. Nahar R, Puri RD, Saxena R, Verma IC. Do parental perceptions and motivations towards genetic testing and prenatal diagnosis for deafness vary in different cultures? *Am J Med Genet A.* 2013;161A:76-81.
75. Puri RD, Kapoor S, Kishnani PS, Dalal A, Gupta N, Muranjan M et al. Diagnosis and Management of Gaucher Disease in India - Consensus Guidelines of the Gaucher Disease Task Force of the Society for Indian Academy of Medical Genetics and the Indian Academy of Pediatrics. *Indian Pediatr.* 2018;55:143-53.
76. Ankleshwaria C, Mistri M, Bavdekar A, Muranjan M, Dave U, Tamhankar P, et al. Novel mutations in the glucocerebrosidase gene of Indian patients with Gaucher disease. *J Hum Genet.* 2014;59:223-8.
77. Sheth J, Bhavsar R, Mistri M, Pancholi D, Bavdekar A, Dalal A, et al. Gaucher disease: single gene molecular characterization of one-hundred Indian patients reveals novel variants and the most prevalent mutation. *BMC Med Genet.* 2019;20:31.
78. Mishra R, Bijarnia-Mahay S, Sharma N, Dimri N, Saxena R, Puri RD et al. Gaucher Disease in Fetus: The Usual and the Unusual Presentations in a Family. *J. Fetal Med.* 2019;6:147–150.
79. Koprivica V, Stone DL, Park JK, Callahan M, Frisch A, Cohen IJ, et al. Analysis and classification of 304 mutant alleles with type 1 and type3 in Gaucher patients disease, *Am. J. Hum. Genet.* 2000; 66:1777–1786.
80. Huie ML, Tsujino S, Sklower Brooks S, Engel A, Elias E, Bonthron DT, et al. Glycogen storage disease type II: identification of four novel missense mutations (D645N, G648S, R672W, R672Q) and two insertions/deletions in the acid alpha-glucosidase locus of patients of differing phenotype. *Biochem Biophys Res Commun.* 1998;244:921-7.
81. Ngiwsara L, Wattanasirichaigoon D, Tim-Aroon T, Rojnuengnit K, Noojaroen S, Khongkraparn A, et al. Clinical course, mutations and its functional characteristics of infantile-onset Pompe disease in Thailand. *BMC Med Genet.* 2019;20:156.
82. Gupta N, Kazi ZB, Nampoothiri S, Jagdeesh S, Kabra M, Puri RD et al. Clinical and Molecular Disease Spectrum and Outcomes in Patients with Infantile-Onset Pompe Disease. *J Pediatr.* 2020;216:44-50.e5.
83. Jaworek TJ, Kausar T, Bell SM, Tariq N, Maqsood MI, Sohail A, et al. Molecular genetic studies and delineation of the oculocutaneous albinism phenotype in the Pakistani population. *Orphanet J Rare Dis.* 2012;7:44.
84. Williams EL, Acquaviva C, Amoroso A, Chevalier F, Coulter-Mackie M, Monico CG, et al. Primary hyperoxaluria type 1: update and additional mutation analysis of the AGXT gene. *Hum Mutat.* 2009;30:910-7.
85. Chanchlani R, Sinha A, Gulati A, Agarwal V, Bagga A. Common mutation underlying primary hyperoxaluria type1 in three Indian children. *Indian J Nephrol.* 2012;22:459-61.
86. Bijarnia S, Kohli S, Puri RD, Jacob RJ, Saxena R, Jalan A, et al. Molecular characterisation and prenatal diagnosis of Asparto-acylase deficiency (Canavan disease)-report of two novel and two known mutations from the Indian subcontinent. *Indian J Pediatr.* 2013;80:26-31.
87. Leal J, Ades AE, Wordsworth S, Dezateux C. Regional differences in the frequency of the c.985A>G ACADM mutation: findings from a meta-regression of genotyping and screening studies. *Clin Genet.* 2014;85:253-9.
88. Gorospe JR, Singhal BS, Kainu T, Wu F, Stephan D, Trent J, et al. Indian Agarwal megalencephalic leukodystrophy with cysts is caused by a common MLC1 mutation. *Neurology.* 2004 Mar 23;62:878-82.
89. Bijarnia-Mahay S, Movva S, Gupta N, Sharma D, Puri RD, Kotecha U, et al. Molecular Diagnosis of Hereditary Fructose Intolerance: Founder Mutation in a Community from India. *JIMD Rep.* 2015;19:85-93.

## Figures



**Figure 1**

Study flow chart

## Supplementary Files

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

- [Supplementaryfile1.docx](#)
- [Supplementaryfile2.docx](#)
- [Supplementaryfile3.docx](#)