Genetics in Medicine Open (2024) 2, 100840

ELSEVIER

Genetics in Medicine OPEN
An Official Journal of the ACMG

www.journals.elsevier.com/genetics-in-medicine-open

ARTICLE

Prevalence estimate of sphingosine phosphate lyase insufficiency syndrome in worldwide and select populations

Joni C. Sedillo[1,2], Chansonette Badduke[3], Steven J. Schrodi[1,2], Vinod Scaria[4], Onur Emre Onat[5], Majid Alfadhel[6,7], Carole Ober[8], William Wentworth-Sheilds[8], Robert D. Steiner[9], Julie D. Saba[10,*]

ARTICLE INFO

Article history:
Received 25 May 2023
Received in revised form
18 October 2023
Accepted 23 October 2023
Available online 30 October 2023

Keywords:
Adrenal insufficiency
Nephrotic syndrome
SGPL1
Sphingosine phosphate lyase insufficiency syndrome
Sphingosine-1-phosphate

ABSTRACT

Purpose: Sphingosine phosphate lyase insufficiency syndrome (SPLIS) is a rare, often fatal, metabolic disorder and monogenic form of steroid-resistant nephrotic syndrome. Other manifestations include primary adrenal insufficiency, ichthyosis, and neurological defects. SPLIS is caused by biallelic pathogenic variants in *SGPL1*, encoding sphingosine-1-phosphate lyase, a pyridoxal 5'-phosphate-dependent enzyme that catalyzes the final step of sphingolipid metabolism. Treatment is primarily supportive, but pyridoxine supplementation may be therapeutic in some cases, and gene therapy is being explored. We sought to determine the prevalence of SPLIS globally and among different populations to facilitate patient finding in anticipation of SPLIS clinical trials.

Methods: Using publicly available genomic data sets, including Genome Aggregation Database (gnomAD) v.2.1.1 and gnomAD v3.1.2, Iranome, IndiGen, and private genomic data sets from Israeli, Saudi, South Dakota Hutterite, and Turkish populations, we estimated SPLIS prevalence based on cumulative variant allele frequencies for high-confidence pathogenic variants. SPLIS prevalence estimates were adjusted by the level of inbreeding when the inbreeding coefficient was known. A Bayesian point estimate and 95% credible interval for worldwide SPLIS were calculated based on gnomAD v2.1.1 (GRCh37).

Results: The SPLIS prevalence estimate based on the total number of samples included from gnomAD v.2.1.1 (n = 141,430) was 0.015/100,000 (95% CI: 0.010 to 0.021). Using additional population data sets, we calculated SPLIS prevalence ranging from 0.046/100,000 to 0.078/100,000 in Turkish and Iranian populations, respectively.

Conclusion: The estimated worldwide number of SPLIS individuals is 11,707. Individuals with East Asian, Finnish, Turkish, and Iranian ancestries have an especially high estimated prevalence.

The Article Publishing Charge (APC) for this article was paid by Julie D. Saba.

Joni C. Sedillo and Chansonette Badduke contributed equally.

Current affiliation: Chansonette Badduke, PreventionGenetics, Marshfield, WI.

*Correspondence and requests for materials should be addressed to Julie D. Saba, 5700 Martin Luther King Jr. Way, Oakland, CA 94609. *Email address:* julie.saba@ucsf.edu

Affiliations are at the end of the document.

doi: https://doi.org/10.1016/j.gimo.2023.100840

Introduction

Sphingosine phosphate lyase insufficiency syndrome (SPLIS), also known as nephrotic syndrome type 14 (OMIM# 617575), is an ultra-rare autosomal recessive disorder of sphingolipid metabolism.[1] SPLIS is caused by inactivating pathogenic variants in *SGPL1* (OMIM#603729; HGNC:10817), which encodes sphingosine-1-phosphate (S1P) lyase, a pyridoxal 5'-phosphate dependent enzyme that catalyzes the final step of sphingolipid metabolism.[2] Loss of function (LOF) is a well-established mechanism of SPLIS based on animal models and experimental analysis of enzyme activity in patient fibroblasts.[3–7] For example, skin fibroblasts derived from children with SPLIS exhibit less than 10% of the S1P lyase activity observed in healthy control fibroblasts.[7] However, other mechanisms could potentially also be at play.

S1P lyase activity results in the irreversible cleavage of S1P, a bioactive sphingolipid metabolite with diverse functions in development and physiology.[8] By controlling intracellular S1P levels and generating extracellular S1P chemotactic gradients, S1P lyase plays a critical role in the regulation of lymphocyte trafficking and other cell migratory events, angiogenesis, pro-inflammatory and pro-fibrogenic cytokine signaling, calcium homeostasis, mitochondrial function, and epigenetic regulation of gene transcription.[9] Hexadecenal and ethanolamine phosphate, the 2 products generated by S1P lyase-mediated cleavage of the S1P substrate, are needed for autophagy, apoptosis, olfaction, and other functions. Additionally, by guarding the only exit point of sphingolipid metabolism, S1P lyase activity indirectly controls the levels of other upstream sphingolipid intermediates, including ceramides and sphingoid bases, which regulate cell-cycle progression and cell fate.

Children with SPLIS manifest a wide spectrum of clinical manifestations, including steroid-resistant nephrotic syndrome, adrenal insufficiency, male gonadal dysgenesis, hypothyroidism, peripheral neuropathy, ichthyosis, failure to thrive, and lymphopenia with or without severe functional immunodeficiency.[2] When nephrotic syndrome is present, it usually progresses rapidly to end-stage renal disease. Rarely, retinal, brain, and cardiac abnormalities have been described in SPLIS patients. The most severe cases present prenatally with hydrops fetalis, adrenal calcifications, and a uniformly fatal outcome. On the other end of the spectrum, some individuals present late in childhood with an isolated peripheral neuropathy/axonopathy.[10] The wide phenotypic spectrum likely reflects the differing amounts of residual enzyme activity among patients.

More than 2 dozen high-confidence pathogenic variants distributed across the length of the *SGPL1* gene, including most exons, have been reported in association with SPLIS.[1,7,10–22] To date, the most frequently reported variant is the c.665 G>A; p. Arg222Gln substitution, which is found in the homozygous state in about 25% of SPLIS cases.[7,11,12,20] Other missense variants, along with a small deletion, truncating variants, and canonical and non-canonical splice site variants, have also been reported in individuals with SPLIS (Supplemental Table 1). To our knowledge, no variants have been reported to occur de novo, and parents, when studied, were shown to be healthy heterozygotes. *SGPL1* genotype/phenotype correlations are not yet well understood, and major phenotypic differences even among SPLIS patients within the same family have been observed, indicating the existence of significant inherited and/or environmental disease modifiers.[23]

Unlike many sphingolipid degrading enzymes, which are found in the lysosome, S1P lyase is an integral membrane protein of the endoplasmic reticulum. Thus, SPLIS is representative of a growing list of atypical, non-lysosomal sphingolipid disorders that are being discovered through diagnostic next-generation sequencing.[24] Treatment for SPLIS includes renal transplantation, hormone replacement, and other supportive measures. Currently, there is no targeted therapy for SPLIS. However, pyridoxine cofactor supplementation and gene therapy show promise as potential strategies to increase S1P lyase activity and resolve the root cause of the condition.[6,7]

Less than 100 individuals with biallelic *SGPL1* pathogenic variants and 1 or more of the typical SPLIS features have been reported worldwide since the syndrome was first described in 2017.[11,12] Case clusters are observed in the Middle East, North Africa, Turkey, Pakistan, and the Hutterite colonies of North America. However, it is reasonable to assume that the most severe cases of SPLIS resulting in fetal loss, as well as the least severe cases with isolated peripheral neuropathy, never receive a genetic diagnosis. Thus, the total reported cases are likely a significant underrepresentation of the true prevalence of SPLIS. A reliable estimate of the prevalence of SPLIS in different geoancestral populations is a crucial step in guiding patient finding efforts in anticipation of conducting clinical trials.

The purpose of the current study was to provide prevalence estimates for SPLIS by determining high-confidence pathogenic *SGPL1* variant allele frequency rates in various populations and to use this information to calculate the expected prevalence of individuals with biallelic *SGPL1* pathogenic variants. Using both public and private population data sets, we have calculated the SPLIS prevalence in multiple populations and provide the first estimate of the number of SPLIS patients worldwide. Because additional pathogenic variants are likely present in populations, this estimate is expected to be an underestimate of the true number of SPLIS patients.

Abbreviations

LOVD – Leiden Open Variation Database
pLOF – predicted loss of function
S1P – sphingosine-1-phosphate
SPL – sphingosine phosphate lyase
SPLIS – sphingosine phosphate lyase insufficiency syndrome

Materials and Methods

General methodology

To estimate the prevalence of SPLIS, we searched for SPLIS cases in the literature and queried *SGPL1* (NM_003901.3) variants in 2 publicly available clinical databases (Leiden Open Variation Database [https://databases.lovd.nl/shared/variants/SGPL1/unique] and ClinVar [https://www.ncbi.nlm.nih.gov/clinvar/]).[25,26] We also searched Human Gene Mutation Database Professional, a curated database of disease-causing variants, and included variants from affected individuals shared with us via personal communications. The resulting list contained 46 variants: 1 small deletion, 16 truncating variants (frameshift and nonsense), 4 canonical and 1 non-canonical splice site variants, 22 missense variants, and 2 in-frame deletions of the same amino acid. We curated the list by applying modified American College of Medical Genetics and Genomics (ACMG)/Association for Molecular Pathology (AMP) codes and identified a subset (pathogenic/likely pathogenic) that we considered high-confidence disease-causing variants (Supplemental Table 1). We used Human Genome Variation Society nomenclature to describe the sequence variants and Mutalyzer (https://mutalyzer.nl/) to determine chromosomal positions in both GRCh37 and GRCh38. We then checked public (Genome Aggregation Database [gnomAD] v.2.1.1 and gnomAD v3.1.2, Iranome, and IndiGenomes) and private (Saudi, Turkish, Israeli, and South Dakota Hutterite) population data sets to determine the allele frequency of our high-confidence disease-causing variants (listed in Supplemental Table 1). In addition, because LOF is a well-established mechanism of disease, we included all predicted LOF (pLOF) variants found in each population data set.[27] Our pLOF list includes all exonic variants annotated as stop gained, frameshift, and splice acceptor sites. This list of variants requires validation for disrupted protein function because coding alterations that result in a functional protein would lead to an overestimate in our prevalence calculations.

Data sets

The publicly available gnomAD v2.1.1 data set (GRCh37/hg19) contains 125,748 exome sequences and 15,708 genome sequences, whereas the gnomAD v3.1.2 data set (GRCh38/hg38) contains 76,156 genomes from unrelated individuals around the world (https://gnomad.broadinstitute.org/).[27] The IndiGenomes resource encompasses the genomic data from more than 1000 genome sequences sequenced from across India and represents diverse geographies and ancestries (https://clingen.igib.res.in/indigen/).[28] The Iranome data set (http://www.iranome.ir/) contains exome sequences of 800 individuals, including 100 healthy individuals from 8 major Iranian ethnic groups: Arabs, Azeris, Balochs, Kurds, Lurs, Persians, Persian Gulf Islanders, and Turkmen.[29] Several additional private data sets (Turkish, Saudi, Israeli, and South Dakota Hutterite), described below, were also utilized. A recently reported Turkish variome, including 2589 exomes and 773 genomes from 3362 unrelated Turkish participants, was shared by (anonymized).[30] A data set of 1563 genomes and 1563 exomes from a mixed cohort of participants suspected of having a genetic disease and unaffected participants of Saudi ancestry was provided by (anonymized).[31] An Israeli data set comprising approximately 10,200 exomes, including 3250 of Ashkenazi Jewish origin and the remainder being non-Ashkenazi Jewish or Arab ancestries (https://hadassahinternational.org/database-at-hadassah-unlocks-the-door-to-disease-detection/), was shared by (anonymized). Genome sequencing for 121 genomes from South Dakota Hutterites was shared by (anonymized).

Variant curation

We compared the classifications from Human Gene Mutation Database (27 disease-causing variants), Leiden Open Variation Database (4 likely pathogenic and 1 uncertain significance), and ClinVar clinical submissions (7 pathogenic, 4 likely pathogenic, 3 uncertain significance, and 1 likely benign) for all the variants in our list (46 variants, Supplemental Table 1). We added pathogenicity calls from the automated ACMG/AMP algorithm, VarSome, scores from the REVEL in silico predictor, and used SpliceAI scores to predict pathogenicity for the non-canonical splice site variants (c.395A>G and c.1298+6T>C). Although some variants had consistent classifiers from multiple sources, several variants were not classified or had discrepant classifiers. We therefore decided to have 2 curators apply ACMG/AMP codes using available data from all sources to classify the variants in our list.[32] We modified the codes as follows: (1) we applied PS1 and PM3 for pathogenic/likely pathogenic variants, (2) we applied PP3 for REVEL scores >0.6 based on the scores of the 2 missense variants (p.Arg222Trp and p.Arg222Gln) at the most frequently reported amino acid in SPLIS patients, (3) we reviewed functional studies and applied PS3_supporting or moderate based on suggestions in Brnich et al (2019),[33] (4) we collapsed PP4 and PS4 and used PP4 to determine which cases to count for PS4, we applied PS4_supporting for (2-3), PS4_moderate (4-5), and PS4_strong (>5), (5) we used the ClinGen Sequence Variant Interpretation recommendation for in trans criterion Version 1.0 to apply PM3 (https://clinicalgenome.org/working-groups/sequence-variant-interpretation/), and (6) we did not apply PP5 or BP6 per recommendations from Biesecker, Harrison, and ClinGen Sequence Variant Interpretation Working Group (2018).[34] Each curator independently evaluated the evidence and applied ACMG codes; discrepancies between application of codes was resolved through discussion. Applied ACMG codes with modifications and final variant classifications are listed in Supplemental Table 1. Only variants that were classified as pathogenic or likely pathogenic in Supplemental Table 1 or pLOF variants from

population data sets were considered high-confidence disease-causing variants and were thus included in our calculations.

Prevalence calculations

Within each population, the frequencies of high-confidence pathogenic variants were combined assuming linkage equilibrium. To obtain the point estimate and 95% credible set for the birth prevalence of SPLIS (denoted by K), a Bayesian approach was applied to these data.[35] Given that pathogenic allele frequencies are rare, under linkage equilibrium the probability of sampling a pathogenic allele-free chromosome from the general population is $\prod_{i=1}^{m}(1-p_i) \cong 1-\sum_{i=1}^{m} p_i$; in which p_i is the frequency of the i^{th} pathogenic allele from a total of m pathogenic variants segregating in the population. Treating the allele frequencies as fixed, the birth prevalence calculation leverages Hardy-Weinberg equilibrium to obtain homozygous frequencies from the sum of allele frequencies, assuming that the penetrance values for each pathogenic variant is 1.0 in homozygous configuration:

$$K = \left[1 - \prod_{i=1}^{m}(1-p_i)\right]^2 \cong \left(\sum_{i=1}^{m} p_i\right)^2$$

To estimate the birth prevalence from population-based sequence data, we explicitly model the sampling properties of chromosomes from the general population and treat the allele frequencies as random variables within a Bayesian estimation procedure. The point estimate was calculated as the mathematical expectation of the posterior density of the frequency of individuals within the population with a homozygous combination of pathogenic variants at *SGPL1* (eqn 1).

$$\widehat{K} = E\left(q^2 \middle| x;2n\right) = \frac{(2n-1)!}{2(2n-x-1)!(x-1)!}\int_0^1 q^{\frac{x}{2}}\left(1-\sqrt{q}\right)^{2n-x-1} dq; \quad \text{(eqn 1)}$$

where q is the sum of the population-specific frequencies of the fully penetrant pathogenic *SGPL1* variants ($q = \sum_{i=1}^{m} p_i$), x is the number of chromosomes harboring a pathogenic variant, and n is the number of diploid samples. The 95% credible set was calculated through numerical integration of the posterior density such that the density mass was 2.5% below the lower value and 2.5% above the upper value.[35] Incorporation of population-specific estimated inbreeding coefficients to the point estimate was accomplished through the adjustment in (eqn 2) which inflates the point estimate of the birth prevalence (K_F):

$$\widehat{K_F} = E\left(q^2 \middle| x;2n\right) + \frac{1}{4n^2}[x(2n-x)]\widehat{F}; \quad \text{(eqn 2)}$$

where $\widehat{F}$ is an estimate of the population-specific inbreeding coefficient from published reports. Prevalence estimates and expected number of cases were calculated using the Bayesian point estimates and 95% credible sets along with estimated population size data for both the United States and globally.

Results

We searched exome and genome data from gnomAD v2.1.1 (GRCh37) for variants from our curated list and identified 31 of our high-confidence pathogenic variants within the *SGPL1* gene (Table 1). These variants occurred in 108 samples out of the 141,391 gnomAD genome and exome samples queried, with a resulting estimate of 0.015/100,000 (95% CI: 0.010/100,000 to 0.021/100,000) individuals at risk for SPLIS globally.

We performed the same calculations for variants from our curated list that are present in gnomAD v3.1.2 (GRCh38) because it includes individuals of Middle Eastern and Amish ancestry not included in gnomAD v2.1.1. Within this data set, 19 of our high-confidence pathogenic variants were identified (Table 1). These variants occurred in 51 samples out of a total 76,136 samples queried, resulting in an estimate of 0.011/100,000 (95% CI: 0.006/100,000 to 0.019/100,000) individuals at risk for SPLIS globally.

The expected number of individuals with SPLIS was calculated for both the United States and globally using Bayesian estimator results for ancestral groups and the 2021 US (United Stated Census Bureau 2022) and global population data.[36,37] Calculations were performed for both the gnomAD v2.1.1 (GRCh37) and gnomAD v3.1.2 (GRCh38), as v2.1.1 contains a much larger sample size, whereas v3.1.2 includes individuals of both Middle Eastern and Amish ancestry.[27] Expected number of SPLIS cases for each ancestral group for both the United States and globally are presented in Table 2. The expected total number of SPLIS cases within the United States is 71 (35 to 125) for gnomAD v2 and 76 (19 to 195) for gnomAD v3. Calculations using global population estimates predict a total of 11,707 (95% CI: 6288 to 19,699) cases of SPLIS using gnomAD v2.1.1 and 11,770 (95% CI: 2941 to 30,965) using gnomAD v3.1.2. It is important to note these are very rough estimates because the ancestral composition of the gnomAD samples may not accurately reflect the ancestral compositions of the US and global populations. These estimates also do not account for consanguinity, which may thereby underestimate the true prevalence of SPLIS. Many variants in the *SGPL1* gene were only recently identified and have very little information available, thus future studies are likely to identify pathogenic variants not included within this study. Because of the highly conservative nature of our variant curation, our values likely underestimate true SPLIS prevalence. Lastly, not all variants included in the calculations have been verified as pathogenic. Although our curated list thoroughly evaluated available data to identify high-

Table 1 Variants in the *SGPL1* gene identified in the gnomAD data set v2.1.1 (GRCh37) and v3.1.2 (GRCh38), the Iranome and IndiGenomes data sets, as well as private data sets (Turkish and Israeli)

SNP	v2	v3	Ir	In	T	Is	Position (GRCh37.p13 chr 10)	Position (GRCh38.p14 chr 10)	Transcript Consequence	Protein Consequence	Annotation	ACMG Classification	Source
rs138122500	1						NC_000010.10:g.72576635T>A	NC_000010.11:g.70816879T>A	c.26T>A	p.(Leu9Ter)	Nonsense	LP [a]	pLOF
rs1196497635	1						NC_000010.10:g.72604336G>A	NC_000010.11:g.70844579G>A	c.134G>A	p.(Trp45Ter)	Nonsense	LP	PC, CV
rs779596408	1						NC_000010.10:g.72604368G>T	NC_000010.11:g.70844611G>T	c.166G>T	p.(Gly56Ter)	Nonsense	LP [a]	pLOF
rs1326452400		1					NC_000010.10:g.72604384del	NC_000010.11:g.70844627del	c.184del	p.(Gln62SerfsTer31)	Frameshift	LP [a]	pLOF
rs1268871343	1						NC_000010.10:g.72614467del	NC_000010.11:g.70854711del	c.265del	p.(Gln89LysfsTer4)	Frameshift	LP [a]	pLOF
rs1383984196		1					NC_000010.10:g.72614500dup	NC_000010.11:g.70854741dup	c.297dup	p.(Ser100Ter)	Frameshift	LP [a]	pLOF
rs763301254	1						NC_000010.10:g.72614521del	NC_000010.11:g.70854767del	c.321del	p.(Val108TrpfsTer6)	Frameshift	LP [a]	pLOF
rs184929689	44	15		1	2		NC_000010.10:g.72619288T>G	NC_000010.11:g.70859531T>G	c.388T>G	p.(Leu130Ter)	Nonsense	LP [a]	pLOF
rs1564626153	1						NC_000010.10:g.72614598A>G	NC_000010.11:g.70854841A>G	c.395A>G	p.(Glu132Gly)	Missense	LP	Pub[11]
rs1450370338	1		1				NC_000010.10:g.72617368C>T	NC_000010.11:g.70857611C>T	c.410-3C>T	NA	Splice	LP [a]	pLOF
rs1282580893	1						NC_000010.10:g.72617369A>G	NC_000010.11:g.70857612A>G	c.410-2A>G	NA	Splice	LP [a]	pLOF
rs867880706	2	1					NC_000010.10:g.72619158dup	NC_000010.11:g.70859401dup	c.517dup	p.(Leu173Profs*55)	Frameshift	LP	PC
rs1460166813	1						NC_000010.10:g.72620150G>T	NC_000010.11:g.70860393G>T	c.565G>T	p.(Glu189Ter)	Nonsense	LP [a]	pLOF
rs1183269280	2						NC_000010.10:g.72620153del	NC_000010.11:g.70860397del	c.569del	p.(Ala190AspfsTer?)	Frameshift	LP [a]	pLOF
rs1168934353	1						NC_000010.10:g.72620159del	NC_000010.11:g.70860407del	c.584_588del	p.(Cys195PhefsTer2)	Frameshift	LP [a]	pLOF
	3	1					NC_000010.10:g.72619246C>T	NC_000010.11:g.70859489C>T	c.605C>T	p.(Ser202Leu)	Missense	LP	Pub[7,11,42]
rs771620665	1	1					NC_000010.10:g.72628110dup	NC_000010.11:g.70868353dup	c.629dup	p.(Thr211AsnfsTer17)	Frameshift	LP [a]	pLOF
rs1131692255	1						NC_000010.10:g.72628150C>T	NC_000010.11:g.70868393C>T	c.664C>T	p.(Arg222Trp)	Missense	LP	Pub[11], CV
rs769259446	2				1		NC_000010.10:g.72628151G>A	NC_000010.11:g.70868394G>A	c.665G>A	p.(Arg222Gln)	Missense	P	Pub[7,11,12,18,20,43–45], CV
rs775261141	24	10					NC_000010.10:g.72628188dup	NC_000010.11:g.70868431dup	c.704+7_704+10dup	NA	Frameshift	LP [a]	pLOF
rs370799576	2	1					NC_000010.10:g.72629548G>A	NC_000010.11:g.70869791G>A	c.705-1G>A	NA	Splice	LP	CV
		2					NC_000010.10:g.72630826del	NC_000010.11:g.70871069del	c.832del	p.(Arg278Glyfs*17)	Frameshift	P	Pub[11]
rs1131692235	3	2					NC_000010.10:g.72631618del	NC_000010.11:g.70871861del	c.934del	p.(Leu312Phefs*30)	Frameshift	P	Pub[13], CV, LOVD
rs1446015195		1					NC_000010.10:g.72631679del	NC_000010.11:g.70871919del	c.995del	p.(Pro332HisfsTer10)	Frameshift	LP [a]	pLOF
rs757042460	1				2		NC_000010.10:g.72631677C>G	NC_000010.11:g.70871920C>G	c.993C>G	p.(Tyr331Ter)	Nonsense	LP	Pub[7,43,46], CV
rs771565079	1	1					NC_000010.10:g.72633106A>T	NC_000010.11:g.70873349A>T	c.1060-2A>T	NA	Splice	LP [a]	pLOF
rs985654475	1						NC_000010.10:g.72633127G>T	NC_000010.11:g.70873370G>T	c.1079G>T	p.(Gly360Val)	Missense	LP	Pub[7,15,18,43], CV
rs1196613157		1					NC_000010.10:g.72633161C>G	NC_000010.11:g.70873404C>G	c.1113C>G	p.(Tyr371Ter)	Nonsense	LP [a]	pLOF

(continued)

Table 1 Continued

SNP	v2	v3	Ir	In	T	Is	Position (GRCh37.p13 chr 10)	Position (GRCh38.p14 chr 10)	Transcript Consequence	Protein Consequence	Annotation	ACMG Classification	Source
rs1233243952		1					NC_000010.10:g.72633235del	NC_000010.11:g.70873478del	c.1192_1222del	p.(Gly398Ter)	Frameshift	LP [a]	pLOF
rs779485098	2	3					NC_000010.10:g.72633295A>G	NC_000010.11:g.70873538A>G	c.1247A>G	p.(Tyr416Cys)	Missense	P	Pub[7,11], PC, CV
rs750709005	1						NC_000010.10:g.72635206C>A	NC_000010.11:g.70875449C>A	c.1346C>A	p.(Ser449Ter)	Nonsense	LP [a]	pLOF
rs371029518	1	3					NC_000010.10:g.72635244C>T	NC_000010.11:g.70875487C>T	c.1384C>T	p.(Arg462Ter)	Nonsense	LP [a]	pLOF
rs1453958211		1					NC_000010.10:g.72636297G>A	NC_000010.11:g.70876540G>A	c.1446-1G>A	NA	Splice	LP [a]	pLOF
rs374024951	3	2					NC_000010.10:g.72636335C>T	NC_000010.11:g.70876578C>T	c.1483C>T	p.(Arg495Ter)	Nonsense	LP	CV
rs746887949	1	3					NC_000010.10:g.72636365C>T	NC_000010.11:g.70876608C>T	c.1513C>T	p.(Arg505Ter)	Nonsense	P	Pub[13], CV, LOVD
rs777196613	1						NC_000010.10:g.72636988del	NC_000010.11:g.70877233del	c.1605del	p.(Asn536IlefsTer26)	Frameshift	LP [a]	pLOF
	1						NC_000010.10:g.72637020_72637022del	NC_000010.11:g.70877263_70877265del	c.1635_1637del	p.(Phe545del)	In-frame Deletion	LP	Pub[12,18]
	1						NC_000010.10:g.72637007del		c.1626_1632del	p.(Val544TrpfsTer16)	Frameshift	LP [a]	pLOF
VUS Variants Present in 1+ Datasets													
rs188194665	116	59	3		25	71	NC_000010.10:g.72604246A>G	NC_000010.11:g.70844489A>G	c.44A>G	p.(Tyr15Cys)	Missense	VUS	Pub[7,21,43], LOVD
	2						NC_000010.10:g.72619152A>G	NC_000010.11:g.70859395A>G	c.511A>G	p.(Asn171Asp)	Missense	VUS	Pub[47]
rs201533115	2	2					NC_000010.10:g.72619192T>C	NC_000010.11:g.70859435T>C	c.551T>C	p.(Ile184Thr)	Missense	VUS	Pub[10]
rs752537084		4					NC_000010.10:g.72631630G>A	NC_000010.11:g.70871873G>A	c.946G>A	p.(Ala316Thr)	Missense	VUS	Pub[11], CV
rs1437439236	2						NC_000010.10:g.72631702C>T	NC_000010.11:g.70871945C>T	c.1018C>T	p.(Arg340Trp)	Missense	VUS	Pub[16,18]
rs752722320	1						NC_000010.10:g.72633274T>G	NC_000010.11:g.70873517T>G	c.1226T>G	p.(Met409Arg)	Missense	VUS	Pub[21], LOVD
Variants From Our List Not Present in the Datasets Queried													
							NC_000010.10:g.72576610_72576636del	NC_000010.11:g.70816854_70816880del	c.1_27del		Start loss	P	Pub[48]
rs1131692254							NC_000010.10:g.72576616dup	NC_000010.11:g.70816860dup	c.7dup	p.(Ser3Lysfs*11)	Frameshift	P	Pub[11,12], CV
							NC_000010.10:g.72604289del	NC_000010.11:g.70844532del	c.87del	p.(Tyr29*)	Nonsense	LP	CV
							NC_000010.10:g.72604367G>A	NC_000010.11:g.70844610G>A	c.165G>A	p.(Trp55*)	Nonsense	LP	CV
rs1131692253							NC_000010.10:g.72610968G>A	NC_000010.11:g.70851211G>A	c.261+1G>A		Splice	P	Pub[12,19], CV
							NC_000010.10:g.72629547A>G	NC_000010.11:g.70869790A>G	c.705-2A>G		Splice	LP	PC

(continued)

Table 1 Continued

SNP	v2	v3	Ir	In	T	Is	Position (GRCh37.p13 chr 10)	Position (GRCh38.p14 chr 10)	Transcript Consequence	Protein Consequence	Annotation	ACMG Classification	Source
							NC_000010.10:g.72629559dup	NC_000010.11:g.70869802dup	c.715dup	p.(Gln239fs*8)	Frameshift	P	Pub[18]
rs1131682256							NC_000010.10:g.72631721G>T	NC_000010.11:g.70871964G>T	c.1037G>T	p.(Ser346Ile)	Missense	LP	Pub[11], CV
							NC_000010.10:g.72631733A>G	NC_000010.11:g.70871976A>G	c.1049A>G	p.(Asp350Gly)	Missense	LP	Pub[44]
							NC_000010.10:g.72633125del	NC_000010.11:g.70873368del	c.1077del	p.(Gly360Alafs*49)	Frameshift	LP	Pub[7]
							NC_000010.10:g.72633127G>A	NC_000010.11:g.70873370G>A	c.1079G>A	p.(Gly360Asp)	Missense	LP	LOVD
							NC_000010.10:g.72633130C>G	NC_000010.11:g.70873373C>G	c.1082C>G	p.(Ser361*)	Nonsense	LP	Pub[10]
							NC_000010.10:g.72633132T>A	NC_000010.11:g.70873375T>A	c.1084T>A	p.(Ser362Thr)	Missense	LP	PC
							NC_000010.10:g.72633281del	NC_000010.11:g.70873524del	c.1233del	p.(Phe411Leufs*56)	Frameshift	P	Pub[14]
							NC_000010.10:g.72633314_72633315del	NC_000010.11:g.70873557_70873558del	c.1266_1267del	p.(Gln422Hisfs*35)	Frameshift	LP	CV
							NC_000010.10:g.72636420T>C	NC_000010.11:g.70876663T>C	c.1566+2T>C		Splice	LP	Pub[7]
							NC_000010.10:g.72629563G>T	NC_000010.11:g.70869806G>T	c.719G>T	p.(Ser240Ile)	Missense	VUS	PC
							NC_000010.10:g.72630848G>A	NC_000010.11:g.70871091G>A	c.854G>A	p.(Cys285Tyr)	Missense	VUS	Pub[7]
							NC_000010.10:g.72630862T>C	NC_000010.11:g.70871105T>C	c.868T>C	p.(Phe290Leu)	Missense	VUS	Pub[7,43,46]
							NC_000010.10:g.72631616C>G	NC_000010.11:g.70871859C>G	c.932C>G	p.(Pro311Arg)	Missense	VUS	Pub[22]
							NC_000010.10:g.72631627G>T	NC_000010.11:g.70871870G>T	c.943G>T	p.(Asp315Tyr)	Missense	VUS	PC
							NC_000010.10:g.72631741A>G	NC_000010.11:g.70871984A>G	c.1057A>G	p. (Lys353Glu)	Missense	VUS	Pub[21]
							NC_000010.10:g.72631742A>G	NC_000010.11:g.70871985A>G	c.1058A>G	p.(Lys353Arg)	Missense	VUS	Pub[7]
							NC_000010.10:g.72633352T>C	NC_000010.11:g.70873595T>C	c.1298+6T>C		Splice	VUS	Pub[19]
rs1131692252							NC_000010.10:g.72637017del	NC_000010.11:g.70877260del	c.1632del	p.(Phe545del)	In-frame Deletion	VUS	CV

VUS variants present in 1+ dataset(s), as well as variants that are not present in the datasets are included for future research but were not used for calculations.

ACMG, American College of Medical Genetics and Genomics; *CV*, ClinVar; *In*, IndiGenomes; *Ir*, Iranome; *Is*, Israeli; *LOVD*, Leiden Open Variation Database; *LP*, likely pathogenic; *P*, pathogenic; *PC*, personal communication; *pLOF*, predicted loss of function; *Pub*, Publication; *v2*, GnomAD v2.1.1 (GRCh37); *v3*, GnomAD v3.1.2 (GRCh38); *T*, Turkish dataset; *VUS*, variants of unknown significance.

[a]Predicted loss of function.[27]

Table 2 Estimation of the expected number of cases of SPLIS within the United States and globally by ancestral group calculated using Bayesian point estimates and 95% credible sets for the gnomAD data set v2.1.1 (GRCh37) and v3.1.2 (GRCh38), the Iranome and IndiGenomes data sets, as well as private data sets (Turkish)

United States Expected SPLIS Cases				
	Population Size	Expected Affected	Credible Interval	Ref
Turkish	203,326	0	(0 to 0)	36
Iranome (Iranian)	465,254	0	(0 to 2)	36
IndiGenomes (Asian Indian)	3,303,512	2	(0 to 11)	36
GRCh37				
Ashkenazi Jewish	7,636,725	NA	(0 to 0)	36
Latino/Admixed American	62,753,963	1	(0 to 3)	36
African American	45,156,291	4	(1 to 12)	36
South Asian	6,308,599	0	(0 to 1)	36
East Asian	8,964,851	62	(34 to 101)	36 [a]
European	196,894,711	3	(1 to 7)	36
Other	4,316,410	0	(0 to 1)	
GRCh37 Total	332,031,554	71	(35 to 125)	
GRCh38				
Ashkenazi Jewish	7,636,725	NA	(0 to 0)	36
Latino/Admixed American	62,753,963	3	(0 to 14)	36
African American	45,156,291	2	(0 to 5)	36
South Asian	6,308,599	1	(0 to 4)	36
East Asian	8,964,851	60	(16 to 145)	36 [a]
European	196,894,711	10	(3 to 23)	36
Middle Eastern	3,320,315	NA	(0 to 0)	36
Other	996,094	0	(0 to 3)	
GRCh38 Total	332,031,554	76	(19 to 195)	
Global Expected SPLIS cases				
	Population Size	Expected Affected	Credible Interval	Ref
Turkish	83,800,000	39	(3 to 136)	37
Iranome (Iranian)	85,000,000	66	(0 to 451)	37
IndiGenomes (Asian Indian)	1,393,000,000	664	(0 to 4516)	37
GRCh37				
Ashkenazi Jewish	10,000,000	NA	(0 to 0)	37
Latino/Admixed American	718,753,964	7	(0 to 30)	36,37
African/African American	1,418,156,291	127	(18 to 388)	36,37
South Asian	1,969,000,000	88	(10 to 286)	37
East Asian	1,650,000,000	11,423	(6256 to 18,624)	37
European	744,000,000	11	(3 to 25)	37
Other	1,327,089,745	51	(0 to 346)	
GRCh37 Total	7,837,000,000	11,707	(6288 to 19,699)	
GRCh38				
Ashkenazi Jewish	10,000,000	NA	(0 to 0)	37
Latino/Admixed American	718,753,964	37	(1 to 160)	36,37
African/African American	1,418,156,291	59	(10 to 171)	36,37
South Asian	1,969,000,000	168	(0 to 1144)	37
East Asian	1,650,000,000	11,070	(2919 to 26,694)	37
European	744,000,000	39	(11 to 89)	37
Middle Eastern	454,400,000	NA	(0 to 0)	37
Other	872,689,745	398	(0 to 2706)	
GRCh38 Total	7,837,000,000	11,770	(2941 to 30,965)	

NA, not applicable; SPLIS, Sphingosine phosphate lyase insufficiency syndrome.
[a]Chinese, Japanese, Korean, Mongolian, and Taiwanese Americans.

confidence variants, further functional analysis is necessary to truly classify variants as pathogenic. We acknowledge that inclusion of any nonpathogenic variants could lead to an overestimate of SPLIS prevalence.

Bayesian estimates for each ancestral group included in both gnomAD v2.1.1 and gnomAD v3.1.2 are presented in Table 3, and global prevalence estimates are presented in Table 4. Most noteworthy, the Turkish population had an

Table 3 Bayesian point estimates and 95% credible sets calculated from the occurrence of potentially pathogenic variants in the *SGPL1* gene for the gnomAD data set v2.1.1 (GRCh37) and v3.1.2 (GRCh38), the Iranome and IndiGenomes data sets, as well as private data sets (Turkish)

Genome Reference Consortium Human Build	Ancestry or Dataset	Samples with Pathogenic Variant	Total Number of Samples	Bayesian Point Estimate	95% Credible Set Lower	95% Credible Set Upper
	Turkish	5	4024	4.63E-07	4.07E-08	1.62E-06
	Iranome (Iranian)	1	800	7.81E-07	2.51E-10	5.31E-06
	IndiGenomes (Asian Indian)	1	1024	4.77E-07	1.53E-10	3.24E-06
GRCh37	African/African American	7	12,480	8.99E-08	1.27E-08	2.74E-07
	Latino/Admixed American	3	17,717	9.56E-09	3.05E-10	4.16E-08
	Ashkenazi Jewish	0	5183	NA	NA	NA
	East Asian	52	9976	6.92E-06	3.79E-06	1.13E-05
	European (Finnish)	24	12,561	9.51E-07	3.75E-07	1.89E-06
	European (non-Finnish)	15	64,563	1.44E-08	4.23E-09	3.31E-08
	Other	1	3612	3.83E-08	1.23E-11	2.61E-07
	South Asian	6	15,308	4.48E-08	5.17E-09	1.45E-07
	GRCh37 Total	108	141,391	1.47E-07	9.82E-08	2.09E-07
GRCh38	Other	1	1047	4.56E-07	1.46E-10	3.10E-06
	Latino/Admixed American	3	7646	5.13E-08	1.64E-09	2.23E-07
	European (Finnish)	10	5314	9.74E-07	2.04E-07	2.58E-06
	Amish	0	456	NA	NA	NA
	East Asian	13	2604	6.71E-06	1.77E-06	1.62E-05
	Middle Eastern	0	158	NA	NA	NA
	African/African American	8	20,739	4.18E-08	6.93E-09	1.21E-07
	South Asian	1	2419	8.54E-08	2.74E-11	5.81E-07
	Ashkenazi Jewish	0	1736	NA	NA	NA
	European (non-Finnish)	15	34,026	5.18E-08	1.52E-08	1.19E-07
	GRCh38 Total	51	76,136	1.14E-07	6.22E-08	1.87E-07

estimate of 0.046/100,000 (95% CI: 0.004/100,000 to 0.162/100,000), and the Iranian population had an estimate of 0.078/100,000 (95% CI: 0.00/100,000 to 5.31/100,000). This finding was not reflected in the gnomAD v3.1.2 Middle Eastern ancestry group; however, this comprised a small number of individuals (158) and may not accurately reflect genetic variants present in those of Turkish and Iranian ancestry. Those of Finnish ancestry had an estimate of 0.095/100,000 (95% CI: 0.037/100,000 to 0.189/100,000) and 0.097/100,000 (95% CI: 0.020/100,000 to 0.258/100,000) for gnomAD v2.1.1 and gnomAD v3.1.2, respectively. Lastly, those of East Asian ancestry had an estimate of 0.692/100,000 (95% CI: 0.379/100,000 to 1.120/100,000) and 0.671/100,000 (95% CI: 0.177/100,000 to 1.618/100,000) for gnomAD v2.1.1 and gnomAD v3.1.2, respectively, indicating that individuals with this ancestry are at highest risk of having SPLIS among all the ancestral groups included in these data sets. Calculated prevalence in additional ancestries were: Latino 0.001/100,000; European (non-Finnish) 0.001/100,000; South Asian 0.004/100,000; African/African American 0.009/100,000. No pathogenic variants were identified in the Amish or Ashkenazi Jewish populations.

Data sets of Saudi, Turkish, Iranian, Indian, Israeli, and South Dakota Hutterite populations were also used.[28–30] No variants from our curated list were identified in the Saudi, Israeli, or South Dakota Hutterite data sets; however, the Saudi and Hutterite data sets had a small sample size, and it is possible other pathogenic variants exist that were not evaluated here; thus, further evaluation of these populations is warranted. High-confidence pathogenic variants (Table 1), Bayesian estimates (Table 3), and prevalence estimates (Table 4) were calculated for each population. Most noteworthy were the Turkish data sets. In the Turkish variome, 2 high-confidence pathogenic variants were identified. These variants occurred in 5 individuals of 4024 total individuals. The Bayesian estimate for individuals of Turkish ancestry at risk for SPLIS was 0.046/100,000 (95% CI: 0.004/100,000 to 0.162/100,000). A reliable inbreeding coefficient for these data was not available. However, Kars et al (2021) determined that a subset of Turkish individuals have a high inbreeding coefficient, suggesting that our estimate likely underestimates the number of individuals at risk for SPLIS.[30] Using the Iranome data set, we determined a SPLIS prevalence of 0.078/100,000 (95% CI: 0.000/100,000 to 0.531/100,000). Using the IndiGenome resource, a prevalence of 0.048/100,000 (95% CI: 0.00/100,000 to 0.324/100,000) was predicted for individuals of South Asian ancestry.

Overall, using gnomAD v2.1.1 we predict 11,707 (95% CI: 6288 to 19,699) cases of SPLIS exist globally, and using gnomAD v3.1.2 we predict 11,770 (95% CI: 2941 to 30,965) cases. Our results also suggest that, of the different ancestral groups evaluated, members of the Finnish, East Asian, Iranian, and Turkish populations are most at risk for having SPLIS. It is worth noting the Iranian data set estimate has a highly variable credible set intervals because of small sample sizes (800 individuals). The Amish and

Table 4 Estimated global prevalence of SPLIS calculated using Bayesian point estimates and 95% credible sets for the gnomAD data set v2.1.1 (GRCh37) and v3.1.2 (GRCh38), the Iranome and IndiGenomes data sets, as well as private data sets (Turkish)

Genome Reference Consortium Human Build	Ancestry or Dataset	Estimated Global Population	Anticipated Affected (per 100,000)	Anticipated Affected Lower (per 100,000)	Anticipated Affected Upper (per 100,000)	Ref
	Turkish	83,800,000	0.046	0.004	0.162	37
	Iranome (Iranian)	85,000,000	0.078	0.000	0.531	37
	IndiGenomes (Asian Indian)	1,393,000,000	0.048	0.000	0.324	37
GRCh37	African/African American	1,418,156,291	0.009	0.001	0.027	36,37
	Latino/Admixed American	718,753,964	0.001	0.000	0.004	36,37
	Ashkenazi Jewish	10,000,000	NA	NA	NA	37
	East Asian	1,650,000,000	0.692	0.379	1.129	37
	European (Finnish)	5,500,000	0.095	0.037	0.189	37
	European (non-Finnish)	738,500,000	0.001	0.000	0.003	37
	Other	1,337,089,745	0.004	0.000	0.026	37 b
	South Asian	1,969,000,000	0.004	0.001	0.015	37
	GRCh37 Total	7,837,000,000	0.015	0.010	0.021	37
GRCh38	Other	1,069,989,745	0.046	0.000	0.310	37 b
	Latino/Admixed American	718,753,964	0.005	0.000	0.022	36,37
	European (Finnish)	5,500,000	0.097	0.020	0.258	37
	Amish	373,620	NA	NA	NA	37
	East Asian	1,650,000,000	0.671	0.177	1.618	37
	Middle Eastern	454,400,000	NA	NA	NA	37 a
	African/African American	1,418,156,291	0.004	0.001	0.012	36,37
	South Asian	1,969,000,000	0.009	0.000	0.058	37
	Ashkenazi Jewish	10,000,000	NA	NA	NA	37
	European (non-Finnish)	738,500,000	0.005	0.002	0.012	37
	GRCh38 Total	7,837,000,000	0.011	0.006	0.019	37

[a]Sum of Egypt, Iran, and Western Asia (minus Armenia, Azerbaijan, and Georgia).
[b]Other - Global minus European, African/African American, Latino, East Asian, and South Asian.

Middle Eastern populations had no occurrences of potentially pathogenic variants within the *SGPL1* gene; however, only 456 Amish and 158 Middle Eastern exomes were present in gnomAD.

Discussion

To the best of our knowledge, this is the first reported estimate of SPLIS prevalence using a method other than counting recognized cases. In contrast to the number of SPLIS cases (<100) that have been reported in the literature or communicated to us privately, we estimate there are 11,707 to 11,770 individuals with biallelic pathogenic *SGPL1* variants and undiagnosed SPLIS globally. Thus, reported cases account for less than 1% of the predicted cases, vastly underestimating the total disease burden associated with this condition. This discrepancy can be explained by a combination of factors. Many individuals on either end of the phenotypic spectrum of SPLIS are not likely to be tested for a genetic etiology. Fetal losses and unexplained infant deaths often go undiagnosed. The mildest cases, such as those manifesting as isolated peripheral neuropathy, may not present until adulthood; a potential genetic diagnosis is less often considered by practitioners of adult patients. In addition, there is a lack of familiarity with SPLIS in the medical community because the condition was just described in 2017 and is not yet in the pediatric or genetic textbooks. Some patients who receive a genetic diagnosis are not reported in the literature. In addition, *SGPL1* is not included on many commercially available disease-focused next-generation sequencing diagnostic panels. Despite these considerations, we expect the rate of SPLIS diagnosis will improve as familiarity with SPLIS increases, and genome and exome sequencing diagnostic investigations continue to move into the mainstream of medical practice.

The prevalence of SPLIS is not uniform around the world. Using gnomAD v2.1.1, we predict SPLIS prevalence in East Asian individuals to be about 0.692/100,000—over 40-fold higher than the worldwide prevalence of SPLIS. Those of Finnish ancestry had a predicted prevalence of 0.095/100,000—roughly 6-fold higher than the worldwide prevalence. Other noteworthy groups are those of Turkish (0.046/100,000) and Iranian (0.078/100,000) ancestries. These findings are consistent with the higher rates of consanguinity in these communities.[30,38] Despite the higher predicted rates of SPLIS in the Iranian people, their smaller global population results in a predicted contribution of less than 1% of the total number of worldwide SPLIS cases. In contrast, East Asian and South Asian ancestries together make up about 60% of the world's total population. Combined, these groups contribute more than 95% of the world's predicted cases. The remaining cases are expected to be found in African, European, and Latino populations in roughly equal numbers. It

should be emphasized that the predictions of SPLIS prevalence in different populations is highly dependent on which variants are included in the calculations. As functional data become available, they could result in the reclassification of variants of uncertain significance with high prevalence in certain populations, which could have a significant impact on the prevalence of SPLIS in these geoancestral groups.

In most cases, we have relied on the gnomAD v2.1.1 (GRCh37) data set for our calculations because this set contains significantly more genomes than the gnomAD v3.1.2 (GRCh38) version. We relied on the gnomAD v3.1.2 (GRCh38) data set for calculation of SPLIS in the Middle Eastern and Amish populations because these groups were not recognized as a separate entity in the gnomAD v3.1.2 (GRCh38) data set. In cases where more than one data set was available for analysis of prevalence in an ancestral group, rates were roughly concordant. Differences in the calculated rates within the same ancestral group could be explained by sampling effects caused by rare variants being present at slightly different rates in 2 data sets from the same population just by chance.

In addition to the major population groups listed above, SPLIS may be more prevalent in smaller communities where inbreeding rates are high. In our analysis, no pathogenic or likely pathogenic *SGPL1* variants were found in the Amish (from gnomAD v.3.1.2) or Hutterites (private data set). However, we are aware of 3 cases of SPLIS in the Hutterite colony in Alberta, Canada, all of which are homozygous for a previously reported c.1247A>G; p.Tyr416Cys *SGPL1* variant.[11] Similar to the Amish, the Hutterites are a sect of Anabaptist Christians; the latter settled in North America in the 1800s.[39] They comprise 3 groups or "leut" located in Western Canada and the upper Great Plains area of the United States. Because of the presence of 3 cases in this small community, we sought to determine the pathogenic *SGPL1* allele frequency in the population. For, the Hutterites, we used genome sequence data from 121 South Dakota Hutterites. The fact that we did not identify pathogenic *SGPL1* variants in these genomes suggests that the variant responsible for the 3 known cases of SPLIS among the Hutterites is isolated to 1 leut or may even have arisen within 1 isolated colony.[11]

One significant reason for establishing SPLIS prevalence is to design clinical trials. We recently reported the efficacy of adeno-associated virus-mediated *SGPL1* gene therapy (AAV-SPL) in proof-of-concept studies conducted in a robust knockout mouse model of SPLIS.[6] Significant improvements in potency of AAV-SPL have now been achieved, setting the stage for further development of AAV-SPL as a potentially lifesaving treatment for SPLIS (our unpublished results). Similar gene therapy approaches could be considered for other disorders of the sphingolipid catabolic pathway. Thus, our methodology for estimating the prevalence of SPLIS could be similarly applied to other conditions of sphingolipid metabolism in anticipation of gene therapy trials in those conditions. Additionally, we have reported the potential efficacy of an alternative therapeutic strategy in SPLIS, namely, the use of pyridoxine cofactor supplementation, which may overcome the poor cofactor binding affinity and poor protein stability of the R222Q variant and possibly others.[7] In regard to clinical trials, we have estimated the number of SPLIS cases in the United States, where clinical trials are likely to be carried out first. In the United States, using the overall rate of 0.015/100,000, an estimated 71 cases should exist. However, extrapolating from the worldwide SPLIS estimate of ~11,770 and with the United States representing 4.3% of the world's population, more than 500 individuals with SPLIS in the United States, and more than a 1000 in the United States and European Union, would be predicted. Even using the lower estimates, our results indicate that clinical trials in SPLIS should be feasible, provided awareness and patient finding improve the rate of detection. Further, considering that the majority of variants used in our calculations are pLOF variants, most individuals with SPLIS predicted by these calculations would be expected to be on the severe end of the phenotypic spectrum.

One major limitation of our study is that ancestral groups are broadly defined, whereas each of the groups listed in our study and defined by gnomAD represent many subgroups separated by national borders, geography, language, and culture and which are not genetically homogeneous. Another caveat is that our predictions are only as good as our variant curation. Although it is reasonable to suggest that *SGPL1* biallelic variants in individuals with SPLIS are putatively pathogenic, the inclusion of missense *SGPL1* variants without functional studies to show they result in LOF could introduce errors in our calculations if any of the variants are proven benign in further studies. Using the conservative approach of leaving out variants of unknown significance makes the possibility of underestimating SPLIS prevalence more likely. Our results could be refined in the future as clinical and/or experimental evidence for a larger number of likely pathogenic variant alleles becomes available and additional databases, particularly those with detailed clinical information, become accessible. It should be noted that, although it is possible that individuals may exist with biallelic combinations of pathogenic *SGPL1* variants and not express SPLIS because of unknown compensatory mechanisms, this possibility is considered unlikely because it is rare to find such individuals for the large majority of similar early-onset, severe monogenic diseases.[40] If there is incomplete penetrance for 1 or more of the pathogenic variants investigated in this study, then the SPLIS prevalence estimates would represent moderately inflated values over the true disease rate. In addition, we have assumed linkage equilibrium between the pathogenic variants in the prevalence estimates and an over-representation of chromosomes carrying 2 or more of the pathogenic variants would also produce mild overestimates in the SPLIS prevalence. This is also not likely because rare alleles are predominantly more recent variants, and the probability of the occurrence of a new pathogenic variant on a chromosome already carrying a pathogenic allele would be

very small given that *SGPL1* pathogenic allele-carrying chromosomes are rare in all populations. On the other hand, additional disease-causing variants are likely to exist that we have not included. For example, we are aware of multiple cases of SPLIS associated with the homozygous c.665 G>A; p.Arg222Gln variant in individuals of Pakistani and Middle Eastern ancestry. However, this variant was not detected in the private population studies we used in our analysis. The existence of known "missed" cases suggests that our calculations underestimate the number of true cases. On the other hand, some databases such as gnomAD tend to be biased toward disease cohorts, which could result in our overestimating SPLIS prevalence. Further, the accuracy and quality control of variant calling and classification can differ significantly between platforms, and the higher detection rates in Turkish and Iranian cohorts could reflect the smaller representation of subjects from these areas. Thus, these rates may need to be interpreted with a degree of caution. We also have not taken into account pathogenic variants occurring within non-coding regulatory regions of *SGPL1*.[19] In fact, ENSEMBL predicts 18 different *SGPL1* transcripts, which suggests the expression of this gene is highly regulated.[41] We do not know the full spectrum of disease in SPLIS and have assumed the recessive combinations of variants are fully penetrant, which may not be true. The shortened lifespan of individuals with SPLIS and the lack of specific treatments would be expected to have a substantial influence on disease prevalence. A prospective SPLIS natural history study and newborn screening for SPLIS would be helpful in providing more information about these factors, which could then be applied to revise our estimates.

Data Availability

All data generated in this project, except *SGPL1* variants from private databases, are provided within the manuscript which will be open access and without embargo. *SGPL1* variants from private databases will be made available upon request. Variants described in the manuscript will be deposited in ClinVar.

Acknowledgments

The authors are grateful to Vardiella Meiner (Department of Genetics and Metabolic Diseases, Hadassah University Medical Center, Jerusalem, Israel) for sharing information for the study.

Funding

J.D.S. is supported by a grant from the UCSF InVent Catalyst Program and NIH R21TR004262; J.C.S. is supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison. Article processing charges are paid by a University of California San Francisco faculty academic enrichment stipend to J.D.S.

Author Information

Conceptualization: J.D.S., R.D.S., S.J.S.; Data Curation: C.B.; W.W.-S., J.C.S.; Formal Analysis: J.C.S., S.J.S.; Funding Acquisition: J.D.S.; Investigation: J.D.S., R.D.S., C.B., J.C.S., S.J.S., C.O., M.A., O.E.O., V.S.; Methodology: S.J.S.; Project Administration: J.D.S.; Resources: V.S., C.O., M.A., O.E.O.; Supervision: S.J.S.; Writing-original draft: J.D.S., J.C.S., C.B.; Writing-review and editing: J.D.S., R.D.S., S.J.S., C.B., J.C.S., V.S., C.O., M.A., O.E.O.

ORCIDs

Julie Saba: http://orcid.org/0000-0002-4894-919X
Robert Steiner: http://orcid.org/0000-0003-4177-4590
Steven J. Schrodi: http://orcid.org/0000-0003-2304-8528

Ethics Statement

All institutions involved in human participant research received local Institutional Review Board (IRB approval), and informed consent was obtained from all participants. SPLIS patients provided consent to share medical information in accordance with an approved University of California San Francisco IRB protocol. The Institutional Ethics Committees (IEC) of Bilkent and Koç Universities approved the Turkish variome study. The IEC of The King Abdullah International Medical Research Center approved the Saudi Arabian genomic database. The IRB of the University of Chicago approved the Hutterite genome study. The Israeli study was approved by the IEC of Hadassah Medical Center.

Conflict of Interest

R.D.S. has equity interest in and has received consulting fees from Acer Therapeutics and PTC Therapeutics. He has received consulting fees from Aeglea BioTherapeutics; Best Doctors/Teladoc; Health Advances LLC; Leadiant, and Travere Therapeutics, Inc; and honoraria from Medscape/WebMD LLC and The France Foundation. He is an employee of and has equity interest in PreventionGenetics, part of Exact Sciences. He received research funding from Alexion Pharmaceuticals, Children's Mercy Research Institute, and the Smith Lemli Opitz Foundation. J.D.S. is an author on patent "AAV-SPL as a treatment for SPLIS:

International Application Serial No. PCT/US2021/018613," "Adeno-Associated Viral (Aav)-Mediated Sgpl1 Gene Therapy for Treatment of Sphingosine-1-Phosphate Lyase Insufficiency Syndrome (SPLIS)" published on 08/26/2021 and assigned publication number WO 2021/168140. All other authors declare no conflicts of interest.

Clinical Trial Information

Not applicable.

Additional Information

The online version of this article (https://doi.org/10.1016/j.gimo.2023.100840) contains supplemental material, which is available to authorized users.

Affiliations

[1]Department of Medical Genetics, University of Wisconsin-Madison, Madison, WI; [2]Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI; [3]Department of Pathology, University of California San Francisco, San Francisco, CA; [4]CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), South Campus, Mathura Road, New Delhi; [5]Bezmialem Vakif University, Institute of Health Sciences and Biotechnology, Department of Molecular Biology, Istanbul, Turkey; [6]Genetics and Precision Medicine Department (GPM), King Abdullah International Medical Research Center (KAIMRC), King Saud Bin Abdulaziz University for Health Sciences (KSAU-HS), Ministry of National Guard Health Affairs (MNG-HA), Riyadh, Saudi Arabia; [7]Genetics and Precision Medicine Department (GPM), King Abdullah Specialized Children's Hospital (KASCH), King Abdulaziz Medical City, Ministry of National Guard Health Affairs (MNG-HA), Riyadh, Saudi Arabia; [8]Department of Human Genetics, University of Chicago, Chicago, IL; [9]Department of Pediatrics, University of Wisconsin-Madison, Madison, WI; [10]Department of Pediatrics, University of California San Francisco, San Francisco, CA

References

1. Pournasiri Z, Madani A, Nazarpack F, et al. Sphingosine phosphate lyase insufficiency syndrome: a systematic review. *World J Pediatr*. 2023;19(5):425-437. http://doi.org/10.1007/s12519-022-00615-4
2. Weaver KN, Sullivan B, Hildebrandt F, et al. Sphingosine phosphate lyase insufficiency syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews*. Seattle, (WA); 2020.
3. Schmahl J, Raymond CS, Soriano P. PDGF signaling specificity is mediated through multiple immediate early genes. *Nat Genet*. 2007;39(1):52-60. http://doi.org/10.1038/ng1922
4. Allende ML, Bektas M, Lee BG, et al. Sphingosine-1-phosphate lyase deficiency produces a pro-inflammatory response while impairing neutrophil trafficking. *J Biol Chem*. 2011;286(9):7348-7358. http://doi.org/10.1074/jbc.M110.171819
5. Bektas M, Allende ML, Lee BG, et al. Sphingosine 1-phosphate lyase deficiency disrupts lipid homeostasis in liver. *J Biol Chem*. 2010;285(14):10880-10889. http://doi.org/10.1074/jbc.M109.081489
6. Zhao P, Tassew GB, Lee JY, et al. Efficacy of AAV9-mediated *SGPL1* gene transfer in a mouse model of S1P lyase insufficiency syndrome. *JCI Insight*. 2021;6(8):e145936. http://doi.org/10.1172/jci.insight.145936
7. Zhao P, Liu ID, Hodgin JB, et al. Responsiveness of sphingosine phosphate lyase insufficiency syndrome to vitamin B_6 cofactor supplementation. *J Inherit Metab Dis*. 2020;43(5):1131-1142. http://doi.org/10.1002/jimd.12238
8. Dixit D, Okuniewska M, Schwab SR. Secrets and lyase: control of sphingosine 1-phosphate distribution. *Immunol Rev*. 2019;289(1):173-185. http://doi.org/10.1111/imr.12760
9. Cartier A, Hla T. Sphingosine 1-phosphate: lipid signaling in pathology and therapy. *Science*. 2019;366(6463). http://doi.org/10.1126/science.aar5551
10. Atkinson D, Nikodinovic Glumac J, Asselbergh B, et al. Sphingosine 1-phosphate lyase deficiency causes Charcot-Marie-Tooth neuropathy. *Neurology*. 2017;88(6):533-542. http://doi.org/10.1212/WNL.0000000000003595
11. Lovric S, Goncalves S, Gee HY, et al. Mutations in sphingosine-1-phosphate lyase cause nephrosis with ichthyosis and adrenal insufficiency. *J Clin Invest*. 2017;127(3):912-928. http://doi.org/10.1172/JCI89626
12. Prasad R, Hadjidemetriou I, Maharaj A, et al. Sphingosine-1-phosphate lyase mutations cause primary adrenal insufficiency and steroid-resistant nephrotic syndrome. *J Clin Invest*. 2017;127(3):942-953. http://doi.org/10.1172/JCI90171
13. Janecke AR, Xu R, Steichen-Gersdorf E, et al. Deficiency of the sphingosine-1-phosphate lyase *SGPL1* is associated with congenital nephrotic syndrome and congenital adrenal calcifications. *Hum Mutat*. 2017;38(4):365-372. http://doi.org/10.1002/humu.23192
14. Bamborschke D, Pergande M, Becker K, et al. A novel mutation in sphingosine-1-phosphate lyase causing congenital brain malformation. *Brain Dev*. 2018;40(6):480-483. http://doi.org/10.1016/j.braindev.2018.02.008
15. Saygili S, Canpolat N, Sever L, Caliskan S, Atayar E, Ozaltin F. Persistent hypoglycemic attacks during hemodialysis sessions in an infant with congenital nephrotic syndrome: answers. *Pediatr Nephrol*. 2019;34(1):77-79. http://doi.org/10.1007/s00467-018-3982-7
16. Linhares N, Arantes R, Araujo S, Pena S. Nephrotic syndrome and adrenal insufficiency caused by a variant in *SGPL1*. *Clin Kidney J*. 2017:1-6.
17. Tran P, Jamee M, Pournasiri Z, Chavoshzadeh Z, Sullivan KE. *SGPL1* deficiency: nephrotic syndrome with lymphopenia. *J Clin Immunol*. 2023;43(1):72-75. http://doi.org/10.1007/s10875-022-01348-9
18. Tastemel Ozturk T, Canpolat N, Saygili S, et al. A rare cause of nephrotic syndrome-sphingosine-1-phosphate lyase (*SGPL1*) deficiency: 6 cases and a review of the literature. *Pediatr Nephrol*. 2023;38(3):711-719. http://doi.org/10.1007/s00467-022-05656-5
19. Yang S, He Y, Zhou J, Yuan H, Qiu L. Steroid-resistant nephrotic syndrome associated with certain *SGPL1* variants in a family: case report and literature review. *Front Pediatr*. 2023;11:1079758. http://doi.org/10.3389/fped.2023.1079758

20. Settas N, Persky R, Faucz FR, et al. *SGPL1* deficiency: a rare cause of primary adrenal insufficiency. *J Clin Endocrinol Metab*. 2019;104(5):1484-1490. http://doi.org/10.1210/jc.2018-02238
21. Spizzirri AP, Cobeñas CJ, Suarez ADC. A rare cause of nephrotic syndrome – sphingosine-1-phosphate lyase (*SGPL1*) deficiency: 2 cases. *Pediatr Nephrol*. 2023;38(1):307-308. http://doi.org/10.1007/s00467-022-05716-w
22. Ron HA, Scobell R, Strong A, Salazar EG, Ganetzky R. Congenital adrenal calcifications as the first clinical indication of sphingosine lyase insufficiency syndrome: a case report and review of the literature. *Am J Med Genet A*. 2022;188(11):3312-3317. http://doi.org/10.1002/ajmg.a.62956
23. Saba JD, Keller N, Wang JY, Tang F, Slavin A, Shen Y. Genotype/phenotype interactions and first steps toward targeted therapy for sphingosine phosphate lyase insufficiency syndrome. *Cell Biochem Biophys*. 2021;79(3):547-559. http://doi.org/10.1007/s12013-021-01013-9
24. Dunn TM, Tifft CJ, Proia RL. A perilous path: the inborn errors of sphingolipid metabolism. *J Lipid Res*. 2019;60(3):475-483. http://doi.org/10.1194/jlr.S091827
25. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32(5):557-563. http://doi.org/10.1002/humu.21438
26. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:D980-D985. http://doi.org/10.1093/nar/gkt1113. database issue.
27. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. http://doi.org/10.1038/s41586-020-2308-7
28. Jain A, Bhoyar RC, Pandhare K, et al. IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. *Nucleic Acids Res*. 2021;49(D1):D1225-D1232. http://doi.org/10.1093/nar/gkaa923
29. Fattahi Z, Beheshtian M, Mohseni M, et al. Iranome: a catalog of genomic variations in the Iranian population. *Hum Mutat*. 2019;40(11):1968-1984. http://doi.org/10.1002/humu.23880
30. Kars ME, Başak AN, Onat OE, et al. The genetic structure of the Turkish population reveals high levels of variation and admixture. *Proc Natl Acad Sci U S A*. 2021;118(36):e2026076118. http://doi.org/10.1073/pnas.2026076118
31. Aloraini T, Alsubaie L, Alasker S, et al. The rate of secondary genomic findings in the Saudi population. *Am J Med Genet A*. 2022;188(1):83-88. http://doi.org/10.1002/ajmg.a.62491
32. Richards S, Aziz N, Bale S, 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(5):405-424. http://doi.org/10.1038/gim.2015.30
33. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med*. 2019;12(1):3. http://doi.org/10.1186/s13073-019-0690-2
34. Biesecker LG, Harrison SM. ClinGen Sequence Variant Interpretation Working Group. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med*. 2018;20(12):1687-1688. http://doi.org/10.1038/gim.2018.42
35. Schrodi SJ, DeBarber A, He M, et al. Prevalence estimation for monogenic autosomal recessive diseases using population-based genetic data. *Hum Genet*. 2015;134(6):659-669. http://doi.org/10.1007/s00439-015-1551-8
36. Bureau USC. 2022. Accessed March 6, 2023. https://www.census.gov/quickfacts/fact/table/US/PST045221
37. Bureau PR. 2023. Accessed March 6, 2023. https://www.prb.org/news/2021-world-population-data-sheet-released/
38. Tadmouri GO, Nair P, Obeid T, Al Ali MT, Al Khaja N, Hamamy HA. Consanguinity and reproductive health among Arabs. *Reprod Health*. 2009;6:17. http://doi.org/10.1186/1742-4755-6-17
39. Boycott KM, Parboosingh JS, Chodirker BN, et al. Clinical genetics and the Hutterite population: a review of Mendelian disorders. *Am J Med Genet A*. 2008;146A(8):1088-1098. http://doi.org/10.1002/ajmg.a.32245
40. Chen R, Shi L, Hakenberg J, et al. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol*. 2016;34(5):531-538. http://doi.org/10.1038/nbt.3514
41. Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. *Nucleic Acids Res*. 2022;50(D1):D988-D995. http://doi.org/10.1093/nar/gkab1049
42. Mathew G, Yasmeen MS, Deepthi RV, et al. Infantile nephrotic syndrome, immunodeficiency and adrenal insufficiency-a rare cause: answers. *Pediatr Nephrol*. 2022;37(4):817-819. http://doi.org/10.1007/s00467-021-05377-1
43. Martin KW, Weaver N, Alhasan K, et al. MRI spectrum of brain involvement in sphingosine-1-phosphate lyase insufficiency syndrome. *AJNR Am J Neuroradiol*. 2020;41(10):1943-1948. http://doi.org/10.3174/ajnr.A6746
44. Alrayes LAF, Alotaibi M, Alsagheir A. A Saudi child with sphingosine phosphate lyase insufficiency syndrome. *JBCgenetics*. 2021;4(1):48-50. http://doi.org/10.24911/JBCGenetics/183-1606918375
45. Bertoli-Avella AM, Beetz C, Ameziane N, et al. Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort. *Eur J Hum Genet*. 2021;29(1):141-153. http://doi.org/10.1038/s41431-020-00713-9
46. Taylor VA, Stone HK, Schuh MP, Zhao X, Setchell KD, Erkan E. Disarranged sphingolipid metabolism from sphingosine-1-phosphate lyase deficiency leads to congenital nephrotic syndrome. *Kidney Int Rep*. 2019;4(12):1763-1769. http://doi.org/10.1016/j.ekir.2019.07.018
47. Maharaj A, Theodorou D, Banerjee II, Metherell LA, Prasad R, Wallace D. A sphingosine-1-phosphate lyase mutation associated with congenital nephrotic syndrome and multiple endocrinopathy. *Front Pediatr*. 2020;8:151. http://doi.org/10.3389/fped.2020.00151
48. Najafi M, Riedhammer KM, Rad A, et al. High detection rate for disease-causing variants in a cohort of 30 Iranian pediatric steroid resistant nephrotic syndrome cases. *Front Pediatr*. 2022;10:974840. http://doi.org/10.3389/fped.2022.974840