Published in final edited form as:

Pediatr Neurol. 2020 October ; 111: 66–69. doi:10.1016/j.pediatrneurol.2020.06.005.

Elevated Leukodystrophy Incidence Predicted From Genomics Databases

Haille E. Soderholm¹, Alexander B. Chapin^{2,3}, Pinar Bayrak-Toydemir^{2,3}, Joshua L. Bonkowsky^{1,4,5,*}

¹Division of Pediatric Neurology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, Utah

²ARUP Institute for Clinical and Experimental Pathology, ARUP Laboratories, Salt Lake City, Utah

³Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah

⁴Brain and Spine Center, Primary Children's Hospital, Salt Lake City, Utah

⁵Primary Children's Center for Personalized Medicine, Salt Lake City Utah

Abstract

Background—Leukodystrophies are genetic diseases affecting the white matter and leading to early death. Our objective was to determine leukodystrophy incidence, using genomics sequencing databases allele frequencies of disease-causing variants.

Methods—From 49 genes, representing the standardly defined group of leukodystrophies, we identified potential disease-causing variants from publications in the Human Genetic Mutation Database and from predictions in the Genome Aggregation Database (gnomAD). Allele frequencies were estimated from gnomAD. Allele frequencies for each gene were summed to generate a super allele frequency (SAF) and we used the Hardy-Weinberg equation to calculate overall expected live birth incidence associated with the gene in question.

Results—We identified 4,564 pathogenic variants for 25 discrete leukodystrophies. The largest effect was from *GALC* variants (Krabbe disease), which had a predicted incidence of 1 in 12,080 live births, 8.3 times higher than published estimates. The second most frequently predicted leukodystrophy were the POLIII-related disorders, which had an incidence of 1:26,160. Overall, we found a leukodystrophy incidence of 1 in 4,733 live births, significantly higher than previous estimates.

* Address correspondence to: Josh Bonkowsky, Department of Pediatrics, 295 Chipeta Way/Williams Building, Salt Lake City, Utah 84108, joshua.bonkowsky@hsc.utah.edu, Phone: 801-581-6756, Fax: 801-581-4233.

AUTHOR CONTRIBUTIONS

HS, AC, PBT, and JLB analyzed the data; drafted and edited the manuscript for intellectual content. JLB conceptualized the study. HS and JLB designed the study.

Potential Conflicts of Interest

JLB has served as a consultant to Bluebird Bio, Inc; to Calico, Inc; to Neurogene, Inc.; to Enzyvant, Inc.; and owns stock in Orchard Therapeutics. HS, AC, and PBT report no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—Our data are consistent with a significant underdiagnosis of leukodystrophy patients. An intriguing additional consideration is that there may be genetic modifiers that lead to weaker, absent, or adult-onset disease phenotypes.

Keywords

Incidence; leukodystrophy; gnomAD; genetic modifiers; Krabbe disease

INTRODUCTION

Leukodystrophies are genetic diseases affecting the white matter and leading to early death [Bonkowsky et al., 2010]. Determining incidence of leukodystrophies is critical for guiding diagnosis, screening, and clinical trial design. However, overall leukodystrophy incidence has been challenging to establish. A population-based approach found a birth estimate of 1 in 7,663 live births [Bonkowsky et al., 2010], but this determination was dependent upon clinical identification. Other estimations have been from centers that are referral-based and hence make incidence determination problematic.

Our objective was to determine leukodystrophy incidence using an approach based on disease allele frequencies in genomic sequence databases. Using curation of known disease-causing variants, as well as of variants predicted to be deleterious from *in silico* predictions, we sought to determine a revised estimate of leukodystrophy incidence.

METHODS

Editorial Policies and Ethical Considerations

The Institutional Review Board of the University of Utah approved this study as exempt as non-human research.

From a starting group of 30 standardly-defined leukodystrophies [Vanderver et al., 2015], we identified potential disease-causing variants in 25 diseases (see further explanation below) from publications in the Human Genetic Mutation Database (HGMD) and from predictions in Genome Aggregation Database (gnomAD) [Stenson et al., 2003; Lek et al., 2016]. gnomAD only includes sequence data from individuals without known disease, thus limiting analysis to conditions that are autosomal recessive [Lek et al., 2016]. We included only HGMD variants classified as disease-causing mutations “DM”. Possible/probable disease-causing mutations were marked “DM?” and manually curated using American College of Medicine Guideline [Richards et al., 2015]. 40 variants marked DM? had additional references supporting pathogenicity, and so were therefore classified DM (Supplemental Table 1) including with ClinVar notes of pathogenicity [Richards et al., 2015]. We removed any duplicated variants.

After determination of alleles for inclusion, allele frequencies for autosomal recessive leukodystrophies were determined by finding the average allele number per gene (“ANG”) for genome and exome sequencing in gnomAD. The allele count was divided by the ANG to calculate individual allele frequencies and mitigate overestimation of true allele frequencies due to the fact that gnomAD has separate files for exome and genome data sets. This method

was not used for the most common Krabbe disease mutation seen in patients, a large 31.7 kb deletion. The raw allele number for this deletion was used in place of the *ANG* for this variant since only genome sequencing data is used for large insertions and deletions in gnomAD. Overall population incidence was calculated by summing allele frequencies for the gene to obtain a super allele frequency (SAF). The SAF was used as the q value and population incidence was calculated using the Hardy-Weinberg equation solving for q^2 . This procedure to generate an SAF was repeated for each gene. Individual incidences of genes known to cause the same form of leukodystrophy were summed independently (i.e., each individual *PEX* gene incidence was found, then *PEX* genes 1–26 SAFs were summed to find overall Zellweger incidence).

For autosomal dominant leukodystrophies, such as caused by mutations in *TUBB4A*, Alexander disease (*GFAP*), or Autosomal Dominant Leukodystrophy (ADLD) (*LMNB1*), we were unable to include these in our calculations since allele counts for affected individuals would not be found in gnomAD. Similarly, Pelizaeus-Merzbacher Disease (PMD) was excluded since it is most often caused by a duplication of the gene *PLP1* and none of these mutations are listed in gnomAD. We also excluded consideration of X-linked adrenoleukodystrophy (ALD), caused by mutations in *ABCD1*, since adult males, and often adult females, would be symptomatic, and would not be included in gnomAD.

Variants marked DM with a contradicting reference were excluded, except one pathogenic GALC Krabbe disease variant associated with adult onset (rs183105855) (Supplemental Table 1). Five variants (rs118204450, rs764863416, rs1489711400, rs758055753, 2:219679730:C:T) were reported as DM, but were classified as low confidence loss of function (LoF) in gnomAD. Three variants (rs72549405, rs1342624234, rs515726131) labeled DM? in HGMD but classified LoF in gnomAD were included (Supplemental Table 1). 203 low-confidence LoF variants predicted from gnomAD were removed. DM variants with homozygous allele counts in gnomAD > 0 were analyzed to ensure pathogenicity. 11 were kept and 7 were excluded based off ClinVar notes of pathogenicity (Supplemental Table 1) [Landrum et al., 2017]. There were five variants representing adult-onset/mild allele forms of Krabbe disease with homozygous individuals present in gnomAD; three of these variants were included (Supplemental Table 1).

Data Availability Statement

All data are included in the main manuscript or in the supplemental table.

RESULTS

Starting from a standardly accepted definition of 30 leukodystrophies [Vanderver et al., 2015], we identified 4,564 pathogenic variants in 49 genes associated with 25 different leukodystrophies (Table 1, Supplemental Table 1). Of these, 3,335 were identified in primary publications listed in HGMD whereas 1,229 were identified as truncating variants in gnomAD. Allele frequencies were estimated using sequence data available from gnomAD. gnomAD only includes sequence data from individuals without known disease, thus limiting analysis to conditions that are autosomal recessive [Lek et al., 2016]. Next, we summed allele frequencies for each gene to generate a super allele frequency (SAF) and used the

Hardy-Weinberg equation to calculate overall expected live birth incidence associated with the gene in question.

By summing all individual gene incidences, we determined a leukodystrophy incidence at 1 in 4,733 (Table 1). The most significant effect was contributed by mutations in *GALC* (Krabbe disease), which had a predicted incidence of 1 in 12,080 live births (Table 1). This is 8.3 times higher than published estimates of 1 in 100,000 live births [Wasserstein et al., 2016]. We examined all of the alleles identified in *GALC* and determined they had all been previously identified in patients with symptomatic Krabbe disease. However, some variants have a milder effect and must be present as a trans-heterozygote with a more severe variant to be pathogenic. The second most frequently predicted leukodystrophy were the POLIII-related disorders, which had an incidence of 1:26,160 (Table 1). The prevalence of POLIII-related disorders has not been previously reported.

DISCUSSION

We have determined a revised estimate of leukodystrophy incidence, 1 in 4,733 live births (Table 1). This is higher than our previous estimation using a population-based determination [Bonkowsky et al., 2010]. Further, our revised estimate only includes data on 49 genes (25 leukodystrophy diseases), and could not include data from dominant mutations, which are excluded from the gnomAD database. Our approach avoids biases based on referral center-based determinations, on requiring clinician identification, and from under-reporting of minority patients [Bonkowsky et al., 2018].

The most common leukodystrophy identified from this study was Krabbe disease, which was unexpected given its infrequent identification in newborn screening programs, which use a biochemical metabolite for detection [Orsini et al., 2016]. Previous studies have shown that two mild bi-allelic *GALC* variants may not lead to disease, however, when a mild variant is paired with a severe *GALC* variant, this will lead to disease [Saavedra-Matiz et al., 2016].

Limitation of this study include its reliance upon calculated disease incidence, as compared to a prospective determination. Another limitation is that sequence databases do not accurately reflect population demographics and may incorrectly classify variants in individuals from different racial backgrounds. Globally, there is an under-representation of data from African, Indian, and Chinese populations. These limitations in sequence databases also include that rare, “private,” intronic, and/or *de novo* mutations will typically not be included for calculations.

A complex and potential additional limitation is that allele variants can have variable penetrance or effects on viability. For example, the severity of the variants may affect whether the patient has clinical symptoms, such as can occur with *GALC* variants. Additionally, for some of the rarer leukodystrophies such as Vanishing White Matter Disease, severe bi-allelic loss-of-function variants may be incompatible with life. However, for the most prevalent leukodystrophies in our study, Krabbe disease and POLIII related disorders, severe bi-allelic loss of function variants are compatible with life and result in leukodystrophy [Wasserstein et al., 2016]. Therefore, we did not exclude severe bi-allelic

variants from our incidence calculations since the overall effect would be minor. It is also not known for each gene whether all allelic combinations are pathogenic. Importantly, we also did not consider the disease phenotype. That is, we were agnostic to whether the variants caused leukodystrophy or a different disease condition (for example, Vanishing White Matter Disease could be symptomatic as a leukodystrophy or as ovarian failure). Our approach also did not take into account if there are two pathogenic variants in *cis*; however, given the overall rarity of variants, this possibility would be extremely rare. Reassuringly, our predicted incidence of MLD, an extensively studied leukodystrophy, is similar to previous estimates.

In addition, 25% (n=1,208) of all variants were reported as loss of function in gnomAD, but there are no published reports in patients (Supplemental Table 1). However, these variants only made up 18% (n=2,945) of the total allele counts (n=16,672) (Supplemental Table 1). While these variants seem likely to cause disease, their actual contribution to pathogenicity is unknown.

In summary, we found a significantly increased leukodystrophy incidence of 1 in ~4,700 live births. This determination is likely an underestimation since we only included 25 leukodystrophies in the calculations, as well as because of limitations as discussed above with the sequence databases used for calculations. Further, while 30 different diseases have been defined using a strict definition as leukodystrophies [Vanderver et al., 2015], not only did we exclude from calculations dominantly inherited and/or genes with heterozygous or X-linked phenotypes (and which are not included in the gnomAD database), there is also evidence suggesting that many more genes in fact cause disease that would be considered a leukodystrophy [van der Knaap and Bugiani, 2017; Urbik et al., in press]. Also, because the Hardy-Weinberg equation used for calculations assumes random interbreeding in a population, in fact some leukodystrophies may be more prevalent, particularly in certain populations, due to founder effects or from increased marriages among relatives. An unexpected finding was that we noted an increase in predicted incidence of Krabbe disease when compared to previous estimates. Overall, our data suggests a significant underdiagnosis of leukodystrophy patients. An intriguing additional consideration is that there may be genetic modifiers that lead to weaker, absent, or later adult-onset disease phenotypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

JLB was supported by NIH grant 3UL1TR002538, and by the Bray Presidential Chair in Child Neurology research.

REFERENCES

Bonkowsky JL, Nelson C, Kingston JL, et al. The burden of inherited leukodystrophies in children. *Neurology* 2010;75:718–725. [PubMed: 20660364]

- Bonkowsky JL, Wilkes J, Bardsley T, et al. Association of Diagnosis of Leukodystrophy With Race and Ethnicity Among Pediatric and Adolescent Patients. *JAMA Netw Open* 2018;1:e185031. [PubMed: 30646379]
- Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2017;46:D1062–D1067.
- Lek M, Karczewski KJ, Minikel EV, et al.; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–291. [PubMed: 27535533]
- Orsini JJ, Saavedra-Matiz CA, Gelb MH, Caggana M. Newborn screening for Krabbe's disease. *J Neurosci Res* 2016;94:1063–1075. [PubMed: 27638592]
- 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;17:405–24. [PubMed: 25741868]
- Saavedra-Matiz CA, Luzi P, Nichols M et al. Expression of individual mutations and haplotypes in the galactocerebrosidase gene identified by the newborn screening program in New York State and in confirmed cases of Krabbe's disease. *J Neurosci Res* 2016;94:1076–83. [PubMed: 27638593]
- Stenson PD, Ball EV, Mort M, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;21:577–81. [PubMed: 12754702]
- Urbik VM, Schmiedel M, Soderholm H, Bonkowsky JL. Expanded Phenotypic Definition Identifies Hundreds of Potential Causative Genes for Leukodystrophies and Leukoencephalopathies. *Child Neurol Open* 2020; in press.
- Vanderver A, Prust M, Tonduti D, et al. Case definition and classification of leukodystrophies and leukoencephalopathies. *Mol Genet Metab* 2015;114:494–500. [PubMed: 25649058]
- van der Knaap MS, Bugiani M. Leukodystrophies: a proposed classification system based on pathological changes and pathogenetic mechanisms. *Acta Neuropathol* 2017;134:351–382. [PubMed: 28638987]
- Wasserstein MP, Andriola M, Arnold G, et al. Clinical outcomes of children with abnormal newborn screening results for Krabbe disease in New York State. *Genet Med* 2016;18:1235–1243. [PubMed: 27171547]

Table 1:

Leukodystrophies, associated genes, disease super allele frequencies, individual disease incidence (1 in × births), and predicted disease incidence per million births.

Disease	Gene	SAF	1 in × births	per million
Aicardi-Goutieres syndrome	ADAR	0.0029	73,494	13.61
	RNASEH2A	0.0005		
	RNASEH2B	0.0017		
	RNASEH2C	0.0002		
	SAMHD1	0.0003		
	TREX1	0.0014		
Adult polyglucosan body disease	GBE1	0.0033	912,000	1.096
Canavan disease	ASPA	0.0016	376,500	2.656
Cerebrotendinous xanthomatosis	CYP27A1	0.0019	287,000	3.484
CIC-2 Chloride Channel Deficiency	CLCN2	0.0021	235000	4.255
D-bifunctional protein deficiency	HSD17B4	0.0006	2,580,000	0.388
Fucosidosis	FUCA1	0.0002	30,500,000	0.033
Hereditary diffuse leukoencephalopathy with spheroids and pigmentary leukodystrophy	CSF1R	0.0003	15,400.000	0.065
Hypomyelination with brainstem and spinal cord involvement and leg spasticity	DARS	0.0006	594.082	1.683
	DARS2	0.0012		
Hypomyelination with congenital cataract	FAM126A	0.0001	65,000,000	0.015
Krabbe Disease	GALC	0.0091	12,080	82.78
Leukoencephalopathy with thalamus and brainstem involvement and high lactate	EARS2	0.0011	812,000	1.232
Metachromatic Leukodystrophy	ARSA	0.0030	112,700	8.873
Megalencephalic Leukodystrophy	HEPACAM	0.0002	1,133,321	0.882
	MLC1	0.0009		
Oculodentodigital Dysplasia	GJA1	0.0007	2,080,000	0.481
Pelizaeus-Merzbacher-Like Disease	GJC2	0.0042	57,000	17.54
Peroxisomal acyl-CoA oxidase deficiency	ACOX1	0.0001	61.000,000	0.016
POLIII-related disorders	POLR3A	0.0033	26,160	38.23
	POLR3B	0.0056		
RNAse T2 deficient leukoencephalopathy	RNASET2	0.0001	109,500,000	0.009
Salla disease	SLC17A5	0.0013	567,000	1.764
Sjogren-Larsson syndrome	ALDH3A2	0.0005	3,500,000	0.286
SOX10-associated disorders	SOX10	0.0001	83,000.000	0.012
Sterol Carrier Protein X	SCP2	0.0003	1,700,000,000	0.001
Vanishing White Matter Disease	EIF2B1	0.0004	714.594	1.399
	EIF2B2	0.0008		
	EIF2B3	0.0003		
	EIF2B4	0.0000		

Disease	Gene	SAF	1 in × births	per million
	EIF2B5	0.0008		
Zellweger spectrum disorder	PEX1	0.0017	48.691	20.54
	PEX2	0.0003		
	PEX3	0.0001		
	PEX5	0.0004		
	PEX6	0.0040		
	PEX7	0.0010		
	PEX10	0.0003		
	PEX12	0.0005		
	PEX13	0.0001		
	PEX14	0.0000		
	PEX16	0.0001		
	PEX19	0.0001		
	PEX26	0.0002		
Overall Leukodystrophy Incidence			1 in 4,733	211 per million