Population-based Incidence Estimates of Classical Homocystinuria Using the Genome Aggregation Database (gnomAD)

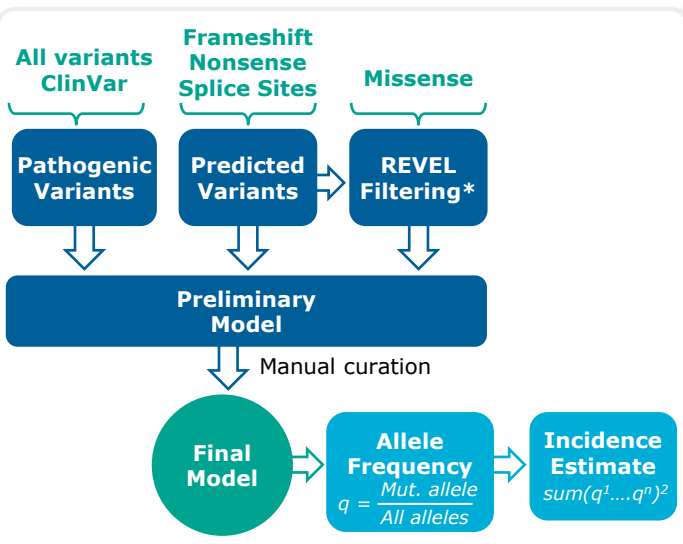
Ida Vanessa D. Schwartz[1], Tiziano Pramparo[2], Henk J. Blom[3], Steve Rodems[2], Frederick P. Roth[4]

[1]Medical Genetics Service/Reference Center for Rare Disorders, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil & Department of Genetics, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil; [2]Travere Therapeutics, San Diego, CA, USA; [3]Center for Lysosomal and Metabolic Diseases, Erasmus MC, Rotterdam, The Netherlands; [4]Donnelly Centre, Molecular Genetics & Computer Science, University of Toronto, Toronto, Canada & Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada

METHODS

- Variants' information was obtained from querying the ClinVar, HGMD, LOVD, and VarSome databases and literature resources **(Table 1)**
- The final list of variants in the model calculation included three groups:
  - Variants with a ClinVar pathogenic label
  - Frameshift/nonsense/canonical splice variants considered pathogenic based on sequence consequence
  - Missense variants predicted to be pathogenic by the REVEL (rare exome variant ensemble learner) software after passing a disease-specific cut-off **(Figure 1)**
- HCU incidence estimates by population were calculated according to the Hardy-Weinberg principle $(p^2 + 2pq + q^2)$[9] under the assumption of mutual independence of rare variants and complete penetrance

Figure 1. Variant Selection and Analysis Strategy



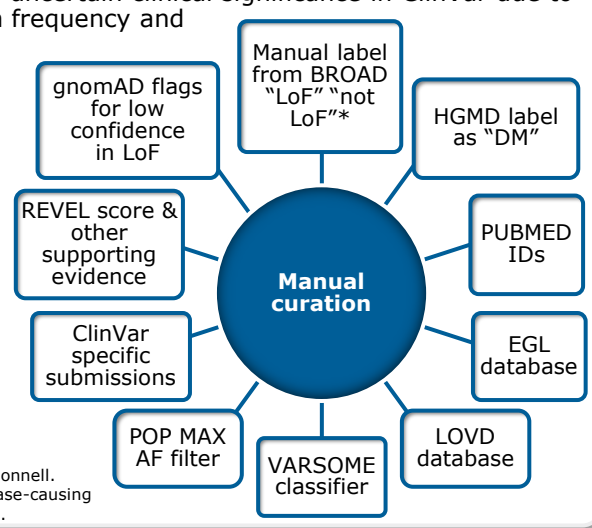
*HCU-specific cut-off was calculated as the REVEL mean score of high-confidence pathogenic missense variants (common pathogenic variants between ClinVar and HGMD + VarSome classification confirmation) after removing outliers with the 1.5 IQR formula. The identified cut-off was 0.90, thus gnomAD missense variants with a REVEL score equal to or greater than 0.9 were retained in the model. IQR, interquartile range; REVEL, rare exome variant ensemble learner.

Table 1. Databases Used for the Study

Genome Aggregation Database (gnomAD) v2.1.1	gnomad.broadinstitute.org
ClinVar	ncbi.nlm.nih.gov/clinvar
Human Gene Mutation Database (HGMD)	www.hgmd.cf.ac.uk/ac/index.php
Mutalyzer	mutalyzer.nl
VarSome	varsome.com
LitVar	ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/LitVar
PubMed	pubmed.ncbi.nlm.nih.gov
Google Scholar	scholar.google.com
Leiden Open Variation Database (LOVD)	databases.lovd.nl/shared/genes
EGL EmVClass	egl-eurofins.com

Figure 2. Manual Curation

- Manual curation was performed to remove low-confidence loss of function variants (lc_lof) and/or minimize possible false positive/negative from the final model (**Figure 2**)
- Incidence was calculated as the squared sum of the carrier AF of the mutant alleles (q) with p approximated to 1
- Variant c.1105C>T (p.Arg369Cys) was not considered pathogenic since it remains of uncertain clinical significance in ClinVar due to its high population frequency and inconclusive segregation with disease
  - Similarly, two variants that fall in the regulatory domain of the *CBS* gene were not considered pathogenic as they are known to have less impact on CBS activity



*Kindly provided by Dr. A. O'Donnell.
AF, allele frequency; DM, disease-causing mutation; LoF, loss of function.

CONCLUSIONS

- Estimates of classical homocystinuria incidence varied widely across gnomAD populations, with the largest estimates in ancestries with best-characterized variation
- Incidence ranged from 1:65,000 in the European non-Finnish population to 1:4,000,000 in the East Asian population
- In populations for which clinical variation has been most intensively studied, the calculated incidence is much higher than historical estimates and in line with real-world healthcare evidence

RESULTS

- We identified 1,294 total variants, of which 358 were missense. Fifty-five (55) variants were designated pathogenic in ClinVar, 23 were considered pathogenic based on sequence consequence after filtering and manual curation, and 38 missense variants were predicted to be pathogenic based on REVEL in-silico analysis and ClinVar or literature and functional evidence **(Table 2)**
- The 116 pathogenic variants identified were used to calculate incidence estimates (**Table 3**) and distribution is depicted in **Figure 3**

Table 2. 116 Pathogenic Variants Were Identified*

Variant Type	ClinV-P	SeqCon-P	REVEL-P	Model
Missense	41	–	38	**79**
Splice region	2	–	–	**2**
Frameshift	6	9	–	**15**
Stop gained	2	4	–	**6**
Splice donor	2	2	–	**4**
Stop lost	–	2	–	**2**
Splice acceptor	2	3	–	**5**
Start lost	–	3	–	**3**
Total	**55**	**23**	**38**	**116**

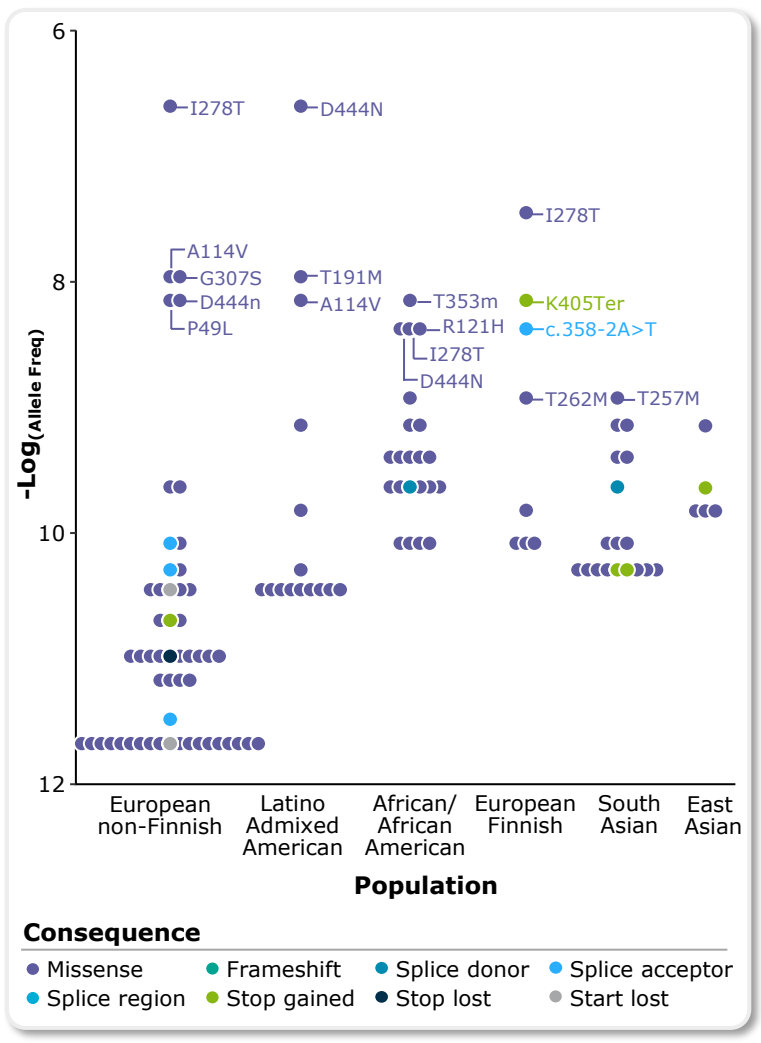
***CBS* gnomAD variants.** Variant count breakdown by type: designated pathogenic based on ClinVar clinical significance label (ClinV-P); predicted pathogenic based on sequence consequence (SeqCon-P); predicted pathogenic based on REVEL analysis (REVEL-P), and final list of variants selected for the incidence calculation (Model).

Table 3. HCU Incidence

	Rate per 100,000	1 per Estimate	Pooled AF
European non-Finnish	1.54	65,012	0.00392
Latino Admixed American	0.60	166,689	0.00245
African/ African American	0.54	183,742	0.00233
European Finnish	0.19	513,120	0.0014
South Asian	0.15	653,536	0.00124
East Asian	0.02	4,027,114	0.0005

The gnomAD v2.1.1 database was accessed on January 04, 2022, to retrieve variants' AFs across six main populations. Other populations were excluded from the analysis due to smaller sample size representation in gnomAD. HGVS, Human Genome Variation Society.

Figure 3. HCU Pathogenic Alleles Distribution from gnomAD



Distribution of the HCU pathogenic alleles with labels for the top alleles (Allele_Freq >1e-04) and color-coded by sequence consequence. Variant names follow the HGVS transcript or protein consequence. Amino acids are abbreviated by their one letter code.

DISCLOSURES

IVDS: none; **TP:** Employee and stockholder, Travere Therapeutics, Inc.; **HJB:** none; **SR:** Employee and stockholder, Travere Therapeutics, Inc.; **FPR:** Shareholder, Ranomics, Inc. Scientific advisory board member and shareholder, SeqWell, Constantiam Biosciences, Inc. and BioSymetrics, Inc. Sponsored research, Alnylam, Inc., Biogen Inc., Deep Genomics, Inc., Beam Therapeutics, Inc.

ACKNOWLEDGMENTS

This study was supported by Travere Therapeutics, Inc. (San Diego, CA). Medical writing and layout assistance was provided by Heather Hartley-Thorne, Sephirus Communications (Los Angeles, CA) and was funded by Travere Therapeutics, Inc.

REFERENCES

1. Sacharow SJ, et al. In: Adam MP, Mirzaa GM, Pagon RA, et al., eds. *GeneReviews®*. Seattle (WA)1993. **2.** Moorthie S, et al. *J Inherit Metab Dis.* 2014;37(6):889-898. **3.** Morris AA, et al. *J Inherit Metab Dis.* 2017;40(1):49-74. **4.** Gaustadnes M, et al. *N Engl J Med.* 1999; 340(19):1513. **5.** Refsum H, et al. *J Pediatr.* 2004;144(6):830-832. **6.** Janosik M, et al. *J Pediatr.* 2009;154(3):431-437. **7.** Sellos-Moura M, et al. *Mol Genet Metab.* 2020;130(1):36-40. **8.** Sellos-Moura M, et al. *BMC Health Serv Res.* 2020;20(1):183. **9.** Weber GR, et al. *Mol Genet Genomic Med*. 2020;8(6): e1214.

To obtain a PDF of this poster, scan the QR code OR visit:

www.traverepublications.com/SSIEM2022/PO10-2316/

No personal information is stored.

#PO10-2316

TRAVERE THERAPEUTICS

INTRODUCTION

- Classical homocystinuria (HCU) is a slowly progressive genetic disease due to pathogenic variants in the *cystathionine beta-synthase* (*CBS*) gene, which leads to elevated homocysteine in the body[1]
- HCU is characterized by developmental delay/intellectual disability, *ectopia lentis* and/or severe myopia, skeletal abnormalities (excessive height and length of the limbs), and thromboembolism[1]
- Literature reviews based both on improved newborn screening and clinical data[2,3] suggest that the prevalence of HCU, historically estimated at ~1:200,000 to 1:335,000 worldwide,[1] is greatly underestimated
- Calculations of the expected population frequency of (likely) pathogenic variants of *CBS* in European populations have also suggested that the prevalence of HCU is underestimated[4-6]
- Consistent with these findings, real-world healthcare claims data, derived from insurance claims databases, indicated a prevalence of symptomatic cases of HCU to be as high as 1:10,000 in the United States[7,8]

Objective

- To leverage the gnomAD variation database as an orthogonal approach to estimate the incidence of HCU based on population allele frequencies (AFs) across six main ancestries

LIMITATIONS

- The Hardy-Weinberg principle in our genetic risk calculations assumes that heterozygous individuals are not subject to selection, populations are at equilibrium with respect to AFs and genotypes, and random mating is observed
- AFs are strictly related to the size of the population; thus, with the availability of larger population-based datasets, estimates may change and become more accurate
- Although the adoption of in-silico tools to predict pathogenic missense with unknown or uncertain clinical significance are commonly used, there is currently no gold-standard and these strategies could lead to the over- or underestimation of disease frequency
- Estimated risk of pathogenic variants differed based on ancestry, however, incidence estimates were likely underestimated for populations that are less well represented amongst characterized variants submitted to ClinVar
- Additional genetic variation that has not been accounted for in our calculation may be conferred by in-frame deletions or insertions, intronic, non-canonical splice sites, and structural variants
- Our approach assumed that all variants in the final model contribute to the risk of disease with 100% penetrance

Contact information: Medinfo@Travere.com

Presented at the Society for the Study of Inborn Errors of Metabolism (SSIEM) Annual Symposium 2022; August 29-September 2, 2022; Freiburg, Germany