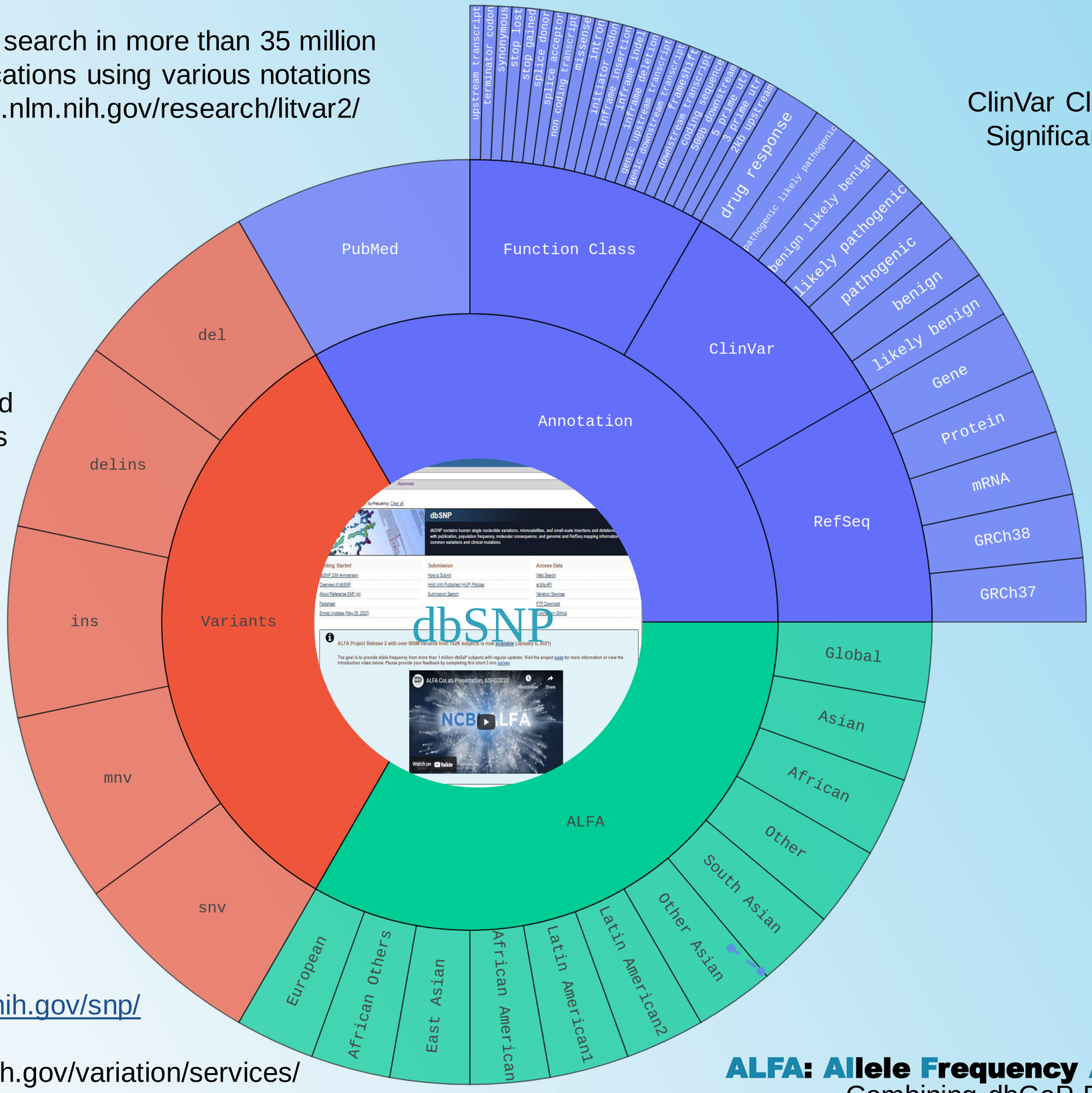
Enhancing dbSNP and ALFA for Genomic Research

L. Phan, Y. Jin, H. Zhang, W. Qiang, E. Shekhtman, D. Shao, D. Revoe, R. Villamarin, E. Ivanchenko, M. Kimura, Z. Y. Wang, L. Hao, N. Sharopova, M. Bihan, A. Sturcke, M. Lee, N. Popova, W. Wu, C. Bastiani, M. Ward, J. B. Holmes, V. Lyoshin, K. Kaur, E. Moyer, M. Feolo, and B. L. Kattman.

U.S. National Library of Medicine
National Center for Biotechnology Information
8600 Rockville Pike, Bethesda MD, 20894 USA

ncbi.nlm.nih.gov

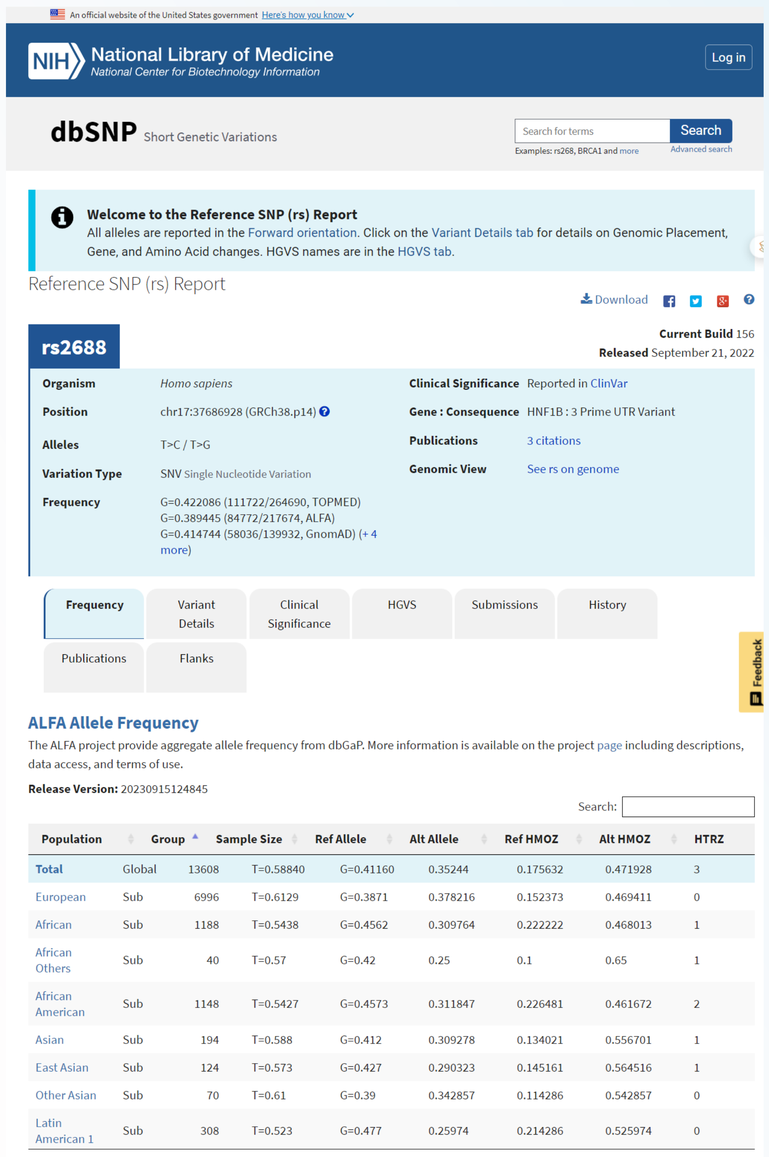
LitVar2 - variant search in more than 35 million biomedical publications using various notations
https://www.ncbi.nlm.nih.gov/research/litvar2/



- >3 billion variants aggregated from thousands of submitters world-wide
- 1.1 Billion unique reference variants (RefSNP)
- 95% RefSNP will population MAF
- Includes common and rare variants and ClinVar mutations

Data Access

Web: https://www.ncbi.nlm.nih.gov/snp/
FTP: ftp.ncbi.nih.gov/snp/
API: https://www.ncbi.nlm.nih.gov/variation/services/



Population	Group	Sample Size	Ref Allele	Alt Allele	Ref HMOZ	Alt HMOZ	HTRZ
Total	Global	13608	T=0.58840	G=0.41160	0.35244	0.175632	0.471928	3
European	Sub	6996	T=0.6129	G=0.3871	0.378216	0.152373	0.469411	0
African	Sub	1188	T=0.5438	G=0.4562	0.309764	0.222222	0.468013	1
African Others	Sub	40	T=0.57	G=0.42	0.25	0.1	0.65	1
African American	Sub	1148	T=0.5427	G=0.4573	0.311847	0.226481	0.461672	2
Asian	Sub	194	T=0.588	G=0.412	0.309278	0.134021	0.556701	1
East Asian	Sub	124	T=0.573	G=0.427	0.290323	0.145161	0.564516	1
Other Asian	Sub	70	T=0.61	G=0.39	0.342857	0.114286	0.542857	0
Latin American 1	Sub	308	T=0.523	G=0.477	0.25974	0.214286	0.525974	0

ALFA: Allele Frequency Aggregator

Combining dbGaP Data

dbGaP studies
Millions of subjects
Billions of variants
Trillions of genotypes

ALFA pipeline
Release 3:
Aggregated >5 Trillion genotypes from ~204K subjects
Computed frequency for >904 Million Variants
Minor Allele and Genotype Frequencies for 12 ancestry populations

Open-Access dbSNP
RefSNP
Variation Services API
Trackhubs and FTP Downloads

More coming soon!!!

Improved GrafPop for ALFA genetic ancestry inference