ClinVar Comparison (release 7th November 2017)

Missense variants = 122,416 / 325,448 entries

ClinVar Pathogenic not specified as DNM: 25,357 distinct missense variants
where "ClinSigSimple" eq "1"
and "ClinicalSignificance" ne "Conflicting Evidence*"
and "Origin" ne "*de novo*"

ClinVar Pathogenic "de novo" specified: 855 distinct missense variants
where "ClinSigSimple" eq "1"
and "ClinicalSignificance" ne "Conflicting Evidence*"
and "Origin" eq "*de novo*"

ClinVar Benign: 15,048 distinct missense variants
where "ClinSigSimple" eq "0"
and "ClinicalSignificance" eq "*Benign*"
***note* provided to evidence inappropriateness of using ClinVar Benign as comparator given this classification is influenced by presence of variant in population controls used to generate regional scores.**

Independent population control missense variants using the DiscovEHR cohort (release GHS_Freeze_50.L3DP10.pVCF.frq)

Population Control: Novel Missense (DiscovEHR): 858,306 missense variants. This is on average 17 novel missense variants per individual (crude estimate as relatedness structure exists within DiscovEHR). *Novel = not reported in gnomAD, thus have not been adopted in construction of regional intolerance scores*.

Population Control: Novel Missense in ClinVar Genes (DiscovEHR): 179,062 / 858,306 novel missense variants occur in genes that are represented among the ClinVar Pathogenic missense variant collections. This subset better controls for contamination from non-disease genes in the assessments of regional intolerance utility.