

The Genome Aggregation Database (gnomAD)

Konrad Karczewski

March 4, 2019



broad.io/gnomad_lof



@konradjk

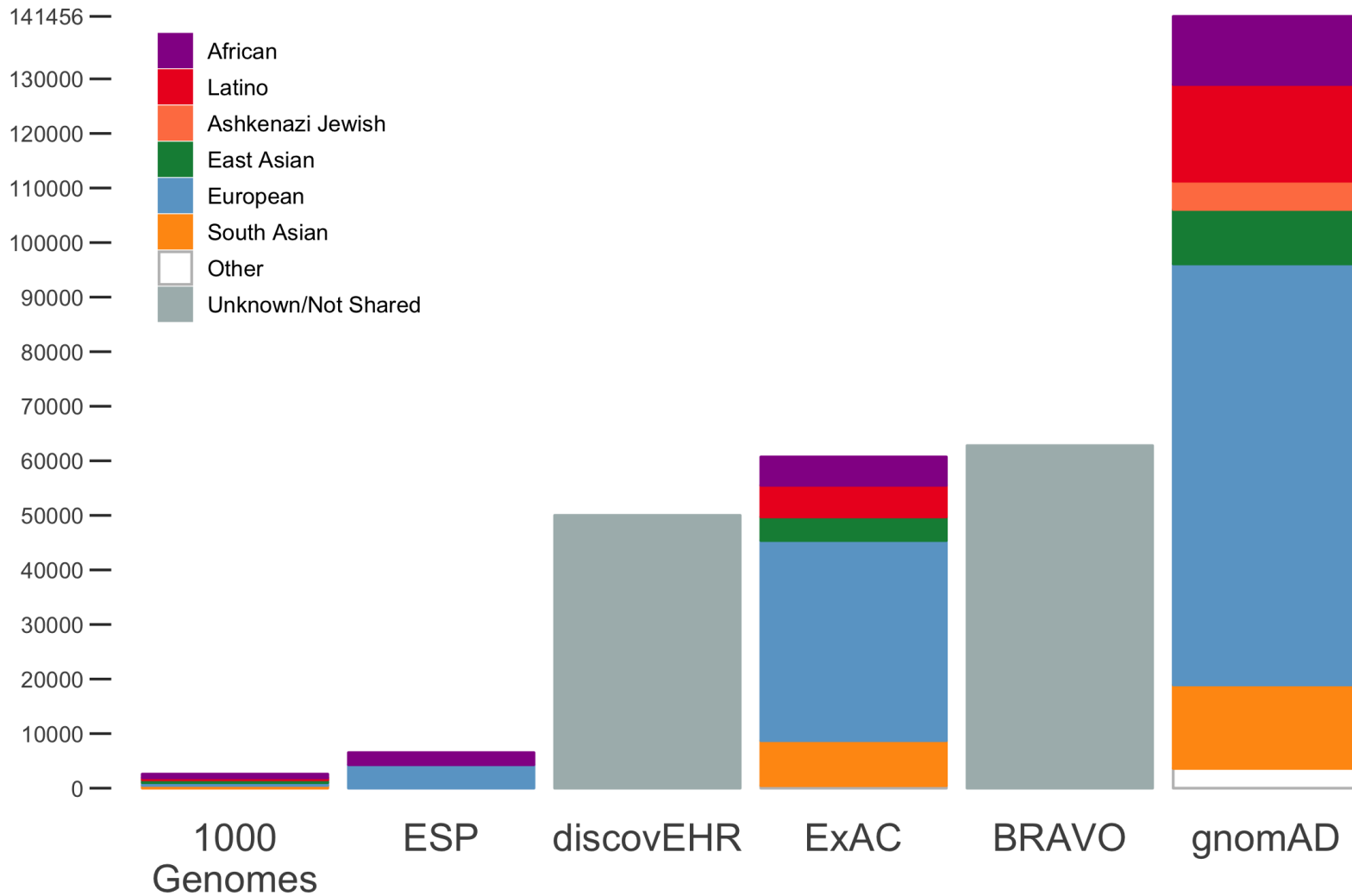
Identifying true LoF variants is challenging

- LoFs are rare
- LoFs are enriched for artifacts

Identifying true LoF variants is challenging

- LoFs are rare
- LoFs are enriched for artifacts

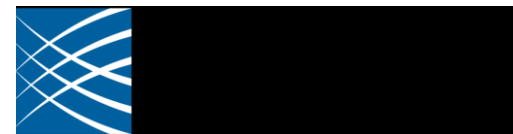
Increasing the scale of reference databases



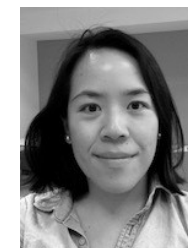
- **gnomAD**: 125,748 exomes and 15,708 whole genomes

gnomAD 2.1.1

- Data provided by 109 PIs
 - 1.3 and 1.6 petabytes of BAM files
- Uniformly processed and joint called
 - 12 and 24 terabyte VCFs
- Developed a novel QC pipeline
 - Complete pipeline publicly available: broad.io/gnomad_qc
- All QC and analysis performed using Hail: hail.is
 - Scalable to thousands of CPUs
 - Enabled rapid iteration (few hours for each component, few days for entire process)



Broad Genomics Platform
Broad Data Sciences Platform



Grace
Tiao



Laurent
Francioli



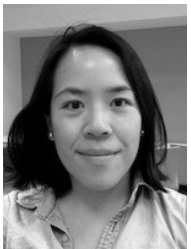
Cotton Seed



Tim Poterba

gnomAD 2.1.1

- Sub-continental ancestry
- Subsets:
 - controls-only
 - non-neuro/non-psychiatric
 - non-cancer
 - non-TOPMed Bravo



Grace
Tiao



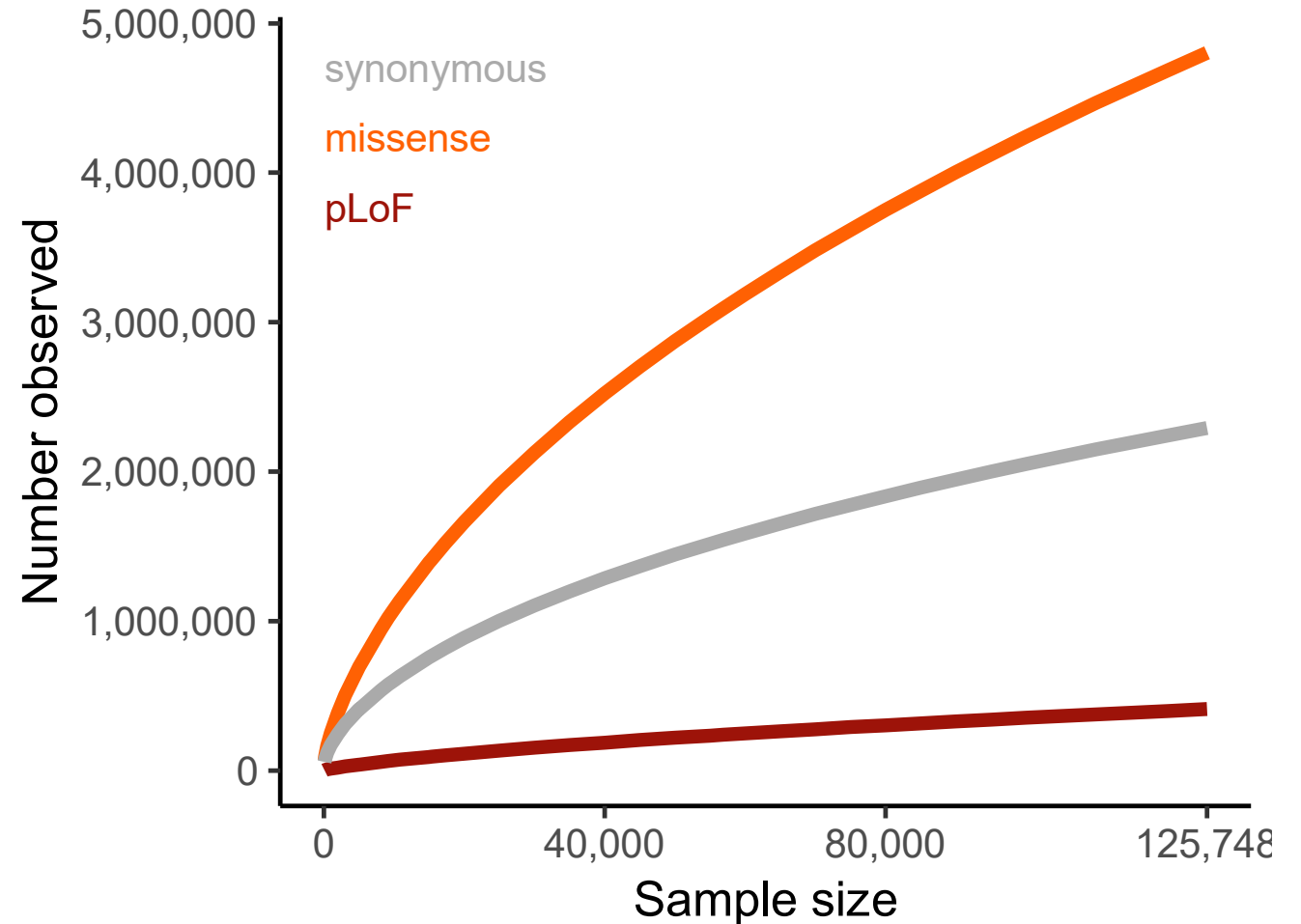
Laurent
Francioli

<http://gnomad.broadinstitute.org>



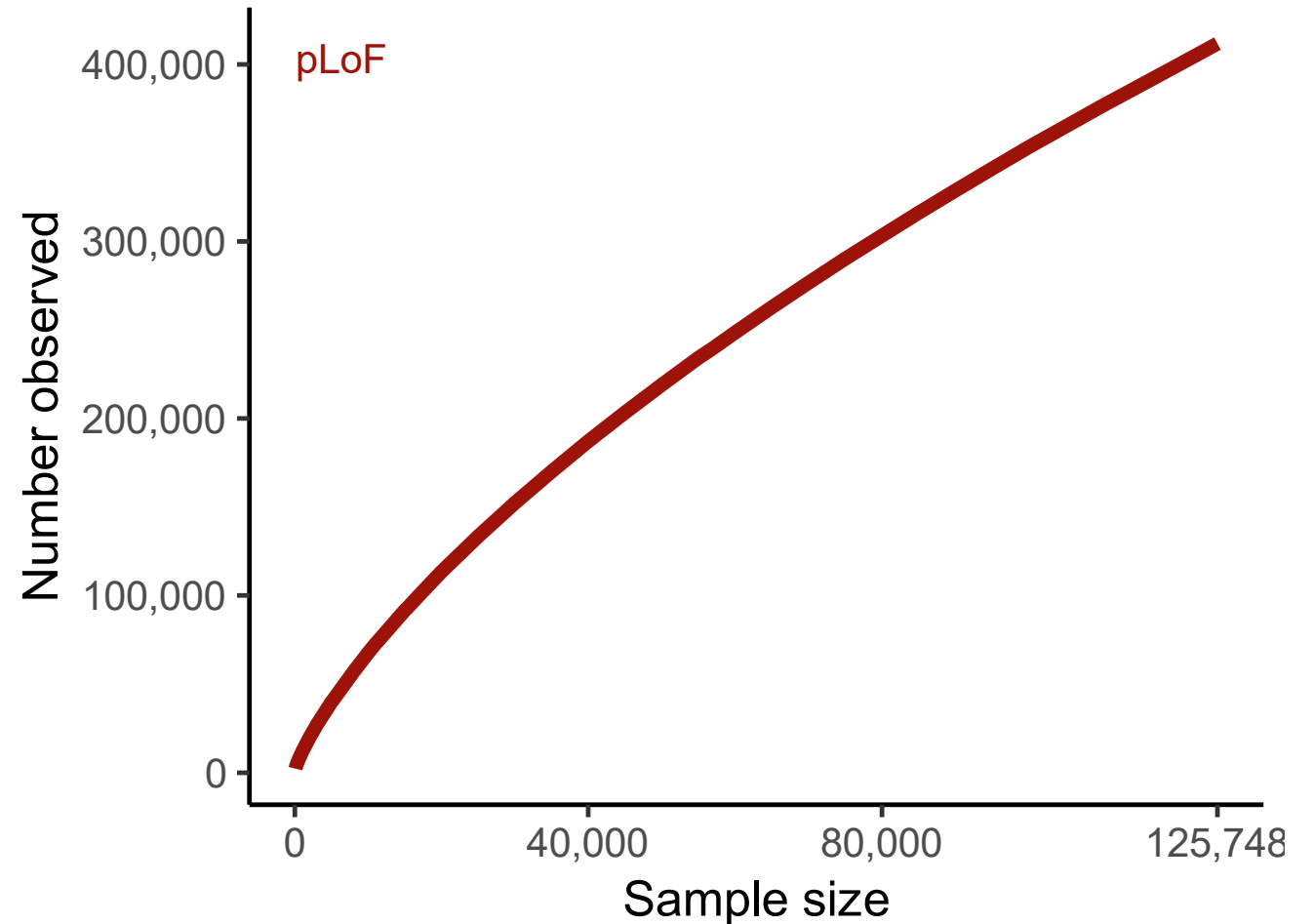
Staggering amounts of variation

- gnomAD 2.1 contains:
 - 230M variants in 15,708 genomes
 - 15M variants in 125,748 exomes



Staggering amounts of LoFs

- gnomAD 2.1 contains:
 - 230M variants in 15,708 genomes
 - 15M variants in 125,748 exomes
- Of these, we observe 515,326 loss-of-function (LoF) variants
 - Stop-gained
 - Essential splice
 - Frameshift indel

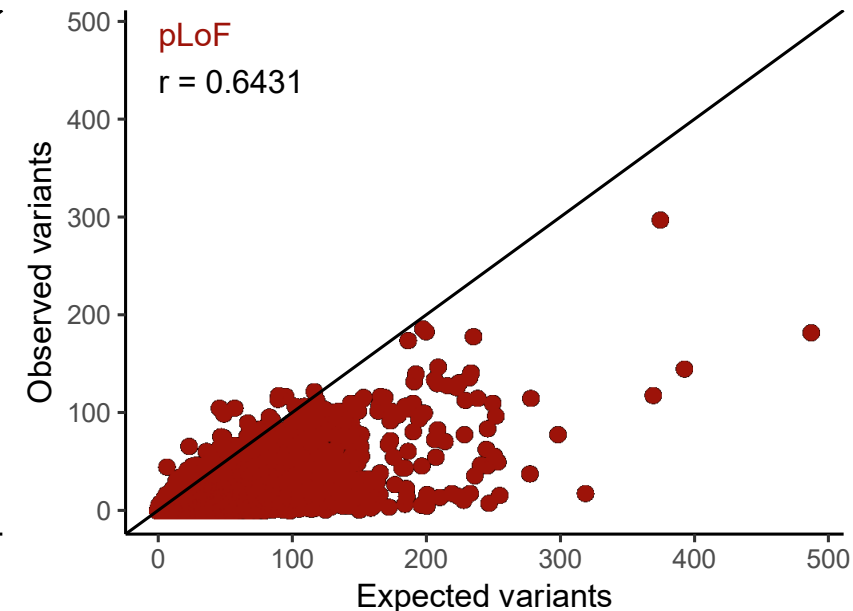
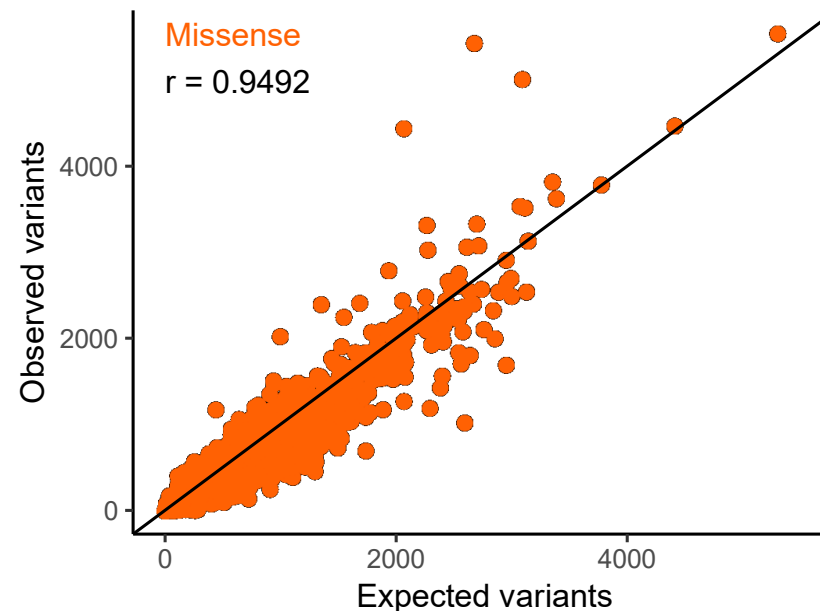
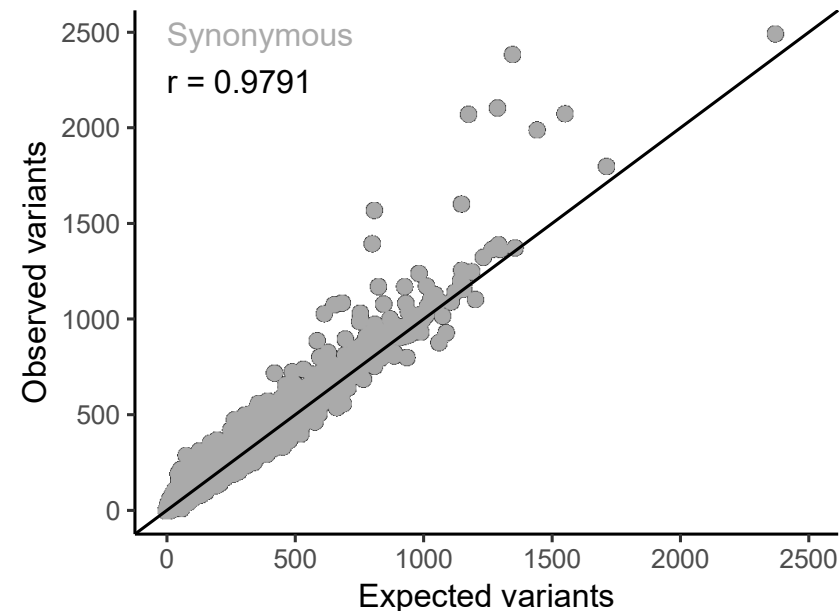


Detecting genes depleted for LoFs

- Mutational model that predicts the number of SNVs in a given functional class we would expect to see in each gene in a cohort

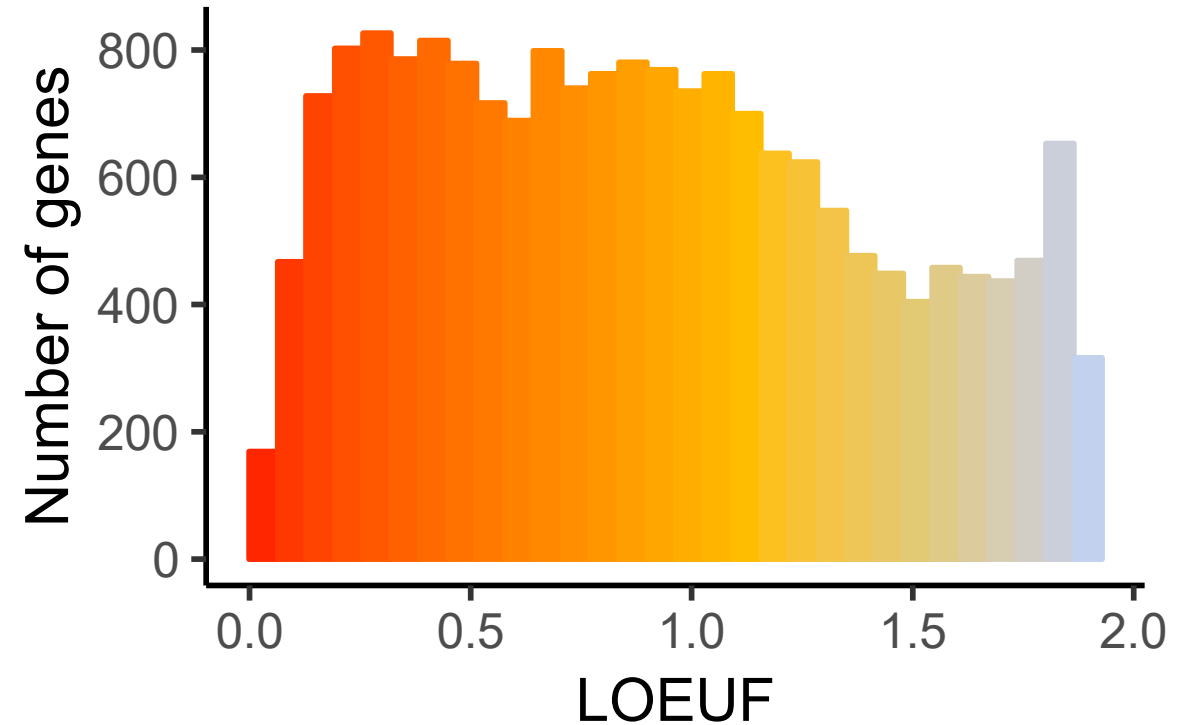


Kaitlin Samocha



Most genes are depleted of LoF variation

- Many are extremely depleted (<20% observed compared to expected)
 - Including most known dominant Mendelian genes
- Using upper bound of confidence interval corrects for small genes



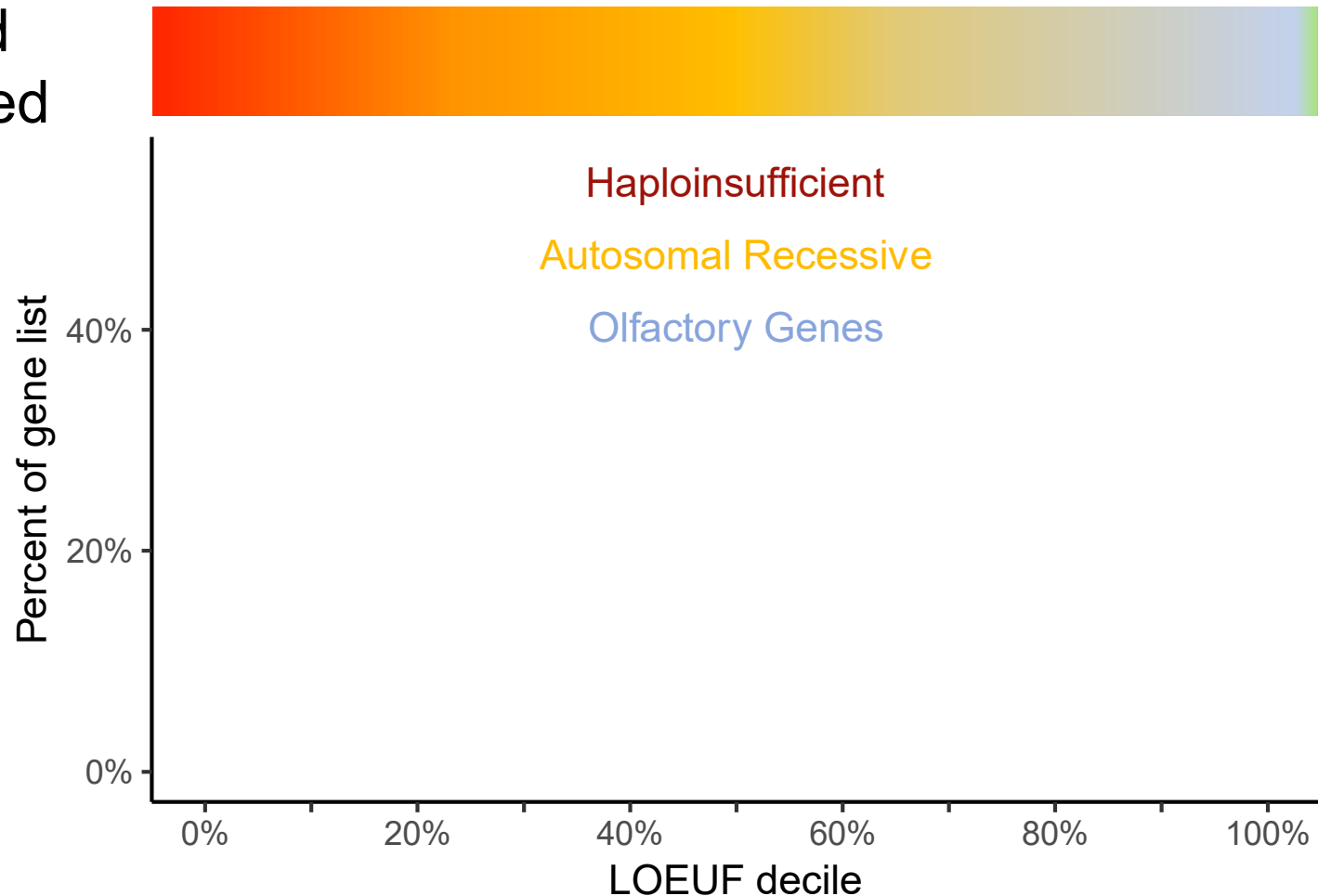
MED13L			
Phenotype	Severe Intellectual Disability		
	Observed	Expected	Obs/Exp (CI)
Synonymous	462	465	0.993 (0.92-1.07)

Resolving the spectrum of LoF intolerance

- Binning this spectrum into deciles

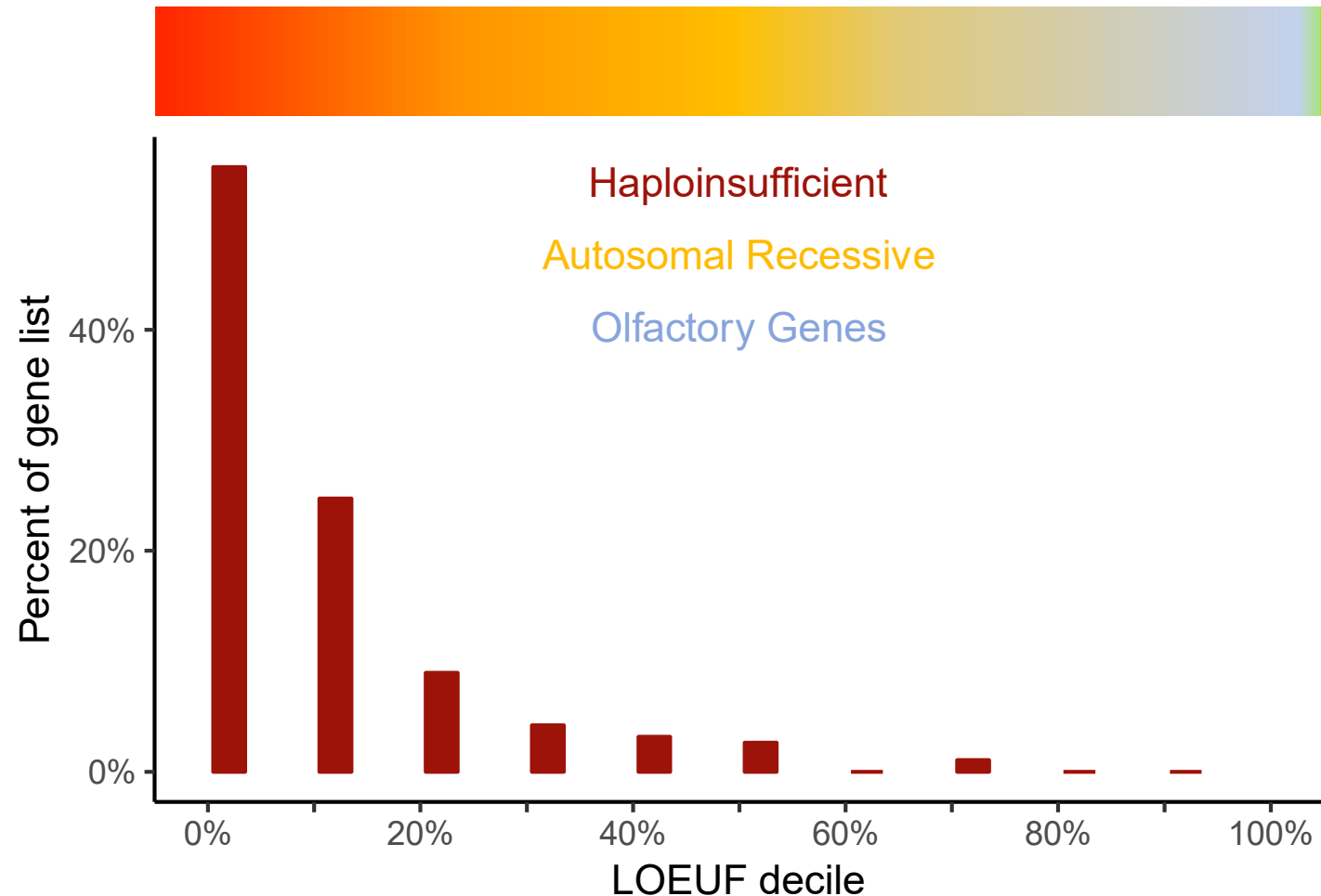
More depleted
More constrained

More tolerant
Less constrained



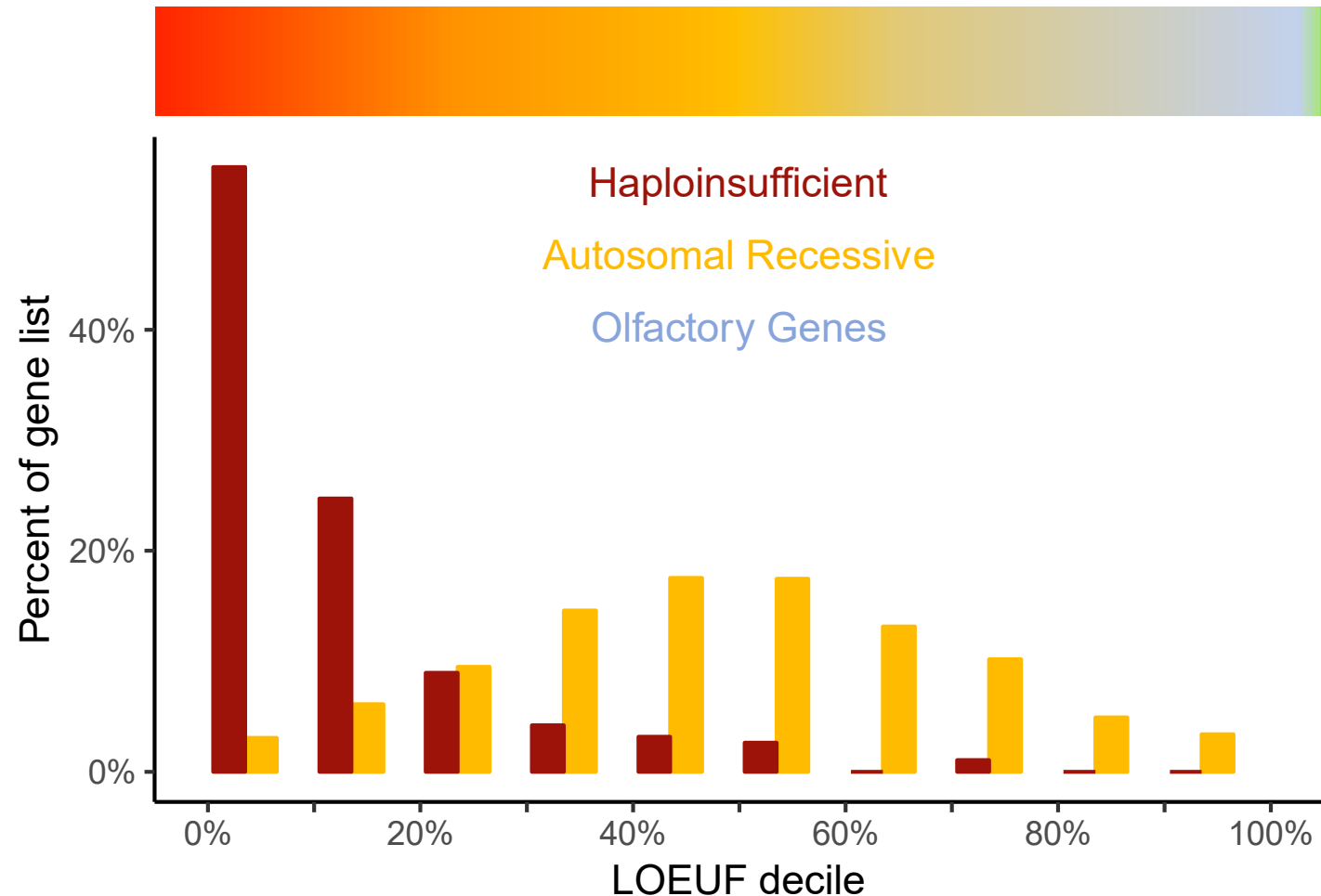
Resolving the spectrum of LoF intolerance

- Known haploinsufficient genes have ~10% of the expected LoFs



Resolving the spectrum of LoF intolerance

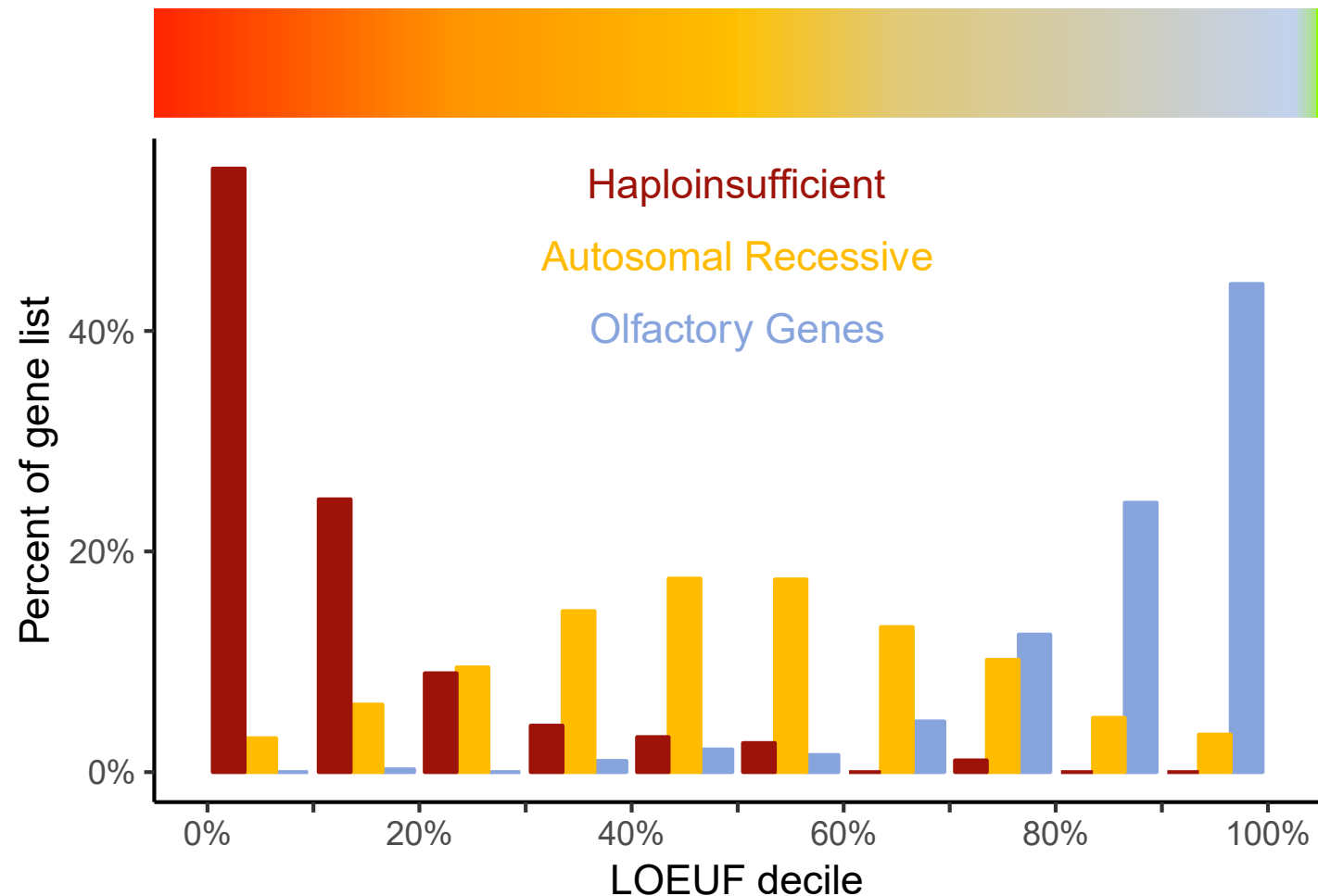
- Autosomal recessive genes are centered around 60% of expected



Gene list from:
Blekhman et al., 2008
Berg et al., 2013

Resolving the spectrum of LoF intolerance

- Some genes, e.g. olfactory receptors, are unconstrained



Data publicly released with no publication restrictions

gnomad.broadinstitute.org



Matt Solomonson



Nick Watts

Gene model with transcripts

Pathogenic Clinvar Variants

gnomAD browser [About](#) [Downloads](#) [Terms](#) [Contact](#) [Jobs](#) [FAQ](#)

CHD8 chromodomain helicase DNA binding protein 8

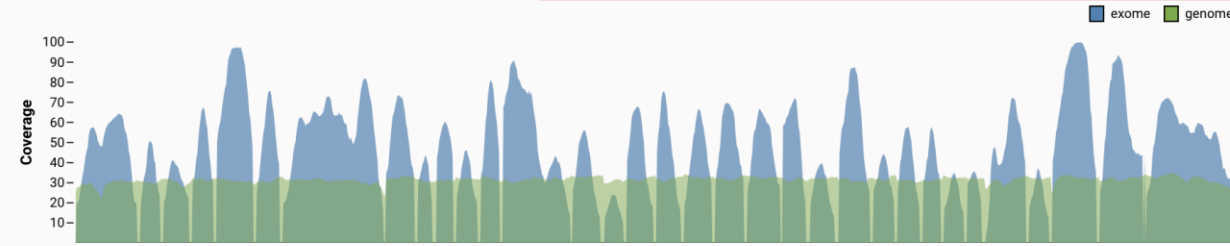
Current Dataset:

Dataset selection box

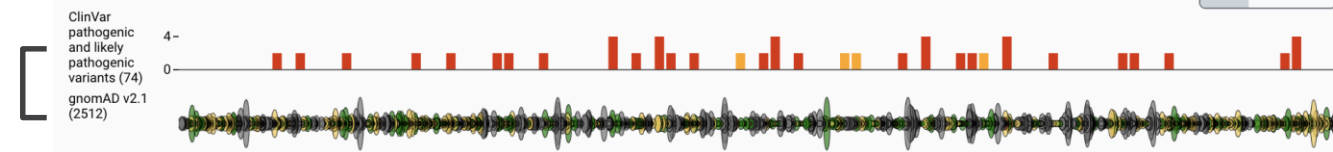
Ensembl gene ID: [ENSG00000100888](#)
 Ensembl transcript ID: [ENST00000399982 \(canonical\)](#)
 Number of variants: 2,674 (including filtered variants)
 UCSC Browser: [14:21853354-21924286](#)
 GeneCards: [CHD8](#)
 OMIM: [610528](#)

Constraint metrics

Category	Exp. no. variants	Obs. no. variants	Constraint metrics		
Synonymous	505.4	468	Z = 1.30	$o/e = 0.93$ (0.86 - 1.00)	0 — 1
Missense	1450.1	792	Z = 6.27	$o/e = 0.55$ (0.51 - 0.58)	0 — 1
LoF	127.0	5	pLI = 1.00	$o/e = 0.04$ (0.02 - 0.08)	0 — 1



Tissue isoform expression



Viewing in table LoF only Missense only Synonymous only Other only Include filtered variants

Variant ID	Source	Consequence	Annotation	Flags	Allele Count	Allele Number	Allele Frequency	Number of Homozygotes
14-21853734-ACAG-A		c.*35_*37d...	3' UTR		1	31272	3.198e-5	0
14-21853735-C-T		c.*37G>A	3' UTR		3	168770	1.778e-5	0
14-21853735-C-G		c.*37G>C	3' UTR		1	137456	7.275e-6	0