

# Ubiquity vs specificity of gene and isoform expression across populations

Population	Individuals	BAM files
CEU	92	188
FIN	94	128
GBR	93	137
TSI	65	81
YRI	89	141

Pipeline: GEM mapper + MISO (isoform quantitation)

Subset: protein coding genes with +1 isoform (n=17,413)

On average, 10,700 genes expressed in each sample

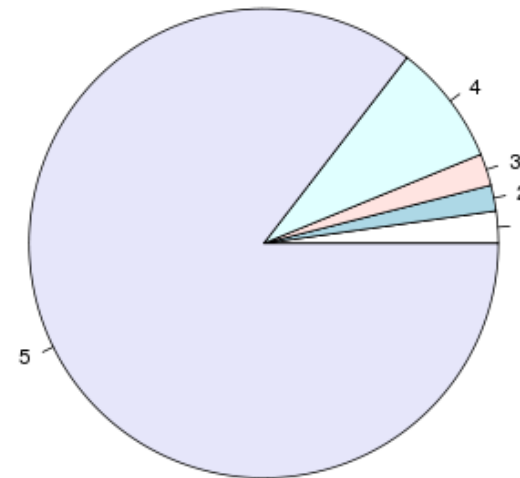
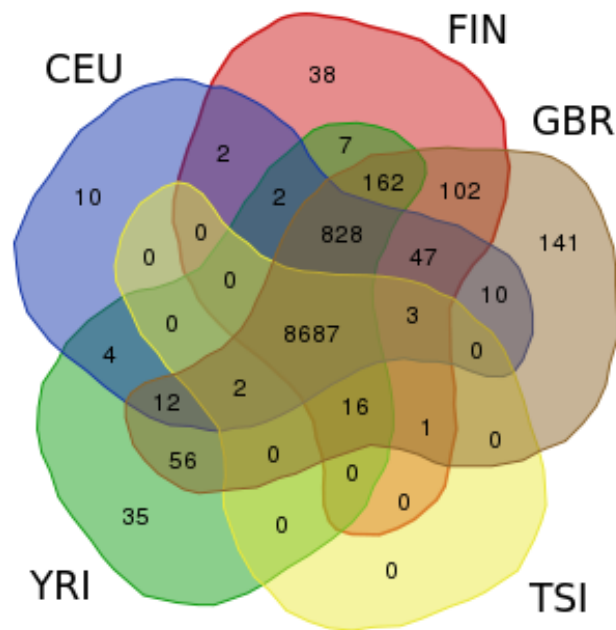
Criteria for ubiquity: 90% of the population

# Ubiquity vs specificity of gene and isoform expression across populations

Population	Individuals	BAM files	Ubiquitous genes
CEU	92	188	9608
FIN	94	128	9895
GBR	93	137	10067
TSI	65	81	8709
YRI	89	141	9811

On average, 10,700 genes expressed in each sample

In each population, ~90% of the expressed genes are ubiquitously detected

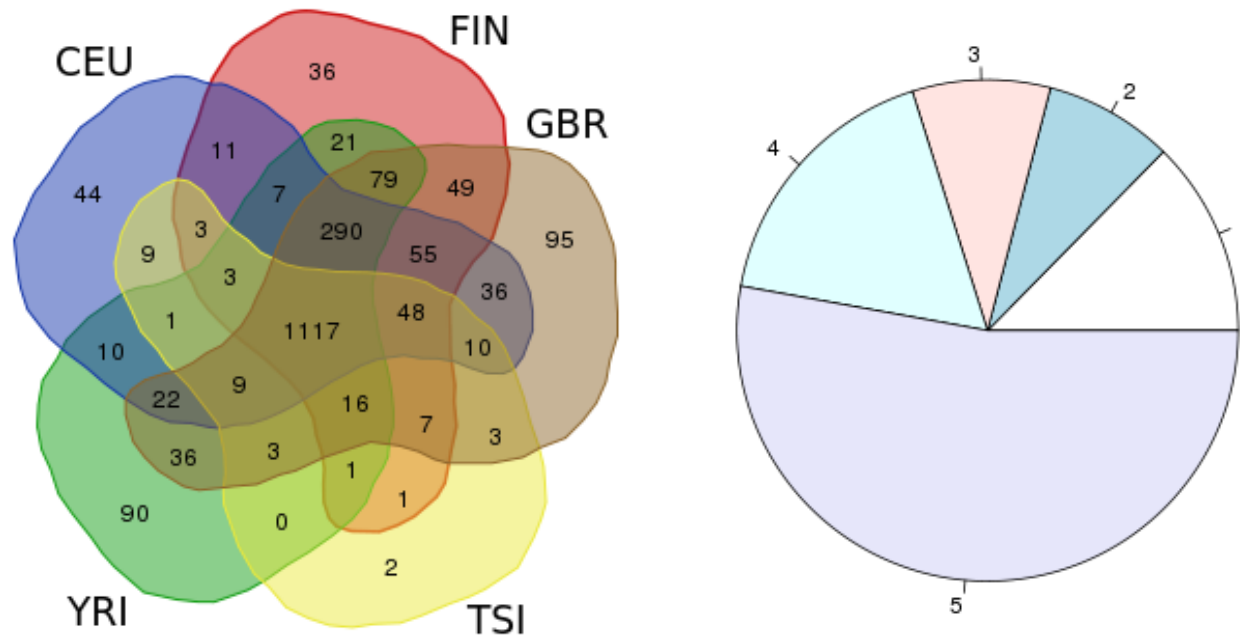


- In each population, ~90% of the expressed genes are ubiquitously detected
- 85% of those genes are consistently ubiquitous across populations
- 4% are specific from 1 or 2 populations:  
*plasma membrane, cell surface, cell junction, signal peptide, transmembrane, alternative splicing*

# Ubiquity vs specificity of gene and isoform expression across populations

Population	Individuals	BAM files	Ubiquitous genes	Common predominant isoform	%
CEU	92	188	9608	1675	17.43
FIN	94	128	9895	1744	17.63
GBR	93	137	10067	1875	18.63
TSI	65	81	8709	1233	14.16
YRI	89	141	9811	1705	17.38

In each population, ~17% of the ubiquitous genes have consistent predominant isoforms



- In each population, ~17% of the ubiquitous genes have consistent predominant isoforms
- 55% of those are also consistently predominant across populations: *protein transport, translation, RNA binding, RNA processing, cell cycle, protein degradation*
- 25% are specific from 1 or 2 populations: *mutagenesis site, response to virus, mRNA splicing*