1. Look at crappy ones (VHL, MLH1, DSC2) and correlation to size
2. Anaconda
3. Automate ExAC
4. Clinvar??

Plot 1000G correlations as histogram

Plot number of variants per individual across all genes

Barplot of individuals who have a variant in each gene, reverse descending (fraction of genes = y, gene = x)

Table (gene specific) for ASW, CEU, YRI (pop = column, gene = row), entry = proportion with a mutation in that gene

Use largest RefGene entry

Create 3 tables:

ACMG\_Disease: Disease/Gene/Chrom/Start/End/Refgene entry

**ACMG\_Lit: Disease/Prev/Cit**