SNPs: DNA polymorphisms in which each possible sequence is present in at least 1% of the population, so that not all individuals have the same sequence (majority of these are 1 nt)

- Coding Region, change AA sequence: a physical variation where some may cause genetic disorders
- Coding Regions, doesn't change AA sequence: may affect splice sites
- Promoter/Regulatory region: may affect gene expression
- Other Regions: no known impact, but can be used as markers

- Medical Implications: SNPs can cause diseases, LDL Receptor SNP can reduce efficiency of the gene and lead to inc. risk of higher cholesterol; CYP2D8 SNP can disrupt debrisoquine breakdown which is used to treat high blood pressure.

- Synonymous SNP: coding region, but AA seq stays the same
- Nonsynonymous SNP: coding region, but the AA sequence is changed
- Noncoding SNP: found in a noncoding region

- dbSNP: hosted by NCBI, have access to both a primary db or submitted SNPs and a curated db of reference SNPs(includes both MNPs and indels)

- SNPEffect: dev'd 2012, focuses on effect SNPs have on prots and their function, includes predictors of protein stability, folding, and chaperones effect predictions

- SNPedia : wiki that collects information from publications about SNPs and their clinical effects

- 1000 Genomes Exon Project: NGS inc. discovery of SNPS; 1000 Genome had an exon pilot project found 12,000 SNPs, with 70% determined as novel, and 74% seen with an allele frequency of less than 1%.

- NCBI 1000 Genome Browser: Search for genomic regions, tracks from dbs like ClinVar and dbSNP. Genotypic information from various ethnicities can be found.

ClinVar: curated database for possible clinical/medical relevance of SNPs, released 2013 and only focuses on humans.

- Searching ClinVar: use tags [dis] for disease or [gene] for genes to get more specific results.
- Results: CFTR[gene] returns all gene records for that gene, and disease records that are associated with those
- Filters: extensive and descriptive, filters like Clinical significance, review status, molecular consequence, and variation type

- ClinVar can be selected as a track in UCSC Genome/Table Browser, where it overlaps with dbSNPs variants that are flagged as clinically significant

- Ensembl provides indirect access to Clinvar, uses data from Provean(swift), PolyPhen

Copy Number Variation (CNV):

- Sebat et al(2004) showed 76 CNVs and concluded that individuals in a population can differ based on their CNVs (though no CNVs found on the X chr. at this time)
- Wong et al(2007) found 3654 autosomal CNVs, where 800 of those existed in more than 3% of the population of study. 77% of these were novel, and many were sensory related which could indicate a relation to different populations having different taste/smell

- Hastings et al examined how novel CNVs form, suggested a model of Microhomology-mediated Break Induction Replication(MMBIR), could contribute to reason behind CNV hotspots in the genome
- Kidd et al found multiple structural CNVs, and variants in inbred mice, lending to the theory that CNVs can occur spontaneously.
- Su et al developed polyHap 2.0 to examine CNV haplotypes, predicts using SNP/CNV genotypes

CNV Tools and Databases:

- Database of Genomic Variants(DGV): maps to hg19 default but hg38 available; DGV structural variation tracs at the UCSC Genome browser can be used.
- DECIPHER(DatabasE of Chr. Imbalance and Phenotype in Humans using Ensembl Resources)
- dbVar incorporates multiple dbs, including DGV.

- cn.mops (R) analyzes NGS data to find CNVs in individual genomes by examining depth of coverage, which is higher in regions that are CNVs.
- CNVtools (R) designed to look for CNV association in case v control studies

- IGV, Cancer, and CNVs: the TCGA runs Firehose(hg19) which shows choices for dif forms of cancer, and with those you can choose CNVs.
- CNV in variation viewer: Includes exon pattern, reveals alternative transcripts in their models
- Gnomad(Genome Aggregation Database): search by gene name, ensembl transcript ID, dbSNP ID, or chromosomal region. Graphics view shows peaks that ID how many protein coding SNPs are in that region