Ensembl:

- Provides genome browser for sequenced genomes of chordates and model organisms, updated around every 4 months, data is stored in MySQL relational databases (species-specific & multi-species
- Redundant: Ensembl automated annotations AND VEGA genes manually annotated.
- Gene Identifiers: human genes begin with ENSG, end with unique 11-digit ID for each gene (other species like *Mus muculus* = ENSMUSG, have their own IDs)
    - Location tab = chr location with genomicalignments and synteny maps
    - Gene tab = all known transcripts/proteins with info on ortho/paralogs, also genetic variation data
    - Tscript tab = all info about particular tscript, exons, cDNA, and protein from that specific one.
    - Variation tab = ACCNsof SNPs for that gene, with info on which tscriptscontain which SNPs

Biomart:

- Allows user to download sets from Ensembl, can access biomart at the top of the homepage.
- Filters after species is selected:
    - Region: areas of the chromosome of interest
    - Gene: restrict by genes associated with disease
- Attributes: categories of features(gene, external, prot. domains/family), and different than filters because it adds parameters of other database IDs to the search
- Preview shows how many results from a query
- Sequences will be in FASTA format

Bioconductor & biomaRt

- Bioconductor: open source to use R to mine data from various databases and manipulate that data in order to visualize high throughput data. Useful as genome scale research impractical to use repeated lookups
- biomaRt: pkg in R to access data from Ensembl & other sources from Biomart(req. internet connection).
    - getBm command: four major arguments
        - Attributes: columns in output, can list with "listAttributes()"
        - Filters: similar to the web form, can limit to specific ACCN in gene databases
        - Values: specific to those filters, refers to gene values by which to limit
        - Mart: the data source for this object