# Finding data: Publically available data sources

Sequence databases:

* GenBank [www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/) genetic sequence database, an annotated collection of all publicly available DNA sequences
* ENA European Nucleotide Archive <http://www.ebi.ac.uk/ena> comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation, contained in GenBank
* SRA Sequence Read Archive <https://www.ncbi.nlm.nih.gov/sra> stores raw sequencing data and alignment information from high-throughput sequencing platforms
* dbGaP database of Genotypes and Phenotypes <https://www.ncbi.nlm.nih.gov/gap> archives and distributes the data and results from studies that have investigated the interaction of genotype and phenotype in Humans
* dbSNP Short Genetic Variation <https://www.ncbi.nlm.nih.gov/SNP/> archive for genetic variation within and across different species

Human/vertebrate specific genome databases

* 1000 Genomes <http://www.internationalgenome.org/> The goal of the 1000 Genomes Project was to find most genetic variants with frequencies of at least 1% in the populations studied
* ExAC Exome Aggregation Consortium <http://exac.broadinstitute.org/about> aggregates and harmonizes exome sequencing data
* GENCODE <https://www.gencodegenes.org/> high quality reference gene annotation and experimental validation for human and mouse genomes
* ENSEMBL <http://www.ensembl.org/index.html> genome browser for vertebrate genomes
* UCSC genome browser <https://genome.ucsc.edu/> broad collection of vertebrate and model organism assemblies and annotations, along with a large suite of tools for viewing, analyzing and downloading data

Cancer related databases

* COSMIC Catalogue Of Somatic Mutations In Cancer <http://cancer.sanger.ac.uk/cosmic> resource for exploring the impact of somatic mutations in human cancer
* TCGA The Cancer Genome Atlas <https://cancergenome.nih.gov/> multi-dimensional maps of the key genomic changes in 33 types of cancer
* ICGC International Cancer Genome Consortium <http://icgc.org/> comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types
* CCLE Cancer Cell Line Encycolpedia <http://www.broadinstitute.org/ccle> detailed genetic characterization of a large panel of human cancer cell lines

Gene expression

* GEO gene expression omnibus [www.ncbi.nlm.nih.gov/geo/](https://www.ncbi.nlm.nih.gov/geo/) microarray, next-generation sequencing, and other forms of high-throughput functional genomics data
* GTEx Genotype-Tissue Expression (GTEx) project <http://www.gtexportal.org> human gene expression and regulation and its relationship to genetic variation in human tissues

Transcriptomes

* SLE data [https://www.ebi.ac.uk/arrayexpress/](https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-65391/?query=SLE&sortby=assays&sortorder=descending) compare transcriptomes of SLE patients at various stages of the disease, as well as to evaluate the transcriptome for individual patients over time

Pathways

- REACTOME <http://www.reactome.org/> tools for the visualization, interpretation and analysis of pathway knowledge

Protein/molecule Structures, chemical data

* ChEMBL <https://www.ebi.ac.uk/chembl/> manually curated [chemical database](https://en.wikipedia.org/wiki/Chemical_database) of [bioactive](https://en.wikipedia.org/wiki/Bioactive) molecules with drug-like properties
* CSD Cambridge Structural Database, <https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/> repository for small-molecule organic and metal-organic crystal structures
* SureChEMBL <https://www.surechembl.org/search/> provides free access to chemical data extracted from the patent literature

HIV Cohort data in Switzerland

* SHCS Swiss HIV Cohort Study [www.shcs.ch](http://www.shcs.ch) systematic longitudinal study enrolling HIV-infected individuals in Switzerland
* ZPHI Zurich Primary HIV infection study [www.viralinfectiousdiseases.uzh.ch/en/ZPHI.html](http://www.viralinfectiousdiseases.uzh.ch/en/ZPHI.html) data of early HIV infections

HIV specific antibody data:

* [www.bnaber.org](http://www.bnaber.org)

Biologically interesting parameters

* [www.bionumbers.hms.havard.edu](http://www.bionumbers.hms.havard.edu)