# ZH BIO392 - Learning Goals

This page list the “core learning goals” in preparation for the exam at the end of the course. These my not cover 100% of possible questions, but rather be considered representative.

## General questions about the topic of the course

You should be able to demonstrate an understanding of the relationships between inherited and acquired genome variants and their possible implications for understanding phenotypic human variation. What are problems encountered there, and why do we think we need many more genomes, to be available for comparative analyses? Also, examples of data types beyond genome data relevant for understanding genomic variation should be provided. You should know some disease examples for which a genomic contribution could be described.

### Some factlets:

* approximate size of human genome
* size of largest human chromosome
* example(s) for sequencing “depth/coverage” in standard analysis scenarios, and the impact this has to the different genome file formats
* (dis)advantages of WES & WGS (and what those acronyms stand for)
* What are “genome reference assemblies”, and can you name (some of) them?
* Structuring of HGVS annotations (and - possibly made up - example)
* Basic understanding of cytogenetic banding annotation, and (approximate) spatial resolution of such annotations
* “1000 genomes” - what are they, and advantages vs. problems associated with using them in genomics workflows

## General computing & science questions

There can be some non-technical questions on e.g. best software practices (OpenSource vs. “black box” software, choice of operating system…). Here, it may be more about justifying an opinion vs. providing a “true answer”.

* reproducibility (OpenSource, versioning, standard formats, programmatic data manipulation, stable resources/repositories)
  + In order to increase reproducibility, written scripts are preferred to “Clicks” in a program.
* general workflow steps from reads to variants

## Resources

Some familiarity with selected genome & molecular knowledge resources, their primary goals and example use cases is expected.

ClinGen

ClinVar

UCSC genome browser

files used to represent features, and load them into custom tracks (BED, bedGraph, Wiggle, BAM, pgSNP…)

importance to select the right genome edition - Why?

## File types in genomics and their use

You should be able to list at least 2-3 core features (main use cases, type, core strength, core weakness).

FASTA

FASTQ

SAM/BAM/CRAM

BED

VCF (more extensive understanding of file structure expected)

“segment files”

## Protein representation of variants

protein sizes

resource(s) for 3D protein structures and other protein information resources

types of genome variants with respect to their impact on protein composition

amino acid physicochemical properties (size, charge) and effect of variation due to amino acid properties

* Sickle cell anemia! Single change in DNA leads to single change from GLU to VAL in hem protein, which hinders normal tetramer formation and instead causes clumps.

conservation state of a given AA and its relation to mutation frequency and functional importance

## Tutorials

What is “liftover” being used for?

Linkage disequilibrium and population genetics

What do you analyse with PLINK?

Examples for filters/parameters used in linkage analysis

## Genomic privacy

genome “Beacons”

concept

“unbreakable”?

de-identification attacks

genomic privacy, research, comparable risks (opinions)