Genome Data Analysis (LVA-Nr. 320.301 and 320.004)

Lecturers: Assoc. Prof. Irene Tiemann-Boege, Angelika Heißl and Theresa Schwarz

The following questions are just a guideline to the types of topics and questions that will be examined in the test.

Lecture 1:

- What is the difference between Genomics and Bioinformatics?
- What is the advantage of studying DNA over proteins?
- What is the difference between the mRNA and the open reading frame coding for a protein?
- What types of DNA comprise the genome of a cell?
- How many nucleotides is the human genome?
- What areas were developed in genomics?
- What were the two strategies to sequence the human genome? What are the differences, what the similarities?
- Principles of Sanger sequencing;
- How can Sanger sequencing be automated?
- How can an unknown sequence be sequenced (how is the primer problem solved)?
- What is shotgun sequencing ?
- What is pairwise end sequencing?
- How can a sequence be reconstructed by sequencing only paired ends?
- Was the human genome the first genome to be sequenced?
- What is coverage in the context of sequencing?
- What kind of problems can be encountered when sequencing repetitive regions with shot-gun sequencing? How can repetitive regions be identified?
- If humans have about the same number of genes than other animals, what could explain our complexity?
- What were the main features discovered in the human genome?
- What is non-coding DNA?
- What are CpGs and what function do CpG have? What are CpG islands?

- What are the main differences between DNA and RNA?
- Why do we need DNA databases?
- What is a genome assembly? Do the gene coordinates vary between different genome assemblies or do they stay the same? Why?
- What is an accession number? What are the accession number prefixes for genomic DNA, mRNA and protein?
- What is a reference sequence? What are common identifiers for reference sequences?
- What information can you retrieve in the Gene database of NCBI?
- What does the Gene Ontology tell you?
- How can a nucleotide search be restricted to reference sequences only? Or to a certain organism?

- What is a FASTA sequence? How is a FASTA sequence characterized? What is the difference to a normal nucleotide or amino acid sequence?
- What is a codon? What is an ORF (open reading frame)? What is an untranslated region in the mRNA? What is a Start-/Stop-codon?
- What is an EST (Expressed sequence tag)?
- What does an EST profile tell you?
- In which database can you find information about genetic diseases and their associated genes?

Lecture 2:

- What is genomic variation? What types of genomic variation are there?
- What are SNPs? What are multiple nucleotide polymorphisms?
- What is a transition/transversion? What are synonymous, missense, frameshift and non-sense mutations?
- What is a microsatellite? How do you write a microsatellite in the databases?
- Compare microsatellites with minisatellites
- What is a copy number variants (CNV)? Do CNVs have a phenotype?
- What are transposons? What are the most frequent transposons in the human genome? What proportion of the human genome are transposons?
- In what area are microsatellites particular useful?
- What are segmental duplications?
- What is the most common type of mutation resulting in the change of a single nucleotide in the human genome?
- Why are minisatellites more variable in the genome compared to SNPs? What is the main use of microsatellites?
- What are retrotransposons?
- What are CpG islands? Where in the genome are they found? What is their potential function?
- What is a segmental duplication?
- What is an allele, genotype, haplotype? What does homozygous mean, what heterozygous?
- In which database would you find polymorphism data in the human population?
- What kind of search limits can you use in the database?
- Where can you find information about the allele frequency and genotype frequencies of a particular SNP in different human populations?
- How do you estimate the frequency of the heterozygotes from a given allele frequency?
- What is the relationship between allele frequencies and genotype frequencies?
- Why are polymorphisms useful? What kind of biological phenomena can be studied with polymorphic information?
- Are allele frequencies constant in a population? What forces/processes can change allele frequencies? What processes change the allele frequency in the population?
- What is the HapMap project? How was it designed? What is the importance (uses) of HapMap?
- How can SNPs be used? Can SNPs be used to infer our ancestry?

- What database is used for literature searches? How can you restrict the searches? What restrictions can you use?
- What is a 'Boolean search'? Which search terms can you use to restrict your search?
- What is a polymorphism? What kind of polymorphisms do you know? What is a SNP?
- ?
- What is a genotype/haplotype?
- What does it mean when you have a R, Y, M,K,… in your nucleotide sequence?
- What is an alleleWhat is the Hardy-Weinberg Equilibrium (HWE)?
- Calculate heterozygous and homozygous genotype frequencies using the HWE.
- What is PCR? What components are necessary? How do you design primers for a PCR?
- What is an amplicon, what is a template, what is a primer? What does reverse complement mean?
- Which end of the primer (5' or 3') is extended by the polymerase? The synthesis therefore goes in 5'->3' or 3'->5' direction?
- Which program can you use to design PCR primers?
- How do you set a target region in Primer3Plus? Why is this useful?
- Where can you set Product Size Ranges in Primer3Plus and why is this useful?
- What is the NEB-cutter used for?
- What is a restriction enzyme?
- How does a simple genotyping assay work when using a PCR product and restriction enzymes?

Lecture 3:

- What are the main sources of DNA for analyzing human variation?
- How does genotyping using RFLP work?
- How does real time PCR work? How can it be used for genotyping?
- What is a high-throughput approach for SNP Genotyping? How does it work?
- How are differences in microsatellites measured?
- What is comparative genomics or functional genomics?
- What is the use of the genome sequence of model organisms?
- Why are dog genomics important?
- What is metagenomics?
- What information can a genome provide in cases of epidemics?
- What are the two main approaches to identify a coding sequence (gene) in a genome?
- What is EST sequencing?
- What is ab initio gene discovery? What elements does it use?
- What information is obtained in Gene Ontology? What is the use of Gene Ontology?
- How are regulatory sequences identified?
- What kinds of functional non-coding RNA have been identified?
- What structural features are found in the genome?
- What is microRNA?

- What types of repetitive sequences are found in the genome?

- What is a genome browser? What genome browsers do you know?
- Where would you look up the exon –intron boundaries of a specific gene?
- Where can you find the transcripts of a gene?
- How can you find chromosome and nucleotide coordinates of a gene?
- What does it mean that the genetic code is redundant?
- What are homologs, paralogs, and orthologs?
- Which database can be used for a pairwise alignment? What is a pairwise sequence alignment used for? What kind of information is obtained in a pairwise alignment?
- What is more informative concerning biological similarity: amino acid or nucleotide sequence? Why?
- What is the use of BLAST? What are the steps of a BLAST search?
- What is the query coverage? Why is it important to examine the query cover in a BLAST result?
- What is the Expect (E)-value in a BLAST result?
- What is similarity and identity in a BLAST result? What is meant by percent positives in an alignment (what kind of amino acids are included in percent positives)?
- Which program can you use to validate your designed primers for a PCR?
- How tolerable are mismatches in the primer sequence for a PCR reaction?
- What information can you get when using PrimerBLAST?

Lecture 4:

- What is the main difference between Sanger sequencing and next generation sequencing (NGS) in terms of amounts of produced data and costs?
- Why can the sequencing costs be dramatically reduced with NGS compared to Sanger sequencing?
- Can a whole human genome be sequenced in a few days with NGS?
- How does NGS work? What different solid surfaces are used in NGS?
- What is single molecule amplification? What are the advantages of single molecule amplification?
- Compare old vs. new sequencing technologies (e.g Ilumina vs. Sanger sequencing)?
- What is bridge amplification?
- What is emulsion PCR?
- How does Illumina NGS work?
- How does Ion Torrent work?
- What technology can directly sequence single molecules without previous amplification steps?
- How does PacBio work? What are its main uses compared to other NGS methods?
- What are the uses/applications of NGS?
- How can the transcriptome be analyzed with NGS?
- How can genomic sequences be targeted for NGS?
- What is exome sequencing?
- What is individual genome sequencing? What are its uses/benefits?
- How can personal genomes be helpful in medicine?
- What is the 1000 Genomes Project?

- What are the four levels of a protein structure?
- How are amino acids classified?
- Which chemical reaction leads to peptide bond formation? Why is a peptide bond said to have a partial double bond character?
- What is meant by angle of rotation “phi” and “psi” in terms of a peptide bonds? What is a Ramachandran plot? Why are glycine and proline residues often found in beta-turns?
- What is a domain? What is a motif?
- How can proteins be classified?
- Name 3 post-translational modifications of proteins?
- What experimental methods are used to determine the 3D structure of a protein? Which one has the highest resolution?
- What does ‘resolution’ mean? What resolution is necessary to visualize individual atoms?
- What is the R-value?
- Which tools do you know from the ExPASy website? What information do you get using those tools?
- What information do you get when using the PDB database?
- What tool can you us to visualize 3D protein structures? What analysis can you do with these tools?

Lecture 5:

- How can mutation causing a rare disease be identified in the genome with next generation sequencing?
- What are the different functional validation steps necessary after identifying a mutation in sequencing data?
- What is targeted gene editing? How does it work?
- How can a minimal genome be built from synthetic components (synthetic biology)?
- How can the minimal set of genes necessary for life be identified?
- What is the Gibson assembly and what is it used for?
- How can the comparison of different primate genomes tell us what makes us human?
- Can ancient genomes be sequenced? (eg. Neanderthal)
- How can the evolution of species be inferred from sequencing data?
- What species is our closets relative?
- Which gene group have the most differences between humans and other primate species?
- Explain how the role FOXP2 gene makes us unique to other primate species?
- What is the evolutionary relationship between Neanderthals, Denisovans and modern humans? Has there been an exchange of genetic materials between these three groups?
- How can genomics be commercialized?
- Does public research or the general public benefit from genomic service companies like 23and Me or Ambry Genetics?
- • How has genomics impacted society and law?
- What is ELSI? And what is ELSI’s main purpose?
- What are the different programs regarding that ELSI is concerned with?

- Why is it important to have programs like ELSI in place for the regulation and use of genetic information?