**Lecture 1.**

Linkage-mapping 1980

* One nuclear genome
* Many mitochondrial genomes (liver cells)

Humans are diploid – genome is comprised of a paternal and maternal haplotype.

Homozygous and heterozygous

Difference 1 in 700 base pairs between two humans.

Chromosome are ordered according to the size: 1 is the biggest, one exception

There are about

* 20 000 protein coding genes (how many non coding) – 2% of the genome
* 3 billion base pairs

Karyotype – set of chromosomes

Chromosomes differ in size, banding pattern, centromere position

Any change of centromere position has huge consequences

Nohup – run in the background after closing terminal

& run in the background

Illumina sequencing

**Lecture 2.**

Fasta file is around 3Gb since 3 billion base pairs – one base pair – one byte

PCR - polymerase chain reaction -method to multiply a DNA sequence. Needed

* 2 Primers
* Taq Polimarase
* DNA Nucleotides

curl https://hgdownload.cse.ucsc.edu/goldenpath/hg38/chromosomes/chr22.fa.gz > chr22.fa.gz

gzip -d chr22.fa.gz

mv chr22.fa hg.b37.chr22.fa

grep -v ">" hg.b37.chr22.fa | wc -c # number of characters

grep -v ">" hg.b37.chr22.fa | wc -l # number of newline characters

The difference is the number of nucleotides

Low case a – part of a repeat, upper case A – not part pf a repeat.

grep -v ">" hg.b37.chr22.fa | grep -o -n ["A|a"] | wc -l # counts A and a

**Lecture 3. Genetic Variation**

Humans: 99.5% DNA is identical (4-5 Mio differences) -

E.g. Drosophila – more diverse 1/180 bp

There are 100 Mio genetic variants in humans

Types:

* Single-nucleotide polymorphisms (SNPs, SNVs or simply polymorphisms ) -Spelling mistakes
* Insertion-deletion polymorphisms (INDELs) – extra or missing DNA
* Structural variants (SVs) - large blocks of extra, missing or rearranged DNA up to Mega bases. 5 to 10 (20) thousands SV in genome (earlier wrongly thought only tumors)

Thousand genomes project (2504 genomes)

Nucleotide diversity (Pi): 1/756 bp to 1/620 bp

Why do we care

* Understand relationship between genetic variation and traits or diseases
* Autosome recessive inheritance diseases cannot be studied using common disease methods (Cases vs. Control groups). (?)
* Evolution – how species evolved

Mutation!=Polymorphism (or SNP) (eg. many mutations in sperms that never will be polymorphism since will not fertilize an egg).

Mutations are private to the gamete/chromosome/ individual if they are never passed to the next generations - they are not polymorphisms!

Reference genome has one haplotype (until some time ago -weeks)

De novo mutations (DNMs) occur during fertilization (sperm or egg mutations)

DNMs are rare: 30-40 DNM per haploid genome, 80 per diploid genome. Protein coding genes have low rate of mutations.

Most developed male disorders are caused by germline mutations aroused spontaneously

DNMs are more frequent in paternal haplotype.

Correlated with the paternal age – one year adds 3 DNV mutations (0.4 maternal)

Mutations:

1. G, T-C, C-T, G-A are frequent and less deleterious (methylated cytosine)

Germline mutations are heritable, somatic mutations (cancer)are not (non-germline tissues)

Types:

Positive – adaptive- mutations are getting more frequent – confers fitness advantage

Negative - purifying – mutations– eliminates an allele - confers fitness disadvantage

Neutral – genetic drift – nothing to do with natural selection – frequency remains same

Genetic bottleneck: mixed population > bottleneck event (e.g. drought) > reduced population diversity (don’t know what type of mutations causes that)

Sub-Saharan population has a higher rate of polymorphism – genetic diversity – less prone to bottlenecks

Site frequency spectrum – most variants are rare, private, specific to eg family

10-20 T variants in the reference genome are not most frequent in the population

Meiotic recombination shuffles alleles and generates new haplotypes (a child produces a mixture of parental haplotypes). Frequency of recombination depends on two markers (polymorphic sites): Alleles that are close are less likely to be recombined (TH Morgan). Centimorgan – 1 percent of recombination.