BIMM-194: GENOMICS, BIG DATA AND HUMAN HEALTH

Glossary of Common Terms and Abbreviations

https://bioboot.github.io/bimm194_W18/

Dr. Barry Grant

Jan 2018

The glossary below provides a list of key terms (Table 1) and abbreviations (Table 2) used throughout this course. You do not need to read them all now. However you may find it useful to refer back to this document if you are unsure of the terminology being used in class or in your assigned primary literature reading.

Table 1. Common terms and their definitions.

Term	Definition
Alignment	The process of matching reads back to their original position in the reference genome.
Allele	An allele is one of a number of alternative forms of the same gene or genetic locus. We inherit one copy of our genetic code from our mother and one copy of our genetic code from our father. Each copy is known as an allele.
Array CGH	Microarray based genomic comparative hybridisation. This is a technique used to detect chromosome imbalances by comparing patient and control DNA and comparing differences between the two sets. It is a useful technique for detecting small chromosome deletions and duplications which would not have been detected with more traditional karyotyping techniques.
Base	A unit of DNA. There are four bases which form the cross links (or rungs) of the DNA double helix: adenine (A), thymine (T), guanine (G) and cytosine (C).

Capture	see Target enrichment.
Cell differentiation	The process by which a cell becomes specialized in order to perform a specific function.
Centromere	The point at which the sister chromatids are joined.
Chromosome	A structure located in the nucleus all living cells, comprised of DNA bound around proteins called histones. The normal number of chromosomes in each human cell nucleus is 46 and is composed of 22 pairs of autosomes and a pair of sex chromosomes which determine gender: males have an X and a Y chromosome whilst females have two X chromosomes.
Chromatid	Two identical copies from the replication of a single chromosome. Therefore a sister chromatid refers to either of the two identical copies.
Codon	A sequence of three adjacent nucleotides constituting the genetic code that determines the insertion of a specific amino acid in a polypeptide chain during protein synthesis.
Clinical exome sequencing	Differs per laboratory, but usually involves sequencing of all genes known to be associated with human disease.
Coverage	The number of reads giving information about the base present at a set position in the reference sequence.
Crystallography	The experimental science of determining the arrangement of atoms in crystalline solids.

DNA	Deoxyribonucleic acid. DNA is a molecule consisting of two long chains of nucleotides twisted together to form a double helix. Genes are made from DNA.
DNA sequencing	The process of identifying the order of a variable number of adjacent nucleotides in a strand of DNA.
Epigenome	Chemical marks on the DNA, regulating whether the gene is turned “on” or “off”.
Exome	The coding portion of the genes. The exome constitutes 1-2 % of the genome.
Gene	A portion of DNA that serves as the basic unit of heredity.
Genetic code	The DNA or RNA sequence that determines the amino acid sequence used in the synthesis of an organism’s proteins.
Genome	The entirety of an individual’s genetic material including ≅20 000 genes and the genetic material between genes.
Gene panel	A collection of genes to be sequenced together, which are usually linked by common biological pathways, or known disease associations.
Germline mutations	Where a genetic error occurs in the egg or sperm pre fertilization. Therefore, the genetic mutation is passed onto the offspring.

Haploid	A cell having a single set of unpaired chromosomes.
Histones	Histones are basic proteins which function as spools for thread-like DNA to wrap itself around, allowing DNA to be packaged efficiently.
Homologous	Corresponding in structure and in origin, but not necessarily in function.
Homologous chromosomes	A pair of chromosomes which have the same genes at the same loci.
Indel	An insertion or deletion.
Karyotype	The number and appearance of the chromosomes when viewed down a microscope.
Mapping	See Alignment.
Meiosis	Occurs in the production of sperm and eggs. Four daughter cells are produced from the original parent cell. There are two stages to meiosis, meiosis 1 and meiosis 2. At the end of meiosis, each daughter cell has only 23 chromosomes.
Missense mutation	A missense mutation describes where a base substitution or change results in a codon which causes the insertion of a different amino acid into a protein. [see non-synonymous change]

Mitosis	Occurs at the end of normal cell division. Two daughter cells are produced from one parent cell, both with the same number of chromosomes (46) as the parent cell.
Monosomy	Where only one chromosome from a pair is present in a cell.
Mosaicism	Where a genetic error occurs after fertilisation, resulting in two distinct genetic cell lines.
Next generation sequencing	High-throughput DNA sequencing where millions of DNA bases are sequenced in parallel.
Non-pathogenic	Not disease-causing.
Non-synonymous change	A non-synonymous change describes where a base substitution or change results in a codon which causes the insertion of a different amino acid into a protein. [see missense mutation]
Nucleotide	A nucleotide is composed of a DNA base, a phosphate and a pentose sugar.
Pathogenic	Disease-causing.
Pharmacogenomics	The branch of genetics concerned with determining the likely response of an individual to therapeutic drugs.
Phenotype	The set of observable characteristics of an individual as a result of their genotype/environment.

Pull down	see Target enrichment.
Read	A computer generated sequence of bases representing the sequenced code from an original DNA fragment.
Reference sequence/ genome	An assembled version of a genome that can be used to make comparisons to the genomes from other individuals.
RNA	Ribonucleic acid, a nucleic acid present in all living cells. Its principal role is to act as a messenger carrying instructions from DNA for controlling the synthesis of proteins.
Single nucleotide polymorphism	A single base substitution occurring at high frequency (more than 1%) in the general population.
Somatic mosaicism	The existance of genetically distinct cells lines in the body of an individual, but not the sex cells (germline)
Spindle	Fibres which draw the chromosomes to either end of the poles during mitosis and meiosis.
Target enrichment	A method for selecting a specific portion of the genome to undergo sequencing.
Transcription	Transcription is the first step of gene expression, in which a particular segment of DNA is copied into RNA by the enzyme RNA polymerase.

Translation	Translation is the step after transcription, in which cellular ribosomes use the RNA to produce a specific protein.
Uracil	A pyrimidine base that is a component of RNA. It forms a base pair with adenine during the generation of messenger RNA. Uracil is therefore structurally analogous to thymine in molecules of DNA.
Variant of uncertain significance	An alteration to the DNA sequence where it is unclear (on the basis of the available evidence) whether it is disease-causing or not.
Whole (human) exome sequencing	Sequencing the portion of the genome which codes for proteins.
Whole (human) genome sequencing	Sequencing of the entire length of the human genome.

Table 2. Common abbreviations and their definitions.

Abbreviations	Meaning
ABL	Abelson murine leukemia viral oncogene homolog 1
Array CGH	Array comparative genomic hybridization
BCR	Breakpoint cluster region
BRCA1 gene	Breast cancer 1, early onset gene

BRCA2 gene	Breast cancer 2, early onset gene
CML	Chronic myelogenous leukaemia
CNV	Copy number variation
DDD	Deciphering Developmental Disorders
DNA	Deoxyribonucleic acid
DS	Down syndrome
GWAS	Genome-wide associated studies
HD	Huntington’s Disease
MED	Multiple epiphyseal dysplasia
mRNA	Messenger RNA
NGS	Next generation sequencing
PGD	Pre-implantation genetic diagnosis
RNA	Ribonucleic acid

SNPs	Single nucleotide polymorphisms
TPMT	Thiopurine S-methyltransferase
tRNA	Transfer RNA
VUS	Variant of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing