GLOSSARY OF TERMS

Table of Contents

A

Amino acids: any of a class of 20 molecules that are combined to form proteins in living things. The sequence of amino acids in a protein and hence protein function are determined by the genetic code. From http://www.geneticalliance.org.uk/glossary.htm#C

- The building blocks of proteins, there are 20 different amino acids. From https://www.yourgenome.org/glossary/amino-acid

Antisense: Antisense nucleotides are strings of RNA or DNA that are complementary to "sense" strands of nucleotides. They bind to and inactivate these sense strands. They have been used in research, and may become useful for therapy of certain diseases (See Gene silencing). From http://www.encyclopedia.com/topic/Antisense_DNA.aspx. Antisense and RNA interference are referred as gene knockdown technologies: the transcription of the gene is unaffected; however, gene expression, i.e. protein synthesis (translation), is lost because messenger RNA molecules become unstable or inaccessible. Furthermore, RNA interference is based on naturally occurring phenomenon known as Post-Transcriptional Gene Silencing. From http://www.ncbi.nlm.nih.gov/probe/docs/applsilencing/

B

Biobank: A biobank is a large, organised collection of samples, usually human, used for research. Biobanks catalogue and store samples using genetic, clinical, and other characteristics such as age, gender, blood type, and ethnicity. Some samples are also categorised according to environmental factors, such as whether the donor had been exposed to some substance that can affect health. Biobanks play a crucial role in biomedical research, such as in genomics and personalised medicine. Researchers access biobanks when they need samples with similar characteristics for their research studies (from https://www.eupati.eu/glossary/eurobiobank/).

Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyse and integrate biological and genetic information. RD-Connect aims to develop highly sophisticated systems able to combine data from the different omics technologies in order to facilitate gene and biomarker discovery through efficient annotation systems and expert systems able to extract knowledge from data.

Biomarker or biological marker, generally refers to a measurable indicator of some biological state or condition. From Wikipedia

C

Chaperones - **Molecular chaperones** refer to specialised proteins that assist with the folding/unfolding and assembly/disassembly of other proteins. These processes are critical for the protein to achieve its native structure and perform its normal biological function.

Chromosomes are structures found in the nucleus of cells composed of DNA and proteins. Normally humans have 46 chromosomes in each cell, 23 from each parent. Of these, 1 is a sex chromosome and 22 are autosomes (i.e. not a sex chromosome). Adapted from http://www.geneticalliance.org.uk/glossary.htm#C

- A threadlike structure in our cells, made of a long DNA molecule, wrapped around a protein scaffold. Humans have 23 pairs of chromosomes. Other organisms also have chromosomes, most bacterial chromosomes are loops or circles of DNA. Adapted from https://www.yourgenome.org/glossary/chromosome

Cohort studies are used to study diseases, their causes, and their prognoses. Cohorts are groups of people who are selected based on certain characteristics. For example, if exposure to a risk factor such as cigarette smoke is suspected to cause a disease, a cohort can be selected in which one group has been exposed and another group has not. Both groups are then studied for signs or symptoms of disease. Cohort studies can be prospective (cohorts are identified before any signs of disease and are followed up over time) or retrospective (data is used that has already been collected, possibly over a long period). Cohort studies are a kind of observational study, in which the researcher does not perform any intervention (such as administering a medicine). Cohort studies are useful when it would be unethical to carry out a randomised controlled trial (RCT). Adapted from https://www.eupati.eu/glossary/)

Consent - **Informed consent:** Participants of clinical trials and studies always retain ownership of their data and hence control over how it can be used. Access to this data is given by way of informed consent whereby potential participants are made aware of various aspects of the research objectives, procedures, risks and benefits and then decide if they are happy for their data to be used. From RD Connect

CRISPR/Cas9: **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (CRISPR) associated with an enzyme called **Cas9** constitute a new genome editing tool which research scientists use to cut the genetic code at a specific place, like DNA scissors. This tool could potentially be used in several applications. For example, the cut DNA could then be stuck back together after deleting faulty genes causing genetic conditions and replacing them with the correct "working" version of the genes. Research projects using the CRISPR system are currently underway to try and correct genetic defects causing rare genetic diseases including cystic fibrosis, Duchenne muscular dystrophy, retinitis pigmentosa or beta-thalassemia. This new genome editing technique is still in the research phase and is not yet ready for clinical applications.

Cytoplasm is the gelatinous liquid that fills the inside of a cell. It is composed of water, salts, and various organic molecules. Some intracellular organelles, such the nucleus and mitochondria, are enclosed by membranes that separate them from the cytoplasm. From be https://www.genome.gov/genetics-glossary/Cytoplasm

D

Diagnosis is the clinical identification of a condition, disease or disorder by systematic analysis of the background or history, examination of the signs or symptoms, evaluation of the research or test results and investigation of the assumed or probable causes.

DNA (DeoxyriboNucleic Acid) is the long molecule that contains our unique genetic code. Like a recipe book, it contains the instructions for making all the proteins in our bodies. The four Nucleotides in DNA contain the bases: adenine (A), guanine (G), cytosine (C), and thymine (T).

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. It includes any method or technology that is used to determine the order of the four bases in a strand of DNA. See also genome sequencing.

E

Enzyme: a protein catalyst which is essential to the correct functioning of biochemical reactions. From http://www.geneticalliance.org.uk/glossary.htm#E

- Biological molecules, usually proteins that are responsible for thousands of metabolic processes essential to life. Adapted from https://www.yourgenome.org/glossary/enzyme

Enzyme Replacement Therapy (ERT) is a medical treatment replacing an enzyme in patients in whom that particular enzyme is deficient or absent. The replacement enzyme is referred to as a "recombinant" enzyme produced through gene engineering**.** Usually, this is done by giving the patient an intravenous (IV) infusion containing the enzyme.

Exome: although the human genome consists of 3 billion Nucleotides or "letters" of DNA, only a small percentage — around 1 to 2 % — of those letters are actually translated into proteins, the functional players in the body. The exome consists of all the genome's exons, which are the coding portions of our genes.

Exome sequencing also known as whole exome sequencing (WES) is the technique used to identify the genetic variants in all of an individual's expressed genes. It consists of first selecting only the subset of DNA that encodes proteins (known as exons), and then sequencing that DNA using any high throughput DNA sequencing technology. There are 180,000 exons, which constitute about 1% of the human genome, or approximately 30 million base pairs. Exome sequencing is increasingly used for the clinical evaluation of genetic disease.

F
G

Gene is an ordered sequence of Nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule). From http://www.geneticalliance.org.uk/glossary.htm#G

- Section of DNA within the genome that carries the information to make a molecule, usually a protein. They contain the instructions for our individual characteristics, like eye and hair colour. In humans and other complex organisms, genes are split into coding

(exons) and non-coding sequences (introns). These split sections allow some genes to make more than one type of protein. Adapted from https://www.yourgenome.org/glossary/gene

Gene modifier refers to a gene that influences the expression of another gene.

Genetic disease is a disease caused in whole or in part by a change in the DNA sequence away from the normal sequence. Changes in the DNA sequence can be as small as a single-base mutation in just one gene, or they can involve the addition or subtraction of entire chromosomes

Genetic Counselling: Information and support provided by a specialist doctor, usually a geneticist, to parents who have known conditions in their families or who are concerned about the future possibility of genetically transmitted conditions. From http://www.geneticalliance.org.uk/glossary.htm#G

- Advice given to patients, prospective parents or their relatives before and after being tested for a genetic disease. Adapted from https://www.yourgenome.org/glossary/genetic-counselling

Genetic Engineering is a set of technologies used to change the genetic makeup of cells.

Genetics is the study of genes and heredity in living organisms.

Gene silencing is the interruption or suppression of the expression of a gene at transcriptional or translational levels. See also Silencing

Gene therapy: Insertion of normal DNA directly into cells to correct a genetic defect. From http://www.geneticalliance.org.uk/glossary.htm#G

- The treatment of a disease by introducing modified DNA into the cells of the patient. Adapted from https://www.yourgenome.org/glossary/gene-therapy

Genome consists of all the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs. From http://www.geneticalliance.org.uk/glossary.htm#G

- An organism's complete set of genetic instructions. Each genome contains all of the information needed to build that organism and allow it to grow and develop. Our genome is approximately 3,000,000,000 base pairs long and is packaged into 23 pairs of chromosomes. Adapted from https://www.yourgenome.org/glossary/genome

Genome editing is a type of genetic engineering in which DNA is inserted, deleted or replaced in the genome of an organism using engineered nucleases, or "molecular scissors."

Genome sequencing is the process of identifying the order of DNA Nucleotides, or bases, in a genome—the order of As, Cs, Gs, and Ts that make up an organism's DNA. The human genome is made up of over 3 billion of these genetic letters. See also Exome sequencing.

Genomics refers to the study of genes and their function. Genomics aims to understand the structure of the genome and examines the molecular mechanisms and the interplay of genetic and environmental factors in disease.

Genotype is an individual's collection of genes, including which genetic variants they have. It is the entire complex of genes inherited from both parents. The genotype determines the hereditary characteristics of an individual. A genotype can be determined by sequencing an individual's genome. From https://www.eupati.eu/glossary/genotype/

Germ cells are cells that create reproductive cells called gametes and are found in the gonads. In males, the gonads are the testes which produce the gametes called sperm whilst in females gonads are the ovaries which produces gametes called eggs. The entire process from germ cell to gamete is called a germline.

H

Heredity: the passing of traits from parents to their children.

I

Incidental findings: discoveries made as a result of genetic testing that are unrelated to the indication for the test.

Induced pluripotent stem cells, commonly abbreviated as iPS cells or iPSCs are a type of pluripotent stem cell artificially derived from a non-pluripotent cell, typically an adult somatic cell, by inducing a "forced" expression of certain genes and transcription factors (Wikipedia).

Infusion refers to the process of introducing a fluid into the body, usually through a vein (intravenous infusion).

Inherited: an inherited trait is one that is genetically determined. Inherited traits are passed from parent to offspring according to the rules of Mendelian genetics. Most traits are not strictly determined by genes, but rather are influenced by both genes and environment. https://www.genome.gov/genetics-glossary/Inherited

J

K

L

M

Messenger RNA (mRNA): Messenger RNA (mRNA) is a single-stranded molecule used for protein production at the ribosome. Because its sequence is used for translation, mRNA is called a "sense" strand or sense sequence. From http://www.encyclopedia.com/topic/Antisense_DNA.aspx

Metabolic Disorder occurs when abnormal chemical reactions in the body disrupt metabolism. When this process is disrupted, you might have too much of some substances or too little of other ones that you need to stay healthy.

Metabolism is the process the body uses to get or make energy from the food people eat. Chemicals in the digestive system break the food parts down into sugars and acids, the body's fuel. The body can use this fuel right away, or it can store the energy in the body tissues, such as the liver, muscles, and body fat.

Metabolite is any substance/small molecule produced during metabolism (digestion or other bodily chemical processes). The term may also refer to the product that remains after a drug is broken down (metabolised) by the body.

Metabolomics refer to the scientific study of metabolites. Using sophisticated analytic tools, it aims to identify, quantify and analyse cellular metabolites.

Microarray: DNA microarray analysis is a technique that scientists use to determine whether genes are switched on or off. If a gene is switched on, it is known as gene expression. Scientists use this technique to measure the expression levels of thousands of genes at the same time. The result is known as an expression profile. This technique is used in many areas of biological and medical research. It can give valuable information about, for example, what genetic changes are responsible for tumour growth in specific individuals, or whether the expression profile of an individual makes them suitable for a specific treatment. Fromhttps://www.eupati.eu/glossary/microarray/

Micro-RNAs (miRNAs) constitute a class of non-coding RNAs that play key roles in the regulation of gene expression. Acting at the post-transcriptional level, these molecules may fine-tune the expression of as much as 30% of all mammalian protein-encoding genes.

Mitochondria are the cell's power sources, compartments within the cell that provide its energy. They contain their own DNA, called mitochondrial DNA, a very small amount of DNA inherited only from your mother. From http://www.geneticalliance.org.uk/glossary.htm#M

- The power plants of our cells. Membrane-bound organelles found in most eukaryotic cells, they generate the cell's source of energy. Adapted from https://www.yourgenome.org/glossary/mitochondria

Mutation - Genetic mutation: A change that occurs in a DNA sequence. Mutations are relatively common in our DNA, but most have no detectable effect. Adapted from https://www.yourgenome.org/glossary/mutation

Missense mutation: when the change of a single base pair causes the substitution of a different amino acid in the resulting protein. This amino acid substitution may have no effect, or it may render the protein non-functional.

N

Neurodegenerative diseases is an umbrella term for a range of conditions which primarily affect the neurons in the human brain. Examples of neurodegenerative diseases include Parkinson's, Alzheimer's, Huntington's disease, Motor Neuron Disease, Spinal muscular atrophy, Spinocerebellar ataxia. Neurodegenerative diseases result in progressive degeneration and / or death of nerve cells. This causes problems with movement (called ataxias), or mental functioning (called dementias). From http://www.neurodegenerationresearch.eu/about/what/

Neurons are the building blocks of the nervous system which includes the brain and spinal cord. Neurons normally do not reproduce or replace themselves, so when they become damaged or die the body cannot replace them. Adapted from http://www.neurodegenerationresearch.eu/about/what/

Next generation sequencing (NGS) also known as high-throughput sequencing, is the catch-all term used to describe a number of different modern sequencing technologies.

Non-coding RNA are RNA molecules that are not translated into proteins. Different classes of non-coding RNAs participate in different cellular processes. For example in gene expression regulation, miRNAs (micro), piRNAs (piwi-interacting), lncRNAs (long non-coding) intervene; in RNA maturation, snRNAs (small nuclear) and snoRNAs (small nucleolar) participate whereas in protein synthesis, it is the rRNAs (ribosomal) and tRNAs (transfer) that play a major role.

Nuclease is an enzyme that cleaves the bonds between Nucleotides in nucleic acids.

Nucleic acid: A large molecule composed of nucleotide subunits. http://www.geneticalliance.org.uk/glossary.htm#N

Nucleotide: a subunit of DNA or RNA consisting of a nitrogenous base (adenine, guanine, thymine, or cytosine in DNA; adenine, guanine, uracil, or cytosine in RNA), a phosphate molecule, and a sugar molecule. Thousands of nucleotides are linked to form a DNA or RNA molecule. From http://www.geneticalliance.org.uk/glossary.htm#N

- Building blocks of nucleic acids such as DNA and RNA. These are adenine, cytosine, guanine and thymine in DNA, and adenine, cytosine, guanine and uracil in RNA. Adapted from https://www.yourgenome.org/glossary/nucleotide

Nucleus is a structure found in all eukaryotic cells that contains most of the cell's genetic material (DNA).

O

Oligonucleotides are short DNA or RNA molecules, oligomers, that have a wide range of applications in genetic testing and research. They are characterized by the sequence of nucleotide residues that make up the entire molecule. From Wikipedia

Omics technologies refer to a field of study in biology ending in *-omics*, such as genomics, proteomics or metabolomics. The related suffix **-ome** is used to address the objects of study of such fields, such as the genome, proteome or metabolome respectively.

P

Phenome:

Phenotype: A phenotype is an individual's observable traits, such as height, eye color, and blood type. The genetic contribution to the phenotype is called the genotype. Some traits are largely determined by the genotype, while other traits are largely determined by environmental factors. Fromhttps://www.genome.gov/genetics-glossary/Phenotype

Phosphate molecule: As a biological molecule, it is composed of phosphorus and oxygen and plays a major role in biological processes of many organisms, e.g. as chemical component of nucleic acids (DNA and RNA), Nucleotides, etc**.** From http://www.biology-online.org/dictionary/Phosphate

Pluripotent stem cells are able to make cells from all three basic body layers, so they can potentially produce any cell or tissue the body needs to repair itself**.**

Polynucleotide molecule is a biopolymer (large molecule composed of many repeated subunits) composed of 13 or more nucleotides covalently bonded in a chain. DNA and RNA are examples of polynucleotides with distinct biological function. DNA consists of two chains of polynucleotides, with each chain in the form of a helical spiral. Adapted from Wikipedia.

Prognosis is the prediction of the likely course of a medical condition.

Proteins are an important class of molecules found in all living cells. A protein is composed of one or more long chains of amino acids, the sequence of which corresponds to the DNA sequence of the gene that encodes it. Proteins play a variety of roles in the cell, including structural (cytoskeleton), mechanical (muscle), biochemical (enzymes), and cell signaling (hormones). Proteins are also an essential part of diet. Fromhttps://www.genome.gov/genetics-glossary/Protein

Protein encoding gene (PEG), protein coding sequence (CDS), and open reading frame (ORF) are nearly synonymous terms. Protein encoding genes are commonly described as *structural* or *regulatory*. Structural genes encode enzymes and structural proteins. Regulatory genes encode proteins that function to regulate the expression of other genes, often by binding specifically to short sequences of the DNA. From http://www.nmpdr.org/FIG/wiki/view.cgi/FIG/ProteinEncodingGene

Proteomics is the large scale study of proteins, particularly their structures and functions.

Q

R

Ribosomes are the cellular structure responsible for decoding the DNA and are the protein "builders" of the cell. They connect one amino acid at a time and build long chains.

RNA (RiboNucleic Acid): A chemical found in the nucleus and cytoplasm of cells; it plays an important role in protein synthesis and other chemical activities of the cell. The structure of RNA is similar to that of DNA. There are several classes of RNA molecules, including messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), and other small RNAs, each serving a different purpose (*from Genetic Alliance UK http://www.geneticalliance.org.uk/glossary.htm#R*). See also Non-coding RNAs.

- A nucleic acid similar in structure and properties to DNA, but it only has a single strand of bases and instead of the base thymine (T), RNA has a base called uracil (U). https://www.yourgenome.org/glossary/rna

RNA interference - see Antisense RNA

Rare Kidney Diseases comprise a group of rare genetic diseases affecting the kidneys. This group includes for example Alport Syndrome, Steroid Nephrotic Resistant Syndrome, Membranous Nephropathy, Tubulopathies and Cystinosis. To find out more about Renal Diseases, please see http://federg.org/

Registry A patient registry is a collection of information about individuals, usually those with a specific diagnosis or with specific risk factors for a disease. Some patient registries seek people with varying health levels who may be willing to take part in research about a particular disease.

Patient registries have multiple uses. For example, registries for rare diseases can be used to establish the basic characteristics of the disease, how it is managed in clinics, and what outcomes people experience. Other uses include helping to measure clinical effectiveness of treatments in 'real world' settings, and investigating quality of patient care. Clinical trial registries collect basic health information from people who agree to be contacted about taking part in future clinical trials. Volunteering for a registry does not mean a person has signed up for a clinical trial. Volunteering for a disease registry can sometimes become a first step toward taking part in a

clinical trial, but registries and specific trials are not directly linked. From https://www.eupati.eu/glossary/patient-registry/

S

Silencing refers to the ability of a cell to prevent the expression of a certain gene. Methods used to silence genes are increasingly being used in the laboratory to produce therapies against diseases, such as cancers, infectious diseases, and neurodegenerative disorders by selectively turning off specific genes in diseased tissues. From https://www.eupati.eu/glossary/silencing/

Somatic cells: any cell in the body others than the reproductive cells (i.e. gametes and their precursors). See also germ cells.

Stem cell therapy, also known as regenerative medicine, is the use of stem cells to treat or prevent a disease or condition. Stem cells grown in a lab are manipulated to specialise into specific types of cells, such as heart muscle cells, blood cells or nerve cells. The specialised cells can then be implanted into a person. For example, if the person has heart disease, the cells could be injected into the damaged heart muscle. The healthy transplanted heart cells could then contribute to repairing defective heart muscle. From https://www.eupati.eu/glossary/stem-cell-therapy/

Splicing: In some genes the protein-coding sections of the DNA ("exons") are interrupted by non-coding regions ("introns"). RNA splicing removes the introns from pre-mRNA to produce the final set of instructions for the protein. From https://www.dnalc.org/resources/3d/rna-splicing.html

T

Transcription The first step during protein synthesis when a sequence of DNA in a gene is copied to produce an RNA transcript called messenger RNA (mRNA). Adapted from https://www.yourgenome.org/glossary/transcription

Transcription Factors can turn genes on and off. It is possible that genes can be turned on and turned off by stress or exposure to environmental events. These events could potentially turn on transcription factors in the genes of interest, thus leading to increased gene expression of these genes. Similarly, genes can be blocked or the transcription of genes can be blocked by certain events and that is also possible. From https://www.dnalc.org/view/2299-Transcription-Factors-Can-Turn-Genes-On-and-Off.html

Transcriptomics is the study of the transcriptome— To produce proteins, genes are first 'transcribed' into messenger RNA (mRNA). The transcriptome represents the whole set of mRNA molecules present in a specific cell or tissue at a certain time. By analysing the transcriptome, researchers can determine when each gene is turned on or off in a cell or tissue, how that type of cell normally functions, and how changes in the normal level of gene activity may be altered by or contribute to disease. From https://www.eupati.eu/glossary/transcriptome/

Translation: The second step during protein synthesis where the genetic code for making a protein has been taken to the ribosome by mRNA (messenger RNA) to direct the synthesis of proteins from amino acids. Adapted from https://www.yourgenome.org/glossary/translation

U

V

W

Whole genome sequencing: Whole genome sequencing (WGS) is a laboratory process that determines the complete DNA sequence of an organism's genome at a single time. Great progress in the speed at which genomes can be sequenced, in the number of genomes that can be sequenced at the same time, and in the reducing the cost of sequencing has had a huge impact on medical research and medicines development. High-throughput genome sequencing technologies have largely been used as a research tool and are currently being introduced into clinical practice. In the future of personalised medicine, whole genome sequencing will be an important tool to guide treatments. From https://www.eupati.eu/glossary/whole-genome-sequencing/

X

Y

Z