Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

Biological data glossary

Key words and terms that you need when using biological data.

Adenine	One of the four 'letters' in the DNA code (A). See 'base'
Allele	A genetic variant, for example at the SNP named 'rs268' on could either have the 'A' allele or the 'G' allele. The 'A' allele is more common, so it is termed the 'major' allele, making 'G' the 'minor' allele.
Allostatic load	Dysregulation of bodily systems caused by chronic psychosocial stress. https://www.sciencedirect.com/topics/neuroscience/allostatic-load
Ancestry	In genetics this refers to the population(s) from which one's genome is derived. It is a more precise term favoured over 'race' or 'ethnicity'. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7082057/
Array	Synonymous with 'microarray'. https://www.nature.com/scitable/definition/microarray-202/
Base	In the context of genetics, refers to the specific letters in the DNA code. The four bases are adenine, guanine, cytosine and thymine (A,G,C,T).
Biological ageing	The progressive decline in physiological ability to meet demands that occur over time. It is due to the accumulation of damage at the cellular level. https://doi.org/10.1093/eurpub/14.3.331
Biomarker	Broad sense: in social science often used to denote any measurement derived from the human body which might relate to health, including grip strength, waist circumference, lung function, etc. Narrow sense (as defined by the National Institute for Health): "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenetic processes, or pharmacological responses to a therapeutic intervention".
Cell	The basic building block of all living things. The human body is composed of trillions of cells. Each cell (except red blood cells) contains a nucleus which contains our DNA. https://medlineplus.gov/genetics/understanding/basics/cell/
Chromosome	Made of DNA tightly coiled many times around proteins called histones that support the structure. Humans have 23 pairs of chromosomes. https://medlineplus.gov/genetics/understanding/basics/chromosome/
Coverage	In microarrays, coverage refers to the proportion of known SNPs that are included in the array, or which can be reliably imputed from the included SNPs.

CpG	A pair of DNA nucleotides (C followed by G) which may be subject to methylation (which takes place in the C portion).
Cytosine	One of the four 'letters' in the DNA code (C). See 'base' and 'CpG'.
DNA	Deoxyribonucleic acid. The chemical in our cells which comprises our genome.
Epigenetic clock	Epigenetic clocks use algorithms for calculate biological age on the basis of a read out of the extent to which to dozens, or even hundreds, of sites across an individual's genome are bound by methyl groups – a form of epigenetic modification. https://www.nature.com/articles/d41586-022-00077-8
Epigenetics	The study of mechanisms that affect gene expression by altering the DNA in a way that does not change its code. This term is often used to indicate 'epigenomics'. Encompasses several mechanisms, one of which is DNA methylation.
Epigenome-wide association study (EWAS)	A study in which an exposure or an outcome is tested for association with a large number of epigenetic marks across the genome, using e.g. empirical Bayes method.
Epigenomics	The study of epigenetic marks across the genome.
Gene	A section of a DNA molecule which the cell can read and translate to produce a protein.
Gene expression	The "use" of a gene to make RNA and proteins. Often measured by the presence of RNA copies of genes (mRNA transcripts) in a biological sample, though the presence of protein is also evidence of gene expression (albeit a more distal one).
Genetics	The study of how the DNA code relates to traits (genetically determined characteristics) and health conditions. This term often indicates approaches that focus on a narrow set of genes or markers, but is often used more broadly to include 'genomics'.
Genome	Can be considered at the species or individual level. All of the genetic material that an individual (or species) possesses. In a human, this usually includes 46 chromosomes (23 pairs) plus the mitochondrial DNA. A 'reference genome' for a species or population should include information regarding variation across the genome.
Genome-wide association study (GWAS)	A study in which a phenotype is tested for association with a large number of genetic marks (especially SNVs) across the genome, using linear or logistic regression.

Genomics	The study of variation across the genome. The term can be used to include epigenomics and other DNA-related "omics" that measure features other than DNA sequence.
Genotype	An individual's unique genetic profile. In the context of microarrays, a person has a genotype for each SNP which can be expressed either as two letters, e.g. AA, AT or TT, or a number: 0, 1 or 2.
Guanine	One of the four 'letters' in the DNA code (G). See 'base' and 'CpG'.
Haplotype	"A set of DNA variations, or polymorphisms, that tend to be inherited together. A haplotype can refer to a combination of alleles or to a set of single nucleotide polymorphisms (SNPs) found on the same chromosome." https://www.genome.gov/genetics-glossary/haplotype
Hardy-Weinburg equilibrium	"When mating is random in a large population with no disruptive circumstances, [Hardy-Weinburg equilibrium] predicts that both genotype and allele frequencies will remain constant". https://www.nature.com/scitable/definition/hardy-weinberg-equilibrium-122/
Heterozygous	Having one copy each of two different alleles at a specific genetic locus.
Homozygous	Having two copies of one allele at a specific genetic locus.
Imputation	Estimating a missing value in a dataset. In the context of genomics, this can be done for missing SNPs that are in linkage disequilibrium with a genotyped SNP.
Limit of Detection (LOD)	"The lowest measurable level of an individual protein" (or other substance). https://www.olink.com/faq/what-does-lod-mean/
Linkage disequilibrium	The state of multiple alleles at different loci being randomly co-inherited. (see also 'haplotype') https://www.sciencedirect.com/topics/neuroscience/linkage-disequilibrium
Locus	The specific point on the genome where a variant or gene can be found. (Plural: loci)
Methylation	The chemical addition of a 'methyl group' to a cytosine, having an effect on the expression of a gene.
Microarray	A microarray is a technology used to test thousands or millions of specific genomic loci simultaneously. It is a slide, or 'chip', with short, manufactured strands of DNA (probes) attached to it which are designed to detect SNVs or methylated/un-methylated CpG sites. https://www.nature.com/scitable/definition/microarray-202/
Minor allele	The minor allele is the least common allele for a given SNV.

Frequency (MAF)	The MAF is the prevalence of the minor allele in a given population, expressed as a proportion (0.0 – 1.0).
Multi-omics	Analytical approaches that incorporate different types of biological datasets, e.g. genomic, transcriptomic, proteomic, etc.
Mutation	A rare genetic variant. https://www.nature.com/scitable/topicpage/genetic-mutation-1127/
Nucleotide	The building blocks of DNA and RNA. Each nucleotide consists of one 'base' plus a sugar and a phosphate. In DNA, the four bases are adenine, guanine, cytosine and thymine (A, G, C and T). Nucleotides are also referred to as 'residues'. https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/nucleotide
Nucleus	The organelle (sub-compartment) of a cell that contains DNA.
Omics	The suffix (sometimes used as a word) that denotes an approach wherein a broad, comprehensive set of molecules of a certain class are assayed (measured) or analysed simultaneously.
Phenotype	An individual's observable traits, such as height, eye colour, and blood type. This could also be broadened to include disease states, disease risk factors and behaviours, etc. https://www.genome.gov/genetics-glossary/Phenotype
Plasma	The liquid portion of an unclotted blood sample that has been separated in a centrifuge
Pleiotropy	"The phenomenon in which a single gene contributes to multiple phenotypic traits." https://www.nature.com/scitable/topicpage/pleiotropy-one-gene-can-affect-multiple-traits-569/
Polygenic score	The sum of an individual's alleles which may contribute to a given phenotype, usually weighted by GWAS effect size.
Population stratification	"Refers to differences in allele frequencies between cases and controls due to systematic differences in ancestry rather than association of genes with disease" (or other phenotype). https://doi.org/10.1038/ng1333
Principal component analysis	A technique used to calculate latent variables (principal components) which explain the variance in a dataset. In genomics, SNP microarray data can be used in this way to produce principal components which reflect ancestry. These can be used to identify population outliers and correct for population stratification.
Probe	A short piece of manufactured DNA affixed to a microarray chip, used to detect variants or methylation levels.

Protein	An extremely diverse class of biological molecule, each protein is composed of amino acids and is encoded by a gene. Proteins carry out every process in our bodies.
Proteomics	The study of a broad range of proteins.
Recombination	The exchange of genetic material between chromosomes. Also known as "crossing over". https://www.genome.gov/genetics-glossary/homologous-recombination
RNA	Ribonucleic acid. A molecule similar to DNA. There are different types of RNA. "Messenger RNA" (mRNA) is essentially a short-lived copy of a gene used to create a protein. mRNA can be quantified to measure gene expression.
Serum	The liquid portion of a clotted blood sample that has been separated in a centrifuge.
Single nucleotide polymorphism (SNP)	A single nucleotide variant (SNV) which has a minor allele frequency of at least 1%.
Single nucleotide variant (SNV)	A specific point in the DNA code where one nucleotide may be substituted with another. https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/single-nucleotide-variant
Thymine	One of the four 'letters' in the DNA code (T). See 'base'.
White blood cells (WBCs)	The blood consists of plasma, red blood cells and white blood cells. As red blood cells do not contain DNA, this is extracted from white blood cells. There are several distinct types of white blood cells with different roles and different DNA methylation profiles.