

GLOSSARY

Basic molecular genetics for epidemiologists

F Calafell, N Malats

This is the first of a series of three glossaries on molecular genetics. This article focuses on basic molecular terms.

A general increase in the number of epidemiological research articles that apply basic science methods in their studies, resulting in what is known as both molecular and genetic epidemiology, is evident. Actually, genetics has come into the epidemiological scene with plenty of new sophisticated concepts and methodological issues.

This fact led the editors of the journal to offer you a glossary of terms commonly used in papers applying genetic methods to health problems to facilitate your "walking" around the journal issues and enjoying the articles while learning.

Obviously, the topics are so extensive and innovative that a single short glossary would not be sufficient to provide you with the minimum amount of molecular and genetic concepts to range over the whole field. Hence, we have organised the manuscript in three short glossaries that will try to guide you from the most basic molecular terms (the first glossary, published in this issue) to the most advanced genetic terms, most of them related to new study designs and laboratory techniques (the last glossary).

We have attempted to provide concise definitions and some examples of the most used concepts and designs in genetic epidemiology articles. Nevertheless, we are aware that the glossaries are not exhaustive and we refer the reader to other texts.¹⁻⁴

This initiative does not pretend to cover concepts in molecular epidemiology as this would require a list of terms as large as the one presented here. However, as the two areas are related, some of the concepts used by molecular epidemiology are defined here, too. In some cases, a single term may be used in both scenarios with slightly different meanings (for example, *marker*).

ALLELE

Each of the different states found at a *polymorphic* site. Different alleles and their combinations may result in different *phenotypes*. For example, the ABO gene contains three major alleles, A, B, and O; AA and AO individuals express the A blood group; BB and BO express B; AB appear as AB, and only OO individuals express the O blood group.

AUTOSOME

Non-sex chromosome.

J Epidemiol Community Health 2003;**57**:398-400

CHROMOSOME

Linear or (in bacteria and organelles) circular DNA molecule that constitutes the basic physical block of heredity. Chromosomes in diploid organisms such as humans come in pairs; each member of a pair is inherited from one of the parents. Humans carry 23 pairs of chromosomes (22 pairs of *autosomes* and two *sex chromosomes*); chromosomes are distinguished by their length (from 48 to 257 million base pairs) and by their banding pattern when stained with appropriate methods.

Homologous chromosome

Each of the chromosomes in a pair with respect to the other. Homologous chromosomes carry the same set of *genes*, and *recombine* with each other during *meiosis*.

Sex chromosome

Sex determining chromosome. In humans, as in all other mammals, embryos carrying XX sex chromosomes develop as females, whereas XY embryos develop as males. The X and Y chromosomes contain different, partly overlapping sets of genes.

CODON

Each of the 64 different nucleotide triplets in DNA that, when transcribed into RNA, are then translated into an amino acid in a protein. For example, the β haemoglobin gene starts with the DNA sequence ATGGTG... (that is, with the ATG GTG ... codons), which is then transcribed into the *messenger RNA* sequence AUG GUG..., which means that the haemoglobin protein sequence will start with amino acids MetVal... Codon ATG always corresponds to amino acid methionine in the corresponding protein, GTG to valine, and so the 64 different codons map to the 20 different amino acids. This correspondence table is called the genetic code. Often, all four codons that differ only in their third nucleotide code for the same amino acid; thus, most DNA sequence changes affecting the third position in a codon do not change the resulting protein.

Stop codon

Codon signalling the end of the coding portion of a gene. In mammals, stop codons are TGA, TAA, and TAG.

DNA (DESOXYRIBONUCLEIC ACID)

Macromolecule that constitutes the basis of heredity. It is a double helix made up of four different types of subunits or nucleotides: adenine, guanine, cytosine, and thymine (or A, G, C, and T). Each nucleotide is made of a different base, plus phosphate and the desoxyribose sugar. Nucleotides in each strand of the helix face

See end of article for authors' affiliations

Correspondence to:
Francesc Calafell, Carrer
del Dr Aiguader 80,
E-08003 Barcelona, Spain;
francesc.calafell@cexs.upf.es

nucleotides in the other in a complementary way: A bonds with T and G with C; the sequence in one strand of the double helix effectively determines the sequence in the other strand. DNA is replicated semi-conservatively by enzymes known as DNA polymerases that open the double helix and bind together two new strands by inserting the appropriate complementary nucleotides. Sections of DNA (see *genes*) are transcribed into RNA, which is then used as a template to build proteins: the DNA sequence is effectively decoded and translated into a protein.

Coding DNA

DNA that actually carries genetic information. It is just 3% of the total DNA.

Junk DNA

DNA that does not seem to have any function. In fact, the human genome is riddled with sequences that derive from non-pathogenic viruses that inserted their DNA into the human genome, and that have been inadvertently copied ever since.

Mitochondrial DNA (mtDNA)

Small circular DNA molecule contained in the mitochondria. mtDNA is 16 500 basepairs long, just a small fraction of the 3200 million bp in the nuclear genome. Each mitochondrion in a cell carries tens of mtDNA copies, usually identical (a situation called *homoplasmy*) but not always so (*heteroplasmy*). Some disease causing mutations in mtDNA are only found in heteroplasmy as they would be lethal in homoplasmy. mtDNA codes for some of the proteins in the respiratory chain, the core of the energy producing cellular machinery that resides in mitochondria. It seems that mtDNA from the sperm cells does not penetrate the ovum, being mtDNA inherited solely from the maternal line.

Non-coding DNA

DNA that is not transcribed into RNA, and, thus, not translated into protein. Non-coding DNA can have other functions, such as acting as a signal to modulate the expression of a particular gene.

Nuclear DNA

DNA contained in the nucleus of the cell; in fact, all but the *mitochondrial DNA* is nuclear.

EPIGENETIC EFFECT

Change in the outcome of a particular gene that is not controlled genetically. DNA methylation is one such change, which can turn off the expression of some genes.

EXON

Each of the segments in a *gene* that are transcribed, and whose transcripts are spliced together to form the *messenger RNA*. In some cases, different proteins can be coded by the same gene by alternative splicing, that is, by different combinations of exons forming different messenger RNAs, and, therefore, being translated into different proteins.

GENE

DNA segment that is transcribed into *messenger RNA* and translated into a protein. Genes comprise the exons that are actually translated plus the intervening *introns*.

GENOME

Whole set of the *DNA* of a species. The human genome is made of 23 pairs of *chromosomes* plus *mtDNA*, for a total of over 3200 million base pairs.

GERM LINE

Cell lineage that, after a number of divisions and *meiosis*, leads to the production of the gametes (sperm or ova). Mutations in the germline can be passed on to the offspring.

HETEROZYGOTE

Individual that carries two different *alleles* at the same site in the two *homologous chromosomes* of a given pair.

HOMOZYGOTE

Individual that carries two copies of the same *alleles* at the same site in the two *homologous chromosomes* of a given pair.

INTRON

Each of the segments of a *gene* that are not transcribed into *messenger RNA* and that are found between *exons*.

LOCUS

Any given genome region

MICROSATELLITE (SYNOBYN, SHORT TANDEM REPEAT, STR)

DNA segment consisting in the repetition 5–50 times of a motif 1–6 basepairs long. Microsatellites tend to be polymorphic in their number of repetitions because of a high mutation rate. DNA polymerases tend to “slip” when copying microsatellite tracts, adding or subtracting repeat units. Given their high polymorphism, microsatellites are widely used in mapping genetic diseases, in genetic counselling, in forensic genetics, and in population genetics.

MUTATION

Any change in a *DNA* sequence arising from an error in the duplication process. In a clinical sense, any such change that disrupts the information contained in DNA and leads to disease. The mechanisms leading to mutations are diverse: from exogen and endogen carcinogens to DNA repair defects.

Frameshift mutation

Indel mutation that disrupts the reading frame within a *gene*. For example, ATG GTG CAC CTG ACT translates into protein sequence MetValHisLeuThr, whereas, if a C is inserted in the fourth position, the reading frame becomes ATG CGT GCA CCT GAC T, which reads MetArgAlaProAsp—that is, a completely different protein and likely to be non-functional.

Gain of function mutation

Mutation resulting in a protein having a different function from the original.

Germline mutation

Any mutation occurring in the *germ line* and transmitted to the offspring.

Indel mutation

Mutation that consists in the insertion (addition) or deletion of one or a few nucleotides

Missense mutation

Nucleotide substitution that changes one codon for another resulting in a single amino acid change, as in ATG GTG CAC CTG ACT to ATG GTG CAC GTG GCT, that is, from MetValHisLeuThr to MetValHisValThr. The phenotypic severity of such a mutation depends on the relative functional importance of the amino acid position mutated and on the chemical similarity between the original and the new amino acids.

Nonsense mutation

Nucleotide substitution that creates a *stop codon*. ATG GTG AAA GTA... (MetValLysVal...) to ATG GTG TAA GTA would result in a truncated protein (MetVal), most likely to be non-functional

Null mutation

Mutation leading to the complete abolition of the expression of a *gene*.

Regulatory mutation

Mutation affecting the regulatory region of a *gene*. Although it does not change the protein sequence coded by the *gene*, it may affect its levels of expression and cause a recognisable *phenotype*.

Silent mutation

Mutation that does not change the genetic information, either because it lies in a non-coding region, or because it changes a *codon* into another coding for the same amino acid. The second case is called a *synonymous mutation*.

Somatic mutation

Mutation happening in any non-germ line cell and affecting the cells descending from it, but not the offspring of the individual. Somatic mutations can cause cancer.

REPEAT EXPANSION

Mutation in a repeat tract that increases the number of repeats by a large amount and that may cause a phenotypic effect. The molecular mechanism causing repeat expansions is different from that of ordinary, single repeat mutations in *microsatellites*. Diseases such as myotonic dystrophy and Huntington's disease are caused by repeat expansions.

RESTRICTION ENZYME

Any enzyme, usually found in bacteria, that cuts DNA when it finds a given four nucleotide or six nucleotide target sequence. Restriction enzymes are widely used in molecular biology. See also *RFLP*.

RNA (RIBONUCLEIC ACID)

Macromolecule that, with *DNA*, constitute the nucleic acids. RNA does not form a double helix, although it can take complex three dimensional structures. Chemically, RNA nucleotides contain ribose rather than desoxyribose, and uracil instead of thymine. Different RNA forms exist with specific functions (see below).

Messenger RNA (mRNA)

Any RNA molecule that results from the transcription of a particular gene. mRNA takes the genetic information from the cell nucleus to the cytoplasm, where it will be translated into proteins in the ribosomes.

Ribosomal RNA (rRNA)

Any of a number of different RNA molecules that have structural functions in the ribosome, the cell organelle where translation occurs.

Transfer RNA (tRNA)

Small RNA molecule involved in protein synthesis that contains an anticodon (a three nucleotide sequence complementary to a given *codon*) and that carries at one end the amino acid that corresponds to that *codon*.

SOMATIC CELL

Any non *germ-line* cell.

WILD TYPE

Applied to the normal, non-altered sequence of a gene, as compared with any mutated sequence.

Contributors

Both authors have contributed equally to the manuscript.

Authors' affiliations

F Calafell, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain
N Malats, Institut Municipal d'Investigació Mèdica, Barcelona, Spain

REFERENCES

- 1 **Various authors**. Genetic epidemiology. *Epidemiol Rev* 1997;19. (Monographic issue on genetic epidemiology).
- 2 **Hartwell LL**, Hood L, Goldberg ML, et al. *Genetics, from genes to genomes*. Boston: McGraw Hill, 2000. (includes a glossary)
- 3 **King RC**, Stansfield WD. *A dictionary of genetics*. Oxford: Oxford University Press, 1997. (extensive but needs updating).
- 4 **Strachan T**, Read AP. *Human molecular genetics*. Oxford: Bios, 2001. (includes an extensive glossary).