GLOSSARY

Basic molecular genetics for epidemiologists

F Calafell, N Malats

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

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.*

See end of article for authors' affiliations

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

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 aminoacid 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 aminoacids MetVal... Codon ATG always corresponds to aminoacid methionine in the corresponding protein, GTG to valine, and so the 64 different codons map to the 20 different aminoacids. This correspondence table is called the genetic code. Often, all four codons that differ only in their third nucleotide code for the same aminoacid; 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

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 (SYNOMYN, 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 **C**GT 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 **C**TG ACT to ATG GTG CAC **G**TG GCT, that is, from MetValHis-**Leu**Thr to MetValHis**Val**Thr. 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.

CORRECTION

An error ocurred in this article by Calafell and Malats (2003;**57**:398–400). On page 399 in the section "Missense mutation" the first sentence should have read ATG GTG CAC CTG ACT to ATG GTG CAC GTG ACT ATG [not GTG CAC CTG ACT to ATG GTG CAC GTG GCT].

Nonsense mutation
Nucleotide substitution that creates a *stop codon*. ATG GTG **A**AA GTA... (MetValLysVal...) to ATG GTG **T**AA 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 aminoacid. 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 is 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).

GLOSSARY

Basic glossary on genetic epidemiology

N Malats, F Calafell

J Epidemiol Community Health 2003;**57**:480–482

This is the second of a series of three glossaries on genetic concepts used in epidemiological research that the journal is publishing with the objective of helping the reader "walk" around the journal.

The first glossary, on basic molecular genetic terms,[1] provided the basis to understand the concepts presented here.

In this section, we refer to the most basic and commonly used genetic terms that epidemiologists interested in this kind of research need to know. The concepts defined here are the pillars on which genetic epidemiology builds up its methodology.

Again, the list is not exhaustive and an attempt has been made to provide you with concise definitions. Hence, we encourage the interested reader to further "explore" the classic bibliography on genetics and genetic epidemiology[2–5] to complete their knowledge on this topic.

CANDIDATE GENE

A known *gene* suspected to be associated with the disease of interest on the basis of the biological function of its protein.

COMPLEX TRAIT (POLYGENIC TRAIT, MULTIFACTORIAL TRAIT)

Any *phenotype* that results from the effect of multiple *genes* at two or more loci, with possible environmental influences too. Examples are: obesity, hypertension, hypercholesterolaemia, skin pigmentation, cancer, etc.

Discontinuous trait

Trait that is either present or absent, such as birth defects and common behavioural disorders. The threshold model is used to explain discontinuous traits: a protein level has a continuous distribution but the *phenotype* does not appear until a certain threshold is reached.

Continuous trait

Measurable trait that is always present and that follows a normal distribution in the population. For example: height, weight, and blood pressure.

See end of article for authors' affiliations

Correspondence to: Dr N Malats, Institut Municipal d'Investigació Mèdica, Carrer del Dr Aiguader 80, E-08003, Barcelona, Spain; nuria@imim.es

CO-SEGREGATION

The tendency of two traits to be jointly inherited.

DNA REPAIR

Major cell defence system against DNA damage produced by environmental and endogenous compounds. There are several different repair pathways and several enzymes (some of them *polymorphic*) involved in each way. Abnormalities in these processes have been implicated in cancer and aging.

EPISTASIS

Gene interaction and, particularly, interaction between different alleles at different genes. Epistasis can occur at the same step or at different stages of the same biochemical pathway.

FAMILIAL AGGREGATION

A tendency of a disease to cluster in families, which is generally taken as evidence for the existence of a genetic aetiological mechanism, or environmental factors common to family members, or a combination of both. Ascertainment bias should be seriously considered.

FAMILIAL ANTICIPATION

The younger age of appearance of a late onset trait in successive generations. A typical effect in *repeat expansions*, in which severity of the disease is proportional to repeat length, which tends to grow in each transmission.

FOUNDER EFFECT

A change in the population *allele* frequency that occurs when a subpopulation is established by a small number of individuals. The change occurs only by chance because the members of the new population are a random subsample that may deviate from the overall allele frequencies. Such changes are stronger in smaller founder populations, given the higher sampling variance.

GENETIC HETEROGENEITY

Process by which a phenotype can be caused by different loci. A complex example is epilepsy, which may be attributable to different causes in different individuals: single *gene* disorders, *multifactorial inheritance*, chromosomal disorders, or even brain injuries. The last case is a *phenocopy*.

GENOTYPE

The genetic constitution of an organism, which is modulated by the environment before being expressed as a phenotype.

HAPLOTYPE

Set of allelic states found at neighbouring loci in a chromosome, as inherited from a parent. Haplotypes can be broken down by *recombination*. A haplotype shared among unrelated individuals affected with a genetic disease may indicate that a gene causing the disease maps to that genomic region.

HARDY-WEINBERG EQUILIBRIUM
State in which the *allele* and *genotype* frequencies do not change from one generation to the next in a population. It requires random mating and the absence of selection, mutation, migration, and genetic drift. In Hardy-Weinberg equilibrium, allele and genotype frequencies are related through the **Hardy-Weinberg law**: for a locus with two alleles P, Q at frequencies p and q respectively, homozygotes for P are found at frequency p^2, homozygotes for Q have a frequency q^2, and heterozygotes are found at a frequency $2pq$. Although conditions for Hardy-Weinberg equilibrium are seldom strictly met, genotype frequencies are usually consistent with the Hardy-Weinberg law. Some useful software packages to test whether a set of genotypic frequencies conforms to Hardy-Weinberg are Arlequin (http://anthropologie.unige.ch/arlequin/) and Genepop (http://wbiomed.curtin.edu.au/genepop/), among others.

INHERITANCE
Pattern followed by the transmission from generation to generation of a given *phenotype*, usually a disease.

Complex inheritance (non-Mendelian inheritance)
Variability in *phenotype* expression that is attributed both to the inheritance of combinations of alleles at multiple loci and to environment exposures.

Multifactorial inheritance
Complex inheritance in which multiple genes are involved jointly with environmental influences.

Polygenic inheritance
Complex inheritance in which multiple genes but no environmental factors are involved.

Mendelian inheritance
Simple pattern of inheritance that follows the rules set out by Mendel. Mendelian traits are determined by just one genetic locus, with complete *penetrance* and no *phenocopies.* Mendelian inheritance can be *dominant*, *recessive*, or *sex linked*.

Dominant inheritance
Type of inheritance in which one copy of an abnormal gene is sufficient to cause disease (for example, Huntington's disease). If *penetrance* is complete, the abnormal gene is inherited from a parent who also has the disorder and every generation in the family has members with the disorder.

Recessive inheritance
Type of inheritance in which two abnormal copies of the gene must be present for the individual to be affected (for example, cystic fibrosis). Each parent contributes one abnormal copy of the gene to the child who has the disorder. *Heterozygous* individuals (such as the parents of the affected) are called carriers of the disorder because they have one normal and one abnormal copy of the gene, but they do not show symptoms of the disorder.

Sex linked inheritance
Type of inheritance followed by the traits caused by genes located on the X or (rarely) on the Y chromosomes. X linked disorders can also be recessive or (very rarely) dominant. When the abnormal gene that is responsible for a recessive disorder is located on the X chromosome (for example, haemophilia) usually only males are affected because they do not possess a second, normal, copy of the gene. Such males are called hemizygous. X linked dominant inheritance (for example, Rett syndrome) follows a pattern similar to autosomal dominant inheritance except that more females are affected than males.

LINKAGE
The phenomenon whereby *phenotypes* and *alleles* at one or more marker *alleles* tend to be inherited together more often than expected. Linkage usually means that a gene contributing partially or completely to the *phenotype* (a genetic disease, for instance) maps in the vicinity of the *markers*.

MARKER
Any neutral *polymorphism* used in *linkage* or *association analysis*.

MEIOSIS
Process whereby four haploid germ cells (gametes) are produced from a diploid parent cell for sexual reproduction. During meiosis crossovers occur between *homologous chromosomes* so that each chromosome found in the gamete consists of a patchwork of material from both members of the pair.

METABOLISM (OF EXOGENOUS AND ENDOGENOUS CHEMICAL COMPOUNDS)
Cellular system of enzymes (most of them polymorphic) that activates and deactivates chemical compounds through chemical radicals. Metabolic enzymes are classified in two groups according their most important function, activating or deactivating, and in several families.

MITOSIS
Asexual reproduction of a somatic cell in which the two daughter cells each have a genetic makeup that is identical to that of the parent cell.

PENETRANCE
The likelihood, or probability, that a particular genotype will be expressed in the *phenotype*. A penetrance of 100% means that the associated phenotype always occurs when the corresponding genotype is present. Similarly, if only 30% of those carrying a particular allele (such as a disease-causing mutation) exhibit a phenotype (the disease), the penetrance is 30%.

PHENOCOPY
An environmentally caused *phenotype* that mimics a genetic trait. For example, epilepsy can be caused by mutations in single genes (with *genetic heterogeneity*), and, among other causes, by brain injury, which produces a phenocopy of genetic epilepsy.

PHENOTYPE
Expressed traits or characteristics of an organism, regardless of whether or to what extent the traits are the result of genotype or environment, or of the interaction of both. For example, hair colour, weight, or the presence or absence of a disease.

(GENETIC) POLYMORPHISM
Genome segment (locus), within or outside a *gene*, in which alternate forms (*alleles*) are present. In population genetics, variation is polymorphic if all alleles are found at frequencies >1%. In clinical genetics, a polymorphism refers to any genetic variation not known to be a direct cause of disease, in contrast with a *mutation*. However, the distinction between mutation and polymorphism in the latter sense may be rather fuzzy, as the path from genetic variation to disease can be sometimes very complex. In molecular epidemiology, *metabolic* and *DNA repair* gene polymorphisms are some of the markers (indicators) used to explore genetic susceptibility to develop a

disease. They are considered under the hypothesis that they can affect the development of the disease only in the presence of an environmental risk factor.

RECOMBINATION

Presence in the offspring of allelic combinations in a chromosome (that is, *haplotypes*) not present in the parents as a result of **crossing over** (*see meiosis*). The average probability of recombination is 1% per million base pairs, although this figure varies greatly across the genome.

Contributors

Both authors have contributed equally to the manuscript.

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

REFERENCES

1 **Calafell F**, Malats N. Glossary on basic molecular genetics. *J Epidemiol Community Health* 2003;**57**:398–400.
2 **Various authors**. Genetic epidemiology. *Epidemiol Rev* 1997;**19**. (Monographic issue on genetic epidemiology).
3 **Hartwell LL**, Hood L, Goldberg ML, *et al. Genetics, from genes to genomes*. Boston: McGraw Hill, 2000. (Includes a glossary).
4 **King RC**, Stansfield WD. *A dictionary of genetics*. Oxford: Oxford University Press, 1997. (Extensive but needs updating).
5 **Strachan T**, Read AP. *Human molecular genetics*. Oxford: Bios Publishers, 2001. (Includes an extensive glossary).

SPEAKER'S CORNER

The urge for evidence based knowledge

Clearly, the context of health sciences is increasingly been influenced by a general commitment related to interventions intending to improve the health conditions of individuals and populations brought by the so called evidence based knowledge. You can easily notice a proliferation of evidence based issues and debates in scientific health (including community health) journals.

Our view about the challenges to an evidence based knowledge does not refer to the impending difficulties of, sometimes, not having confirmed evidences on specific matters, or, even when they are available, there is uncertainty on which of them would be the best choice. The emphasis here is placed in the debatable conviction of the supremacy of some kind of knowledge over others that does not seem to require a plain belief in the truth itself to support it.

This standpoint needs some kind of a reasonably grounded concept of truth (even if it is a provisional one). Especially when anyone says that they possess the techniques to establish what is the actual best approximation to the truth. None the less, truth as a hierarchical quality of knowledge is, eventually, a social relationship whereby hegemony and domination are acting decisively. Rational truth will, in the end, guarantee the certainty that rational order is the only ticket that should enable mankind to arrive—sooner or later—to the ultimate and universal truth, whatever it may be.

You can notice in language that there are verbal forms for the opposite nouns of truth/veracity: falsehood—to falsify; lie—to lie. But there is not an equivalent verb for "truth". We can only verify the truthfulness of something. Perhaps because it seems that there is always the supposition that its "concrete" existence as a fact can, sooner or later, be established. The truth extracted by scientific means seems to have an ontological dimension in itself. Only, it would be necessary in a due course of time, to discover it, reveal it...

Well, a theory of truth tries to sustain a constant and secure superiority of some set of beliefs over others based upon an idea that these beliefs were achieved through reliable procedures and/or were put forward by special kind of people that can be trusted sufficiently to be followed. Therefore, a rhetoric of power granted by a rationalistic mode of reasoning and its undeniable successes is always at stake.

L D Castiel
National School of Public Health, Oswaldo Cruz Foundation, Rua Leopoldo Bulhões 1480/802, Rio de Janeiro, Brazil 21041-210;
castiel@ensp.fiocruz.br

GLOSSARY

Advanced glossary on genetic epidemiology

N Malats, F Calafell

J Epidemiol Community Health 2003;**57**:562–564

This is the last of a series of glossaries on terms used in genetic epidemiology published by the journal. This glossary covers the most advanced genetic terms, most of which are related to new study designs and laboratory techniques. It provides the reader with examples and references of real studies that applied each of the study designs defined in the glossary. This should help the reader grasp the subtleties of each of these strategies and will allow the reader to research the literature according to their interest.

See end of article for authors' affiliations

Correspondence to: Dr N Malats, Carrer del Dr Aiguader 80, E-08003 Barcelona, Spain; nuria@imim.es

The previous glossaries on basic molecular and genetic concepts[1 2] gave the basis for the understanding of those included here.

Given the space constraints, we chose not to be exhaustive and be concise. Hence, we again encourage the interested reader to "explore" the classic bibliography on genetics and genetic epidemiology.[3–6]

AFFECTED SIB-PAIR APPROACH

Study design used to find genetic factors contributing to a complex trait. It tests for *linkage* by considering the proportion of shared alleles between affected sib-pairs at *markers* spaced over the whole genome or over a section of it. A null distribution of the expected relative frequencies of sibs sharing zero, one or two alleles at a marker can be derived and tested against the observed data. An excess of allele sharing at a marker may indicate the presence in its vicinity of a gene contributing to the disease. This method also permits testing of gene-environment interaction. This design was applied by Lachmeijer *et al* to assess the involvement of IL1B and IL1RN gene polymorphisms in causing pre-eclampsia.[7] They collected 150 pairs of sisters that had suffered pre-eclampsia while pregnant and typed two polymorphisms at IL1B and one at IL1RN. Unfortunately, the degree of allele sharing among sisters did not suggest that those genes were involved in pre-eclampsia.

ALLELE SPECIFIC AMPLIFICATION

Polymerase chain reaction based methods for detecting disease causing mutations that consist in amplifying specifically one or both *alleles* by using specific primers in one or two independent reactions. If two allele specific primers are used in a single reaction, additional chemistry is needed to determine which primer produced the amplification.

ASSOCIATION ANALYSIS

Comparison of the frequency of *alleles* in *candidate genes* between unrelated affected and unaffected individuals. The alleles analysed may be thought to contribute to the disease or be in *linkage disequilibrium* with any such causative variation. It can provide sufficient power to distinguish slight variations in disease risks being more sensitive than linkage methods when the *genes* of interest contribute to disease susceptibility but are neither necessary nor sufficient to cause disease. The methodology it uses is the same as used by epidemiological studies (cohort and case-control design). For instance, Perikac-Vance *et al* had mapped a gene conferring susceptibility to late onset Alzheimer's disease at chromosome 19q13.2[8]; as apoliprotein E is found bound to the amyloid plaques characteristic of Alzheimer's disease and is also found in that genome region, it became a candidate gene for Alzheimer's disease. This was confirmed by Strittmatter *et al* by typing variants of the ApoE gene in 30 affected individuals and in 91 presumably healthy controls.[9] They found that the frequency of the APOE-ϵ4 allele was significantly higher in the patients than in the controls, which showed that this allele confers susceptibility to Alzheimer's disease.

CASE ONLY DESIGN

Approach to screen *gene*-environment interactions under the assumption of independence between exposure and *genotype* in the population. This design does not require control subjects. Therefore, sample sizes will be less than half than those required in case-control studies and the estimated odds ratios will not suffer from potential biases related to control selection. Cases are distributed in a 2×2 table according to their genetic and environmental exposure status. To further explore the differences between a case only and a case-control design we suggest the reader looks at the study by Bai *et al* that compared both approaches to assess gene-environmental interaction on the disease liability.[10]

CASE-PARENTAL CONTROL DESIGN

Design based on the *TDT test*, which compares the relative frequencies of transmitted and non-transmitted alleles from parents to their affected offspring. It prevents the confounding effects of population stratification and permits testing of gene-environmental interactions by stratifying cases according to their environmental exposure status. For example, in a seminal paper, Spielman *et al* compared the genotypes at the insulin gene of juvenile diabetics and their parents and found that heterozygous parents transmitted to their affected children class 1 more often that other classes of alleles, and therefore concluded that susceptibility to juvenile diabetes is linked to the insulin gene.[11]

GENE MAPPING
Any strategy that permits finding the chromosomal location of one or more genes, often related to a disease. See *affected sib-pair approach*, *case-parental control design*, and *linkage analysis*.

HERITABILITY
Fraction of the total phenotypic variation in a population that is caused by genetic differences between individuals: *genetic variance/total variance*. The *genetic variance* is the part of the total variance that is caused by allelic variations at whatever loci influence the trait. The *total variance* is the amount of variation in *phenotype* in a defined population. It only applies to a population on which observations are made and cannot be extended to other populations that have different allele frequencies or environments. Therefore, it cannot be used to explain differences between populations. Lichtenstein *et al* applied this strategy to assess the effects of heritable and environmental factors in cancers at various sites on the basis of the twin registries from Finland, Sweden, and Denmark.[12]

LINKAGE ANALYSIS
Strategy for *gene mapping* by testing for linkage between *markers* and *phenotypes* using families. In classic linkage analysis the transmission model is fixed (possibly with parameter values obtained from segregation analysis) and the likelihoods (*LOD scores*) of the disease and marker data are compared under the null hypothesis of no linkage and the alternative hypothesis of linkage. Non-parametric linkage analysis avoids fixing an explicit mode of inheritance of the disease. Free application programs for human genetic linkage analysis are listed, classified, and available for downloading (http://linkage.rockefeller.edu/). For instance, using data from 39 families containing individuals affected with cystic fibrosis, Tsui *et al* found that the inheritance of alleles at the D0CRI-917 polymorphism seemed to be linked to cystic fibrosis.[13] Later on, and guided by this discovery, Kerem *et al* found that cystic fibrosis was caused by mutations in the CFTR gene, which is close to the D0CRI-917 polymorphism.[14]

LINKAGE DISEQUILIBRIUM
A condition in which alleles at two *loci* or *genes* are found together in a population at a greater frequency than that predicted simply by the product of their individual allele frequencies. Alleles at markers near disease causing genes tend to be in linkage disequilibrium in the affected individuals. This is particularly the case in isolated, homogeneous populations, in which it can be assumed that most affected individuals carry the same mutation. Thus, Hastbacka *et al* found that diastrophic dysplasia, a rare disease almost confined to Finland, mapped to the genome region *5q32-q33.1* by observing that, in patients, alleles at the polymorphims in that region were in close linkage disequilibrium with each other.[15]

LOD SCORE
A statistical estimate, obtained in *linkage analysis*, which indicates whether alleles at two loci are inherited together more often than expected and are thus likely to be placed near each other on a chromosome. A LOD score is the ratio of two probabilities: (1) the probability of the observed inheritance of a trait (usually a disease) and alleles at a marker in a pedigree if they were linked given a inheritance model for the trait and a recombination probability between marker and disease, and (2) the probability of the observed inheritance of a trait and marker in a pedigree under the assumption that they are not linked. A LOD score is the logarithm of the ratio of those two probabilities. LOD scores can be added across pedigrees, and are usually taken to indicate significant linkage if they are above three. The recombination fraction that gives the highest LOD score from a marker of known genomic location can be used to map a gene.

MICROARRAY
A novel method of studying large numbers of *genes* simultaneously by automating and miniaturising a hybridisation detection system. The method uses a robot to precisely apply tiny droplets containing *DNA* to glass slides. Researchers then attach fluorescent labels to DNA from the cell they are studying. The labelled probes are allowed to bind to complementary DNA strands on the slides. The slides are put into a scanning microscope that can measure the brightness of each fluorescent dot; brightness reveals how much of a specific DNA fragment is present.

MULTIFACTORIAL THRESHOLD (MFT) MODELS OF INHERITANCE
Models that assume the joint effect of multiple *genes* and environmental exposures in determining the liability of an individual to present the trait of interest. A threshold is assumed under which the subject would not present the trait and above it would.

POLYMERASE CHAIN REACTION (PCR)
A procedure for obtaining a large number of copies of a particular segment of *DNA*. The principle depends on the requirement by DNA polymerase of a primer with a 3′ end to which nucleotides can be added. Two such synthetic primers define a segment that is replicated in a thermal cycle of denaturation, reannealing (reformation of complementary primer-DNA structure), and replication. Each cycle, which takes two to three minutes, doubles the amount of DNA between the primer boundaries. Thirty cycles would yield 2^{30} copies. PCR has made it possible to characterise extremely small amounts of DNA.

RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP)
Genetic variation at the site where a *restriction enzyme* cuts a piece of *DNA*. Such variations affect the ability of the restriction enzyme to cut, and therefore, produce different fragment sizes. Most RFLPs are single base pair changes in the 4–6 bp target sequence of the restriction enzyme. Vice versa, many single nucleotide polymorphisms (SNPs) are RFLPs and can be detected with this technique.

SEGREGATION ANALYSIS
Analysis of the inheritance ratios of offspring from a particular parental cross to test for conformity with Mendelian theory. Either *genotypes* or *phenotypes* can be the object of segregation analysis.

SEQUENCING
Determining the exact order of the base pairs in a segment of DNA by biochemical methods. Semiautomated biochemical methods are available for sequencing, which are based in the sequential incorporation of fluorescently labelled nucleotides.

SINGLE STRAND CONFORMATION POLYMORPHISM (SSCP)
Fast and simple technique widely used for mutation detection in various diseases. Basically, a fragment of interest is amplified by *PCR*, followed by electrophoresis in non-denaturing gel. The mutant DNA is separated from the normal due to the difference in mobility in electrophoresis, which is believed to be caused by the conformational change of the

single stranded mutant DNA. Usually the DNA fragment size is restricted to less than 200 bp as the sensitivity of PCR-SSCP decreases with fragment size.

WHOLE GENOME SCAN

Linkage analysis in which *markers* placed at regular intervals and covering the whole *genome* are typed. It is often the first approach when no genetic information is available about a particular *phenotype*. For instance, Stefansson *et al* found that neuregulin-1 is a candidate gene for schizophrenia after typing 950 microsatellite markers covering the whole genome in 110 Icelandic patients for whom they had reconstructed their genealogical relationships.[16]

Contributors

Both authors have contributed equally to the manuscript.

Authors' affiliations

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

REFERENCES

1 **Calafell F**, Malats N. Basic molecular genetics for epidemiologists. *J Epidemiol Community Health* 2003;**57**:398–400 .
2 **Malats N**, Calafell F. Basic glossary on genetic epidemiology. *J Epidemiol Community Health* 2003;**57**:480–2.
3 **Various authors**. Genetic epidemiology. *Epidemiol Rev* 1997;**19**. (Monographic issue on genetic epidemiology).
4 **Hartwell LL**, Hood L, Goldberg ML, *et al*. *Genetics, from genes to genomes*. Boston: McGraw Hill, 2000.
5 **King RC**, Stansfield WD. *A dictionary of genetics*. Oxford: Oxford University Press, 1997.
6 **Strachan T**, Read AP. *Human molecular genetics*. Oxford: Bios Publishers, 2001.
7 **Lachmeijer AM**, Nosti-Escanilla MP, Bastiaans EB, *et al*. Linkage and association studies of IL1B and IL1RN gene polymorphisms in preeclampsia. *Hypertens Pregnancy* 2002;**21**:23–38.
8 **Pericak-Vance MA**, Bebout JL, Gaskell PC Jr, *et al*. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet* 1991;**48**:1034–50.
9 **Strittmatter WJ**, Weisgraber KH, Huang DY, *et al*. Binding of human apolipoprotein E to synthetic amyloid beta peptide: isoform-specific effects and implications for late-onset Alzheimer disease. *Proc Natl Acad Sci USA* 1993;**90**:8098–102.
10 **Bai Y**, Goldin LR, Goldstein AM. Gene x environment interaction from case-control and case-case approaches. *Genet Epidemiol* 2001;**21**:S825–30.
11 **Spielman RS**, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*. 1993;**52**:506–16.
12 **Lichtenstein P**, Holm NV, Verkasalo PK, *et al*. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;**343**:78–85.
13 **Tsui LC**, Buchwald M, Barker D, *et al*. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 1985;**230**:1054–7.
14 **Kerem B**, Rommens JM, Buchanan JA, *et al*. Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;**245**:1073–80.
15 **Hastbacka J**, de la Chapelle A, Kaitila I, *et al*. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 1992;**2**:204–11.
16 **Stefansson H**, Sigurdsson E, Steinthorsdottir V, *et al*. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 2002;**71**:877–92.

SPEAKER'S CORNER

The right to health of the European Union citizens. A strategy for a social European construction

The fundamentals of the European construction were only economic until the Unique Act was passed. Then some social aspects were partially incorporated.

Several advances in the field of common citizens' rights have, indeed, been introduced (free movement and residency, etc). Advances in matters such as education, health care, culture, and the fight against illicit drugs have been quite limited.

In the field of health policies, only a few measures of health protection to prevent diseases by means of research, improvement of information, and health education, etc, have been adopted to date. In summary, the social counterparts of the economic measures are not very concrete.

The absence of a common social policy may create serious problems and imbalances in public health, as a consequence of the influence of the health care expenditure on every country's economic competence ability. Differences in the services offered may attract patients toward the countries with best public services. The free circulation of persons may endanger the persistence of health services as ours, because we are receptors of retired people, whose health care consumption is fourfold that of the younger. Differences in the technological means available, and in professional training, may lead to an attraction of the best professionals by the most developed countries.

A common and homogeneous social policy should be implemented. Nevertheless, many of the reforms developed by the different countries were aimed to reduce the public expenditure, to introduce the market in to health care relations, and to increase the presence of the private sector in it. The results of such a strategy are being catastrophic for the rationality, the efficiency, and the equity of the health care systems, and also for their users' right to health.

Some proposals should, therefore, be advanced to get a Letter of the Rights to Health of the Europeans that ought to be incorporated to the project now debated. It should include the right to health protection for all. A common public health system is needed for that. It must contemplate universal health care provisions, and a homogenous offer in different countries. A public insurance, and a redistributive financing, ensuring a minimal common health care expenditure, must guarantee the equity in access to services. The present existing differences between health systems should not be forgotten. We should be conscious that we live in a progressively more interconnected world. If we wish a really consistent EU, the persons' rights ought to play an increasingly prominent part.

M Garcia, M Sanchez Bayle,
Secretario y Portavoz respectivamente de la Federación de Asociaciones para la Defensa de la Sanidad Publica, Madrid, Spain
Correspondence to: Dr M Sanchez Bayle; fadspu@teleline.es