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