

Principles in genetic risk assessment

Pedro Viana Baptista

Centro de Investigação em Genética
Molecular Humana, Secção Autónoma
de Biotecnologia, Faculdade de
Ciências e Tecnologia – Universidade
Nova de Lisboa, Caparica, Portugal

Abstract: Risk assessment constitutes an essential component of genetic counseling and testing, and the genetic risk should be estimated as accurately as possible for individual and family decision making. All relevant information retrieved from population studies and pedigree and genetic testing enhances the accuracy of the assessment of an individual's genetic risk. This review will focus on the following general aspects implicated in risk assessment: the increasing genetic information regarding disease; complex traits versus Mendelian disorders; and the influence of the environment and disease susceptibility. The influence of these factors on risk assessment will be discussed.

Keywords: genetics, risk assessment, genetic testing, genetic risk, genetic susceptibility

Introduction

The beginning of the 21st century has seen unprecedented advances in genome understanding. Since completing the sequence of the human genome there has been an explosion of available genetic information, providing unparalleled opportunities in understanding disease, drug design and toxicology, ecology, and risk assessment. Risk assessment is an essential part of genetic testing and counseling, and should be calculated as accurately as possible to enable both the clinician and the patient (or his/her family) to make decisions. An individual's genetic risk refers to the probability of the individual carrying a specific disease-associated mutation, or of being affected with a specific genetic disorder. The calculation of genetic risk should incorporate all available information at a particular point in time, such as the results of genetic testing (mutations, polymorphic markers); the presence of an independent risk factor derived from genetic test results; genetic test results on either or both parents, siblings, and close relatives (the probability of carrying a particular mutation or mutations often differs considerably among families and even among individuals within the same family); the ethnic background of each parent; an overall mutation rate for each ethnicity; and, if possible, the frequency of mutation in the population. Risk assessment should be looked at as an ongoing process of analysis of estimates.

Genetic risk data obtained from population studies are commonly used as a starting point in genetic risk assessment (prior probability of the individual having the disease). These estimates are mostly based on published data from a finite number of cases in specific and highly monitored populations (Ioannidis et al 2002). A good example of genetic risk variation between ethnic groups has been described for cystic fibrosis. Cystic fibrosis is caused by mutations in the cystic fibrosis transmembrane conductance regulator gene (*CFTR*) (OMIM nrs 602421; 219700(CF)). The disease-allele distribution of the *CFTR* gene varies greatly among different ethnic groups, leading to different inputs when calculating risk of disease. In *CFTR*, for example, if the individual in question is Caucasian European there are plenty of genomic data, making it easier to calculate the prior probability. Conversely, if little information is available for the population subgroup, several approximations might be done to obtain

Correspondence: Pedro Viana Baptista
CIGMH – SABT, FCT/UNL, 2829-516
Caparica, Portugal
Tel/Fax +351 21 294 8530
Email pmvb@fct.unl.pt

a prior probability (see Bobadilla et al 2002). When looking at the genetic risk for a determined population, attention should be brought to the possible misrepresentation of experimental data such as the allelic bias introduced by migrating populations, the study settings that established the genetic association (moment of diagnosis of a certain condition in a population and time of the genetic study), environmental factors, misclassification of outcome (conditions grouped under the same *outcome* but clearly distinct; for example see Gambaro et al 2000), and others. Nevertheless, researchers have calculated that only in extreme situations (few ethnic groups, great differences in disease and genotype frequencies) will any substantial misinterpretation occur (Wacholder et al 2000).

Another aspect reflects the development of the human-genome map and the advances in molecular technology, together with large-scale population-based studies, requiring close integration of genetic and epidemiologic research (Gambaro et al 2000). The post-genome era has been responsible for a flood of new data with studies reporting an association between genetic variation and disease-related outcomes (Halldorsson et al 2004; Hu et al 2004). With several million single-nucleotide polymorphisms (SNPs) characterized in the human genome, the number of possible genetic associations that can be tested is limited only by the rate at which laboratories can type these polymorphisms. The application of common genetic variations in association with studies to generate potential risk profiling using data from multiple vulnerability genes may play a key role in the early identification of high-risk individuals and groups of patients (The International SNP Map Working Group 2001). Recent studies have shown unprecedented sequence and haplotypic diversity in the genome (Durrant et al 2004; Smith et al 2004; Zhang et al 2004). If more than a polymorphism in a gene shows association with disease it must be established which variant(s) cause(s) predisposition. This is easy to demonstrate in animal models, where the whole genome can be experimentally manipulated to identify all variant(s). In humans, evaluating the association of different but closely related haplotypes may lead to the identification of the ancestral segment(s) that carry the predisposing/protective alleles (for example, see Funke et al 2004; Kokubo et al 2004). It should be noted that different studies on the same genetic association sometimes have discrepant results, each of them needing to be carefully monitored and tailored to the casual assessment (Ioannidis et al 2003).

For any given genotype-disease association, considerable effort is needed to assemble all available studies. As data accumulate, maintaining up-to-date reviews and information of each genotype-outcome association is becoming almost impossible. A more feasible approach may be to establish an automated system, such as a web-based database in which submitted data are classified by polymorphism, and outcome could be constructed. This database would need to be continuously updated, submission of both negative and positive findings in a standard format would be encouraged, and users would be able to assess a systematically up-to-date summary estimate of the effect associated with a certain allele or genotype. For this database to work, some sort of reward for researchers who contribute would need to be created. Also, some sort of critical analysis would need to be performed to revise the data entries.

Including the results from genetic testing can dramatically increase the accuracy of genetic risk assessment. Recent advances in genetic information and technology have led to an explosion in molecular testing for mutations and genome diversity. However, analytical or interpretative laboratory errors may affect risk assessment directly (false detection of mutation or polymorphism) or indirectly by altering estimates of allele distribution, carrier frequencies, and so on. The available technical platforms for use in genetic testing such as PCR, RFLPs, DHPLC, cDNA microarrays, DNA-chips, automated sequencing (which has become increasingly low-cost technology), RNA technology (RNA interference, antisense RNA), nanotechnology-based labeling, among many others, have contributed to a great expansion in the number of DNA tests capable of being used in the clinical environment (see also Amos and Grody 2004; Trent et al 2004). It should be emphasized that when several genes and/or environmental factors are involved in the genesis of a given disorder, the genetic testing for correct risk assessment is far from straightforward.

Genetic risk assessment in the clinical environment

Genetic counseling, developed to address the medical and social consequences of Mendelian disorders, has become an integral part of genetic testing. Genetic counseling as currently practiced is focused on the assessment of genetic risk, education of at-risk individuals and family members about the disease management and manifestation, education about reproductive options, and provision of psychological

and emotional support to cope with mostly untreatable diseases. Analysis based on Baye's theorem is routinely applied to calculate genetic risk in complex pedigrees and to estimate the probability of having or lacking a disease-causing mutation after a negative test result is achieved (Claus 2000). In brief, Baye's theorem states that the probability of a disease or condition after having performed a test is dependent on two things: the specificity and sensitivity of the test used (test characteristics), and the probability of that disease or condition before conducting the test (or prior probability). Whilst the first is rather straightforward as it depends solely on the technology (whether the mutation/alteration in question is susceptible to being detected and characterized with the technological platform in hand; possible bias introduced by a test being more sensitive in detecting one mutation than another), the latter is far more complex as many factors can influence the calculation and assessment of the prior-probability (see above).

Also, the genetic background of a specific disease greatly influences the assessment of risk for that particular disease, whether a single gene disorder or a more complex disease. A clarification of general issues involved in genetic risk assessment for both types of disease will be attempted. A substantial number of common human diseases are determined wholly or in part by genetic factors, usually characterized by the recognition of specific heritable conditions and identification of familial risk factors. Generally, genetic risk assessment has been largely focused on the evaluation of risk in Mendelian disorders, where a disease causing mutation in a single gene has high penetrance, producing an observable, often profound effect on phenotype. However, most common disorders are complex and multifactorial, and are presumed to result from the additive effect of mutations of low penetrance at multiple loci. This way, when considering Mendelian disorders, the assessment of risk is mainly relevant for the individual's own decision making through enhanced impartial information. In mainly complex disorders, however, the estimation of risk assesses the susceptibility of the individual to a certain disease and is essentially used towards risk modification through medical/behavioral intervention. In this case, the primary role of genetic risk assessment is to promote medical interventions (eg, through presymptomatic medication) or behavioral changes (eg, lifestyle and diet) in healthy individuals, which could contribute to a risk modification. Of utmost importance is the actual perception

of risk by the individual. Presumably, a more realistic perceived risk (after a correct risk assessment is performed) will motivate the initiation and maintenance of health-protective behaviors at a level that is appropriate for the individual's level of risk. The time of risk perception depends heavily on psychological and cognitive variables and influences adherence to putative therapeutics (Weitzel et al 2003).

Mendelian disorders

In Mendelian disorders, the affected individuals are perceived due to obvious symptoms or signs, or family history. In these single-gene disorders, risk assessment is primarily based on: (1) the results derived from genetic testing in direct diagnosis of symptomatic individuals (eg, Huntington's disease) or confirmation of diagnosis of a heritable disorder; (2) the assessment of the relatives' genetic risk, as relatives have an a priori higher risk that can be quantified; (3) prenatal diagnosis, where the risk for offspring can be easily calculated; and (4) the predictive evaluation, considering that an at-risk individual harboring the disease-causing mutation/allele, given enough time, will develop symptoms. The following example could help illustrate what has been stated: the carrier of a mutation has a 50% probability of passing the mutation to an offspring, and a 50% probability of having received the mutation from one parent. Based on this, the probability of being a carrier by the relatedness of the individual to an affected family member or obligate carrier can be estimated. For autosomal recessive diseases, provided that the carrier frequency in the general population is sufficiently small, the carrier's risk decreases by half at each "step" across a pedigree from the affected family member to a given individual. Some exceptions exist, as for X-linked recessive diseases, where the carrier's risk decreases by half at each "step" from one female member to another female member in the preceding or successive generation of the pedigree.

It should be noted that in recessive diseases there is a strong possibility that a disease-associated mutation is missed. In contrast, dominant disorders are usually caused by more specific mutations, which are usually easier to identify and follow across a pedigree. In recessive disorders, it is of paramount importance to consider the effect of consanguinity: alleles that are identical because they derive from a single allele present in a common ancestor are defined as being identical by descent. Alleles that are the same but that derive from two different and apparently unrelated

sources are defined as being identical by state. Consanguineous mating increases the frequency of autozygosity for two mutant alleles that are identical by descent.

Common complex disease susceptibilities

The genetic architecture of complex disease is not fully understood (Risch 2000). Most common complex disorders are assumed to result from the additive effect of alterations/mutations of low penetrance at multiple loci (Glazier et al 2002). Normally, when these mutations occur individually they may have no defective phenotypic effect. Then again, these mutations may act in cooperation with each other, usually in the presence of environmental stimuli, leading directly to disease or to an increase in disease risk. Accordingly, these mutations can be regarded as predisposing mutations, as they are necessary to cause disease but alone they are not sufficient. It is possible, that the risk for some common diseases is due to a very large number of loci, with each having a low frequency of disease-predisposing alleles (Pritchard 2001; Wang et al 2003). For example, a disease with 10% incidence in the population might reflect 100 independent monogenic diseases, each with high penetrance. This would, however, imply a higher relative risk to family members than actually observed for complex diseases. Actually, the perceived drop in relative risk for family members suggests (but does not prove) that the biggest fraction of the risk could be credited to a small number of loci with a higher frequency of alleles predisposing to disease, as seen for example for Factor V Leiden in deep venous thrombosis (Alhenc-Gelas et al 2001; Castoldi and Rosing 2004) and ApoE ϵ 4 allele in Alzheimer's disease (Wang et al 2000; Lambert et al 2002).

Non-Mendelian genetics also contribute to genetic risk through a number of different mechanisms, including de novo mutations and mosaicism. A de novo mutation is a mutation that is absent from the somatic cells of the parents, but present in the somatic cells of the offspring. A de novo mutation may cause a subset of germ cells to have the mutation; ie, isolated germline mosaicism. Germline mosaicism may or may not be noticeable as a de novo mutation in offspring. Non-Mendelian genetics also influence genetic risk assessment through a number of other mechanisms, including genomic instability (anticipation), non-paternity, imprinting (epigenetics), positional effects of genes (which can sometimes be viewed as a multifactorial

mechanism), and mitochondrial inheritance (for further reading see Bridge 1997; Young 1999).

One could be led to assume that there is a clear distinction between these two models: one, with the predominance of genetic influence—Mendelian; and the other, with the predominant influence of the environment. The distinction between the two models is far from clear-cut. In reality, they refer to the extremes of the wide spectrum of disease/disorder where genetics plays a significant role, with most disorders usually situated somewhere in the middle of the spectrum. It is increasingly recognized that the so called Mendelian disorders are complex diseases where the environment plays a central role in disease causation. This *one* disease can be decomposed in several simpler disorders for which a risk can be calculated. Variability in some heritable disorders is clearly mediated by the environment, eg, dietary phenylalanine in classic phenylketonuria (PKU). Even in the absence of identifiable environmental influences, genotype often does not completely predict phenotype, eg, Huntington's disease. The presence and/or number of pathogenic CAG repeats does not predict the age of onset or the severity of phenotype. The presence of a major disease-causing allele should be considered more "predisposing" than predictive. Thus far, most efforts that have successfully identified genes associated with common diseases have focused on population groups showing high inheritance. Genes inherited in a Mendelian manner have been identified as a cause of familial breast cancer, eg, Alzheimer's disease, Parkinson's disease, and diabetes. These genes are of great importance for the 5%–10% of cases of high heritability within families with common complex diseases where a Mendelian disorder is identified.

It becomes less clear to determine what is solely genetic and to distinguish inherited disease from acquired disease. To calculate the genetic risk in these complex disorders, one has to take into account all the inputs of risk; ie, ultimately the risk associated with each factor involved in the disease (number of genes causing or involved in the disease and their associated risk, number of environmental factors, etc). Whether all these factors have the same risk input or some are more important to the outcome than others needs to be clearly elucidated. The more factors involved the more complex the models involved to accurately calculate the probability of the individual developing the disease, and for most diseases the factors (genetic, environmental, and others) influencing the outcome are yet to be fully characterized.

In some cases, some simplifications of risk inputs can be made. For example, when the risks conveyed by a single gene are high, the testing for disease susceptibility may rely on the Mendelian model described earlier; eg, the presence of a *BRCA1* mutation can confer a lifetime risk for breast cancer of over 70% (Dunning et al 1999; Evans et al 2001). The major advances in the understanding of the molecular basis of cancer have made it possible to establish clear parameters for genetic risk assessment in cancer patients, especially for those individuals belonging to families in which cancer has a strong inheritance link (Petruccioli et al 2002; Comings et al 2003). In these situations, the likelihood that breast cancer risk is due to specific genetic susceptibility (such as *BRCA1* or *BRCA2* mutations) can be easily determined, and specific management strategies are now defined for *BRCA1* and *BRCA2* mutation carriers (Wenham et al 2003). Risk may be assessed as the likelihood of developing breast cancer or as the likelihood of detecting a *BRCA1* or *BRCA2* mutation. Clinical examples have told us that optimal counseling may require the use of both models, together with clinical judgment, to provide accurate and useful information.

Genetic testing for Mendelian disorders is a complex team approach with inputs from genetic counselors, clinicians, nurses, and genetic researchers. Nevertheless, genetic testing for common complex disease susceptibilities is almost completely absent from medical practice. In Mendelian disorders, the primary role of genetic testing is to give some information that might allow healthcare providers to assign relative risks to otherwise healthy looking individuals, in an attempt to reduce the actual risk of developing the disease (eg, health screening, pre-symptomatic medication aimed at reducing or eliminating one or more risk factors), or through behavioral changes (eg, smoking, diet, and lifestyle). This risk reduction, or more accurately described, risk modification, needs to be a continuous effort aimed at achieving small changes in one of the many small background risk factors involved. In these cases, non- or presymptomatic individuals will most likely be identified through public health or broad-spectrum screening measures, based on ethnicity or any factor identified in population-based epidemiological studies.

Concluding remarks

The technologies for genetic testing continue to change, and genetic risk assessment will no doubt become increasingly integrated into all aspects of medical care. Advances in gene

identification and characterization, polymorphism association studies, disease classification, and so on, continue to rapidly provide new and clinically relevant information that can contribute to a better detection, evaluation, prevention, and follow-up strategies in human disease. Additionally, prospective clinical trials need to be conducted to define the optimal use of existing management strategies, develop risk assessment instruments that incorporate additional risk-factor information, and evaluate populations for whom validated risk assessment approaches do not yet exist.

It is becoming increasingly frequent that “healthy” people seek predispositional genetic testing, and clinicians should incorporate genetic risk assessment and management into their routine screening and health maintenance appointments. Attention should be brought to the current limitations, uncertainties, legal, and psychological implications of genetic risk assessment, eg, the right to not know. As mentioned earlier, incorporating genetic data from a close relative may be extremely useful for assessing the genetic risk of an individual. But the relative may not wish to know whether he/she is at risk and will not be tested for the disease.

The powerful new methods and expansive scope combine to project us into ethically unfamiliar territory. It has been observed that discrimination or fear of discrimination for medical, life, and disability insurance, as well as employment discrimination, could undermine the acceptance of genetic testing by individuals who otherwise would choose to receive such information. These features lend special urgency to questions of genetic ownership and privacy, disease and normalcy, identity and genetic determinism, and early diagnosis and therapy.

Abbreviations

ApoE4, apolipoprotein ε4 allele; *BRCA1*, breast cancer 1 gene; *BRCA2*, breast cancer 2 gene; cDNA, complementary DNA; CF, cystic fibrosis; CFTR, cystic fibrosis transmembrane conductance regulator; DHPLC, denaturing high-performance liquid chromatography; DNA, deoxyribonucleic acid; OMIM, Online Mendelian Inheritance in Man (<http://www.ncbi.nlm.nih.gov>); PCR, polymerase chain reaction; RFLPs, restriction fragment length polymorphisms; RNA, ribonucleic acid; SNPs, single nucleotide polymorphisms.

References

- Alhenc-Gelas M, Aiach M, de Moerloose P. 2001. Venous thromboembolic disease: risk factors and laboratory investigation. *Semin Vasc Med*, 1:81–8.
- Amos J, Grody W. 2004. Development and integration of molecular genetic tests into clinical practice: the US experience. *Expert Rev Mol Diagn*, 4:465–77.

- Bobadilla JL, Macek Jr M, Fine JP, et al. 2002. Cystic fibrosis: a worldwide analysis of CFTR mutations: correlation with incidence data and application to screening. *Hum Mutat*, 19:575–606.
- Bridge PJ. 1997. The calculation of genetic risks: worked examples in DNA diagnostics. Baltimore: John Hopkins Univ Pr.
- Castoldi E, Rosing J. 2004. Factor V Leiden: a disorder of factor V anticoagulant function. *Curr Opin Hematol*, 11:176–81.
- Claus EB. 2000. Risk models in genetic epidemiology. *Stat Methods Med Res*, 9:589–601.
- Comings DE, Gade-Andavolu R, Cone LA, et al. 2003. A multigene test for the risk of sporadic breast carcinoma. *Cancer*, 97:2160–70.
- Dunning AM, Healey CS, Pharoah PD, et al. 1999. A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev*, 8:843–54.
- Durrant C, Zondervan KT, Cardon LR, et al. 2004. Linkage disequilibrium mapping via cladistic analysis of single-nucleotide polymorphism haplotypes. *Am J Hum Genet*, 75:35–43.
- Evans JP, Skrzynia C, Burke W. 2001. The complexities of predictive genetic testing. *BMJ*, 322:1052–6.
- Funke B, Finn CT, Plocik AM, et al. 2004. Association of the DTNBP1 locus with schizophrenia in a US population. *Am J Hum Genet*, 75: 891–8.
- Gambaro G, Anglani F, D'Angelo A. 2000. Association studies of genetic polymorphisms and complex disease. *Lancet*, 355:308–11.
- Glazier AM, Nadeau JH, Aitman TJ. 2002. Finding genes that underlie complex traits. *Science*, 298:2345–9.
- Halldorsson BV, Bafna V, Lippert R, et al. 2004. Optimal haplotype block-free selection of tagging SNPs for genome-wide association studies. *Genome Res*, 14:1633–40.
- Hu X, Schrodi SJ, Ross DA, et al. 2004. Selecting tagging SNPs for association studies using power calculations from genotype data. *Hum Hered*, 57:156–70.
- Ioannidis JP, Trikalinos TA, Ntzani EE, et al. 2003. Genetic associations in large versus small studies: an empirical assessment. *Lancet*, 361:567–71.
- Ioannidis JPA, Rosenberg PS, Goedert JJ, et al. 2002. Meta-analysis of individual participant's data in genetic epidemiology. *Am J Epidemiol*, 156:204–10.
- Kokubo Y, Kamide K, Inamoto N, et al. 2004. Identification of 108 SNPs in TSC, WNK1, and WNK4 and their association with hypertension in a Japanese general population. *J Hum Genet*, 49:507–15.
- Lambert JC, Araria-Goumidi L, Myllykangas L, et al. 2002. Contribution of APOE promoter polymorphisms to Alzheimer's disease risk. *Neurology*, 59:59–66.
- Petrucelli N, Lazebnik N, Huelsman KM, et al. 2002. Clinical interpretation and recommendations for patients with a variant of uncertain significance in BRCA1 or BRCA2: a survey of genetic counseling practice. *Genet Test*, 6:107–13.
- Pritchard JK. 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*, 69:124–37.
- Risch NJ. 2000. Searching for genetic determinants in the new millennium. *Nature*, 405:847–56.
- Smith MW, Patterson N, Lautenberger JA, et al. 2004. A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet*, 74:1001–13.
- The International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409:928–33.
- Trent RJ, Yu B, Caramins M. 2004. Challenges for clinical genetic DNA testing. *Expert Rev Mol Diagn*, 4:201–8.
- Wacholder S, Rothman N, Caporaso N. 2000. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst*, 92:1151–8.
- Wang JC, Kwon JM, Shah P, et al. 2000. Effect of APOE genotype and promoter polymorphism on risk of Alzheimer's disease. *Neurology*, 55:1644–9.
- Wang WY, Cordell HJ, Todd JA. 2003. Association mapping of complex diseases in linked regions: estimation of genetic effects and feasibility of testing rare variants. *Genet Epidemiol*, 24:36–43.
- Weitzel JN, McCaffrey SM, Nedelcu R, et al. 2003. Effect of genetic cancer risk assessment on surgical decisions at breast cancer diagnosis. *Arch Surg*, 138:1323–8.
- Wenham RM, Schildkraut JM, McLean K, et al. 2003. Polymorphisms in BRCA1 and BRCA2 and risk of epithelial ovarian cancer. *Clin Cancer Res*, 9:4396–403.
- Young ID. 1999. Introduction to risk calculation in genetic counseling. Oxford: Oxford Univ Pr.
- Zhang W, Collins A, Morton NE. 2004. Does haplotype diversity predict power for association mapping of disease susceptibility? *Hum Genet*, 115:157–64.