How to identify the genetic basis of gastrointestinal and liver diseases?

P Ferenci

New insights into the genetic basis of disease are being generated at an ever increasing rate. This explosion of information was ignited by technological advances, such as the polymerase chain reaction and automated DNA sequencing. Although its promise is great, the integration of genetics into the everyday practice of medicine remains challenging. This review discusses the application of molecular genetics in general with a specific focus on hereditary diseases of the digestive organs. The application of molecular genetics in everyday clinical routine is hampered by the difficult interpretation of test results. These difficulties include the prediction of disease penetrance, the presence of multiple mutations of a particular gene with varying functional consequences, and the importance of exogenous factors modulating disease expression. To date, the most significant impact of genetics has been to increase our understanding of disease aetiology and pathogenesis and to reliably identify siblings of affected patients with the risk to develop symptomatic disease.

The era of genetics began with the observations of Gregor Mendel that changes in the colour of flowers and shape of the seeds followed a clear pattern over the years. His fundamental rules of inheritance thus were based on easily recognisable signs. His work preceded the discovery of DNA as carrier of the genetic information. An observed trait is referred to as a phenotype; the genetic information defining the phenotype is called the genotype. With more advanced understanding of the function of DNA phenotypic genetics were replaced by molecular genetics. In contrast with phenotypic genetics, which assumes that gene products are either fully functional or devoid of function as consequence of a mutation, molecular genetics describe variations in the base sequence of gene. Such changes are not always associated with impaired functions of the gene product. Even gene products of mutated genes may still have some residual function. Thus, the presence of a change in the base sequence does not necessarily imply the presence of phenotypic disease. These fundamental differences to phenotype based genetics limit the presence of phenotypic disease. Base variations of the “wild type” gene in healthy subjects are named “DNA polymorphisms”. These alternative forms of a gene or a genetic marker are referred to as alleles. Alleles have no apparent effect on gene expression or function. In other instances, these variants may have subtle effects on gene expression, thereby conferring the adaptive advantages associated with genetic diversity. On the other hand, allelic variants may reflect mutations in a gene that clearly change its function.

What constitutes a normal gene?

A normal gene is defined by the base sequence that is observed in most healthy subjects in a given population and is called the “wild type”. Definition of “healthy” requires the presence of a functionally normal gene product and the absence of phenotypic disease. Base variations of the “wild type” gene in healthy subjects are named “DNA polymorphisms”. These alternative forms of a gene or a genetic marker are referred to as alleles. Alleles have no apparent effect on gene expression or function. In other instances, these variants may have subtle effects on gene expression, thereby conferring the adaptive advantages associated with genetic diversity. On the other hand, allelic variants may reflect mutations in a gene that clearly change its function.

What is a mutation?

A mutation is a base sequence that differs from the “wild type” in a patient presenting with a phenotypic disorder but is never observed in healthy subjects. Thus the definition whether this variation in the base sequence is a disease causing mutation requires testing of healthy subjects. Several disease causing mutations may be present within the same gene. The functional consequences of a mutation are manifold. Mutations may result in the complete absence of gene products (“null” mutations) or in proteins devoid of any function. Such mutations are associated with severe diseases occurring at birth or early childhood. They are mostly attributable to large deletions or insertion in the DNA or to mutations that result in the occurrence of stop codons (“nonsense” mutations) or of frame shifts attributable to deletion or insertion of one or two or a small number of nucleotides. Some mutations affect messenger RNA splicing mechanisms.

Functional consequences of a mutation

Functionally, mutations can be broadly classified as gain of function and loss of function mutations. Gain of function mutations are typically dominant; that is, they result in phenotypic alterations when a single allele is affected. Inactivating mutations are usually recessive, and an affected person is homozygous or compound heterozygous.

DEFINITIONS

To understand the implications of molecular genetics, several basic definitions are needed:

1 What constitutes a normal gene?

A normal gene is defined by the base sequence that is observed in most healthy subjects in a given population and is called the “wild type”. Definition of “healthy” requires the presence of a functionally normal gene product and the absence of phenotypic disease. Base variations of the “wild type” gene in healthy subjects are named “DNA polymorphisms”. These alternative forms of a gene or a genetic marker are referred to as alleles. Alleles have no apparent effect on gene expression or function. In other instances, these variants may have subtle effects on gene expression, thereby conferring the adaptive advantages associated with genetic diversity. On the other hand, allelic variants may reflect mutations in a gene that clearly change its function.

2 What is a mutation?

A mutation is a base sequence that differs from the “wild type” in a patient presenting with a phenotypic disorder but is never observed in healthy subjects. Thus the definition whether this variation in the base sequence is a disease causing mutation requires testing of healthy subjects. Several disease causing mutations may be present within the same gene. The functional consequences of a mutation are manifold. Mutations may result in the complete absence of gene products (“null” mutations) or in proteins devoid of any function. Such mutations are associated with severe diseases occurring at birth or early childhood. They are mostly attributable to large deletions or insertion in the DNA or to mutations that result in the occurrence of stop codons (“nonsense” mutations) or of frame shifts attributable to deletion or insertion of one or two or a small number of nucleotides. Some mutations affect messenger RNA splicing mechanisms.

3 Functional consequences of a mutation

Functionally, mutations can be broadly classified as gain of function and loss of function mutations. Gain of function mutations are typically dominant; that is, they result in phenotypic alterations when a single allele is affected. Inactivating mutations are usually recessive, and an affected person is homozygous or compound heterozygous.

Abbreviations: CF, cystic fibrosis; RFLP, restriction fragment length polymorphism; PCR, polymerase chain reaction
mutations, like well based on the genotype. Homozygous carriers of severe particular CFTR mutation. With respect to the sweat gland, the predict the severity of disease in specified organs from a par-
misense mutations are the most informative class of than 70% of cases of clinically overt CF. Other mutations are
function is mutation specific with five basic classes of
insufficiency. CFTR mutations may be classified in another
absence of a functionally intact CFTR protein (class I). Class II
processing steps.
the entire genetic locus for mutations, as these can differ in
each patient.
The difficulties of understanding the role of mutation can be best described in cystic fibrosis (CF). Today more than 850 mutations of the CFTR gene were reported. Some mutations like the ΔF508 mutation are common and account for more than 70% of cases of clinically overt CF. Other mutations are rare and occur sometimes in single families. By far, the missense mutations are the most informative class of mutation in the CFTR gene and account for 40% of the CF mutations. These mutations result in important alterations in the structure and function of their encoded protein. Certain clinical predictions that can be made from the analysis of the mutations in theCFTR gene that the patient may be carrying. This “genotype-phenotype” analysis explores the feasibility to predict the severity of disease in specified organs from a particular CFTR mutation. With respect to the sweat gland, the sweat chloride concentrations can be predicted reasonably well based on the genotype. Homozygous carriers of severe mutations, like ΔF508, will routinely have severe pancreatic insufficiency. CFTR mutations may be classified in another way and that is by their molecular consequence. Channel function is mutation specific with five basic classes of mutations recognised. Some mutations result in a complete absence of a functionally intact CFTR protein (class I). Class II refers to those mutant proteins that have blocked in the processing steps. ΔF508 is an example of a protein that is made but that cannot mature properly; and at the end, there is no functional molecule on the apical membrane. Class III refers to mutant proteins that are blocked in regulation; the protein can get to the apical membrane but cannot be opened by cAMP. CFTR-class IV gene mutations result in proteins that can get to the apical membrane, but when they open, their conductance has changed and the amount of chloride ion that can get through the apical membrane has changed. Class V, is a combination of different types of mutations that mainly reduce the total amount of functional CFTR protein on the apical membrane because of reduced synthesis; either at the messenger RNA level, or at the protein maturation level. Over all, Class IV and V have a milder consequence than the class I-III and do not cause pancreatic insufficiency.

Genetic deficiency of α1 antitrypsin provides a prototype for the diseases associated with conformational instability. The most common mutation is the S mutation. In homozygotes plasma α1 antitrypsin concentrations are decreased by 40%. This by itself poses a negligible threat to health, but the S variant becomes important if it is co-inherited with the more severe Z mutation, which is present in 4% of northern Europeans. In homozygotes plasma α1 antitrypsin concentrations are decreased by 85%. Consequently, the plasma concentrations of α1 antitrypsin in both ZZ homozygotes and SZ compound heterozygotes are insufficient to ensure lifetime protection of the lungs from proteolytic damage, especially in smokers. The low plasma α1 antitrypsin concentrations result not from a lack of synthesis but from a blockage of its processing and secretion. The retained α1 antitrypsin aggregates in the endoplasmic reticulum of hepatocytes as inclusions that are readily recognisable on periodic acid Schiff staining. Z mutant of α1 antitrypsin forms long polymers in the endoplasmic reticulum of hepatocytes, which are resistant to the usual degradative processes.

TOOLS OF MOLECULAR GENETIC ANALYSIS
Molecular genetics require the visualisation of sequence differences directly in DNA. DNA polymorphisms in coding regions (exons) or non-coding regions of genes (for review see Housman) are inherited according to the Mendelian rules. The value of highly variable DNA sequences as genetic markers rests on straightforward principles. Every person carries two copies of each chromosome except the sex chromosomes. If a DNA polymorphism is to be useful in analysing the transmission of the two chromosomes in a family, then the DNA copies at the polymorphic site of the person under study must be different in the two chromosomes. The likelihood that a given person will have different DNA sequences at the polymorphic site directly determines the usefulness of that site in genetic studies. Chromosomal sites at which the DNA sequences can have many alternative forms are thus ideal sites for genetic markers. At these sites, a person is most likely to carry two alternative DNA sequences, accurately marking the two alternative chromosomes. In the human genome, the sites
that have the properties most favourable to such extensive
variation include a repetition of the same short DNA sequence
a variable number of times. Such sequences are called tandem
repeat sequences (microsatellites). A DNA sequence with such
variation may be as short as two base pairs or as long as sev-
eral hundred base pairs. Highly variable sequences of this type
are well distributed throughout the length of every human
chromosome. When tandemly repeated sequences are repli-
cated during cell division, the number of repeats can change.

METHODS TO DETECT DNA POLYMORPHISMS
Restriction fragment length polymorphism (RFLP) analysis

DNA polymorphisms can be detected by variations in the size
of DNA fragments obtained after digestion with restriction
enzymes. Restriction enzymes cut DNA strands at highly
specific sites (restriction site). A variation in the nucleotide
sequence may result in the loss or the creation of a new
restriction site or in the length of the DNA fragment between
existing restriction sites. Thus, the length (and eventually also
the number) of the restriction fragment(s) will be different by
Southern blot analysis. The RFLP pattern is specific for every
individual tested. Other methods to study DNA polymor-
phisms include the detection of the altered mobility of the
PCR product of DNA segment (single strand conformational
polymorphism or denaturing gradient gel electrophoresis)
and WAVE-DNA fragment analysis, which is based on
temperature modulated liquid chromatography and a high
resolution matrix. If the gene is unknown, polymorphic
markers flanking the unknown gene can be used to construct
haplotypes for DNA linkage analysis. A haplotype refers to a
group of alleles that are closely linked together at a genomic
locus. Haplotype analysis is useful for tracking the transmission
of genomic segments within families and for detecting evidence
of genetic recombination. By using various restriction
enzymes and DNA probes, multiple RFLPs for a given gene
can be obtained. By this approach, both the paternal and the
maternal gene can be “reconstructed”. If both genes have dif-
fere...
is, more than 800 in CF). Mutations may reflect a common ancestor and are enriched in certain populations but may be absent in others.

2 Haploinsufficiency
Mutation in a single allele can result in a situation in which one normal allele is not sufficient for a normal phenotype. This phenomenon applies, for example, to expression of rate limiting enzymes in heme synthesis that cause porphyrias. A single allele mutation in a single allele can also result in loss of function because of a dominant negative effect.

3 Loss of heterozygosity
Subjects with a normal and an abnormal gene without any apparent disease may undergo somatic mutations of the normal gene later in life. Such an event may result in overt dysfunction of the gene product in the affected cells. This loss of heterozygosity is assumed to be one important event in cancerogenesis. 

Subjects not carrying the mutation
A negative finding does not exclude phenotypic disease, as other mutations of the gene may be present. Furthermore, gene defects may be attributable to mutation of other genes (that is, mutation of a promoter of mismatch repair genes results in hypermethylation of their genes products with impaired functional capacity).

TARGET POPULATIONS FOR MOLECULAR GENETIC TESTING
1 Patients with symptomatic phenotypic disease
In patients with hereditary diseases (diagnosed by phenotypic criteria; for example, polyposis coli), DNA analysis strengthens the final diagnosis. In diseases with only few mutations (like in HFE associated haemochromatosis) mutation analysis can replace invasive diagnostic tests. Because of the large number of mutations in most diseases DNA analysis cannot be used as a diagnostic test.

2 Family screening
Mutation analysis is the state of the art approach for screening the family of index patients and can replace other diagnostic tests to identify subjects at risk to develop the disease. A negative test result in a family member of a patient with a disease related mutation indicates a low risk of the disease. This can decrease anxiety and, for some diseases, reduce the frequency of monitoring for early signs of the disease.

3 Population screening
Mutation analysis to detect presymptomatic disease in the general population has not been tested so far. Beyond the discussed difficulties of interpretation of test results several factors limit the use of genetic tests for population screening. Firstly, screening is only appropriate, if a validated treatment to prevent occurrence of phenotypic disease is available for asymptomatic subjects. Secondly, other screening strategies may be more cost effective or straightforward than mutation analysis. For colorectal cancer, DNA based mutation analysis cannot replace endoscopy, because a colonoscopy is needed whether a mutation is present or absent. Furthermore, formation of cancer can be prevented by endoscopic polypectomy. Thus, endoscopy in combination with testing for occult blood in stool will remain the standard for the foreseeable future.

REFERENCES