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

P Ferenci

Gut 2003;52(Suppl II):ii6–ii9

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 role of molecular genetics in clinical medicine. This review discusses the application of molecular genetics in general with a specific focus on hereditary diseases of the digestive organs (table 1).

Correspondence to:
Professor P Ferenci,
Department of Internal
Medicine IV,
Gastroenterology and
Hepatology, University of
Vienna Allgemeines
Krankenhaus, Waehringer
Guertel 18–20, A-1090
Vienna, Austria;
peter.ferenci@akh-wien.ac.at

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

Abbreviations: CF, cystic fibrosis; RFLP, restriction fragment length polymorphism; PCR, polymerase chain reaction

Table 1 Selected genetic diseases in gastroenterology and hepatology

Diseases	Gene	References
Cholestatic liver diseases		21
Byler's disease, Summerskill syndrome	FIC1	
Progressive familial intrahepatic cholestasis-2	ABCB11	
Progressive familial intrahepatic cholestasis-3	ABCB4	
Dubin-Johnson syndrome	ABCB2	
Hepatic storage diseases		
Wilson disease	ATP7B	22
Haemochromatosis	HFE	23
Colon cancer		24
Familial polyposis coli	APC	25
Hereditary non-polyposis colon cancer	MSH2, MLH1	26, 27
Peutz-Jegher syndrome	STK11	28
Idiopathic pancreatitis		
Cystic fibrosis	CFTR	2
Hereditary pancreatitis	Trypsinogen, SPINK	29, 30
Crohn's disease	NOD2	31, 32

heterozygous (that is, carrying two different mutant alleles) for the disease causing mutations. Other mutations result in less pronounced functional consequences. A single amino acid change may result in an altered, but still functional protein. The mutation may affect the tertiary structure of the gene product, its assembly, inactivation, secretion, or conformational stability.

Allelic heterogeneity refers to the fact that different mutations in the same genetic locus can cause an identical or similar phenotype. Inactivating mutations in genes usually show a near random distribution. Exceptions include a "founder effect", in which a particular mutation that does not affect reproductive capacity can be traced to a single person; "hot spots" for mutations, in which the nature of the DNA sequence predisposes to a recurring mutation; and localisation of mutations to certain domains that are particularly critical for protein function. Allelic heterogeneity creates a practical problem for genetic testing because you must often examine 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).¹ Today more than 850 mutations of the CFTR gene were reported. Some mutations like the $\Delta F508$ 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 the CFTR 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 $\Delta F508$, 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.² 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. $\Delta F508$ 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 α_1 antitrypsin provides a prototype for the diseases associated with conformational instability.³ The most common mutation is the S mutation. In homozygotes plasma α_1 antitrypsin concentrations are decreased by 40%. This by itself poses a negligible threat to health, but the S variant becomes important if it is coinherited with the more severe Z mutation, which is present in 4% of northern Europeans. In homozygotes plasma α_1 antitrypsin concentrations are decreased by 85%. Consequently, the plasma concentrations of α_1 antitrypsin in both ZZ homozygotes and SZ compound heterozygotes are insufficient to ensure lifetime protection of the lungs from proteolytic damage, especially in smokers.^{4,5} The low plasma α_1 antitrypsin concentrations result not from a lack of synthesis but from a blockage of its processing and secretion.⁶ The retained α_1 antitrypsin aggregates in the endoplasmic reticulum of hepatocytes as inclusions that are readily recognisable on periodic acid Schiff staining. Z mutant of α_1 antitrypsin forms long polymers in the endoplasmic reticulum of hepatocytes,⁷ which are resistant to the usual degradative processes.⁸

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 Housman⁹) 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 several hundred base pairs. Highly variable sequences of this type are well distributed throughout the length of every human chromosome. When tandemly repeated sequences are replicated 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.^{10, 11} 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 polymorphisms 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. Haplotypes are 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 different allele patterns that are also different within members of the family, the pattern of its inheritance can be traced within the family.

By haplotype analysis inheritance of a disease can be studied even if the gene/and or the mutation are unknown.¹² Precondition is the availability of an index patient (in whom the disease was diagnosed by standard phenotypic criteria) and both of his/her parents for testing. Limitations are the lack of informative allelic markers within the family and the presence of crossovers within the region of interest. Haplotype analysis is time consuming and can only be applied in selected families. Haplotype analysis is also useful to investigate the origin and geographical distribution of a particular mutation.¹³ The importance of RFLP analysis is the localisation of an unknown gene to a distinct part of a chromosome. This information is needed for the identification of a disease gene by a variety of methods.

Direct mutation analysis

1 Direct sequencing

New technologies permit automated sequence analysis of large portions of a gene to detect point mutations, deletions, inversions, and other changes in the nucleotide sequence. However, direct sequencing of the whole to date does not play a part in clinical medicine. A more practical approach is first to screen the gene of interest for possible mutation by haplotype analysis or by single strand conformation polymorphism analysis. Those samples showing a shift of one or both bands or unusual haplotypes can then be sequenced to identify the exact mutation. This approach is quite useful as research tool, but impractical for clinical diagnosis.

2 Polymerase chain reaction (PCR) based detection of known mutations

A variety of approaches are commonly used to detect mutations. The simplest takes advantage of the base sequence specificity of restriction endonucleases. These enzymes recognise precise sequences of four to eight bases and cut double stranded DNA only at these sites. A mutation at such a site will prevent the enzyme from cutting there; conversely, a mutation may result in the creation of a new enzyme recognition site and lead to cutting where it normally should not occur. To detect the mutation, DNA surrounding the site of potential mutation is amplified by the PCR, the product is incubated with the restriction enzyme, and then the DNA is analysed by electrophoresis. If the enzyme cuts, two fragments will result; otherwise there will be a single fragment. The presence or absence of the mutation can be inferred, depending on whether the mutation creates or destroys an enzyme recognition site.

The other scheme for detecting mutations, is based on the specificity of the PCR reaction itself. A PCR primer is designed that ends right at the site of a potential mutation. If the primer is homologous to the wild type sequence, it will amplify only the wild type sequence in conjunction with another primer some distance away in the gene. The wild type primer will not amplify mutant DNA, however. Conversely, a primer that is homologous to the mutant sequence will amplify only mutant DNA. If the PCR is carried out with both sets of primers in separate reactions, the presence or absence of mutant and wild type sequences can easily be determined. Here multiple sequences can be assayed simultaneously in a single sample, as long as the sizes of the PCR products from each segment differ.

The direct determination of mutations is independent of family analysis. There is no need to test an index patients with the disease. PCR based mutation assays can be automated and thus permit mass screening. Multiplex PCR strips can detect the most common mutations of a particular disease simultaneously. Several mutation assays are commercially available now (for example, for CF, haemochromatosis, familial adenomatous polyposis).

INTERPRETATION OF TEST RESULTS

Molecular genetic analysis can yield three possible findings: the tested subject is either a homozygous or a heterozygous carrier of the mutation, or does not carry the mutation at all.

Homozygous mutation carrier

In contrast with the phenotypic diagnosis of a disease, genotypic diagnosis in a healthy subject raises the question, whether tested subject will ever develop the disease. In most hereditary diseases there is no complete penetrance of the disease. In genetic haemochromatosis for example, a large proportion of C282Y (the typical mutation) homozygotes have no evidence of iron overload.^{14, 15}

Heterozygous mutation carrier

In this setting the most important question is, whether the tested subject is (and will remain) free of a disease or not. According Mendelian rules, subjects carrying a "wild type" and a disease causing gene with autosomal recessive inheritance are healthy. This statement is only valid if the other gene not carrying the mutation is also functionally intact.

1 Compound heterozygotes

The gene not having the mutation may have a different (disease causing) one, which is not detected by the assay. Such compound heterozygotes may suffer from the disease (diagnosed by phenotypic criteria). Unfortunately this is not an exception but a general rule. In most inherited disease multiple different mutations of the affected gene are present (that

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. 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.¹⁶

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 promoters 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 effective¹⁷ or straightforward than mutation analysis. For colorectal screening, DNA based mutation analysis¹⁸ 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 stool¹⁹ will remain the standard for the foreseeable future.²⁰

REFERENCES

- 1 **Noone PG**, Knowles MR. 'CFTR-opathies': disease phenotypes associated with cystic fibrosis transmembrane regulator gene mutations. *Respir Res* 2001;**2**:328–32.
- 2 **Wilschanski M**, Zielinski J, Markiewicz D, *et al*. Correlation of sweat chloride concentration with classes of the cystic fibrosis transmembrane regulator gene mutations. *J Pediatr* 1995;**127**:705–10.
- 3 **Carrell RW**, Lomas DA. Alpha-1-antitrypsin deficiency—a model for conformational diseases. *N Engl J Med* 2002;**346**:45–53.
- 4 **Brantly M**, Nukiwa T, Crystal RG. Molecular basis of alpha-1-antitrypsin deficiency. *Am J Med* 1988;**84** (suppl 6A):13–31.
- 5 **Turino GM**, Barker AF, Brantly ML, *et al*. Clinical features of individuals with PI*SZ phenotype of α_1 -antitrypsin deficiency. *Am J Respir Crit Care Med* 1996;**154**:1718–25.
- 6 **Foreman RC**, Judah JD, Colman A. Xenopus oocytes can synthesise but do not secrete the Z variant of human α_1 -antitrypsin. *FEBS Lett* 1984;**168**:84–8.
- 7 **Elliott PR**, Lomas DA, Carrell RW, *et al*. Inhibitory conformation of the reactive loop of α_1 -antitrypsin. *Nat Struct Biol* 1996;**3**:676–81.
- 8 **Cabral CM**, Choudhury P, Liu Y, *et al*. Processing by endoplasmic reticulum mannosidases partitions a secretion-impaired glycoprotein into distinct disposal pathways. *J Biol Chem* 2000;**275**:25015–22.
- 9 **Housman D**. Human DNA-polymorphism. *N Engl J Med* 1995;**332**:318–19.
- 10 **Korf B**. Molecular diagnosis (First of 2 parts). *N Engl J Med* 1995;**332**:1218–21.
- 11 **Korf B**. Molecular diagnosis (Second of 2 parts). *N Engl J Med* 1995;**332**:1499–502.
- 12 **Maier-Dobersberger Th**, Rack S, Granditsch G, *et al*. Diagnosis of Wilson's disease in an asymptomatic sibling by DNA linkage analysis. *Gastroenterology* 1995;**109**:2015–18.
- 13 **Firneisz G**, Lakatos PL, Szalay F, *et al*. Common mutations of ATP7B in Wilson disease patients from Hungary. *Am J Hum Genet* 2002;**108**:23–8.
- 14 **Crawford DH**, Jazwinska EC, Cullen LM, *et al*. Expression of HLA-linked hemochromatosis in subjects homozygous or heterozygous for the C282Y mutation. *Gastroenterology* 1998;**114**:1003–8.
- 15 **Olynyk JK**, Cullen DJ, Aquilia S, *et al*. A population-based study of the clinical expression of the hemochromatosis gene. *N Engl J Med* 1999;**341**:718–24.
- 16 **Jen J**, Kim H, Piantadosi S, *et al*. Allelic loss of chromosome 18q and prognosis in colorectal cancer. *N Engl J Med* 1994;**331**:213–21.
- 17 **Hickman PE**, Hourigan LF, Powell LW, *et al*. Automated measurement of unsaturated iron binding capacity is an effective screening strategy for C282Y homozygous haemochromatosis. *Gut* 2000;**46**:405–9.
- 18 **Traverso G**, Shuber A, Levin B, *et al*. Detection of APC mutations in fecal DNA from patients with colorectal tumors. *N Engl J Med* 2002;**346**:311–20.
- 19 **Rex DK**, Johnson DA, Lieberman DA, *et al*. Colorectal cancer prevention 2000: screening recommendations of the American College of Gastroenterology. *Am J Gastroenterol* 2000;**95**:868–877.
- 20 **Ransohoff DF**, Sandler RS. Screening for colorectal cancer. *N Engl J Med* 2002;**346**:40–4.
- 21 **Ferenci P**, Zollner F, Trauner M. Hepatic transport systems. *J Gastroenterol Hepatol* 2002;**17**:S106–13.
- 22 **Ferenci P**. Wilson's disease. *Ital J Gastroenterol Hepatol* 1999;**31**:416–25.
- 23 **Powell LW**, Yapp TR. Hemochromatosis. *Clin Liver Dis* 2000;**4**:211–28.
- 24 **Jass JR**. Familial colorectal cancer: pathology and molecular characteristics. *Lancet Oncol* 2000;**1**:220–6.
- 25 **Fearnhead NS**, Britton MP, Bodmer WF. The ABC of APC. *Hum Mol Genet* 2001;**10**:721–33.
- 26 **Lynch HT**, Smyrk T. Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review. *Cancer* 1999;**78**:1149.
- 27 **Wijnen JT**, Vasen HF, Khan PM, *et al*. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. *N Engl J Med* 1998;**339**:511–18.
- 28 **Jenne DE**, Reimann H, Nezu J, *et al*. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet* 1998;**18**:38–43.
- 29 **Whitcomb DC**. Hereditary pancreatitis: new insights into acute and chronic pancreatitis. *Gut* 1999;**45**:317–22.
- 30 **Etemad B**, Whitcomb DC. Chronic pancreatitis: diagnosis, classification, and new genetic developments. *Gastroenterology* 2001;**120**:682–707.
- 31 **Hampe J**, Cuthbert A, Croucher PJ, *et al*. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001;**357**:1925–8.
- 32 **Judge T**, Lichtenstein GR. The NOD2 gene and Crohn's disease: another triumph for molecular genetics. *Gastroenterology* 2002;**122**:826–8.