Human mutations and their detection by gene and linkage analysis, allele sharing and association methods

J.A. Phillips III[1] and R. Hamid[1]

SUMMARY It has been 20 years since DNA analysis was first used in the detection of sickle-cell anaemia. Here, techniques for detecting human mutations are reviewed. We describe direct detection of mutations using restriction enzyme analysis and polymerase chain reaction amplification to detect gene deletions, rearrangements and point mutations. Indirect detection of mutations include the use of DNA polymorphisms in linkage analysis.

Introduction

Some of the many types of mutations that can cause inherited disorders are:

- single base changes that can alter and disrupt regulatory sequences, cap sites, initiation, coding or termination codons, splicing or polyadenylation signals;
- deletions or insertions of one, a few, or many bases;
- rearrangements, such as inversions or fusions.

Any of these mutations can perturb expression of the altered gene and cause disease. Detection and characterization of mutations in human genes provide insights into the pathophysiology of hereditary disorders. Diagnostic tests can be developed and new therapeutic approaches designed.

The basic tools used to detect gene mutations include restriction endonucleases, oligonucleotides and specific DNA polymerases. Restriction endonucleases can be used to produce fragments from genomic DNA, their size reflecting their nucleotide sequence. Gene segments and oligonucleotides can be labelled and used as probes that anneal to, and detect specific genomic fragments in Southern blots. Oligonucleotides and DNA polymerases can be used to synthesize complementary genomic segments using polymerase chain reaction (PCR) amplification. The PCR products can then be cleaved by restriction endonucleases, or annealed to complementary sequences to enable detection of mutations using a variety of techniques. Applications of these techniques for directly or indirectly detecting gene alterations are illustrated in the following examples [*1–3*].

Direct detection of mutations

Restriction enzyme analysis

Restriction enzyme analysis can be used to detect chromosome fragments and a wide

[1]*Division of Genetics, Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America.*

Received: 10/09/98; accepted: 05/09/99

variety of mutations, including gene deletions, insertions, rearrangements and point mutations. Each restriction enzyme has a unique cutting (restriction) site, giving in turn a unique pattern when used to cut a piece of DNA. The most common technique using this property of restriction enzymes is called Southern blot analysis.

Restriction enzymes are used to fragment genomic DNA into reproducible sizes. These fragments are subjected to gel electrophoresis to separate them by size [*1–3*]. The fragments are then transferred to a membrane and hybridized to a radioactively-labelled probe. The number and size of DNA fragments are analysed, based on appropriate imaging. Mutations or deletions in a gene are picked up if they involve a particular restriction site, thus giving a pattern that is different from the control. In the following section, some examples of restriction enzyme analysis are given.

Detection of gene deletions

Familial isolated growth hormone (GH) deficiency type 1A (IGHD 1A) is an endocrine disorder caused by gene deletions. IGHD 1A has an autosomal recessive mode of inheritance. Affected individuals have severe growth retardation due to complete deficiency of GH. Most individuals respond only briefly to GH replacement therapy because of their tendency to develop high titres of anti-GH antibodies [*4*]. Restriction analysis of the structural gene for GH (GH1) is complicated by there being 5 GH related genes (*5'-GH1:CSHP1:CSH1:GH2:CSH2-3'*) in the GH gene cluster. Although these related genes share extensive sequence homology, only the *GH1* locus encodes GH. The *GH1* gene is flanked by consistent *Bam*HI sites approximately 3.8 kb apart. While the *CSHP1*, *CSH1*, *GH2*, and *CSH2* genes are sufficiently homologous to hybridize to the *GH1* probe, they are all contained in *Bam*HI derived fragments that differ in size from that of *GH1*. Autoradiograms of DNA from IGHD 1A individuals lack the 3.8 kb fragments that normally contain the *GH1* genes. In addition, the intensity of the 3.8 kb hybridizing fragments from DNA of the heterozygous (GH/del) parents is intermediate between that of controls (GH/GH) and their affected children, who are homozygous (del/del) for *GH1* gene deletions. Since homozygosity for these *GH* deletions precludes production of any GH, IGHD 1A individuals tend to be immunologically intolerant to, and develop antibodies to exogenous GH.

Detection of gene rearrangements in neoplasias

Acquired neoplasias can be associated with specific chromosome rearrangements. For example, in chronic granulocytic leukaemia (CGL) the Philadelphia (Ph^1) chromosome, a 9q34;22q11 reciprocal translocation, is consistently found. The breakpoints producing Ph^1 are within the *abl* proto-oncogene on chromosome 9q34, and a gene referred to as the breakpoint cluster region (*BCR*) on 22q11 [*5,6*]. DNA fragments resulting from digestion of the fused or chimeric *abl:BCR* genes can be detected by Southern blots using a probe derived from the *abl* gene. Such studies can detect the relative proportion of cells containing Ph^1 chromosomes to monitor the response to treatment as well as early relapses.

Detection of point mutations

Sickle-cell anaemia is an autosomal recessive disorder characterized by episodic sickle crises caused by irreversible sickling and destruction of red blood cells. This results in anaemia and abdominal and musculoskeletal pain. In the United States, sickle-cell anaemia occurs predominantly

in African–Americans who have a carrier frequency of approximately 1 in 15, and disease due to homozygosity in approximately 1 in 500 births. It is caused by an A to T transversion in the sixth codon of the β-globin gene encoding a GAG (glutamine) to GTG (valine) substitution [*7*]. This transversion destroys an *Mst* II recognition site (CCTGAGG) spanning codons 5–7 of the β-globin gene [*2,3*]. Digestion of genomic DNA containing the sickle mutation yields restriction fragments of 1.35 kb, rather than the 1.15 kb fragments generated from the normal β-globin squence. Detection of this A→T transversion, to diagnose sickle-cell anaemia by Southern blot analysis, was one of the first applications (1982) of DNA techniques to detect an inherited disease prenatally. Subsequently Southern blot analysis for the detection of sickle-cell anaemia has been replaced by PCR-based techniques.

Polymerase chain reaction amplification

PCR amplification is a primer-directed enzymatic amplification of specific DNA sequences [*8*]. The specificity of PCR amplification results from the use of specific primers that flank the DNA segment to be amplified, and a critical annealing temperature during the reaction [9]. For example, after denaturing genomic DNA at 94 °C, the specific primers are annealed to opposite genomic strands at 45 °C. After annealing, the reaction is incubated at 72 °C with a specific DNA polymerase to synthesize a new copy of each strand of the gene. The PCR copies are then denatured at 94 °C. and the next PCR amplification cycle begins. After each cycle, the number of gene copies synthesized doubles, so that after 30 cycles, there are between 10^6 and 10^7 copies. These PCR products, called amplicons, can then be analysed in a variety of ways to detect mutations, as illustrated in the following examples.

Detection of gene deletions

The most common mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) that causes cystic fibrosis (CF) is a 3 bp deletion in exon 10. This deletion (called ΔF508) at codon 508 of the *CFTR* gene, causes loss of phenylalanine and shortens the normal 98 bp amplicon to 95 bp [*10*]. Studies for this and other *CFTR* gene mutations are usually carried out to confirm that a patient has CF. or to determine the carrier status of individuals with a family history of CF. The amplicons derived from the DNA of individuals heterozygous for the ΔF508 mutation form heteroduplexes. These heteroduplexes are sensitive indicators for allelic PCR products that differ in size or sequence. Heteroduplexes arise when the forward and reverse strands from two different alleles that differ in size or sequence anneal, causing "bubbles" to form that alter their migration.

Detection of gene rearrangements

CGL cells containing Ph^1 chromosomes can be detected by PCR amplification using one primer specific for *BCR*, and one specific for *abl* sequences. The *BCR/abl* fusion products can be detected using such a primer pair to amplify cDNAs derived from leukocyte-isolated mRNA. Another application of this approach is the detection of certain types of lymphomas resulting from fusion of specific segments of chromosome 14 and 18. These can be detected by amplification of their chimeric PCR products.

Detection of point mutations

Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency is a rare (approximate-

ly 1 in 50 000 live births) autosomal recessive disorder of fatty acid oxidation. Onset of symptoms (hypoglycaemia, metabolic acidosis, lethargy and coma) usually occurs in infancy or early childhood after prolonged fasting. While MCAD plays a central role in fat metabolism, some symptoms of its deficiency result from secondary carnitine deficiency. The treatment of MCAD deficiency includes not fasting, and supplementing with oral L-carnitine. While several mutations have been identified in the *MCAD* gene, an A to G transition of codon 329 (K329E) represents more than 90% of reported cases. The K329E mutation can be detected by *Nco* I digestion of amplicons containing a segment of the *MCAD* gene. Normal MCAD alleles yield CCATGA PCR products that are not cut by *Nco* I, thus producing 63 bp fragments, while K329E alleles yield CCATGG products that are cleaved into 43 bp and 20 bp fragments.

DNA sequence analysis of amplified DNA

DNA sequence analysis is usually carried out on PCR-amplified products (amplicons) that contain segments of interest. Sequence analysis of amplicons provides data on the sequence of an individual's genomic DNA. Cycle sequencing is the method that is most often used to directly sequence PCR products. This method requires very small (i.e. nanogram) quantities of amplicons. It uses the PCR cycles of denaturation and extension at temperatures that prevent reannealing of the two strands of template DNA. This forces the primer used in sequencing to anneal with its complementary template strand for primer extension. Dideoxy terminators are used and the products are analysed by size to infer their DNA sequence.

Indirect detection of mutations

In most inherited disorders the exact gene mutation is unknown. Despite this, linkage analysis using DNA polymorphisms can be used to determine the transmission pattern of the defective gene.

DNA polymorphisms

Most variations seen between individuals at the DNA level are associated with DNA polymorphisms which occur approximately every 250–500 nucleotides in noncoding regions of the genome. DNA polymorphisms are defined as DNA changes that are seen in the normal population at frequencies of more than 1% [*1–3*]. In some cases these changes alter the number and location of restriction enzyme recognition sites. The resulting differences in DNA fragment sizes, referred to as restriction fragment length polymorphisms (RFLPs), are easily detected by analysis of Southern blots or PCR products.

A second, common type of DNA polymorphisms are the 2 (di), 3 (tri) or 4 (tetra) nucleotide repeats, called microsatellites. Microsatellites are small simple repeats of 1 to 6 bases (i.e. AGAG, CAGCAG, CGGGCGGG), which are found throughout the genome. Because of their frequency and wide distribution, they provide a source of abundant markers. Their uses include facilitating the identification of many genes, determining the order and locations of genes on the chromosomes, and traditional linkage studies for diagnostic analysis. In forensic applications they are used for identity and paternity testing.

Linkage analysis using DNA polymorphisms

Linkage analysis involves proposing a model to explain the inheritance pattern of

phenotypes and genotypes observed in a pedigree. It is the method of choice for simple Mendelian traits, because the allowable models are few (autosomal dominant, recessive or X-linked) and easily tested. In linkage analysis, the transmission patterns of DNA polymorphisms are used to infer the transmission of genes that lie adjacent to, and co-segregate with, these markers. The closer DNA polymorphisms or other markers and genes are physically, the more likely they will co-segregate. Rough estimates indicate that a DNA polymorphism and gene that are approximately 10^6 bp apart have a 99% probability of segregating together. For predetermined recombination frequencies, (θ) probabilities are calculated. Then the logarithm of the ratio of probability (θ)/probability of random assortment ($\theta = 0.5$) = the logarithmic odds score (LOD score) or Z score ($Z = \log_{10}$), is used to determine the most probable θ (method of maximum likelihood). The peak LOD score is used as an approximation of the 95% confidence interval for a given θ. For example, for a simple Mendelian recessive or dominant disease, $\theta = 0\%$ between the disease gene marker with a LOD score of 1, 1.5, 2 or 3 indicates no recombination between the marker and disease phenotype, with probabilities of significant linkage being 1/10, 1/32, 1/100 and 1/1000 respectively. When LOD scores are negative they reflect the probability against linkage. When the recombination frequency (θ) is 0.00, 0.05, 0.10, 0.30, the recombination frequencies are 0%, 5%, 10% and 30%, with 50% being the theoretical maximum that is observed. For rapid calculations of LOD scores for multiple θs, linkage computer programs are used.

A clinical example of the use of DNA polymorphisms and linkage analysis is haemophilia A caused by deficiency of clotting factor VIII. As previously mentioned, under restriction enzyme analysis, an inversion within the factor VIII gene can occur, leading to its inactivation. This rearrangement accounts for the disease in approximately 45% of severe cases. However, many different mutations have been found in the factor VIII genes of different haemophiliacs [*11,12*]. This heterogeneity makes direct detection of non-inversion mutations impractical. Linkage analysis using DNA polymorphisms within the factor VIII gene provides an alternative method to determine if female relatives are carriers. For example, a DNA polymorphism that affects a *Bcl* I recognition site lies within IVS18. Since approximately 50% of females are heterozygous for this *Bcl* I/IVS18 DNA polymorphism, its analysis often enables inference of a factor VIII status for their offspring. For example, a woman, an obligate carrier of haemophilia A, is heterozygous for the IVS18 DNA polymorphisms, and has the 1.2 kb and 0.9 kb fragments associated with the absence and presence of the *Bcl* I site. Her affected sons and grandsons all inherited her factor VIII gene containing the 1.2 kb DNA polymorphism, indicating that her factor VIII gene, "coupled" to the 1.2 *Bcl* I/IVS18 allele, is her mutant factor VIII gene. Thus, her three younger daughters are all carriers, because each inherited her 1.2 kb DNA polymorphism. By contrast, her eldest daughter is a non-carrier because she inherited the 0.9 kb DNA polymorphism, which is in "repulsion" to her mother's mutant factor VIII gene, and is therefore on the chromosome that contains her normal factor VIII gene.

Conclusion

In summary, four different ideas are important in understanding diagnostic applications using DNA analysis. These are:

- When DNA changes in a gene are detected, it must be determined if the

changes represent DNA polymorphisms (changes not associated with disease) or mutations that affect expression of the gene.

- Differences seen in mutations in different patients (heterogeneity) often explain clinical variation at a molecular level.
- Current methods used for gene analysis include restriction enzyme analysis, PCR amplification, allele specific oligonucleotides, denaturing gradient gels and DNA sequence analysis.
- Gene diagnosis is applicable to many clinical disorders, both genetic and acquired. What is required is a portion of the gene involved, or a segment of DNA that lies close to the gene.

It has been 20 years since DNA analysis was first used in the diagnosis of sickle-cell anaemia [*13*]. Since that time, accurate tests using different methods of DNA analysis have been developed for many inherited diseases. With the Human Genome Project, the number of such applications should increase dramatically in the next few years. This initiative in gene mapping and sequencing will identify all DNA markers and genes contained in the human genome. This will provide countless DNA segments, oligonucleotides and PCR primers that could be used to detect mutations underlying many inherited disorders, both single and polygenic, as well as acquired gene rearrangements associated with neoplasias and ageing.

References

1. Antonarakis SE et al. Genetic diseases: diagnosis by restriction endonuclease analysis. *Journal of pediatrics*, 1982, 100:845–56.
2. Phillips 3rd JA. Clinical applications of gene mapping and diagnosis. In: Childs B et al., eds. *Progress in medical genetics*. New York, Elsevier, 1987:68–99.
3. Phillips 3rd JA. Diagnosis at the bedside by gene analysis. *Southern medical journal*, 1990, 83:868–75.
4. Phillips 3rd JA. Inherited defects in growth hormone synthesis and action. In: Scriver CR et al., eds. *The metabolic basis of inherited disease*, 6th ed. New York, McGraw–Hill, 1989:1965–83.
5. Collins SJ. Breakpoints on chromosomes 9 and 22 in Philadelphia chromosome-positive chronic myelogenous leukemia (CML). Amplification of rearranged c-*abl* oncogenes in CML blast crisis. *Journal of clinical investigation*, 1986, 78:1392–6.
6. Kogan SM et al. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences: application to hemophilia A. *New England journal of medicine*, 1987, 317:985–90.
7. Posey YF et al. Prenatal diagnosis of sickle-cell anemia. Hemoglobin electrophoresis versus DNA analysis. *American journal of clinical pathology*, 1989, 92:347–51.
8. Gyllensten UB, Erlich HA. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proceedings of the National Academy of Sciences of the United States of America*, 1988, 85:7652–6.
9. Saiki RK et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 1988, 239:487–91.

10. Campbell 3rd PW et al. Cystic fibrosis: relationship between clinical status and F508 deletion. *Journal of pediatrics*, 1991, 118:239–41.

11. Antonarakis SE. The molecular genetics of hemophilia A and B in man: Factor VIII and Factor IX deficiencies. In: Harris H, Hirschhorn K, eds. *Advances in human genetics*, New York, Plenum Press, 1988:27–59.

12. Gitschier J et al. Genetic mapping and diagnosis of haemophilia A achieved through a *Bcl* I polymorphism in the factor VIII gene. *Nature*, 1985, 314:738–40.

13. Kan YW et al. Antenatal diagnosis of sickle-cell anaemia by DNA analysis of amniotic-fluid cells. *Lancet*, 1978, 2:910–2.

Genetics in Health and Disease — Status, Implicit and Implications for Individuals and Community and 3rd MEGA Meeting

This meeting will be held from 15–18 Muharram 1422 (9–12 April 2001) at the Cultural Palace, Diplomatic Quarters, Riyadh, Saudi Arabia.

The deadline for submission of abstracts is 31 January 2001: the deadline for registration and visa application is 31 January 2001 and the deadline for pre-registration from within the country is 28 February 2001.

The main sponsors are

- King Saud University
- MEGA
- World Health Organization
- Human Genome Organization

Further information can be obtained from: The Secretariat, Department of Medical Biochemistry and WHO Collaborating Centre/Postgraduate Centre, College of Medicine, King Saud University, PO Box 2925, Riyadh 11461, Saudi Arabia. Tel: (966) 1 4670831/4671551; Fax: (966) 1 4672575/4811853; E-mail: mohsen@ksu.edu.sa