Upsala J Med Sci 107: 1–8, 2002

Merging classical and modern genetic tools in the identification of disease genes[1)]

Review based on the doctoral thesis:
"Diamond-Blackfan anemia. Mapping and Identification of the Disease Gene"

Peter Gustavsson

Unit of Clinical Genetics, Department of Genetics and Pathology, Uppsala University, SWEDEN

INTRODUCTION

Identification of genes involved in different diseases has increased our knowledge of molecular pathology, the development of diseases and the normal function of genes and their products – proteins. In addition, the characterization of disease genes has introduced genetic analysis as a diagnostic tool. The OMIM (Online Mendelian Inheritance in Man) database shows that more than 1,000 genes contain variants associated with a specific phenotype (1,2). Different research groups have estimated the total number of genes in the 3,000 million base pairs of DNA in the human genome to be between 34,000 to 140,000 (3).

This review is based on the work presented in the thesis "Diamond-Blackfan anemia. Mapping and Identification of the Disease Gene" and it will focus on three important strategies that were used in cloning the disease gene (4);

1) **Cytogenetic analysis** in a key patient carrying a balanced translocation which lead to the chromosomal localization of the disease gene,
2) **Genetic analysis in familial cases** showing statistically significant linkage between the inherited form of the disorder and the genetic region identified by cytogenetic analysis,
3) **Bioinformatics** used in sequence assembly after large scale sequencing of the candidate region, candidate gene identification and determination of gene structures before mutation analysis.

In the research group of Clinical Molecular Genetics, Department of Genetics and Pathology, methods and approaches briefly presented in this review are applied on different projects aiming to identify disease genes. Several ongoing studies were initiated after identification of key patients carrying cytogenetic abnormalities associated with phenotypes such as mental retardation, urogenital malformations, autism and extreme obesity.

1) Received 20 March 2002
Revised manuscript accepted 17 April 2002

GENERAL BACKGROUND

1) Cytogenetic analysis

Chromosomal abnormalities are present in an estimated 0.6% of live births (5). Chromosomal abnormalities can be divided into numerical or structural abnormalities (5). Numerical abnormalities include monosomies, trisomies, triploidy and mosaicism of chromosomes. Structural aberrations consist of duplications, insertions, chromosomal translocations, inversions and large deletions.

The approach of using chromosomal rearrangements in key patients may be an important first step towards the identification of a disease gene since a chromosomal rearrangement may pinpoint the localization of a disease gene (6). In a balanced reciprocal translocation, breaks occur in two chromosomes and the segments are exchanged and no genetic material is lost or gained (Figure 1). A balanced translocation is present in about 1 in 500 newborns (5). Although certain individuals manifest a phenotype related to disruption of a gene at the site of one of the breakpoints, a carrier of a balanced reciprocal translocation may be asymptomatic (6).

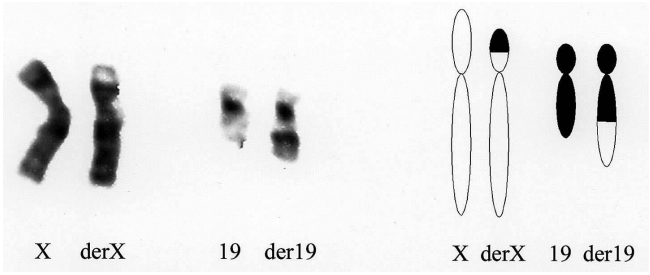

Figure 1. A partial karyotype of G banded chromosomes representing a chromosomal translocation between chromosomes X and 19 (left). A corresponding schematic picture illustrating a balanced translocation with exchange of chromosomal material is shown (right).

The X;19 translocation (left), identified in a girl presenting with Diamond-Blackfan anemia, resulted in two chromosome derivatives; derX (Xpter-p21:19q13-pter) and der19 (Xqter-p21:19q13-qter).

In order to characterize and map chromosomal rearrangements, classical cytogenetic analysis with chromosomal banding is often used. Due to the low resolution after chromosome banding, fluorescent in situ hybridization (FISH) was developed where fluorescent labeled nucleotides are incorporated in probes and physically mapped by hybridization to metaphase chromosomes (Figure 2, ref 7,8).

2) Genetic analysis in families

Without prior knowledge of the function or location of a disease gene, linkage analysis in families can be performed in order to identify a chromosomal region that segregates with a phenotype (9). However, linkage analysis may be time consuming and not always successful in mapping a disease gene. There are several factors that limit the efficiency of a linkage analysis such as small family materials and low

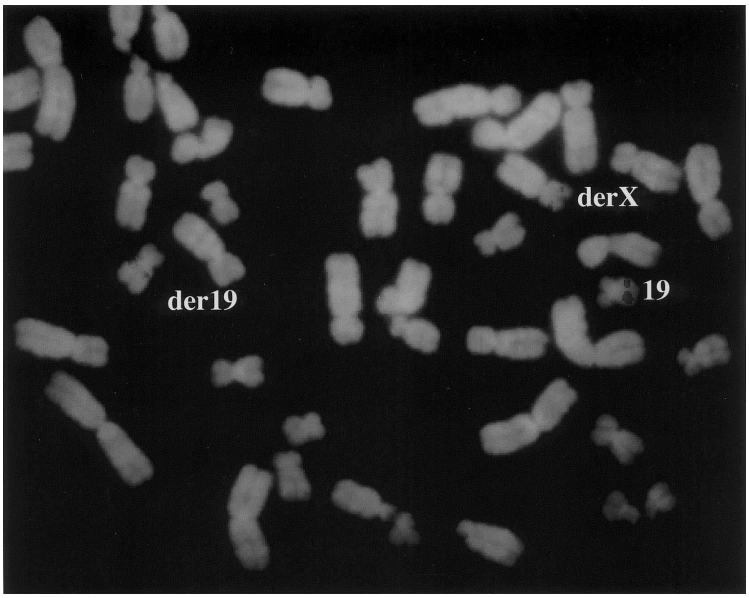

Figure 2. Fluorescent in situ hybridization (FISH) to metaphase chromosomes. Red color represents a DNA probe specific against a 40 kb DNA region on chromosome 19q13.

In this picture, the chromosome 19 translocation breakpoint is identified in the DBA patient with the X;19 translocation. FISH analysis shows that the probe hybridizes to both translocation derivatives (derX and der19) as well as to the normal chromosome 19.

penetrance of the disease. Also, well-defined patient materials are important since specific criteria for a specific disease allow one to avoid overlapping phenotypes in the patient material (9).

Linkage analysis is carried out by analyzing segregation of marker loci spanning the complete genome or a candidate gene region in DNA from familial cases (Figure 3, ref 9). If two loci are located near each other on a chromosome, they are likely to be inherited together. The statistical evaluation of this relationship can be performed by a LOD (logarithm of the odds ratio) score analysis (9).

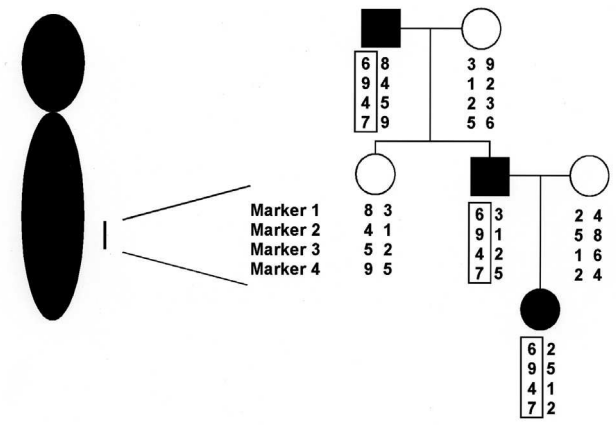

Figure 3. Haplotype analysis in familial cases using polymorphic genetic markers. A candidate gene region on a chromosome (left) is studied in a family presenting with a dominant inherited disorder (right). Filled symbols indicate affected individuals.

The corresponding chromosome haplotypes are indicated below the symbols. Segregation of the four marker alleles is compatible with linkage since affected relatives have inherited identical haplotypes indicated as a squared symbol.

When a chromosomal translocation indicates the localization of a putative disease gene, linkage analysis can be used as a statistic tool to show that inherited forms of the disease also is linked to the gene region identified in the patient(s) carrying the chromosomal translocation. Studies in a family material will not only provide a statistical estimation of linkage but also indicate genetic heterogeneity i.e. identical clinical symptoms caused by mutations in other genes. In addition, if genetic heterogeneity is present, identification of linked families is important for following analyses including mutation detection.

3) Bioinformatics

Linkage analysis and/or chromosomal mapping of a rearrangement can be followed by a positional cloning approach to identify the disease gene (10). This involves the construction of a physical map of the chromosomal region and identification of candidate genes.

The identification of genes or transcripts can be performed by several methods (11,12). Due to the efficacy of sequencing and sequence assembly, large scale sequencing is at present the best approach to generate a map of a candidate gene region (Figure 4, ref 13). Different computer programs are used to assemble sequence data into continuous and longer DNA segments i.e. contigs (14,15).

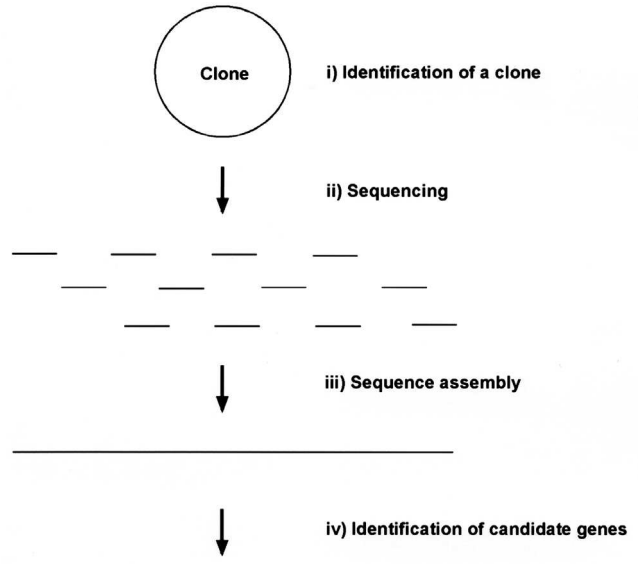

Figure 4. Identification of candidate disease genes based on the identification of a chromosomal rearrangement associated with a disorder. Important steps include i) the identification of a candidate clone, ii) large-scale sequencing (if the corresponding sequence is not available), iii) assembly of sequences into contigs and iv) bioinformatic analyses in order to identify or predict candidate genes.

Genes and transcribed sequences in established contigs can be identified by different bioinformatical tools, such as homology searches in different databases (16). In parallel, different gene prediction programs are used since homology searches rely on previously known genes or transcripts submitted to the database (17).

When candidate genes have been identified, mutation detection is performed in patient samples and normal controls (Figure 5). Sequencing of genomic DNA is a robust technique that results in few false negative results (18).

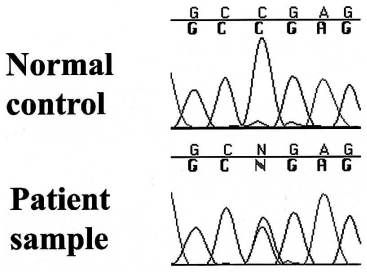

Figure 5. Identification of a DNA mutation after sequence analysis. Sequence chromatograms are shown for a control (top) as well as for an individual with Diamond-Blackfan anemia (below). The mutation corresponds to a C to T transition resulting in a premature stop codon of the RPS19 gene.

Following identification of mutations, bioinformatic analyses can be used to understand and explain the mechanisms behind the effects of mutations in relation to the disease and to examine the function of the geneproduct. These studies include prediction of protein structure, phylogenetic analysis and comparative genomics. Comparative studies are important since homology searches against databases containing findings from other species may result in understanding of the function of a gene (19).

SPECIFIC BACKGROUND AND THESIS

The overall aim of the work presented in the thesis "Diamond-Blackfan anemia. Mapping and Identification of the Disease Gene" was to investigate the genetic basis of Diamond-Blackfan anemia. Various methods were used, including fluorescent in situ hybridization (FISH) to metaphase chromosomes, linkage analysis in families with DBA, large-scale sequencing, bioinformatic analyses, expression studies and mutation detection.

Diamond-Blackfan anemia (DBA) is characterized by a congenital defect in the production of red blood cells with an absence or decreased amounts of erythroblasts in the bone marrow (20-22). The disease is rare, with an estimated incidence of 0.5-1 per 100,000 births in Europe (22).

At least 10% of DBA cases follow an autosomal dominant or recessive inheritance pattern (22). Approximately 50% of DBA patients have congenital abnormalities, including craniofacial anomalies and malformations of the thumb and the heart (20,22). Growth failure and psychomotoric retardation is also present in a proportion of patients with DBA (21).

Treatment of DBA is limited to corticosteroid therapy, red blood cell transfusions or bone marrow transplantation (22). The first line of treatment for DBA consists of corticisteroids, to which more than 60 % of patients respond (22). Transfusions are required for the remaining patients with risk of transfusion hemosiderosis. A proportion of DBA patients may go into spontaneous remission (20).

Patient material

A girl, diagnosed with DBA, was identified carrying a chromosomal rearrangement associated with the disease. Karyotype analysis of Giemsa stained banded chromo-

somes revealed what appeared to be a balanced translocation, involving chromosome Xp and 19q (Figure 1). Previous studies had shown that DBA presented in families with an autosomal dominant inheritance. Therefore, FISH was performed to metaphase chromosomes derived from the girl with the chromosome X;19 translocation in order to further map the chromosome 19q translocation breakpoint.

In order to study the segregation pattern of chromosome 19q markers in families with DBA, we collected blood or DNA samples from approximately 150 DBA patients from 10 countries. The local ethical committee at Uppsala University approved the work in the thesis. The European studies were approved by the European Society of Pediatric Hematology.

Results and Discussion

The thesis resulted in mapping and identification of the gene encoding ribosomal protein S19 that in a mutated form causes Diamond-Blackfan anemia. The RPS19 protein has previously been found to be located on the external surface of the 40S ribosomal subunit using a specific antibody (23). Seventy-nine different proteins are involved in the ribosomal organelle that constitute the protein translation machinery in our cells (24). The RPS19 protein is the first ribosomal protein found to be associated with a disease.

Fluorescent in situ hybridization (FISH) to metaphase chromosomes was found to be a key tool in the identification of the DBA disease gene. LOD score analysis in the candidate gene region, pinpointed by the chromosomal translocation breakpoint, was also important. In the positional cloning of the DBA gene, the RPS19 gene was identified by computer and web based programs after FISH analysis and large-scale sequencing. Sequence analysis of patient genomic DNA revealed RPS19 gene mutations in 10 out of 40 DBA cases.

Other findings that support the involvement of the RPS19 gene in the development of DBA include the fact that mutations in the RPS19 gene segregate with familial cases and are not present in healthy control individuals. In addition, different types of mutations were identified, including the chromosomal translocation breakpoint that resulted in a disrupted RPS19 gene, microdeletions spanning the entire gene, and missense and frame shift mutations.

There are two main explanations to correlate a disease to a mutated ribosomal protein (24, 25); 1) A ribosomal protein insufficiency may result in a reduced amount of intact and functional ribosomes at one or more critical stages of development. A reduced translational capacity, caused by the defect ribosomal protein, would result in a specific phenotype. 2) An extraribosomal function in a disrupted or alternated form during development that causes a pathological condition.

Future studies

Once a gene has been identified, the work still remains to understand the function of the gene and its gene product (26). In order to study the molecular mechanism of the

RPS19 gene product in relation to erythropoiesis we have initiated functional studies of RPS19. We are studying protein interactions by yeast two-hybrid systems where the RPS19 gene product is used as a bait to pull out an interacting molecule (27). The two-hybrid system has been shown to be a successful approach in Saccharomyces cerevisiae where more than 2500 protein-protein interactions are described (28).

We are currently making a mouse model for the disease by homologous recombination of the corresponding mouse gene. Mice have been the most useful organism in biomedical research, both in providing animal models of disease and in the studies of mammalian gene function (29, 30). The DBA mouse model will give opportunities to study the physiology and biology behind the disease as well as the normal erythroid differentiation.

In collaboration with a research group at Lund Hospital University, we will perform retroviral transfer of an intact RPS19 gene to the defective hematopoietic cells in bone marrow cultures from DBA patients. It would then be possible to see whether the effects of the mutant gene, representing low amounts of proerythroblasts and reduced production of red blood cells, could be reconstituted or not (31, 32). This would not only show that the RPS19 gene is involved in the development of DBA, but it would also be a first step towards the treatment of patients with Diamond-Blackfan anemia.

ACKNOWLEDGEMENTS

I would like to thank Professor Niklas Dahl and Dr Edward Davey for comments on the manuscript, the Beijer Foundation and Marcus Borgström's Foundation for financial support.

REFERENCES

1. Antonarakis SE, McKusick VA (2000) OMIM passes the 1,000-disease-gene mark. Nat Genet 25: 11.
2. OMIM Statistics (Jan 17, 2002); http://www.ncbi.nlm.nih.gov/Omim/Stats/mimstats.html
3. Aparicio SA (2000) How to count ... human genes. Nat Genet 25: 129-130.
4. Gustavsson P (2000) Diamond-Blackfan Anemia. Mapping and Identification of the Disease Gene. Acta Universitatis Upsaliensis. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 974, 48 pp.
5. Thompson MW, McInnes RR, Willard HF (1991) Genetics in medicine, WB Saunders, Philadelphia.
6. Tommerup N (1993) Mendelian cytogenetics. Chromosome rearrangements associated with mendelian dis orders. J Med Genet 30: 713-727.
7. Rudkin GT, Stollar BD (1977) High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. Nature 265: 472-473.
8. Trask BJ (1991) Fluorescence in situ hybridization: applications in cytogenetics and gene mapping. Trends Genet 7: 149-154.
9. Ott J (1991) Analysis of Human Genetic Linkage. The Johns Hopkins Press, Baltimore.
10. Collins FS (1995) Positional cloning moves from perditional to traditional. Nat Genet 9: 347-350.
11. Lovett M, Kere J, Hinton LM (1991) Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. Proc Natl Acad Sci USA 88: 9628-9632.
12. Auch D, Reth M (1990) Exon trap cloning: using PCR to rapidly detect and clone exons from genomic DNA fragments. Nucleic Acids Res 18: 6743-6744.
13. Weber JL, Myers EW (1997) Human whole-genome shotgun sequencing. Genome Res 7: 401-409.
14. Staden R (1996) The STADEN sequence analysis package. Molecular Biotechnology 5: 233-241.
15. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8: 186-194.

16. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.
17. Uberbacher EC, Mural RJ (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc Natl Acad Sci USA 88: 11261-11265.
18. Wilson RK, Chen C, Avdalovic N, Burns J, Hood L (1990) Development of an automated procedure for flu orescent DNA sequencing. Genomics 6: 626-634.
19. Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y (1998) Predicting function: from genes to genomes and back. J Mol Biol 283: 707-725.
20. Young NS, Alter BP (1994) Aplastic anemia: acquired and inherited, WB Saunders, Philadelphia.
21. Krijanovski OI, Sieff CA (1997) Diamond-Blackfan anemia. Hematol Oncol Clin North Am 11: 1061-1077.
22. Willig TN, Ball SE, Tchernia G (1998) Current concepts and issues in Diamond-Blackfan anemia. Curr Opin Hematol 5: 109-115.
23. Lutsch G, Stahl J, Kargel HJ, Noll F and Bielka H (1990) Immunoelectron microscopic studies on the location of ribosomal proteins on the surface of the 40S ribosomal subunit from rat liver. Eur J Cell Biol 51: 140-150.
24. Kenmochi N, Kawaguchi T, Rozen S, Davis E, Goodman N, Hudson TJ, Tanaka T, Page DC (1998) A map of 75 human ribosomal protein genes. Genome Res 8: 509-523.
25. Wool IG (1996) Extraribosomal functions of ribosomal proteins, In: (eds Green R and Schroeder R) Ribosomal RNA and Group I Introns, RG Landes and Springer Company, New York, pp 153-178.
26. Fields S (1997) The future is function. Nat Genet 15: 325-327.
27. Chien CT, Bartel PL, Sternglanz R, Fields S (1991) The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc Natl Acad Sci USA 88: 9578-9582.
28. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403: 623-627.
29. Nomura T (1997) Practical development of genetically engineered animals as human disease models. Lab Anim Sci 47: 113-117.
30. The Mouse Genome Database (MGD); http://www.informatics.jax.org/
31. von Kalle C, Glimm H, Schulz G, Mertelsmann R, Henschler R (1998) New developments in hematopoietic stem cell expansion Curr Opin Hematol 5: 79-86.
32. Lutzko C, Dube ID, Stewart AK (1999) Recent progress in gene transfer into hematopoietic stem cells. Crit Rev Oncol Hematol 30: 143-158.

ABOUT THE AUTHOR:

Peter Gustavsson received the Israel Hwasser award from the Upsala medical Association for the best dissertation in medicine in the academic year 2000/2001. He is at present carrying out his postdoctoral education at the Department of Genetics and Pathology, Unit of Clinical Genetics, Uppsala University. He has previously received a prestigous award from the Ronald McDonald Children Foundation and been visiting scientist in molecular genetics in Helsinki.

Address for reprints: Peter Gustavsson
Enheten för klinisk genetik
Institutionen för genetik och patologi
Rudbecklaboratoriet
SE-751 85 Uppsala
Sweden
Telephone +46 18 471 48 51
Fax +46 18 55 40 25
E-mail peter.gustavsson@genpat.uu.se