Clin Chem Lab Med 2000; 38(9):819–825 © 2000 by Walter de Gruyter · Berlin · New York

Hunting for Disease Genes in Multi-Functional Diseases

Chi-Pui Pang, Larry Baum and Dennis S. C. Lam

Department of Ophthalmology & Visual Sciences,
The Chinese University of Hong Kong, Hong Kong, China

Disease genes may be identified through functional, positional, and candidate gene approaches. Although extensive and often labor-intensive studies such as family linkage analysis, functional investigation of gene products and genome database searches are usually involved, thousands of human disease genes, especially for monogenic diseases with Mendelian transmission, have been identified. However, in diseases caused by more than one gene, or by a combination of genetic and environmental factors, identification of the genes is even more difficult. Common examples include atherosclerosis, cancer, Alzheimer's disease, asthma, diabetes, glaucoma, and age-related macular degeneration. There have been conflicting reports on the roles of associated genes. Even with population-based case-control studies and new statistical methods such as the sib-ship disequilibrium test and the discordant alleles test, there is no agreement on whether α_2-macroglobulin *(A2M)* is a gene for Alzheimer's disease. Another example is the inconsistent association between age-related macular degeneration and ATP-binding cassette transporter *(ABCR)*. Ethnic variation causes further complications. In our investigation of *LDL-receptor* variants in familial hypercholesterolemia, and the trabecular meshwork inducible glucocorticoid response protein, or myocillin (*TIGR-MYOC)* mutation pattern in primary open angle glaucoma, we did find dissimilar results in Chinese compared to Caucasians. New information from the Human Genome Project and advancements in technologies will aid the search for and confirm identification of disease genes despite such challenges.

Key words: Gene hunting; Candidate gene; Multi-functional diseases.

Abbreviations: ABCR, ATP-binding cassette transporter; AD, Alzheimer's disease; APP, amyloid precursor protein; A2M, α_2-macroglobulin; CFTR, cystic fibrosis gene; CSGE, conformation sensitive gel electrophoresis; DGGE, denaturing gradient gel electrophoresis; DHPLC, denaturing high performance liquid chromatography; ESTs, expressed sequence tags; FH, familial hypercholesterolemia; FISH, fluoresence in situ hybridization; LDL-R, low-density lipoprotein receptor; LPL, lipoprotein lipase; LRP, LDL-R related protein; SAGE, serial analysis of gene expression; SNPs, single nucleotide polymorphisms; STR, short tandem repeats; SSCP, single strand conformation polymorphism; TIGR-MYOC, trabecular meshwork inducible glucocorticoid response protein, or myocillin.

Introduction

Applying Murphy's law to genetics, any gene that can go wrong, will go wrong. Sequence alterations may, in one person or another, occur in any gene, and could be sporadic or inherited. Some of these changes will not affect function of the gene and have no association with any disease. But others are gene lesions that may be a gain, loss, or change of gene function. Subsequently the expressed gene product could be absent, deficient, wrongly processed or simply functionally defective. They will cause or increase the risk of a disease. How to detect these changes among the roughly 3 billion bases of human DNA, containing 30,000–50,000 genes (1), is a major task of biomedical science. Though the debate over nature *vs.* nurture has swung back and forth over the decades, the search for genes related to disease has steadily accelerated. Identification of disease-causing genes has led to the elucidation of the pathogenetic mechanisms, establishment of molecular diagnostics, and development of novel therapy of many diseases.

Now that the Human Genome Project is nearing completion, the search for disease genes will not end but will enter a new phase. In this phase, conventional techniques for locating disease genes will continue to be used, but they will be complemented by new methods to speed up the process. New or proposed methods of candidate gene identification include whole genome DNA chips for parallel rather than serial mutation detection, functional candidate gene identification by use of database searches, and quantitative trait loci identification through transgenic and gene knockout models (2). With the aid from advances in genome-wide screening methods and data from the Human Genome Project, it is anticipated that more and more disease-causing genes will be identified in the near future, and that candidate gene approaches will predominate. In order to avoid false-positive results in the confirmation of candidate genes as causative for disease, a large sample size, accurate diagnosis of patients, and replication of gene-phenotype associations in more than one independent dataset are preferred (3, 4). In addition, an understanding of the biological function of a candidate gene product should provide affirmative support to its link to the disease.

With these new tools at their disposal, medical scientists will be better armed to hunt more elusive prey, such as genes associated with diseases at low pene-

trance, late onset, or with multiple causative genes or risk factors. Advancement in analytical and information technologies will also help to affirm disease genes. Below, we will review the current techniques, and the potential future methods.

Current Techniques

Methods of gene hunting may be divided into three main categories: functional, candidate, and positional (5).

Functional approach

If knowledge of the function of a disease gene is available, this information may be used to identify the gene. This knowledge may allow identification of the gene's protein product, followed by any of several approaches to find the gene (5). The protein product should be extracted and purified. The purified protein may be then digested with proteases and the fragments partially sequenced. The amino acid sequences correspond to many possible nucleotide sequences since there are multiple codons for most amino acids. But degenerate oligonucleotides, covering all possible sequences, can be synthesized. These can be labeled and used as probes against replica plates of bacterial colonies from a cDNA library. Positive clones are sequenced to confirm whether they may indeed be capable of producing the desired protein. Though this method is prone to false positives, it has been a popular method for identifying novel genes. One Alzheimer's disease gene, amyloid precursor protein (*APP*), was identified this way (6, 7).

A purified protein may alternatively be used to generate an antibody, which can provide a means to screen an expression cDNA library. Positive clones can be sequenced to obtain the gene. A complementation assay may be used in special situations in which, for a recessive disease, a cell line is available which displays an abnormal phenotype. Inserting fragments of chromosomes or DNA into the cells, and then selecting for those whose phenotype is corrected, may allow identification of the chromosomal region or DNA sequence containing the mutated gene. A Fanconi's anemia gene was isolated through this approach (8).

Candidate gene

This approach involves taking an educated guess at what known genes may be factors in the disease, then testing them for mutations (5). The functional approach is a more time-consuming yet more reliable method of gene hunting, while the candidate gene approach is more dependent on chance and requires a good number of patients who must be diagnosed accurately. However, it is usually possible to increase the odds well above the 1 in 30,000–150,000 (1) chance resulting from a random choice among all human genes. Especially as more genes are becoming known and characterized, the information available to make a logical choice is increasing.

Knowledge of the phenotype and pathogenesis of the disease permits the narrowing of the potential candidate pool, or at least of the first focus of attention, to genes expressed in the diseased tissue at the time in development or stage at which the pathology appears. The identification of the cystic fibrosis gene *CFTR* (cystic fibrosis trans-membrane conductance regulator), for example, was helped in part by the screening of genes expressed in disease tissues (9). However, housekeeping genes with universal expression, as well as genes with spatio-temporal expression patterns not correlated with pathogenesis, often turn out to be the culprits. For example, the primary open angle glaucoma associated gene *TIGR-MYOC* (trabecular meshwork inducible glucocorticoid response protein, or myocillin) was found in many tissues throughout the body, not only in the tissue within the eye that regulates intraocular pressure (10, 11). Thus expression patterns, though currently popular subject of study, especially with the availability of microarray technology, should not be overly relied on, but should provide only part of the bait to help fish from the candidate gene pool.

Subtractive hybridization is a method for comparing gene expression between two samples, such as tissue from a disease patient and a normal control individual (12). The mRNA is purified from the two samples, and mRNA from one sample is copied into cDNA. The cDNA is hybridized with an excess of mRNA from the other sample and removed on a hydroxyapatite column. This subtraction can be repeated until the remaining population of cDNAs is sufficiently enriched for genes to be selectively expressed in that sample. These cDNAs can be identified by labeling them and using them to screen a cDNA library. The disadvantages of subtractive hybridization are that a large amount of mRNA is needed, and only genes with very different levels of expression can be detected. The advantages are the technique's low cost and ability to identify genes that are highly expressed in only one of the samples.

Differential display uses PCR to fractionate and amplify mRNA fragments so that they can be further fractionated on a gel, and the bands from different samples compared to identify differently expressed genes (13). First, mRNA is purified from the samples. A set of degenerate primer pairs are used to perform reverse transcription PCR in several tubes. The oligo-dT primers end in a few non-T bases (for two bases there would be 12 different primers). The other primers are arbitrary short sequences (about 10 bases). The PCR products are run on a gel, and each corresponding band of corresponding lanes from the different samples is compared. Bands that differ in intensity between samples are cut out of the gel and sequenced to identify differentially expressed genes. A disadvantage of differential display is that not all genes, only a random subset, are analyzed. Success may be quick, but may instead require a large amount of analysis. Advantages of this method are its low cost and simple procedure. Sophisticated equipment is not required.

Among the better-characterized genes, there may be

sufficient knowledge about their function to select some as candidates for mutation screening. Genes are potential candidates if they display similarity, in sequence, structure, or function, to genes known to cause or play a role in similar diseases. This applies not only to human diseases or genes, but also to animal models of human disease. When the gene mutated in an animal model of a human disease is known, the homologous human gene (or genes) is a prime candidate (14).

Once a candidate gene is selected, it is screened for mutations or polymorphisms that alter disease risk. Candidate gene mutation screening relies on PCR amplification, usually of all protein-coding exons plus nearby splice sites of the gene, followed by one of several methods of detecting the presence of a sequence variant (5, 15). Some of these methods cut the PCR product at a sequence variant: chemical cleavage of mismatches and enzymatic cleavage of mismatches. Other methods alter the conformation or secondary structure of the PCR product that contains a sequence variant: conformation sensitive gel electrophoresis (CSGE), the single strand conformation polymorphism (SSCP) method, denaturing high performance liquid chromatography (DHPLC), and denaturing gradient gel electrophoresis (DGGE). When any of these methods of initial screening suggests a sequence alteration, the PCR product is sequenced to find the exact site and nature of the alteration.

Candidate genes may be screened in a family or a case-control study. In a family with a disease gene mutation that is of high penetrance and low phenocopy (same phenotype but different cause), it would be expected to occur in all family members with disease, but none of the family members without the disease. The mutation should also be absent among the controls. However, for a situation of low penetrance or high phenocopy, correspondence between mutation and disease will no longer be one-to-one. To establish with confidence that a sequence change causes diseases, a very large family or a case-control study may be needed.

Diseases with multiple contributing factors, or polymorphisms that are either rare in occurrence in a particular population or that only modestly alter disease risk also present problems. The use of new statistical methods such as the sib-ship disequilibrium test and discordant alleles test can increase the power of studies. It has led to the observation of association of an α_2-macroglobulin (*A2M*) polymorphism with Alzheimer's disease. However, later studies were not able to clearly confirm the association (16). It may be that the association is population-specific, or requires the presence of secondary factors.

In age-related macular degeneration, studies have produced inconsistent conclusions about the association with *ABCR* (ATP-binding cassette transporter) gene variants (17–20). This gene is large with 50 coding exons encoding 2273 amino acids. It also has many non-disease associated polymorphisms, which may make the clear determination of association with disease more difficult by obscuring potential disease-associated sequence changes. One potential cause of false-positive associations in case-control studies is the screening of the whole gene in cases, and the subsequent screening of controls only for sequence alterations found in the cases. Thus, any rare, non-disease associated sequence changes that may happen to occur only in the controls will not be detected, and the total number of sequence changes may appear artifactually to be higher in cases than controls.

Low-density lipoprotein receptor (*LDL-R*) mutations in familial hypercholesterolemia patients (FH) appear at different sites in Chinese than whites, though the overall frequency of mutations is about the same (21–23). Also in contrast to some other populations such as Caucasians or Japanese, there is no definite common mutation in Chinese FH patients due to a founder effect. Although the majority of FH patients in Caucasian populations is associated with a *LDL-R* mutation in the promoter or protein coding exons, a substantial proportion of Chinese FH patients may be affected by multiple genetic factors, or by sequence aberrations in other lipid-related genes, notably the apolipoprotein B gene (21, 22).

Sometimes, however, ethnic variation in frequency of gene variants may mask or reveal disease genes. In the *TIGR-MYOC* gene, only one nonsense sequence change was detected in studies among whites; this change, Gln368Stop, which truncated part of the C-terminal of TIGR-MYOC protein, was associated with primary open angle glaucoma (24, 25). However, another nonsense change, Arg46Stop, that truncated TIGR-MYOC leaving less than 10% of the wild-type protein length, was found in subsequent studies only in Asians; in contrast to the truncation found in whites, this sequence change was present in normal control subjects as well as glaucoma patients (25–27). One elderly control subject was homozygous for Arg46Stop. Though she is essentially a human TIGR-MYOC "knock out," her vision is normal, implying that absence of TIGR-MYOC does not by itself cause glaucoma, and that TIGR-MYOC is not essential for normal eye function (26). In whites, *TIGR-MYOC* mutations associated with primary open angle glaucoma occur in exon 3, whereas in our study of Chinese most sequence changes are in exon 1, and so far none are clearly associated with disease (Figure 1) (26, 28). These findings would not have been possible if only one ethnic group, in this case whites, had been examined for mutations in this gene. Therefore ethnic genetic variation makes it important to study potential gene-disease associations in multiple ethnic groups.

Polymorphisms in apolipoprotein E and the gene for its receptor, LDL-R related protein (*LRP*), affect risk of Alzheimer's disease (AD) (29–33). A protein which mediates their interaction with each other, lipoprotein lipase (LPL), has polymorphisms that can affect triglyceride levels and risk of cardiovascular disease (34–37). *LPL* therefore appears to be a reasonable candidate gene for AD risk, and indeed, a recent study suggested an association of two LPL polymorphisms, 447Stop and 291Ser, with AD (38, 39). The former may offer a protective effect while the latter may be causative.

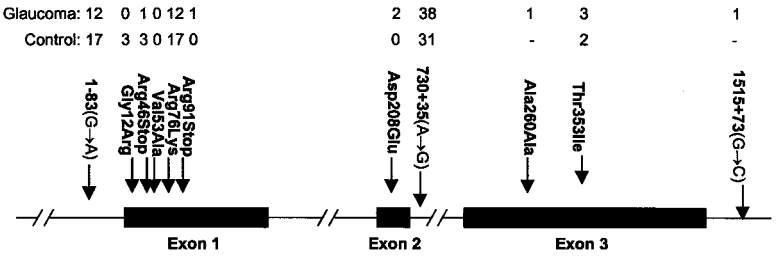

Fig. 1 The *TIGR-MYOC* gene showing sequence alterations and number of subjects displaying the alterations in Chinese with primary open-angle glaucoma (91 subjects) or unrelated normal controls (132 subjects). The black boxes represent coding regions of the 3 exons (26).

Positional

If there are families in which the disease is inherited, and most of the family members are available for DNA sampling and reliable diagnosis, then linkage analysis may narrow the chromosomal region in which the disease gene is likely to locate (5, 40, 41). This approach relies on the fact that chromosomes inherited from each person's parents recombine with each other during meiosis, so that the person's child will receive a chromosome that contains some segments from its parent's grandfather and some pieces from its parent's grandmother. Because each child comes from different germ cells (except for monozygotic siblings), individual child may inherit different combinations of chromosomal segments. These segments can be identified and traced from generation to generation by the use of chromosomal markers, and the inheritance of both the disease and a marker from parent to child suggests that the disease and the marker may be linked, implying that the disease gene is near the marker. Markers are short DNA sequences that have been selected because they are highly polymorphic, and have been mapped to known chromosomal locations (42). Being highly polymorphic increases the reliability of the assignment of an allele to one parent or the other by decreasing the chance that both parents will coincidentally happen to have the same alleles.

Several hundred short tandem repeats (STR), usually dinucleotide, polymorphic markers, are usually sufficient to map a disease gene to a chromosomal region. Denser maps of markers in that region may then help reduce the possible range of the gene locus (43). Because the alleles of STR markers are various multiples of the tandem repeat, a simple measurement of the length of a PCR product spanning a marker region is sufficient to genotype the marker. A marker and disease may often be inherited together by chance from a parent to a child. Therefore many parents and children must be examined to accumulate strong enough statistical evidence to conclude whether there is linkage of a marker with disease. Consequently, the positional approach to gene hunting requires at least one large family or several smaller families. Another important factor for success in positional cloning is that diagnosis of the disease is unambiguous. This approach also assumes that the disease is associated with a disease gene. Thus, diseases with low penetrance or late onset, or diseases with multiple causes, including environmental factors or several genes, may not be amenable to a positional strategy.

Large-scale mutations that are visible cytogenetically as chromosomal abnormalities, when found in patients, can provide valuable clues to the position of a disease gene. These are usually rare, but are relatively more common in severe, early-onset, and dominant diseases. Fluoresence *in situ* hybridization (FISH) mapping, restriction enzyme mapping, or chromosome band staining may be used to detect such large-scale mutations. For example, the muscular dystrophy gene dystrophin and neurofibromatosis gene *NF1* were found partly with the aid of such cytologically visible abnormalities (44, 45).

Once a chromosomal region is identified that is linked to a disease gene, screening of selected candidates within the region is usually necessary to find the gene (9, 44, 45). Such positional candidate gene screening is possible for larger and larger linked regions with the increasing fraction of human genes that have been characterized.

Recently Developed and Potential Future Methods

Bioinformatics

The near completion of the Human Genome Project, along with the expected later completion of the mouse DNA sequence, will enhance the efficiency of current gene hunting methods. A combination of automated and experimental approaches will annotate the raw DNA sequence to identify genes. Knowing the locations of all genes will allow the results of linkage analyses to be used to proceed directly to the screening of candidate genes within the linked region. Accurate information on the sequences of all genes will allow work to begin on the massive task of characterizing the functions of all previously unknown proteins (46). Cross-species comparisons can identify homologous genes and help to fill in clues to protein functions when more is known about a gene in one species than another (47). As protein functions become better understood, the functional and candidate approaches to finding disease genes will become more feasible (48).

Expression analysis

This has become very popular since the tools for performing rapid gene expression analysis have recently become available. Since an enormous amount of data (as well as pretty, color, cover illustrations) can be produced quickly, expression analysis is able to simultaneously measure the relative amounts of mRNA transcribed from a large number of genes (49–53). Different tissues or patients may be compared. By measuring mRNAs in healthy *vs.* diseased tissue, a gene whose expression is increased or decreased in disease may be further examined to see at what point in the chain of events leading from disease initiation does it play a role. Such comparisons are not limited to human tissue. Animal or cell culture models of disease may be analyzed to identify genes for further study.

The emerging standard format for expression analysis is a glass microarray slide on which have been spotted probes, such as cDNAs or expressed sequence tags (ESTs), for a set of genes (50–53). The mRNA to be analyzed is amplified and labeled with a fluorescent dye, then hybridized with the microarray. A microarray scanner reads the fluorescent signal on the slide, allowing computerized analysis of relative expression levels (50–53).

Serial analysis of gene expression (SAGE) is another technique for measuring mRNA levels (54). Its basic strategy is to produce short DNA tags unique for each gene, and then to concatenate many tags together for efficient sequencing in order to generate a digital representation of the relative expression of all genes. In this complex protocol (http://www.sagenet.org), mRNA is copied to cDNA using biotinylated oligo-dT, and then a second DNA strand is synthesized. The double stranded cDNA is cleaved with a restriction endonuclease, and the biotinylated fragments are ligated to a linker which contains the recognition site for a second restriction enzyme. The second enzyme is one which cuts several bases downstream from its recognition site. Cleavage by the second enzyme leaves only short fragments, or tags, of the cDNA. The tags are ligated together, producing ditags. The ditags are amplified by PCR. The amplified ditags are cut with the first restriction enzyme to remove the linker, and then are ligated into long concatamers. These concatamers are cloned into bacteria, and the cloned DNA concatamer from each colony is sequenced. Each sequence contains many tags, and is usually unique to a gene. Each time a tag is sequenced, a tick is added next to that tag's name. As clones are sequenced, the ticks can be counted to give a digital approximation of the relative abundance of all mRNAs. This approximation becomes more and more accurate as the number of sequenced clones increases. Disadvantages of SAGE are its technical complexity and time required to generate and sequence the clones, occasional difficulty associating tags with genes, and poor ability to measure low expressing genes. Advantages are the low cost of equipment, the flexibility to stop sequencing at any desired level of accuracy of the expression approximation, good ability to measure high expressing genes, and ability to detect virtually all genes (whether or not they are yet available on a microarray).

Mendelian genetics deals in categorical traits, but level of expression can manifest in continuous or quantitative traits. Quantitative trait loci identification through transgenic and gene knockout models may help to identify disease genes or genes that contribute to disease risk (2).

Parallel mutation detection

DNA chips, made by Affymetrix Corporation, have arrays of *in situ* chemically synthesized, short DNA strands (55, 56). These may be used for parallel detection of mutations in a single gene or even several genes. The method involves synthesis on the chip of oligonucleotides capable of selectively hybridizing with DNA containing any possible single nucleotide alteration. Thus each chip must have 8 times as many oligonucleotides as there are nucleotides in the gene to detect all four possible nucleotides at each position on both strands (56). Fluorescent labeling allows detection of mutations. This method is thus equivalent to sequencing the gene, but it can be performed in one step. A current drawback of DNA chips is the extremely high costs for the chips and for the chip reader (52).

If the size of the oligonucleotide spots can be dramatically reduced and the resolution of the chip reader increased to match, it is possible to imagine a chip for screening coding regions of all genes. Such a chip must have many millions of spots. If this were possible, all genes could become candidates, and all mutations in all genes could be detected in a case-control or family study. The story would not end there, however, as the problem would then shift to finding the mutant needle in the resulting haystack of polymorphisms and sequence changes unrelated to the disease being studied. Bioinformaticians have some time to solve this problem since, for the near future at least, such a chip is only a fantasy. However, a project has already used this method to screen for sequence variations in 2.3 megabases of human DNA, already around 2.5% of all protein coding regions (57).

Single nucleotide polymorphisms

There are several projects currently in progress to identify many thousands of human single nucleotide polymorphisms (SNPs) (58). This knowledge may advance gene hunting methods in two ways. One is by providing a denser map of markers for linkage analysis. SNPs are the end of the line on the road to denser and denser marker maps. From minisatellites to microsatellites to SNPs, the length of marker polymorphic sites has drastically decreased and the number of markers has increased. A SNP map is thus potentially more informative for linkage analysis. In addition, because a SNP map would be so dense, it may be possible to use whole populations rather than families in association studies, allowing screening for disease-associated

genes when no large disease families are available (58). By using non-gel based methods, SNPs may also be more amenable to rapid genotyping (57).

Another way in which a SNP map may aid disease gene identification relies on the fact that most polymorphisms that alter protein sequences are SNPs. A SNP map will help make possible the simultaneous genotyping of SNPs from multiple candidate genes in order to directly identify the multiple polymorphic risk factors that contribute, sometimes interactively, to many diseases (58). Such common, multi-functional diseases include atherosclerosis, cancer, AD, asthma, schizophrenia, diabetes, glaucoma and age-related macular degeneration.

Conclusion

Applying an anti-Murphy's law to genetics, any method that could work, will work. Our arsenal of weapons for hunting disease genes is ever-expanding, and often more than one will do the job to catch a particular gene. Historically, we have jumped from method to method following the path of least resistance to a gene, more like lightning than a laser (44, 45). In the future, the plethora of techniques and information available will make the hunt faster yet perhaps even more jumpy. To reveal disease genes as quickly as possible, we should consider all possible methods and all available information at all times.

Acknowledgements

The authors thank Mr. Y. F. Leung and Ms. W.S. Lee for excellent technical assistance and Ms. Fion Lau for recruitment of study subjects. This work was supported by a Research Grants Committee earmarked grant for research, CUHK4282/98M, and a grant of the Industrial Support Fund AF/308/97, Hong Kong.

References

1. Aparicio SAJR. How to count…human genes. Nat Genet 2000; 25:129–30.
2. Todd JA. From genome to aetiology in a multifactorial disease, type 1 diabetes. Bioessays 1999; 21:164–74.
3. Altshuler D, Kruglyak L, Lander E. Genetic polymorphisms and disease. N Engl J Med 1998; 338:1626.
4. Todd JA. Interpretation of results from genetic studies of multifactorial diseases. Lancet 1999; 354 (Suppl. I):15–6.
5. Strachan T, Read AP. Human molecular genetics. Oxford: BIOS, 1996.
6. Kang J, Lemaire HG, Unterbeck A, Salbaum JM, Masters CL, Grzeschik KH, *et al*. The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. Nature 1987; 325:733–6.
7. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, *et al*. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 1991; 349:704–6.
8. Strathdee CA, Gavish H, Shannon WR, Buchwald M. Cloning of cDNAs for Fanconi's anaemia by functional complementation. Nature 1992; 356:763–7.
9. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, *et al*. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989; 245:1066–73.
10. Swiderski RE, Ying L, Cassell MD, Alward WLM, Stone EM, Sheffield VC. Expression pattern and in situ localization of the mouse homologue of the human *MYOC* (*GLC1A*) gene in adult brain. Mol Brain Res 1999; 68:64–72.
11. Karali A, Russell P, Stefani FH, Tamm ER. Localization of myocilin/trabecular meshwork-inducible glucocorticoid response protein in the human eye. Invest Ophthalmol Vis Sci 2000; 41:729–40.
12. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual, 2nd ed. Plainview, New York: Cold Spring Harbor Laboratory Press, 1989.
13. Liang P, Pardee AB. Differential display. Mol Biotechnol 1998; 10:261–7.
14. Strachan T, Read AP. PAX genes. Curr Opin Genet Dev 1994; 4:427–38.
15. Van Ommen GJB, Bakker E, Den Dunnen JT. The human genome project and the future of diagnostics, treatment, and prevention. Lancet 1999; 354 (Suppl. I):5–10.
16. Nature Genetics editorial. Freely associating. Nat Genet 1999; 22:1–2.
17. Allikmets R, Shroyer NF, Singh N, Seddon JM, Lewis RA, Bernstein P, *et al*. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. Science 1997; 277:1805–7.
18. Dryja T, Briggs CE, Berson E, Rosenfeld P, Abitbol M. ABCR gene and age-related macular degeneration. Science [online] 1998; http://sciencemag.org/cgi/content/full/279/5354/1107a.
19. Stone EM, Webster AR, Vandenburgh K, Streb LM, Hockey RR, Lotery AJ, *et al*. Allelic variation in ABCR associated with Stargardt disease but not age-related macular degeneration. Nat Genet 1998; 20:328–9.
20. Allikmets R, International ABCR Screening Consortium. Further evidence for an association of *ABCR* alleles with age-related macular degeneration. Am J Hum Genet 2000; 67:487–91.
21. Mak YT, Pang CP, Tomlinson B, Zhang J, Chan YS, Mak TWL, Masarei JRL. Mutations in the low-density lipoprotein receptor gene in Chinese familial hypercholesterolemia patients. Arterioscler Thromb Vasc Biol 1998; 18:1600–5.
22. Mak YT, J Zhang, YS Chan, TWL Mak, B Tomlinson, JRL Masarei, CP Pang. Possible common mutations in the low density lipoprotein receptor gene in Chinese. Hum Mutation 1998; Suppl 1:S320–3.
23. Pang CP. Molecular diagnostics of cardiovascular diseases. Clin Chem Lab Med 1998; 36:605–14.
24. Allingham RR, Wiggs JL, de la Paz MA, Vollrath D, Tallett DA, Broomer B, *et al.* Gln368STOP myocilin mutation in families with late-onset primary open- angle glaucoma. Invest Ophthalmol Vis Sci 1998; 39:2288–95.
25. Fingert JH, Héon E, Liebmann JM, Yamamoto T, Craig JE, Rait J, *et al.* Analysis of myocilin mutations in 1703 glaucoma patients from five different populations. Hum Mol Genet 1999; 8:899–905.
26. Lam DSC, Leung YF, Chua JKH, Baum L, Fan DSP, Choy KW, Pang CP. Truncations in the *TIGR* gene in individuals with and without primary open-angle glaucoma. Invest Ophthalmol Vis Sci 2000; 41:1386–91.
27. Yoon SJK, Kim HS, Moon JI, Lim JM, Joo CK. Mutations of the *TIGR/MYOC* gene in primary open-angle glaucoma in Korea. Am J Hum Genet 1999; 64:1775–8.
28. Pang CP, Leung YF, Chua JKH, Baum L, Fan DSP, Lam DSC.

Novel *TIGR* sequence alteration Val53Ala. Hum Mutat 1999; 15:122.
29. Roses AD, Saunders AM. APOE is a major susceptibility gene for Alzheimer's disease. Curr Opin Biotech 1994; 5:663–7.
30. Kang DE, Saitoh T, Chen X, Xia Y, Masliah E, Hansen LA, *et al.* Genetic association of the low-density lipoprotein receptor-related protein gene (LRP), an apolipoprotein E receptor, with late-onset Alzheimer's disease. Neurology 1997; 49:56–61.
31. Baum L, Chen L, Ng HK, Chan YS, Mak YT, Woo J, *et al*. Low density lipoprotein receptor related protein gene exon 3 polymorphism association with Alzheimer's disease in Chinese. Neurosci Lett 1998; 247:33–6.
32. Kamboh MI, Ferrell RE, DeKosky ST. Genetic association studies between Alzheimer's disease and two polymorphisms in the low-density lipoprotein receptor-related protein gene. Neurosci Lett 1998; 244:65–8.
33. Lambert JC, Wavrant-De Vrièze F, Amouyel P, Chartier-Harlin MC. Association at *LRP* gene locus with sporadic late-onset Alzheimer's disease. Lancet 1998; 351:1787.
34. Zhang Q, Cavanna J, Winkelman BR, Shine B, Bross W, Marz W, *et al.* Common genetic variants of lipoprotein lipase that relate to lipid transport in patients with premature coronary artery disease. Clin Genet 1995; 48:293–8.
35. Jemaa R, Fumeron F, Poirier. Lipoprotein lipase gene polymorphisms: association with myocardial infarction and lipoprotein levels, the ECTIM study. J Lipid Res 1995; 36:2141–6.
36. Galton DJ, Mattu R, Needham EW, Cavanna J. Identification of putative beneficial mutations for lipid transport. Z Gastroenterol 1996; 34 (Suppl):56–8.
37. Wittrup HH, Tybjaerg-Hansen A, Nordestgaard BG. Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease. A meta-analysis. Circulation 1999; 99:2901–7.
38. Baum L, Chen L, Masliah E, Chan YS, Ng HK, Pang CP. Lipoprotein lipase mutations and Alzheimer's disease. Am J Med Genet 1999; 88:136–9.
39. Baum L, Wiebusch H, Pang CP. Roles for lipoprotein lipase in Alzheimer's disease: an association study. Microsc Res Tech 2000; 50:291–6.
40. Http://www.marshmed.org/genetics
41. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, *et al*. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 1995; 375:754–60.
42. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K. A genetic linkage map of the human genome. Cell 1987; 51:319–37.
43. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, *et al*. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 1996; 380:152–4.
44. Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM, Kunkel LM. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature 1986; 323:646–50.
45. Goldberg NS, Collins FS. The hunt for the neurofibromatosis gene. Arch Dermatol 1991; 127:1705–7.
46. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 2000; 403:623–7.
47. Marcotte EM, Pellegrini M, Thompson MJ, Teates TO, Eisenberg D. A combined algorithm for genome-wide prediction of protein function. Nature 1999; 402:83–6.
48. Dryja TP. Gene-based approach to human gene-phenotype correlations. Proc Natl Acad Sci USA 1997; 94:12117–21.
49. Lander ES. Array of hope. Nat Genet supplement 1999; 21:3–4.
50. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. Nat Genet supplement 1999; 21:10–4.
51. Bowtell DDL. Options available – from start to finish – for obtaining expression data by microarray. Nat Genet Supplement 1999; 21:25–32.
52. Kurian KM, Watson CJ, Wyllie AH. DNA chip technology. J Pathol 1999; 187:267–71.
53. Khan J, Bittner ML, Chen Y, Meltzer PS, Trent JM. DNA microarray technology: the anticipated impact on the study of human disease. Biochim Biophys Acta 1999; 1423:M17–28.
54. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995; 270:484–7.
55. Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. Nat Genet supplement 1999; 21:20–4.
56. Hacia JG, Collins FS. Mutational analysis using oligonucleotide microarrays. J Med Genet 1999; 36:730–6.
57. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, *et al*. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 1998; 280:1077–82.
58. Brookes AJ. The essence of SNPs. Gene 1999; 234:177–86.

Received 27 July 2000; accepted 21 August 2000

Corresponding author: Professor Chi-Pui Pang, Department of Ophthalmology & Visual Sciences, Chinese University of Hong Kong, Hong Kong Eye Hospital, 3/F, 147K Argyle Street, Kowloon, Hong Kong, China
Tel.: +852 27623145, Fax: +852 27159490
E-mail: cppang@cuhk.edu.hk