REVIEW

Polymorphic sequence variants in medicine: a challenge and an opportunity

Shirley V Hodgson and Sanjay Popat

Shirley V Hodgson DM D(Obst) RCOG DCH FRCP, Professor of Cancer Genetics, St George's Hospital Medical School; Honorary Consultant Geneticist, St George's Hospital

Sanjay Popat MB PhD MRCP, Institute of Cancer Research, London

Clin Med 2003;**3**:260–4

ABSTRACT – The ability to detect polymorphic DNA sequence variants poses both opportunities for improved healthcare and concerns about the ethical challenges and confidentiality issues involved. Many polymorphisms confer only minor variability in disease susceptibility, which are difficult to detect and quantify, and therefore of minor value for improving healthcare. Important exceptions are high penetrance but uncommon disease susceptibility mutations, and those altering drug metabolism, knowledge of which should influence medical management. The development of Biobank initiatives to promote the detection and evaluation of important polymorphic variants highlights the need to ensure appropriate confidentiality guarantees and continuing debate about the ethical issues.

KEY WORDS: alleles, genetics, penetrance, pharmacogenetics, polymorphism

In 1892, Sir William Osler said: 'If it were not for the great variability among individuals, medicine might as well be a science and not an art'.[1] Over 100 years later this still remains true, as we understand more about the underlying genetic causes of this variability. Although human DNA sequences are 99.9% identical to each other, the remaining 0.1% of variation accounts for the observed phenotypic differences in individuals.[2] Sequencing the human genome has shown up numerous sites of variation in nucleotide sequence scattered throughout the genome, both in coding and in non-coding regions. These polymorphisms include the ubiquitous single nucleotide polymorphism (SNP).[1,3] Such sequence variants generally confer only low to moderate effects and have been termed low penetrance alleles, reflecting the low probability of their effect on phenotype. Study of this variation is giving new insights into a host of medical applications, ranging from the genetic predisposition to disease, to the streamlining of treatment strategies. The scope for this 'new genetics' is great. However, will it be fulfilled? What obstacles will need to be overcome? How will genetic profiling benefit public health? This article discusses the role of polymorphic sequence variation in medicine both today and in the near future.

Ethical issues

With the arrival of high throughput whole genome sequencing, there is increasing concern that the availability of information about an individual's genetic polymorphisms may become a threat to their privacy and security. Important issues of confidentiality and consent are being discussed in response to this concern. Is it necessary, though, to regard such information with suspicion? In the past, ethical, legal and social analyses of the consequences of scientific discoveries were usually relegated to third parties outside science and medicine. However, since its inception in 1990, the Human Genome Project has committed between 3–5% of its annual budget to precisely this cause (http://www.nhgri.nih.gov/ELSI/aboutels.html). Whether this process will appease its critics or result in legislative changes, will become apparent in due course. Certainly, it has stimulated ongoing debate.

Concerns about the use of genotype information include the possibility of litigation against doctors or companies for failure to use this information appropriately. There is already a precedent set by a recent case where a class action lawsuit was filed against Smith Kline Beecham because the company failed to warn that a vaccine against Lyme disease could cause arthritis in individuals with a certain genotype.[4] Another concern is the possibility of discrimination against people with particular genotypes for whom drug treatment would be more expensive. This could potentially promote ethnic discrimination.

The public perception of the impact of this 'new genetics' may, however, be unrealistic and overfearful:[5] the 1999 meeting of the American Society of Human Genetics reported that not a single case of discrimination by health insurers had been identified.[6] Despite this, the *fear* of genetic discrimination remains significant for many.[7] Open rational debate of these issues, engaging the public without sensationalism, can only help to defuse much of the public's fear of the threat of harm.

Much of the argument in favour of genetic

profiling stems from the putative benefits. However, many questions remain to be answered with respect to these issues:

- How important are low penetrance gene profiles in predicting health and disease?
- How useful are these data in tailoring treatment and management strategies for patients?
- Will knowledge of an individual's low-penetrance gene polymorphism profiles influence behaviour towards promoting better health, given that the risks conferred may be small?
- Will doctors be able to use this information to develop useful healthcare strategies in a cost-effective manner?[8]

Undoubtedly, it will be important to provide appropriate education for the public, as well as for healthcare professionals. Although medical practice is currently some way from routinely applying genetic technology in management paradigms, the potential clinical applications of sequence variation are only just being appreciated.

Pharmacogenetics

The most obviously 'useful' information about genetic polymorphisms for an individual is probably in the field of pharmacogenetics. Approximately 5% of all hospital admissions are due to adverse reactions to drugs, and they are the fourth most common cause of death in the USA after heart disease, cancer and stroke.[9] Much individual variability in drug response can be attributed to polymorphic variation in drug metabolising enzymes,[10] about which a great deal is known. There are many examples of this variability, including the relationship between suxamethonium response and serum cholinesterase levels, important because of prolonged anaesthesia secondary to suxamethonium sensitivity; haemolytic anaemia after treatment with primaquine in individuals with a specific G6PD polymorphism (there are over 400 variants of this X-linked gene); and peripheral neuropathy after treatment with isoniazid, due to variations in N-acetyl transferase activity.

The person-to-person variability in drug response is a major problem in clinical medicine, and again a large part of this variation can be explained by polymorphic sequence variants. The best known example of this is at the gene encoding the drug metabolising enzyme cytochrome P450 2D6 (*CYP2D6*). More than 70 allelic variants of *CYP2D6* have been detected (www.imm.ki.se/CYPalleles/cyp2d6.htm). These alleles may cause variation in enzyme activity, some resulting in an increased rate of drug metabolism, others in ultra slow metabolism due to a low level of enzymatic activity. They may also alter disease susceptibility, for example the susceptibility to cancer triggered by a carcinogen.[11]

Cytochrome P450 enzymes not only oxidise chemicals to inactive product,[12] but also activate prodrug to active compound; for example the conversion of codeine to morphine, and paracetamol to hepatotoxic NABQ1. In fact, *CYP2D6* is probably responsible for the metabolism of 25% of all commonly administered drugs.[10]

Key Points

Ethical issues surrounding the use of knowledge of polymorphic DNA variants must be debated and the interests of the public protected

Most such variants will have only minor effects on response to drugs or disease susceptibility, which will be difficult to quantify

Some polymorphic variants do alter disease susceptibility and drug toxicity, and are important for clinical management

Biobanks are being developed to evaluate the clinical effects of such polymorphic variants. Care must be taken to ensure confidentiality and safeguard against discrimination against individuals based upon such information

There is marked ethnic variation in many metabolic polymorphisms. Thus, the PM genotype of *CTPYP2D6* is found in 6% of Caucasians, 2% of American black individuals, and 1% of Oriental people.[12] The last group has other low-activity polymorphisms, and may need altered drug dosing, for example lower doses of antidepressants. Rapid metabolising alleles of *CYP2D6* occur in 20% of Ethiopians, 7% of Spaniards, and 1.5% of Scandinavians. Slow acetylators (alleles of *NAT2*) occur in 13% of Japanese and 60% of Caucasians. *GSTT* null individuals occur in 60% of Asians and 15% of Caucasians. Although this ethnic variation is useful for tailoring drug dosage, it is important that it does not lead to discrimination.[13,14] As yet there is no evidence that such practices occur.

The role of pharmacogenetics in cancer chemotherapy is becoming widely recognised. The need to optimise treatment strategies using cytotoxic chemotherapy is overwhelming. The potential toxicities from most cytotoxic agents and their use at maximally tolerated doses, together with the potential for resistant disease, make some cytotoxic treatments high-risk procedures, particularly for those patients who carry uncommon predisposition genotypes. Genotyping sequence variants at specific gene loci will not only aid in identifying individuals at increased risk of toxicity, but might also give information on treatment outcome, thereby helping to stratify patients into prognostic cohorts and reduce the unpredictability of current treatment paradigms.

An important example of the former is the variability in thiopurine methyl transferase (TPMT) activity. S-methylation of the thiopurines mercaptopurine, thioguanine and azathioprine is catalysed by TPMT. One in 300 Caucasians has no TPMT activity (homozygous null at the *TPMT* locus), and suffers severe reactions to ordinary doses of azathioprine, which can be fatal.[12] Furthermore, around 10% of Caucasians are heterozygous for deficiency of this enzyme, and may tolerate reduced doses, although full doses are toxic. Because of this, consideration is being given to genotyping all individuals prior to commencing azathioprine or 6-mercaptopurine.[15]

Similar variability in response to other drugs is likely to be widespread, and knowledge of pharmacogenetic variability is

likely to have an important effect on both new and conventional drug therapies, aiding identification of non- and toxic responders. The advent of this 'new genetics' will lead to the association of many polymorphic variations in drug metabolising enzymes with drug toxicity[15] and possibly prognosis. Although it is envisaged that molecular profiling will be used to reduce the number of toxic events, concern has been raised that this might lead to the identification of a small proportion of the population for whom drug therapy will be difficult or expensive; economic constraints encourage drug companies to concentrate their efforts on developing treatments for the majority only.

The development of pharmacogenetic technology will raise a number of issues related to commercialisation.[16–18] In the initial phase, pharmacogenomic drugs will be expensive.[19,20] Although individualisation of therapy will be improved, market incentives to invest resources in developing drugs for treatment of subpopulations will be low compared to developing drugs for more prevalent disorders.

Disease predisposition

The role of genetic variability in disease predisposition is becoming better understood, particularly in the susceptibility to cancer.[21,22] Although several high-penetrance alleles predisposing to cancer have been identified (including predisposition alleles at *APC, BRCA1, BRCA2*), these only account for a small proportion of the total familial risk of the particular malignancy. It is likely that a number of low-penetrance alleles exist that confer the greater proportion of disease susceptibility, after heterogeneity and environmental factors. This has been particularly well highlighted in the case of predisposition to colorectal cancer. Here, less than 5% of all colorectal cancers can be ascribed to dominant syndromes (principally familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC)) for which mutations have been shown to be causative. However, some progress has been made in identifying low-penetrance predisposition sequence variants. Over the past five years, two allelic variants have been identified in Ashkenazim that confer modestly increased risks: the I1307K *APC* allele[23] (OR 1.58), and the BLM^{Ash} allele (OR 2.34).[24]

Sequence polymorphism at genes encoding metabolic enzymes may well play a role in predisposition to malignancy. As with drugs, carcinogen metabolism is determined by enzymes, in this case those determining phase 1 and phase 2 pathways. Phase 1 reactions involve the attachment of substrates (generally in oxidative reactions) to chemicals, including carcinogens, to increase their reactivity, and are usually mediated by cytochrome P450 1A2 (CYP1A2). Phase 2 reactions increase the water solubility of xenobiotics by the further addition of substrates resulting in detoxification, breakdown and excretion, and are usually mediated by N-acetyl transferases (NAT1 and NAT2 acetylators) and GST (gluatathione-S transferase) enzymes, which detoxify mutagenic electrophiles.

Many association studies have been published examining the relationship between sequence polymorphism in genes encoding metabolic enzymes and cancer risk. Of the CYP enzymes, CYP1B1 catalyses the conversion of oestrogen to 4-OH oestradiol and may alter the risk of breast cancer; GST enzymes may detoxify activated carcinogen metabolites in tobacco smoke; *GSTM1* null individuals may have an increased colorectal cancer risk.[25] The NAT enzymes can deactivate aromatic amines in tobacco smoke, reducing lung cancer risk, but they also activate heterocyclic amines in cooked meats, which may increase colorectal cancer risk. NAT1-mediated acetylation may be fast or slow. The NAT1*10 polymorphism results in a fast acetylation phenotype, and may be associated with an increased risk of colorectal cancer[26] and lower age at diagnosis of HNPCC. By contrast, slow acetylators may have a greater risk of lung cancer.[27]

As a whole, however, results of studies relating individual polymorphisms to disease outcomes have been disappointing. Too often, strong associations have been reported only to be discounted later in better designed studies using larger sample sizes. Many factors have contributed to the plethora of studies claiming a putative causative association only to be subsequently shown to be insignificant. These include poor study design, using too small a sample size, the use of inappropriate controls, and publication bias in the preferential reporting of studies with a positive effect. Furthermore, a failure to collect environmental data may have contributed to this inconsistency.

Thus, despite many published association studies of polymorphic sequence variants and cancer risk, there still remains conflicting evidence about the magnitude of the effects conferred by some alleles. In the case of colorectal cancer risk, a meta-analysis of polymorphism studies demonstrated that the only significant effects seen in more than one study were at *APC* I1307K (OR 1.58), the *HRAS1*-VNTR (OR 2.50), and the Val667Val polymorphism of the 5,10-methylenetetrahydrofolate reductase enzyme (encoded by *MTHFR*) (OR 0.76).[28]

In addition to the genetic predisposition to cancer, the significance of genetic factors in the aetiology of common multifactorial disorders is becoming better understood. An example of this is in the case of inflammatory bowel disease. Here, a proportion of the familial risk of Crohn's disease is conferred by predisposition alleles at *NOD2*.[29–31] Another example of how sequence variation influences common disease is the relationship between cardiovascular disease and hypercholesterolaemia, underpinned by sequence variation at the apolipoprotein E low density lipoprotein E gene (*APOE LDL*). Alleles at this locus cause variation in serum cholesterol levels (the E2 variant lowers and the E4 raises the level in approximately 10%). Apolipoprotein B (*APOB*) polymorphisms also have an effect, and a defect of the LDL receptor gene is related to increased cholesterol concentration.[32,33]

The recognition that the genetic contribution to most common diseases is mainly attributable to a number of predisposition alleles (many possibly polymorphic) has important clinical implications. By contrast to the rare high penetrance mutations (as exemplified by mutations in *APC* causing FAP), common polymorphic sequence variants are only likely to confer small relative risks. Thus, individual polymorphisms are unlikely to be useful for screening purposes, since relative risks

of at least ten are thought to be required for effective screening.[34] However, it may be possible to identify polymorphisms that individually contribute only weakly to risk of disease but in combination confer much larger risks. Profiling of these polymorphisms in combination might then be used to predict disease with greater efficiency.

In addition to the targeted screening of at-risk individuals, polymorphism profiling will have the potential to stratify individuals for entry into prevention protocols (eg chemoprevention strategies or life-style modification). More interestingly, these sequence variants are now giving tantalising insights into the relationship between genotype and phenotype.[35]

Biobanks

Due to the small relative risks conferred by each polymorphic sequence variant, research to assess the medical implications of genetic polymorphisms will require generic genotype/phenotype correlation studies of large population groups. This, in turn, will require facilities for the storage of large sets of DNA samples under carefully controlled conditions of confidentiality and security with appropriate informed and ethical consent. Development of such facilities will involve validation of susceptibility screening tests, taking into account economic considerations and consent issues, and delivery of such a service will require consideration of insurance, confidentiality and public awareness issues. The essential components of ethical requirements for genetic databases include:[36]

- open-ended, broad informed consent
- a method for interlinking databases with appropriate anonymisation procedures
- an established method for responding to requests for information derived from the databases
- confidentiality and security safeguards
- sound data access and ownership policies
- a method for dissemination of results
- establishment of research ethics supervision.

For the future, it will be important to establish methods for efficient re-contact of patients without breaking confidentiality, should this be clinically indicated or further consent required. It is possible that this could be made easier if DNA and patient contact details were held by an independent third party.

Such a population biomedical DNA sample collection currently being developed is the Wellcome Trust, MRC and DH bank, comprising 500,000 volunteers aged 45–64 years, linking NHS clinical files and questionnaire-based health/lifestyle and environmental data with genetic polymorphism profiles. The ethical issues surrounding the development of this database have been considered by the House of Lords Select Committee on Science and Technology and the Human Genetics Commission enquiry. There is also a cross-industry pharmaceutical group promoting standards for research (http://www.diahome.org/docs/Education/Training_index.cfm), and a group comprising 13 pharmaceutical companies, five academic centres and the Wellcome Trust who have formed an SNP consortium holding details of all known SNPs (http://snp.cshl.org/news/consortium_pr.shtml). These large collaborative enterprises should militate against monopolies being set up, particularly because of the huge financial investment necessary to establish and analyse the very large amounts of data required to evaluate the small effects of low-penetrance genetic variants.

Future perspectives

Now that the human genome has been sequenced, the cataloguing of common polymorphisms has begun in earnest. In the future, screening for genetic profiles is likely to be extensively automated and performed in large laboratories. However, such laboratories cannot be expected to deliver the results to patients. Interpretation of individuals' results may well not be by geneticists, and is likely to be performed increasingly in primary care as well as by specialists and doctors from other medical disciplines. It is important that the process of genotyping and delivery and interpretation of results is managed carefully, and that issues of confidentiality are taken into account.[37] Patients' genetic information may appear on the files of several healthcare providers. This will increase the risk of misuse and misinterpretation of data as well as increasing the risk that information will become available to employers, insurers and others.[38] Possibly such information should be personalised and held by the patient only. Once sharing of genetic information has begun, it will be increasingly difficult to stop. Entry of genetic information onto databases should be considered carefully: once information has been deposited it will prove extremely difficult to remove and relatively easy to disclose, unless stringent safeguards are implemented.[7]

To date, sequence information generated by the Human Genome Project has armed medicine with growing understanding of the role of sequence variation in the risk of common disease and for the development of new pharmacological agents and individualised intervention. This understanding will increase vastly in the near future, providing both a major challenge and a significant opportunity for the enhancement of future healthcare.

Acknowledgement

Sanjay Popat is in receipt of a Clinician Scientist Fellowship from the Department of Health.

References

1 Roses AD. Pharmacogenetics and the practice of medicine. *Nature* 2000;**405**:857–65.
2 Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997;**278**:1580–1.
3 Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, *et al* (the International SNP Map Working Group). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001;**409**:822–3.
4 Otto A. Lyme vaccine linked to autoimmune arthritis. *Pharmacy Today* 2001;**7**:10.

5 Nowlan W. A rational view of insurance and genetic discrimination. *Science* 2002;**297**:195–6.
6 Stephenson J. Genetic test information fears unfounded. *JAMA* 1999;**282**:2197–8.
7 Rothenberg KH, Terry SF. Human genetics. Before it's too late – addressing fear of genetic information. *Science* 2002;**297**:196–7.
8 Rees J. Complex disease and the new clinical sciences. *Science* 2002; **296**:698–701.
9 Caldwell J. The pharmacogenetic basis of adverse drug reactions. *Int J Pharmaceut Med* 2001;**15**:83–4.
10 Meyer UA. Pharmacogenetics and adverse drug reactions. *Lancet* 2000; **356**:1667–71.
11 Roddam PL, Rollinson S, Kane E, Roman E *et al*. Poor metabolizers at the cytochrome P450 2D6 and 2C19 loci are at increased risk of developing adult acute leukaemia. *Pharmacogenetics* 2000;**10**:605–15.
12 Wilson JF, Weale ME, Smith AC, Gratrix F *et al*. Population genetic structure of variable drug response. *Nature Genet* 2001;**29**:265–9.
13 Yancy CW, Fowler MB, Colucci WS, Gilbert EM *et al*. U.S. Carvedilol Heart Failure Study Group. Race and the response to adrenergic blockade with carvedilol in patients with chronic heart failure. *N Engl J Med* 2001;**344**:1358–65.
14 Exner DV, Dries DL, Domanski MJ, Cohn JN. Lesser response to angiotensin-converting-enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction. *N Engl J Med* 2001;**344**:1351–7.
15 McLeod HL, Evans WE. Pharmacogenomics: unlocking the human genome for better drug therapy. *Ann Rev Pharmacol Toxicol* 2001;**41**: 101–21.
16 Holgate ST, Lunnon MW. Pharmacogenetics Working Party: introduction. *Int J Pharmaceut Med* 2001;**15**:1563–6.
17 Shah RR. Regulatory aspects of integration of pharmacogenetics into drug development. *Int J Pharmaceut Med* 2001;**15**:67–69.
18 Rothstein MA, Epps PG. Ethical and legal implications of pharmacogenomics. *Nat Rev Genet* 2001;**2**:228–31.
19 Holmer AF. The pharmaceutical industry – to whom is it accountable? *N Engl J Med* 2000; **343**:1415.
20 Mrazek MF, Mossialos E. Increasing demand while decreasing costs of genetic medicines *Lancet* 2000; **365**:1784–5.
21 Houlston RS, Tomlinson IP. Detecting low penetrance genes in cancer: the way ahead. *J Med Genet* 2000;**37**:161–7.
22 Peto J, Houlston RS. Genetics and the common cancers. *Eur J Cancer* 2001;**37** (Suppl 8):S88–96.
23 Woodage T, King SM, Wacholder S, Hartge P *et al*. The APCI1307K allele and cancer risk in a community-based study of Ashkenazi Jews. *Nature Genet* 1998;**20**:62–5.
24 Gruber SB, Ellis NA, Rennert G, Offit K *et al*. BLM heterozygosity and the risk of colorectal cancer. *Science* 2002;**297**:2013.
25 Cotton SC, Sharp L, Little J, Brockton N. Glutathione S-transferase polymorphisms and colorectal cancer: a HuGE review. *Am J Epidemiol* 2000;**151**:7–32.
26 Moisio AL, Sistonen P, Mecklin JP, Jarvinen H, Peltomaki P. Genetic polymorphisms in carcinogen metabolism and their association to hereditary nonpolyposis colon cancer. *Gastroenterology* 1998;**115**: 1387–94.
27 Mucci LA, Wedren S, Tamimi RM, Trichopoulos D, Adami HO.The role of gene-environment interaction in the aetiology of human cancer: examples from cancers of the large bowel, lung and breast. *J Intern Med* 2001;**249**:477–93.
28 Houlston RS, Tomlinson IP. Polymorphisms and colorectal tumour risk. *Gastroenterology* 2001;**121**:282–301.
29 Todd JA. Human genetics. Tackling common disease. *Nature* 2001;**411**: 537–9.
30 Hampe J, Cuthbert A, Croucher PJ, Mirza MM *et al*. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001;**357:**1925–8.
31 Ogura Y, Bonen DK, Inohara N, Nicolae DL *et al*. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603–6.
32 Day IN, Wilson DI. Science, medicine, and the future: Genetics and cardiovascular risk. *Br Med J* 2001;**323**:1409–12.
33 Sijbrands EJ, Westendorp RG, Defesche JC, de Meier PH, et al. Mortality over two centuries in large pedigree with familial hypercholesterolaemia: family tree mortality study. *Br Med J* 2001;**322**:1019–23.
34 Wald NJ, Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? *Br Med J* 1999;**319**:1562–5.
35 Abreu MT, Taylor KD, Lin YC, Hang T *et al*. Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology* 2002;**123**:679–88.
36 Clarke A, English V, Harris H, Wells F. Ethical considerations. *Int J Pharmaceut Med* 2001;**15:**89–94.
37 Reilly PR, Boshar MF, Holtzman SH. Ethical issues in genetic research: disclosure and informed consent. *Nature Genet* 1997;**15**:16–20.