Clinical review

Science, medicine, and the future

DNA microarrays in medical practice

Timothy J Aitman

The completion of the draft sequence of the human genome[1 2] has raised public awareness of "genomics" and of the ways in which the emerging technologies of the genomics "revolution" will have direct applications to patient care.[3] DNA microarrays, or biochips, are prominent among these new technologies. The past five years have seen a dramatic rise in the use of DNA microarrays for biomedical research, in some cases with immediate applicability to clinical practice. The uses of microarrays for gene expression profiling, genotyping, mutation detection, and gene discovery are leading to remarkable insights into the function of thousands of genes previously known only by their gene sequence. This review discusses ways in which microarrays have started to affect clinical practice and research and how this role will develop in the next five to 10 years.

Anticipated developments

DNA microarrays offer unprecedented opportunities for analysing gene expression, understanding gene function, and detecting gene dysfunction

Microarray expression profiles will define new prognostic subgroups in cancer and other diseases

The potential clinical benefit brought by microarrays will have to be rigorously appraised and weighed against their costs before their introduction into routine clinical use

The increased diagnostic and prognostic information provided by microarrays should assure their entry into clinical practice in specialist centres within 3-5 years and in most large hospitals within 5-10 years

Physiological Genomics and Medicine Group, MRC Clinical Sciences Centre, and Imperial College Genetics and Genomics Research Institute, Hammersmith Hospital, London W12 0NN

Timothy J Aitman
MRC clinical scientist

t.aitman@csc.mrc.ac.uk

BMJ 2001;323:611–5

Methods

The data for this review reflect my personal views from 10 years of research in molecular medicine and five years on DNA microarrays. My views have developed over this period from discussions with colleagues in my own laboratory and at basic science and biotechnology meetings and from regular reading of *Nature Genetics*, *Nature*, *Science*, and other journals.

What are DNA microarrays?

For the past 25 years, the standard techniques used to detect specific sequences of DNA or RNA in the laboratory have depended on the use of a DNA probe labelled with, for example, a radioactive isotope or a fluorescent tag. The probe is complementary in sequence to the fragment of DNA or RNA to be detected, and so it hybridises (or sticks) selectively, by Watson-Crick base pairing, to the correct fragment of DNA or RNA if that fragment is present in the test sample.

The principle of DNA microarrays is that technological advances have made it possible to miniaturise this DNA probe detection method. Instead of detecting and studying one gene at a time, microarrays allow thousands or tens of thousands of specific DNA or RNA sequences to be detected simultaneously on a small glass or silica slide only 1-2 cm square. Although the principles of specific DNA and RNA detection have remained unchanged, the greatly increased scale on which this can be achieved with DNA microarrays has made it possible to tackle qualitatively different questions in biology and medicine.

Microarray manufacture

Two main methods are used to make microarrays. In the first, DNA is spotted onto a glass slide (fig 1); in the second, oligonucleotides of 15-30 nucleic acid base pairs are synthesised on to a silica slide by a process known as photolithography (fig 2).

The technology for producing spotted microarrays (fig 1) is widely accessible, and spotted arrays are now produced in scores of biomedical research institutions. Production and use of custom arrays are now within the grasp of any molecular biology laboratory, and, once the DNA has been prepared, the costs are low. The disadvantage of this technique is that consistency of spotting and reliable annotation of the DNA on the microarray are hard to achieve.

The photolithography technique, initially borrowed from the semiconductor industry, is limited to a small number of manufacturers, of whom the best known is Affymetrix, a California based biotechnology firm that has cornered a large share of the microarray

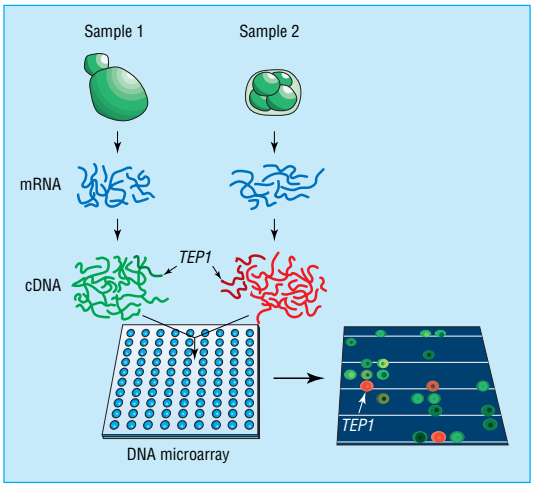

Fig 1 Gene expression analysis in two tissue samples using spotted DNA microarray. RNA extracted from samples 1 and 2 is labelled with red or green fluorescent dyes. The dye labelled RNA populations are mixed and hybridised to the microarray, on which has been spotted cDNA from thousands of genes, each spot representing one gene. The RNA from each sample hybridises to each spot in proportion to the level of expression of that gene in the sample. After hybridisation, the red and green fluorescent signal from each spot is determined, and the ratio of red to green reflects the relative expression of each gene in the two samples. For example, the gene TEP1 is shown to be expressed at a higher level in sample 2 than in sample 1. (Adapted with permission from Brown and Botstein[4])

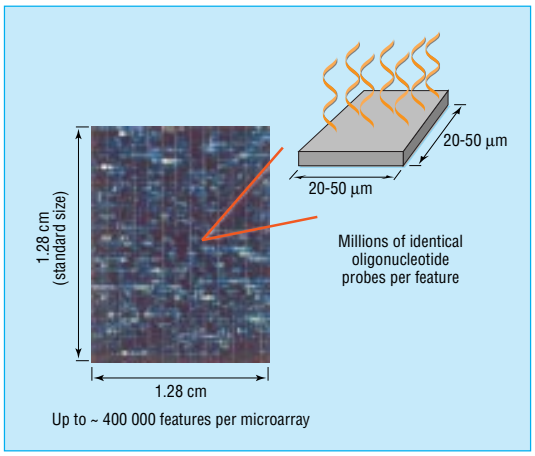

Fig 2 Gene expression analysis using oligonucleotide microarray. Up to half a million distinct oligonucleotides are synthesised on the microarray by photolithography and act as probes in individual "features" on the microarray surface. About 30 distinct oligonucleotides, printed as individual features, represent the partial sequence of one gene. Fluorescent labelled cDNA derived from a single test sample is hybridised to the microarray, allowing the expression level of up to 15 000 genes to be measured in the test sample. (Illustration courtesy of Affymetrix, Santa Clara, CA, USA)

market. The very high spot (or "feature") density possible with photolithography (fig 2) means that the expression of up to 15 000 genes can be measured on a single slide (the signal for each being derived from about 30 spots). Although these commercial arrays have high unit costs, the consistency between arrays is assured by the production process, and microarrays representing most of the genes in the human genome and a range of other genomes can be bought ready for use without the need for development by the user.

Major applications of microarrays

The major applications of microarrays fall into three groups:

Gene expression profiling—RNA extracted from a complex sample (such as body tissues or fluids or bacterial isolates) is applied to the microarray. The result reveals the level of expression of tens of thousands of genes, effectively all the genes in the genome, in that complex sample. This result is known as a gene expression "profile" or "signature."

Genotyping—Genomic DNA, extracted from an individual's blood or saliva, is amplified by the polymerase chain reaction and applied to the microarray. The genotype for hundreds or thousands of genetic markers across the genome can be determined in a single hybridisation. This approach has considerable potential in risk assessment, both in research and clinical practice.

DNA sequencing—DNA extracted from an individual's blood is amplified and applied to specific "re-sequencing" microarrays. Thousands of base pairs of DNA can be screened on a single microarray for mutations in specific genes whose normal sequence is already known. This greatly increases the scope for precise molecular diagnosis in single gene and genetically complex diseases.

Expression profiling in clinical practice

The ability to measure the expression level of thousands of genes in any tissue sample allows exploration of gene function on a scale previously impossible. Through global analysis of gene expression, the function of genes previously identified only by their DNA sequence is now being discovered almost as a matter of routine. The identification of new metabolic pathways and pathogenetic mechanisms, new indicators of disease prognosis, and new drug targets are likely to be realised in several diseases in the next three to five years and will start to enter clinical practice with new treatment strategies within this time frame.

For individual patients, more precise diagnosis and risk assessment based on expression profiles are already achievable for certain conditions, leading to more accurate determination of prognosis and more individually tailored treatment. Oncology has taken the lead in this area, although it is by no means the only discipline to have realised the power of the technology. Expression profiles in acute leukaemia, B cell lymphoma, melanoma, and breast cancer have led to new methods for disease staging and classification, based on knowledge of the expressed genes in individual tumours or cell populations.

Oncology

An example of this diagnostic approach is in acute leukaemia, in which expression profiles of 6817 genes were used to distinguish between acute lymphoid leukaemia and acute myeloid leukaemia (fig 3). Independent of any histological or histochemical diagnosis, the results correctly classified 36 out of 38 "unknown" leukaemia samples derived from either bone marrow or peripheral blood.[5] The remaining two samples could not be assigned to either acute lymphoid or acute myeloid leukaemia. However, the technique successfully divided acute lymphoid leukaemia into T cell derived or B cell derived leukaemia and raised new and intriguing

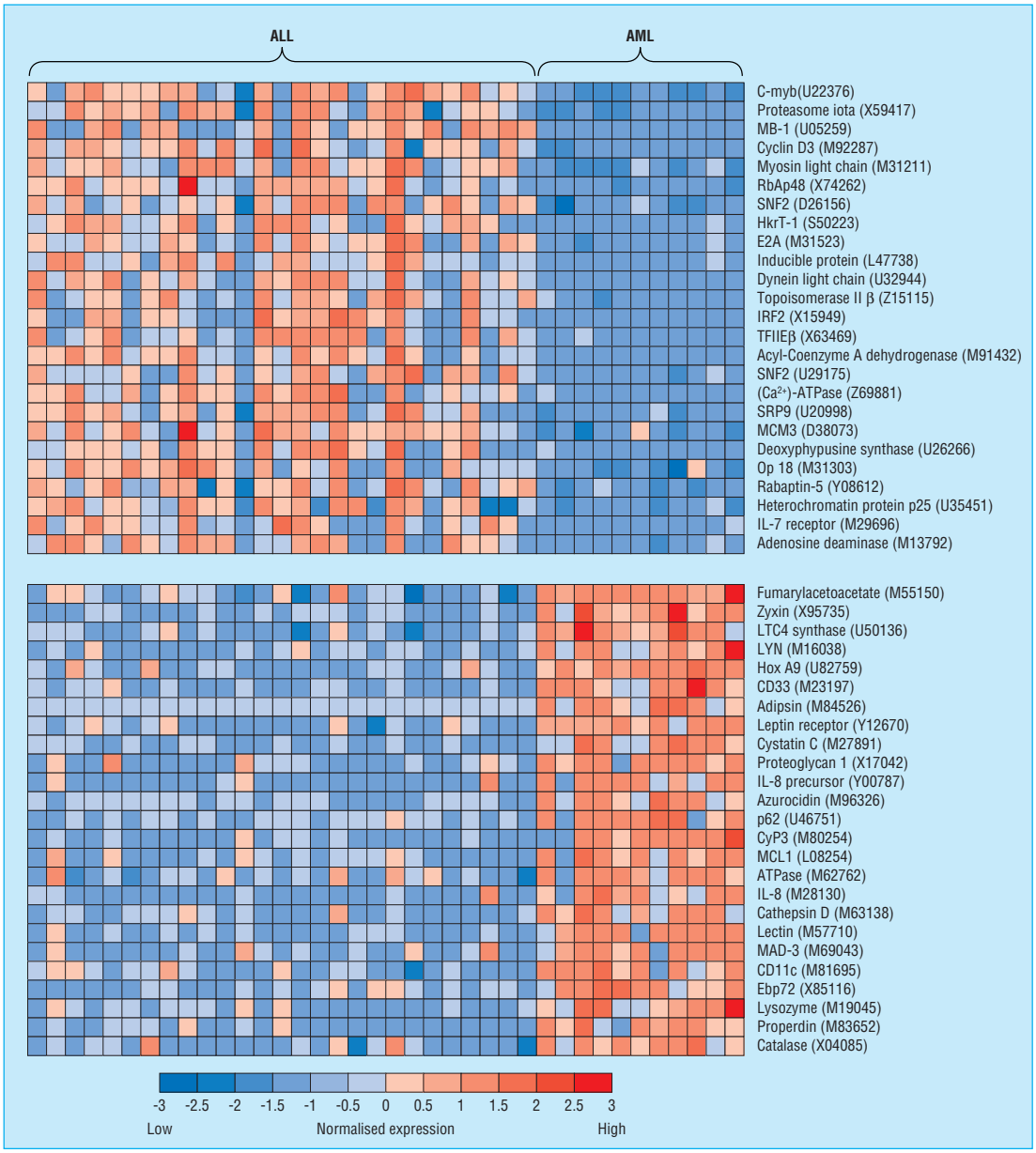

Fig 3 Genes distinguishing acute lymphoid leukaemia (ALL) from acute myeloid leukaemia (AML). The figure shows the 50 genes with the greatest distinction between ALL and AML, the top panel showing genes more highly expressed in ALL and the bottom panel showing genes more highly expressed in AML. Each row corresponds to a gene, with the columns corresponding to expression levels of these genes in different patient samples. Expression levels greater than the mean are red, and those below the mean are blue. The scale indicates standard deviations above or below the mean. (Adapted with permission from Golub et al[6])

insights into the role of the 50 most predictive genes in disease pathogenesis.

A limitation of this study was that the results did not allow new prognostic groups to be defined. This was not the case for a similar study of diffuse large B cell lymphoma.[6] This condition is the commonest subtype of non-Hodgkin's lymphoma and is notoriously difficult to classify prognostically on morphological or clinical grounds. It is therefore of considerable clinical relevance that the gene expression analysis defined new prognostic subgroups in diffuse large B cell lymphoma, one with five year survival of 80% and one with 40% survival (fig 4).

Similar studies have been made of metastatic and non-metastatic melanoma,[7 8] breast cancer,[9] and colorectal cancer.[10] It seems likely that major new insights will be derived for a wide range of cancers, leading to the prospect of better targeted treatment and, in the longer term, new treatments based on improved understanding of the molecular pathogenesis of these diseases.

Infectious disease

Microarrays that detect gene sequences in the genomes of *Mycobacterium tuberculosis*, HIV, and other pathogens have been developed[11 12] with the aim of providing a diagnostic tool that detects expression of antibiotic resistance genes or specifies viral subtypes. A major advantage is that these tests can be undertaken rapidly (in less than 24 hours) without the need for bacterial or viral cultures. If such tests are brought into clinical practice they will lead to earlier, more targeted treatment based on antibiotic or antiviral sensitivities

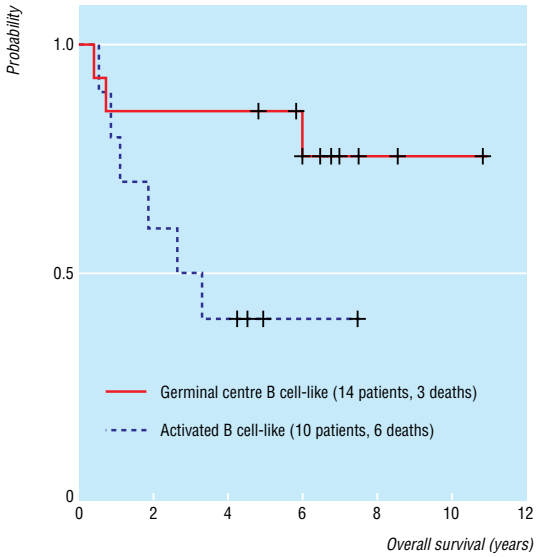

Fig 4 Kaplan-Meier plots of overall survival of patients with diffuse large B cell lymphoma at low clinical risk (international prognostic index score 0-2) who are grouped on the basis of their gene expression profiles. Patients' tumours in the upper plot show a germinal centre-like gene expression profile while those in the lower plot show an activated B cell-like profile. (Adapted with permission from Alizadeh et al[6])

and will be particularly valuable for organisms such as *M tuberculosis* and HIV, for which sensitivity profiles can presently be determined only after lengthy analysis by other methods.

Microarrays for genotyping and re-sequencing

The human genome sequencing projects have, over the past two years, provided more than two million single nucleotide polymorphisms (SNPs) as genetic markers.[2 13] These resources, coupled with improved methods for genetic analysis, mean that systematic identification of the genetic determinants of common, genetically complex disorders is now being realised.[14] As relevant genes are identified, the mutations and polymorphisms that underlie susceptibility to a range of common diseases will become direct predictors of susceptibility. Likely examples include diabetes, hypertension, coronary heart disease, asthma, inflammatory bowel disease, cancer susceptibility, and susceptibility to adverse and favourable drug responsiveness.

Genotyping arrays, or "SNP chips," offer one option for use of these new genetic markers of disease and drug responsiveness. These arrays, capable of genotyping up to 2000 polymorphisms in a single hybridisation,[15 16] will allow the construction of an individual's "genetic fingerprint," that can be related to his or her risk of developing single gene disorders or more common complex diseases.

Clinical pharmacology

The approach is already being applied in clinical pharmacology, with the use of genetic markers across the genome to predict individual drug responsiveness. This new science, pharmacogenomics, aims to predict both good and adverse clinical responses to individual drugs.[17] Since many drugs are metabolised by the cytochrome P450 pathway, polymorphisms within the P450 genes are good candidates for determining some of these responses. P450 genotyping microarrays are already available from Affymetrix and are probably forerunners of arrays that will determine much larger numbers of pharmacologically relevant genotypes.

Mutation detection and genotyping for specific genes

Different types of microarray, used for re-sequencing of known genes, allow massive stretches of an individual's DNA to be screened for mutations. Re-sequencing arrays, based on a similar design, can already be used to screen entire genes for pathogenic mutations such as in the cancer susceptibility genes BRCA1 and p53.[18] Gene specific arrays will also be of value for detecting polymorphisms that induce specific diseases such as factor V Leiden in deep vein thrombosis, apolipoprotein E4 in Alzheimer's disease, and the recently identified susceptibility gene for Crohn's disease, NOD2.[14] Some arrays for this type of clinical use are already under commercial development.

Microarrays and identification of new disease genes

Three recent studies have combined the use of DNA microarrays with genetic linkage analysis, leading to identification of the disease genes under study. In the first report microarrays were used to identify genes that were differentially expressed between a rat strain with insulin resistance and a normal, insulin sensitive control strain.[19] One of the genes identified in this microarray study, Cd36 or fatty acid translocase, mapped to a chromosomal location previously shown to contain an insulin resistance gene.[20] Defects in the gene were then shown to result in glucose intolerance and defective fatty acid metabolism, both in this rat strain[19 21] and in humans.[22] In the other two studies, microarray based expression profiles combined with linkage analysis led to identification of ABC transporters as disease genes in the autosomal recessive conditions Tangier disease[23] and sitosterolaemia.[24]

As yet, it is too soon to know if the combined microarray and linkage approach will be more generally applicable to identification of monogenic or complex trait genes, but adoption of the approach in a growing number of gene mapping laboratories will answer this question in the next two to three years.

Progress towards clinical practice

Most microarray based tests are still in the development stage, though substantial progress towards commercialisation has occurred in some cases. The increased clinical information provided by microarrays should assure their entry into routine clinical practice within the next three to five years, although the added costs will have to be justified by the clinical benefit. Like any new diagnostic tool, microarrays will have to be rigorously appraised for sensitivity, specificity, and predictive value and, in some cases, licensing by regulatory bodies. The high costs of microarray based tests will inevitably limit the speed with which they are introduced into clinical practice and initially restrict

their use to teaching centres and specialised units. However, given the huge potential gain in clinically relevant information for individual patients and their diseases, the technology is likely to reach most large hospitals within the next 10 years.

I thank Professors McCarthy, Scott, and Bangham and Drs Causton, Withers, and Rowley for their constructive comments in the preparation of the manuscript and Anne Glazier for assistance with the figures.

Funding: I have received intramural funding from the MRC Clinical Science Centre.

Competing interests: I have received research funds from Affymetrix.

Educational resources

Relevant websites

- http://cmgm.stanford.edu/pbrown/array.html—The website of the Pat Brown laboratory, where custom arrays were first built. Gives a simple overview of the microarray process and a wealth of protocols for custom array research
- www.gene-chips.com—A comprehensive information source for presently available microarray platforms

Key reviews

Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature* 2000;405:827-36—An in depth description of how microarrays are giving insights into gene function

The chipping forecast. *Nature Genet* 1999;21(suppl):1-60—A series of reviews dealing comprehensively with microarray theory, manufacture, and applications

1 International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.
2 Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304-51.
3 Collins FS. Shattuck lecture—medical and societal consequences of the human genome project. *N Engl J Med* 1999;341:28-37.
4 Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nature Genet* 1999;21(suppl):33-7.
5 Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531-7.
6 Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503-11.
7 Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 2000;406:532-5.
8 Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Handrix M, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 2000;406:536-40.
9 Perou CM, Serlie T, Eisen BB, Van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747-52.
10 St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, et al. Genes expressed in human tumor endothelium. *Science* 2000;289:1197-202.
11 Kozal MJ, Shah N, Shen N, Yang R, Fucini R, Merigan TC, et al. Extensive polymorphisms observed in the HIV-1 cladeB protease gene using high-density oligonucleotide arrays. *Nature Med* 1996;2:753-9.
12 Gingeras TR, Ghandour G, Wang E, Berno A, Small PM, Drobniewski F, et al. Simultaneous genotyping and species identification using hybridization pattern recognition of generic mycobacterium DNA arrays. *Genome Res* 1998;8:435-48.
13 The SNP Consortium. http://snp.cshl.org/ (accessed 30 Jul 2001).
14 Todd JA. Tackling common disease. *Nature* 2001;411:537-9.
15 Wang DG, Fan J-B, Siao C-J, Berno A, Young P, Sapolsky P, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998;280:1077-82.
16 Fan J-B, Chen X, Halushka MK, Berno A, Huang X, Ryder T, et al. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* 2000;10:853-60.
17 Roses AD. Pharmacogenetics and the practice of medicine. *Nature* 2000;405:857-65.
18 To affinity . . . and beyond! [editorial]. *Nature Genet* 1996;14:367-70.
19 Aitman TJ, Glazier AM, Wallace CA, Cooper LD, Norsworthy PJ, Wahid FN, et al. Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet* 1999;21:76-83.
20 Aitman TJ, Gotoda T, Evans AL, Imrie H, Heath K, Trembling PM, et al. Quantitative trait loci for cellular defects in glucose and fatty acid metabolism in hypertensive rats. *Nature Genet* 1997;16:197-201.
21 Pravenec M, Landa V, Zidek V, Musilova A, Kren V, Kazdova L, et al. Transgenic rescue of defective Cd36 ameliorates insulin resistance in spontaneously hypertensive rats. *Nature Genet* 2001;27:156-8.
22 Miyaoka K, Kuwasako T, Hirano K, Nozaki S, Yamashita S, Matsuzawa Y. CD36 deficiency associated with insulin resistance. *Lancet* 2001;357:686-7.
23 Lawn RM, Wade DP, Garvin MR, Wang X, Schwartz K, Porter JG, et al. The Tangier disease gene product ABC1 controls the cellular apolipoprotein-mediated lipid removal pathway. *J Clin Invest* 1999;104:R25-31.
24 Berge KE, Tian H, Graf GA, Yu L, Grishin NV, Schultz J, et al. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science* 2000;290:1771-5.

Medical mishaps

Mistaken identity

The *BMJ* has recently focused on the underreporting of adverse events and near misses in the NHS. The clinical review article entitled "Reporting and preventing medical mishaps" (*BMJ* 2000; 320:759-63) stressed the importance of reporting incidents, as near misses offer numerous benefits over adverse events. In the light of this and recent events we would like to draw attention to the following.

Working in a busy plastic surgery unit, we have noticed that a variety of water, saline, and lignocaine ampoules all look virtually identical in terms of bottle design and labelling (see figure). Our concerns about the potential for a mistake being made were confirmed recently during administration of a local anaesthetic wrist block, when a vial of water was mistaken for lignocaine.

Fortunately, in this case mistaking water for lignocaine had no serious consequences other than requiring re-injection with local anaesthetic. However, the potential consequences of an accidental intravenous injection of lignocaine (such as while preparing an intravenous infusion) are obvious.

We have contacted the manufacturers (B Braun), who are aware of the similarity in labelling and packages of these products. They state that there is nothing they can do about it and referred us to the Medicines Control Agency, which is ultimately responsible for overseeing product labelling.

When we contacted the agency it said that it was aware of the problem. Hopefully, in reporting this near miss, we will have alerted others to this potential danger. The Medicines Control Agency may even look further into its labelling guidelines if others have had similar experiences.

Charles Nduka *specialist registrar in plastic surgery*

Daniel Leff *house officer in plastic surgery, Colchester General Hospital, Essex*