

# On the trail of SNPs

Extracting useful data from the human genome sequence is a major challenge. Lisa Melton examines the early steps towards personal genotyping.

GLAXOSMITHKLINE

One of the first approaches to extracting medical and biological value from the human genome has been the study and exploitation of single-nucleotide polymorphisms (SNPs; pronounced 'snips'), the most abundant type of genetic variation. SNP genotyping promises to reveal why some people are more susceptible to diseases such as cancer or diabetes, and what predisposes others to suffer adverse reactions to drugs. The implications for drug discovery, clinical drug trials, biomedical research, diagnostics and even forensics are considerable (see 'Crime busters', below, and 'Risk assessment', page 919).

To enable this promise to be fulfilled, faster and cheaper ways of screening individuals for SNPs are being developed. For start-up biotech firms and large pharmaceutical companies alike, the search for these genetic variations is now a serious business.

On average, two unrelated people differ at about 1 base in every 1,000 of the 3 billion or so bases in their genome, so any individual will have about 3 million SNPs.



Allen Roses hopes pharmacogenetics will help avoid adverse drug reactions.

What biomedical researchers are looking for are the much smaller subsets of SNPs that might be linked to the predisposition to disease or adverse drug responses. By comparing SNPs in patients and healthy controls it should be possible to track down those genetic differences. Once this

has been achieved, any person can be genotyped for this limited set of SNPs. Genotyping is a booming market: annual expenditure on SNP research is predicted to grow from US\$158 million in 2001 to more than \$1.2 billion in 2005.

As adverse drug reactions are the fifth-commonest cause of death in the United States, most major pharmaceutical companies have research and development (R&D) programmes to identify the SNPs that predispose to these responses. This pharmacogenetic strategy heralds an era in which the choice of drugs for a particular patient will be based on evidence rather than trial and error. "If you are going to sell a drug, what better characteristic to sell it on, than it will work for you and be safe," says Allen Roses, senior vice-president of genetics research at GlaxoSmithKline in Research Triangle Park, North Carolina. "By going along this pharmacogenetics route we will have more drugs out there that work and are safe to take," he says. To this end, around 1.8 million SNPs have been mapped with reference to the human genome sequence and placed in a public database by

## CRIME BUSTERS

The ability to single out the Y chromosome and genotype male-specific markers is proving popular among forensic pathologists, because fully sequencing each DNA sample is prohibitively expensive. "With our Y genotyping kit, it is possible to screen many pieces of evidence at a relatively modest cost," says James Lazar, vice-president of research and development at Marligen Biosciences in Ijamsville, Maryland.

Marligen's Signet Y-SNP Identification System detects 43 single-nucleotide polymorphisms

(SNPs) from the Y chromosome simultaneously. If these fail to match those of a suspect, he can be eliminated from consideration. And in rape cases, the ability to single out male DNA is a boon. Population genetics is another application.

"All SNPs in our panel have been widely studied and there is a lot of literature on the frequency of those SNPs in different populations. So one can figure out the ethnic background in a broad sense

by looking at those SNPs," says Lazar.

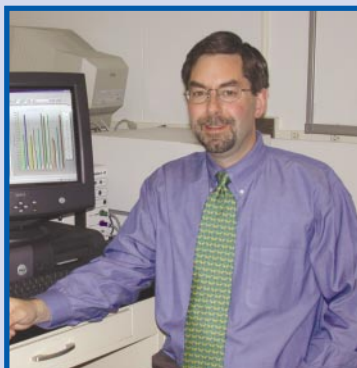
When samples have been out in the woods for weeks, or only a piece of bone is found, nuclear DNA can be degraded, yet the more plentiful mitochondrial DNA remains intact. A full forensic analysis of mitochondrial DNA can cost thousands of dollars, but Marligen provides an inexpensive alternative: the Signet Mitochondrial DNA Screening System, which analyses 30 variations at a cost of less than \$20 per sample.

Marligen's products all use the xMAP technology developed by Luminex in Austin, Texas. This combines microbead-based assays with detection by small lasers. The current platform uses 100 different colour-coded microspheres, to which capture probes are coupled. For many studies an array platform with 100 features is quite enough for large-scale high throughput. "It turns out you are usually only interested in screening a few dozen SNPs rather than tens of thousands," says Jim Jacobson, vice-president of research and development at Luminex.

The small number of features was decisive for Marligen. "In many disease states it is emerging that 20–50 genes are important. That's where we expect science to take us — more than one gene but less than 100," says Lazar.

The Luminex system allows multiplexing of up to 50 SNPs in a single well, saving up to tenfold on reagents and allowing the detection of 10,000 SNPs a day. "It's highly reproducible and very fast. To read a 96-well plate takes 30–40 minutes," says Lazar.

MARLIGEN



James Lazar: genotyping the Y.

the not-for-profit SNP Consortium.

Of the more than 4 million SNPs mapped so far and deposited in public and private databases, only 3% are within genes. "People are not interested in all the polymorphisms; they want to look at those that are functionally relevant," says Dennis Gilbert, vice-president of genomics applications at Applied Biosystems in Foster City, California. This generally means those SNPs within genes and their regulatory regions. The company's Assays-by-Design service offers a route to this goal — it delivers pre-tested, ready-to-use assays for a customer's chosen SNPs.

Customers send in their target DNA sequence, with the SNP locations indicated. Applied Biosystems then develops optimal primers and 'TaqMan' probes to detect both alleles of the selected SNP sequences. Customers receive the probes for each SNP in a single tube with a two-dimensional bar code for sample tracking.

The assay is designed to run on Applied Biosystems' Sequence Detection Systems, real-time quantitative instruments for the polymerase chain reaction (PCR) that use the company's TaqMan probe technology. The use of a non-fluorescent quencher dye increases assay performance to the point at which it is possible to score 1,000 samples for 250 SNPs — that is, 250,000 genotypes — in a day, the company says.

Applied Biosystems also sells complementary off-the-shelf assays and

currently has more than 120,000 validated assays for SNPs in and around genes for linkage disequilibrium mapping studies. "You can go to our public website and type in a gene or region of the genome, and get a whole list of SNP assays for your association studies," says Gilbert.

#### The more the merrier

For those intending to genotype on a massive scale, Illumina of San Diego, California, has recently launched its hybridization-based BeadArray system. Each high-density array of probe-coated microbeads fits into a well of a standard 96-well microtitre plate and can simultaneously detect up to 1,152 different SNPs in a single DNA sample. The fully automated system can process more than a million assays a day. And reduced reagent volumes mean considerable savings in running costs, says Jay Flatley, Illumina's president and chief executive. "The strategy is to bring down the costs of high-throughput and highly accurate genotyping," he says.

The first system was sold to Génome Québec, a member of the International HapMap Project which is mapping SNP haplotypes in the human genome (see 'Haplotypes or pools?', page 921). The UK arm of HapMap, at the Wellcome Trust Sanger Institute in Hinxton near Cambridge, will also use Illumina's technology, and last autumn the US



APPLERA

Dennis Gilbert: online shopping for SNP assays.

National Institutes of Health awarded the company \$9 million to genotype 15% of the HapMap itself.

Illumina's million-dollar system is beyond the means of most academic researchers, but using the company's custom services, a screen for 2,000 SNPs, for example, would cost some 35 cents for each SNP assayed. "As the studies get larger the price drops considerably," says Flatley. "People who run our system can genotype at less than a nickel a data point."

The company has also developed a fine-mapping product — an extensive set of validated SNP assays — from a large genotyping study in collaboration with

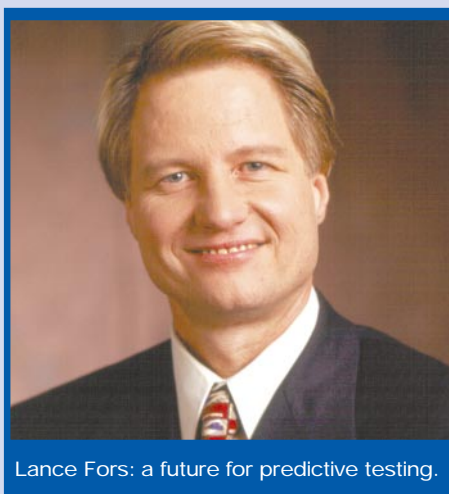
## RISK ASSESSMENT

THIRD WAVE TECHNOLOGIES

Even small numbers of single-nucleotide polymorphisms (SNPs) can work as predictive tools to assist clinical decisions, and this is the area that interests Third Wave Technologies. Based in Madison, Wisconsin, the company has set low-volume SNP-genotyping tests for individual diagnostics as its priority.

"We started out in the cardiovascular arena and helped to create the field for 'economy-class syndrome' mutation detection," explains chief executive Lance Fors. Third Wave developed a reasonably priced test for mutations in the blood-clotting factors V and II that can trigger venous blood clots in the brain and legs. Women who carry the factor V mutation (V Leiden) and are also taking contraceptive pills are at 100 times greater risk of deep-vein thrombosis and stroke, and today an increasing number of women taking contraceptives are tested for these variants. "It's moving towards mainstream medicine," says Fors. "An individual test for a specific SNP may turn into a \$10- or \$30-million market."

Low false negatives and low false positives are the winning solution in diagnostics, Fors adds. Third Wave's platform uses the firm's Invader RNA Assay. Two probes are designed to overlap or 'invade' each other to make a



Lance Fors: a future for predictive testing.

particular three-dimensional structure. If the mutation is present, there is an invasion and the structure is formed. A nuclease then recognizes this overlapping complex and cuts it. With 'normal' DNA, the probes cannot invade and the DNA remains intact, Fors explains. Structure-specific cleavage is far more precise than conventional methods based on the polymerase chain reaction (PCR), so mistakes are very rare, he points out.

The Invader technology is a one-step assay that simply requires the sample to be added. "Some of our clinical customers could never adopt PCR, because they couldn't afford the equipment or the infrastructure required to minimize contamination," Fors says.

Invader technology can scale up to detecting hundreds of thousands of SNPs, as its fluorescence readout is readily adaptable for high-throughput. It was used to screen 768 Japanese cardiovascular patients for the presence of 92,000 SNPs located in genes that are thought to confer a risk of cardiovascular disease, and to identify a candidate gene associated with susceptibility to heart attack on chromosome 6p21 (K. Ozaki *et al.* *Nature Genet.* **32**, 650–654; 2002).

L.M.

GlaxoSmithKline. These consist of about 42,000 SNP positions spread across the genome. Customers can buy them in localized pieces or they can run all 42,000 across their samples to get a broader association study.

### The full monty

"The biggest problem is picking the right genes — it's a bit like playing the lottery," says Brad Margus, chief executive at Perlegen Sciences of Mountain View, California. "The other problem is that even if you find one gene associated with the disease, it may only account for 2% of the disease, so it can be very frustrating."

Whole-genome association studies that identify a set of SNPs linked with a disease are the obvious solution, but only a couple of years ago the cost would have been prohibitive. Today, Perlegen performs whole-genome studies at prices that few could have predicted. "Our people have devised a way to look at the entire genome at very high resolution," says Margus.

Formed in 2000 as an affiliate of DNA-microarray manufacturer Affymetrix in Santa Clara, California, Perlegen has taken the Affymetrix chip and shrunk the features to produce a glass wafer that holds about 60 million DNA probes — in contrast to the 500,000 of a standard Affymetrix array. In addition, Perlegen has developed more than 242,000 different PCR assays that amplify the entire genome.

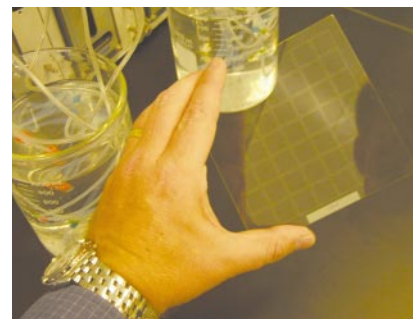
Compared with the few hundred bases usually amplified in conventional PCR reactions, the long-range assays average about 10,000 bases a time. These innovations, coupled with new high-resolution scanners, give a system that has allowed Perlegen to sequence the non-repeat regions of 50 haploid human genomes fully in just 15 months.

"Humans are very similar, so instead of looking at 3 billion bases in people who have a disease and those who don't, we sequenced 50 complete haploid genomes, which yielded 1.7 million SNPs," Margus explains. "We then designed SNP-reading oligonucleotide arrays to genotype those 1.7 million SNPs." With the Perlegen platform, the cost of genotyping these SNPs in 1,000 people is roughly \$2 million.

### Lab in a tube

For academic researchers on a more modest budget, genotyping options are also opening up. ParAllele BioScience of South San Francisco, California, has developed a single-tube, amplification-based methodology, which it claims can detect up to 8,000 SNPs simultaneously. "You have one tube with 8,000 of these probes and you just add genomic DNA," says Tom Willis, ParAllele's chief executive. "The magic happens because we use a pool of molecular 'inversion probes' that target thousands of different SNPs in that one patient's DNA."

ParAllele's inversion probes are



Perlegen's chip holds 60 million probes.

insensitive to cross-reactions — often the limiting factor in highly multiplexed PCR reactions. The probes are incubated with the DNA (only 0.1 ng DNA per SNP is needed) and then all those that have bound are simultaneously targeted by a standard PCR reaction, using a common primer. "It's genotyping first and then amplifying all the SNPs at once," says Willis. The read-out signal is provided when the probe interacts with the genome and rearranges itself to create an amplifiable molecule. Once the assay is performed, the PCR products are loaded onto any chip platform for detection.

ParAllele claims that its technology achieves comprehensive coverage of the genome, or sections of the genome, at affordable prices, by reducing the millions of individual screening reactions typically

## HAPLOTYPES OR POOLS?

As the Human Genome Project winds down, many researchers are stepping over into the next international public endeavour, the haplotype map project, or HapMap. This US\$100-million initiative was launched in October 2002 by the US National Human Genome Research Institute to provide the tools that will speed up the discovery of genetic factors that contribute to common conditions such as diabetes and heart disease. The results will be made quickly and freely available on the Internet.

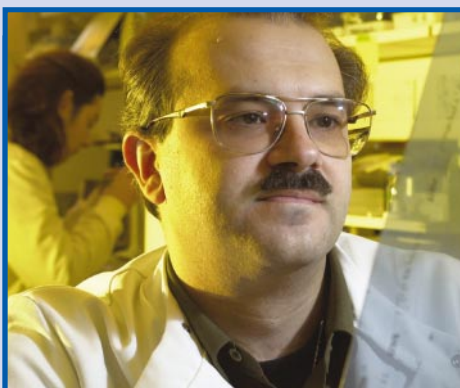
Along the genome, certain combinations of alleles at nearby SNP sites have remained unchanged over many generations and are inherited together as haplotypes. The aim of HapMap is to build a genome map of these haplotype blocks from different ethnic groups, mainly from China, Japan, Nigeria and northern Europe.

Labs in Canada, China, Japan, the United Kingdom and the United States will be involved. "The goal of the project is to validate SNPs across the entire genome and build tools to enable association studies, to compare genetic markers in people who have a disease with markers in a control population," says Panos Deloukas, senior group leader at the human genetics department at the Wellcome Trust Sanger Institute near

Cambridge, UK. The HapMap, set up as an international public-private research consortium, is expected to take three years to complete.

Why bother with a HapMap when whole-genome scanning technologies already exist? "With the technologies we have now, scanning the whole genome gene by gene is very costly. Scanning the genome haplotype by haplotype would be 5% of the cost," points out Julio Licinio at the pharmacogenomics lab of the University of California, Los Angeles.

An alternative way of reducing the costs of whole-genome scanning has just been launched by the discovery-genetics company Sequenom in San Diego, California. Its Allelotyping technology uses high-resolution MALDI-TOF (matrix-assisted laser desorption/ionization-time-of-flight) mass spectrometry to analyse the frequency of a SNP allele in large DNA pools of up to 500 individual samples. Only 50 picograms of DNA from each individual are required, and allelic frequencies as low as 3% can be detected. "Our customers can afford to do whole-genome scans today, because whole-genome scans are affordable on pools but not on individuals," says Charles Cantor, chief scientific officer at Sequenom.



Panos Deloukas: mapping SNP haplotypes.

used by PCR technology to a single reaction for each DNA sample. The system also boasts a high conversion rate — 85% of all SNPs are detected. “It allows researchers to pick which SNPs they want to interrogate, rather than having to use the SNPs that just happen to work,” says Willis.

The system has won the company a \$4-million grant from the US National Human Genome Research Institute to participate in the HapMap project in partnership with the Baylor College of Medicine in Houston, Texas.

### Array-based genotyping

Affymetrix also aims to bring large-scale genotyping by DNA microarray within the reach of individual researchers. Currently in development is the GeneChip Human Mapping 10K Array, which should genotype more than 10,000 SNPs by allele-specific hybridization using only 250 ng of genomic DNA.

Key to this method is the preparation of the sample DNA. Using two standard molecular biology techniques, the genomic DNA is converted to a form that allows multiple loci to be amplified using a single PCR primer. The PCR product contains information for more than 10,000 SNPs which are then genotyped by hybridization to a single array.

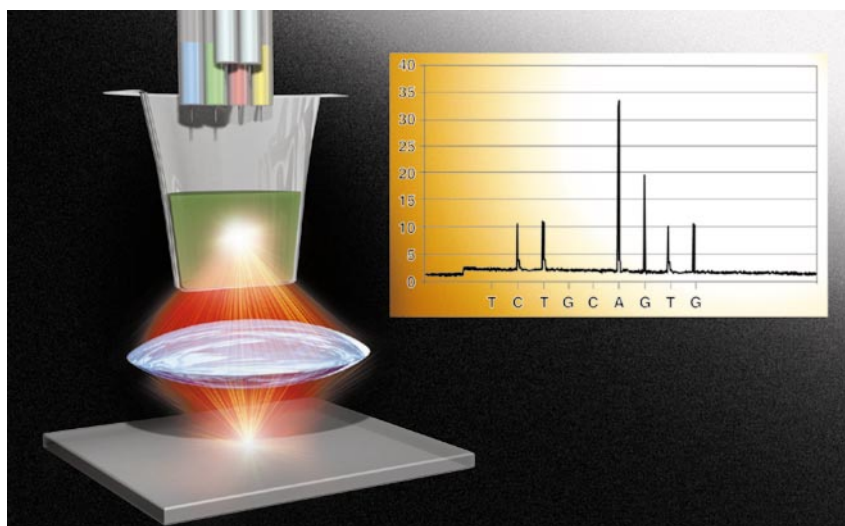
In addition to the SNPs the lab digests throw up, the draft human genome sequence makes it possible to run computer models of digests of total genomic DNA. “This allows us to predict which fragments we can amplify using this method and which SNPs reside in each fragment,” explains Keith Jones, vice-president for molecular genetics at Affymetrix.

The method can be used both for familial-based linkage analysis and for population genetic studies, for example. The assay can scale both downwards and upwards towards 100,000 SNPs, simply by changing the restriction enzyme used to prepare the sample.

### One at a time

Sequencing is an important but often overlooked technology when it comes to scoring SNPs. “Most platforms fail to genotype 10–15% of the polymorphisms you are interested in. To get them typed, the best back-up methodology for us has been DNA sequencing at high reagent dilutions,” says Walt Klimecki, head of genomics at the Arizona Respiratory Center in Tucson.

The ability to sequence SNPs is what many customers of Pyrosequencing in Uppsala, Sweden, find attractive. “The advantage is you have the sequence context so you know you are looking at the right



Pyrosequencing throws light on sequence-based genotyping.

SNP,” says Karen Frost, the company’s market communications director. The technique is simple to use and delivers accurate and consistent analyses of a large number of genetic variations, she says. Because there is no restriction on where the primer can be positioned, it is possible to develop assays for all SNPs. “We have yet to find a SNP that we cannot design an assay for,” claims Frost.

The company’s technology is based on light emission. A sequencing primer is hybridized to the template DNA and the nucleotides are added one by one. If the nucleotide is complementary to the template, it is incorporated and light is emitted. If it is not complementary, there is no incorporation and no light.

Pyrosequencing’s workstations can analyse di-, tri- and tetra-allelic SNPs, insertions and deletions, as well as quantifying alleles in pooled samples. Other applications, such as bacterial typing, use the systems’ capacity for real-time, short-to-medium-length DNA sequencing. The company claims that the high-throughput automated version, with plate stacker and robotic arm, can score around 30,000 polymorphisms in 24 hours.

### High-density maps

In contrast to the fluorescence-based readout systems used by most genotyping platforms, Sequenom, the discovery genetics company based in San Diego, California, uses MALDI-TOF (matrix-assisted laser desorption/ionization–time-of-flight) mass spectrometry in its SNP-discovery system and its Allelotyping whole-genome scanning system (see ‘Haplotypes or pools?’, page 921).

“Fluorescence is a two-digit measurement, whereas spectroscopy is a four- or five-

digit measurement,” says Charles Cantor, Sequenom’s chief scientific officer. “We have very sharp peaks coming up against a relatively flat background, whereas in fluorescence the peaks are broad and tend to overlap.” The samples are presented to the mass spectrometer on a chip. “This lets us use very small samples — 100 times smaller than traditionally used — and also lets us automate,” Cantor explains.

The company has stockpiled 100,000 validated and immediately usable SNP assays through projects with the US National Cancer Institute and GlaxoSmithKline. “These are the gold standard set — the ‘try me first’ set,” says Cantor. “Most of them are near genes or in genes, the polymorphisms are real, the assays work and they are polymorphic enough to be cost-effective.”

With this constellation of ingenious approaches, SNP hunters on both the large and small scale are paving a fast track to personalized medicine. “The technologies have only just started to emerge and we are using them already — we do high-density whole-genome SNP mapping,” says Mihael Polymeropoulos, head of Novartis Pharma’s pharmacogenetics department in Gaithersburg, Maryland. “We now systematically collect DNA from every patient in every clinical trial, analyse that for variations and then at the end of the trial do association studies between the genetic variation, efficacy and the adverse effects.” Detecting SNPs is no longer the technical difficulty, the challenge is now to make sense of all the data. ■

Lisa Melton is science writer at the Novartis Foundation, London.

The SNP Consortium ♦ [snp.cshl.org](http://snp.cshl.org)  
HapMap ♦ [hapmap.cshl.org](http://hapmap.cshl.org)