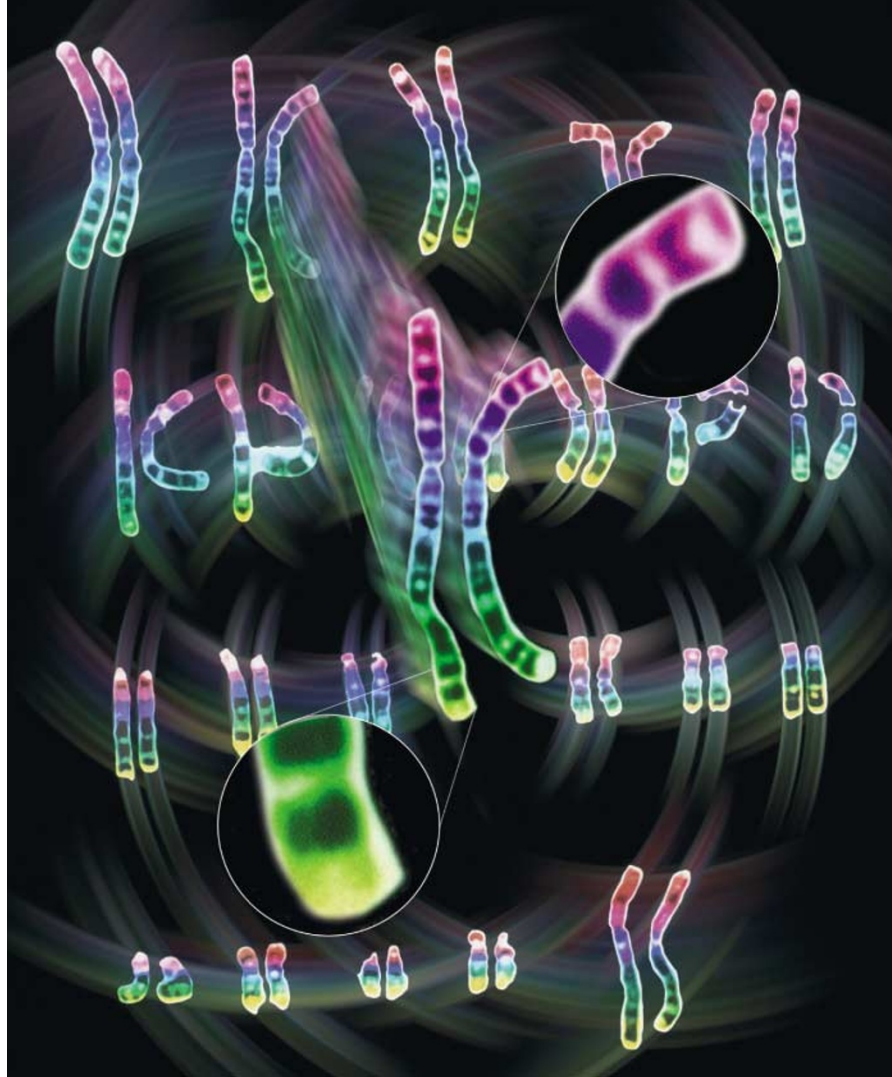


# Faster, better, cheaper genotyping

Scanning the genome for subtle genetic variations across thousands of individuals may help researchers find genes that underpin susceptibility to common diseases. Marina Chicurel considers the technological requirements.



Back to bases: mapping SNPs to locations on chromosomes could help to unravel disease susceptibility.

The publication in February of draft sequences of the human genome<sup>1,2</sup> dominated the news worldwide. But for many of the researchers hunting the genes that underlie conditions such as heart disease and cancer, just as important were the less-trumpeted accompanying announcements on the discovery of single nucleotide polymorphisms (SNPs) — points in the genome at which the genetic code can vary by a single ‘letter’.

SNPs — pronounced ‘snips’ — account for most of the genetic variability across human populations. Because they are simple, abundant and widely dispersed, they make excellent landmarks for navigating the genome. As genetic variants that lie close to each other on a chromosome tend to be inherited together down the generations, monitoring SNPs may help gene hunters to trace sequences associated with the susceptibility to common diseases. Doctors might also in the future routinely test for particular SNPs and so tailor drug treatments to each patient’s individual genetic make-up.

In their February paper, researchers at Celera Genomics of Rockville, Maryland, announced the location of 2.1 million SNPs<sup>2</sup>. The International SNP Map Working Group, a coalition of academic labs backed by leading companies and Britain’s Well-

come Trust, published a map containing 1.4 million (ref. 3). The search continues: in some countries, including Japan and China, efforts are under way to identify SNPs specific to their respective populations.

## The hunt is on

But mapping SNPs is merely the first step in the hunt for genes involved in disease susceptibility. Researchers must then identify which SNPs are most valuable as markers — many show insufficient variability within a given population, and some are found in repetitive regions of the genome and so do not make useful landmarks<sup>4</sup>. Then comes the task of screening for the useful SNPs in large numbers of people to look for those variations that are associated with particular traits, such as susceptibility to coronary heart disease. And this, at present, is where the available technology is falling short.

Genome-wide gene hunts could require the analysis of hundreds of thousands of SNPs from tens, or even hundreds of thousands of individuals. That sends the number of individual SNPs to be genotyped into the billions<sup>5</sup>. Many researchers are focusing on ‘candidate’ genes already suspected of being linked to a particular trait. But even these more limited efforts can require screening tens of thousands of SNPs in thousands of

individuals. To make such studies possible, the throughput of the world’s SNP genotyping labs must increase by one or two orders of magnitude, and costs will need to be brought down at least tenfold. “The ideal assay will be very quick, cheap and easy,” says Pui-Yan Kwok, an expert on SNP discovery and genotyping at Washington University in St Louis, Missouri. “It is not available.”

Ever since the first large-scale attempts at SNP genotyping started three years ago, dozens of alternative techniques have emerged<sup>6</sup>. “But if you look at them, they’re based on a very few experimental concepts,” says Anthony Brookes of the Center for Genomics Research and Bioinformatics at the Karolinska Institute in Stockholm.

These concepts can be divided into three main categories: reactions, detection systems and formats. Reactions are designed to generate specific molecules based on the presence or absence of a particular SNP. The detection systems are coupled to the reactions to reveal these products. And the formats are the conditions under which the reaction and detection steps take place.

One approach under the reaction category is hybridization, first used on a large scale in 1998 by Eric Lander and his colleagues at the Massachusetts Institute of Technology’s Whitehead Institute for Biomedical

Research<sup>7</sup>. Hybridization depends on the pairing of 'complementary' letters in the genetic code, in which adenines (A) bind to thymidines (T), and guanines (G) to cytosines (C). Lander's team used short synthetic DNA sequences complementary to known SNPs. The sequences were immobilized onto glass 'chips', which were then exposed to a chemically tagged sample of an individual's DNA. The researchers looked for the presence of 500 different SNPs simultaneously by detecting where on the chips each sample hybridized. The chemical tags bound to a fluorescent dye, allowing the chips to be scanned using an optical read-out system.

But hybridization can be difficult — it often needs careful calibration to give reliable results. So many researchers are instead using DNA-manipulating enzymes to reveal the presence of particular SNPs. "Enzymes are highly discriminating," says Scott White, a geneticist at the Los Alamos National Laboratory in New Mexico. They also tend to work reliably without the need for extensive optimization of the experimental set-up.

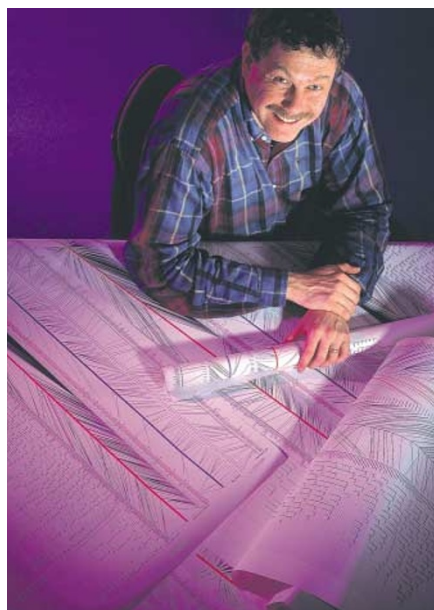
### Prime target

Based on these advantages, researchers have developed SNP assays using enzymes that synthesize, cleave or splice DNA. One popular approach, called primer extension, uses a DNA polymerase enzyme to add individual letters of the genetic code, or nucleotides, to a small piece of synthetic DNA called a primer. The primer is designed to hybridize to sequences immediately adjacent to a particular SNP. Once it is in place, the DNA polymerase reads along the rest of the sequence, building a complementary strand of DNA. Researchers can then identify whether a SNP variant is present by monitoring which nucleotide the polymerase incorporates, or fails to incorporate, as it reads along the DNA sample.

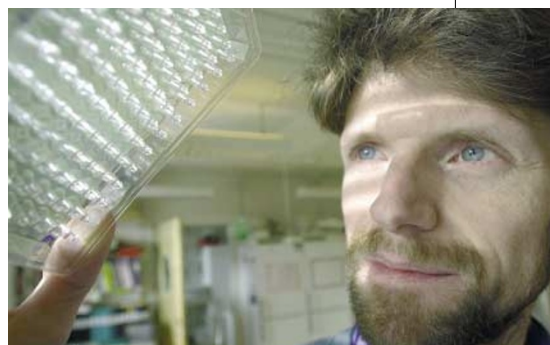
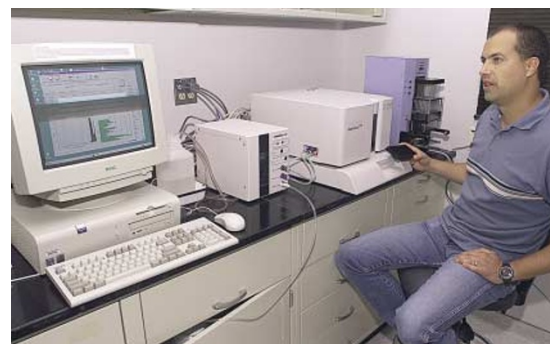
Another assay, called Invader and marketed by Third Wave Technologies in Madison, Wisconsin, relies on an enzyme that cleaves DNA<sup>8</sup>. The assay uses two synthetic pieces of DNA, or probes, designed to hybridize to sequences adjacent to a particular SNP. The probes flank the SNP and overlap precisely at the SNP site. If a particular SNP is not present, the overlapping structure will not form. By adding an enzyme that cleaves DNA only when it encounters such overlaps, researchers can assess whether or not the given SNP is present.

The various approaches can be mixed with different detection systems. In primer extension, the DNA polymerase can be fed fluorescently labelled nucleotides, where each of the four nucleotides produces light of a different colour. Alternatively, the extended primer's mass can be measured using mass spectrometry, which can distinguish between DNA molecules differing by only one nucleotide.

Each reaction and detection technique



The detectives: Eric Lander (above), Scott White (top right) and Anthony Brookes are all interested in improving methods for identifying SNPs.



has its pros and cons. Many researchers working on large-scale SNP genotyping prefer primer extension because it is robust and flexible. It requires few synthetic DNAs, the design of the primers is simple, and similar reaction conditions can be used for many different primers.

On the detection front, mass spectrometry is popular because it is reliable and yields readily quantifiable results that can be scored easily and rapidly by automated computer systems. Sequenom of San Diego, for instance, markets a technology based on primer extension and mass spectrometry. "What most impressed me was how accurately we could genotype with these out-of-the-box assays," says Kenneth Buetow of the National Cancer Institute (NCI) in Bethesda, Maryland, who is developing methods to streamline SNP genotyping.

Mass spectrometry has the added advantage of not depending on fluorescent labels, which can be expensive. But despite its popularity, the technique has its limitations. Until recently, for example, researchers had to spend a lot of time separating their primer extension products from chemical buffers, the sample DNA, as well as removing the DNA polymerase enzyme and the free nucleotides left over from the reaction. This is because

mass spectrometry requires pure products.

Methods such as the one developed by a team led by Ivo Gut of the French National Centre for Genotyping in Evry, near Paris, have helped circumvent this problem. Gut has boosted the sensitivity of detection by adding a chemical group that modifies the charge on the extended primer<sup>9</sup>. Thanks to this increased sensitivity, simple dilution to lower the concentration of leftover reagents will still give a detectable signal.

But the biggest problem with mass spectrometry is that it generally only allows researchers to screen for up to a dozen SNPs at a time<sup>10</sup>. Chip-based hybridization approaches, meanwhile, have advanced to the point at which thousands of SNPs can be screened in parallel. Unless 'multiplex' techniques can be developed for mass spectrometry, argues Michael Boyce-Jacino of Orchid BioSciences in Princeton, New Jersey, the technique ultimately will "hit the wall". Orchid is also marketing a system that relies on primer extension, but offers its clients a variety of detection systems.

Just as detection systems can be mixed and matched with different reactions, the situations under which the reactions occur — the format — can also be varied. When fluorescent tagging is used as a detection system in primer-extension genotyping, many SNPs can be analysed in parallel if the DNA primers for different SNPs are immobilized on a chip. The light given off from each complementary strand built by the DNA polymerase enzyme can then be detected independently. But, compared with assays in which the reagents and products float free in solution, such methods are less flexible. Adding new SNPs to the analysis means that the chips must be redesigned. And the heating and cooling

The throughput of the world's SNP genotyping labs must increase by one or two orders of magnitude.

required for the primer-extension reaction are difficult to achieve on solid surfaces.

Hoping to get the best of both worlds, some researchers, such as John Nolan and Hong Cai, working with White at Los Alamos, are turning to tiny glass beads about 5 micrometres in diameter<sup>11</sup>. The researchers first perform standard primer-extension reactions with fluorescently labelled nucleotides in solution.

### Extended play

In solution-based assays, it is usually only possible to study one SNP at a time. But by using beads to capture and sort the products of their reactions, the Los Alamos team can study many SNPs in parallel. The researchers place dozens of primers specific for different SNPs in a tube with DNA samples and fluorescently tagged nucleotides. At the end of the primer-extension reaction, the tube contains a complex mixture of labelled products. Then the team adds colour-coded beads carrying 'address tags', pieces of DNA that are complementary to portions of the different primers used in the reaction. As a result, all the products built from one type of primer get attached to beads of the same colour.

The beads can be sorted and analysed using a machine called a flow cytometer. This funnels the sample of beads through a very narrow opening to create a stream in which the beads travel in single file. The cytometer has a laser and a light detector facing the stream, so it can detect the fluorescent colour of each bead as it goes by, as well as the colour of its associated fluorescent nucleotides. It can do so extremely quickly — scoring hundreds to thousands of beads per second. "I think that's going to be one of the concepts that takes us to the next generation of methods," says Brookes.

But Brookes and most other researchers suspect that further advances will be needed to achieve the desired breakthroughs in cost

One of the key bottlenecks to high-throughput genotyping of SNPs is DNA amplification.

and speed. "I think we're a long way away from mature technologies," says Mark Lathrop, director of the French National Centre for Genotyping.

One of the key bottlenecks is the amplification of DNA. Most current assays include a step that produces many copies of a short segment of the sample DNA spanning each target SNP. This amplification is usually necessary because only small amounts of DNA can be harvested from typical clinical samples. Also, the amplification improves the signal-to-noise ratio of the assay, increasing the reliability of detection.

Most genotyping techniques accomplish this amplification using molecular biology's workhorse, the polymerase chain reaction, or PCR. Although PCR is very competent at its job, it is expensive. In addition, setting up PCR to amplify more than 10 targets in parallel is extremely difficult. Those researchers who have achieved multiplex PCR have had to work hard to optimize their systems<sup>12,13</sup>.

That is why researchers working on SNP genotyping are watching the progress of a team led by Yusuke Nakamura of the University of Tokyo's Human Genome Center. Nakamura is working with sealed cards in which samples are subjected to 100 parallel PCRs, and claims his team can genotype almost 400,000 SNPs a day<sup>14</sup>. He says the key lies in the design of the PCR primers, the artificial DNA sequences that define the stretch of sample DNA to be amplified. But some experts remain sceptical. "I don't know how they can do it," says Kwok. A paper outlining Nakamura's methods will appear shortly<sup>15</sup>.

### Pool cues

Exploring the flip side of multiplex PCR, some researchers are amplifying and genotyping single SNPs from many individuals at once. Working with researchers at Sequenom, the NCI's Buetow has pooled DNA samples from close to 100 individuals and assessed the presence of thousands of SNPs collectively<sup>16</sup>. Although pooling obscures the presence of rare SNPs and results in the loss of information on how SNPs are arranged on individuals' chromosomes, it speeds up genotyping immensely. It can, for example, allow rapid comparisons of SNPs from a group of individuals suffering from a particular type of cancer with those who are cancer-free.

Predicting the future of SNP genotyping



**Bright idea:** Solexa will use laser optics to speed-read an individual's genome.

technology is not easy. The field is moving rapidly, with new approaches springing up all the time. Among the most ambitious ideas being mooted is a novel DNA sequencing technology from Solexa, a British company based in Saffron Walden, near Cambridge<sup>17</sup>.

Solexa aims to make chips that will contain up to a hundred million immobilized fragments of single-stranded DNA. The chips will be sequentially washed with solutions containing a single type of nucleotide, each bearing a fluorescent tag, in the presence of a DNA polymerase enzyme, which will try to build complementary strands of DNA. After each wash, lasers will be used to record where the tagged nucleotides have been added, before the tags are chemically removed and the process repeated with a different nucleotide. In this way, claims Solexa, it will be possible to speed-read an individual's genome, SNPs and all, in a matter of days without recourse to PCR.

Whether Solexa's technology will provide what researchers working on SNP genotyping are looking for remains to be seen. But most feel that a technique that is similarly ambitious in its scope will probably be required. "The winner may not even be in the race yet," says Buetow.

**Marina Chircurel is a writer in Santa Cruz, California.**

1. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
3. The International SNP Map Working Group *Nature* **409**, 928–933 (2001).
4. Marth, G. *et al.* *Nature Genet.* **27**, 371–372 (2001).
5. Abbott, A. *Nature* **406**, 340–342 (2000).
6. Kwok, P.-Y. *Annu. Rev. Genomics Hum. Genet.* (in the press).
7. Wang, D. G. *et al.* *Science* **280**, 1077–1082 (1998).
8. Lyamichev, V. *et al.* *Nature Biotech.* **17**, 292–296 (1999).
9. Sauer, S. *et al.* *Nucleic Acids Res.* **28**, E13 (2000).
10. Ross, P., Hall, L., Smirnov, I. & Haff, L. *Nature Biotechnol.* **16**, 1347–1351 (1998).
11. Cai, H. *et al.* *Genomics* **66**, 135–143 (2000).
12. Fan, J. B. *et al.* *Genome Res.* **10**, 853–860 (2000).
13. Lindblad-Toh, K. *et al.* *Nature Biotechnol.* **18**, 1001–1005 (2000).
14. Cyranoski, D. *Nature* **410**, 1013 (2001).
15. Ohnishi, Y. *et al.* *J. Hum. Genet.* (in the press).
16. Buetow, K. H. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 581–584 (2001).
17. Adam, D. *Nature* **411**, 402 (2001).



**Fast-track:** Yusuke Nakamura claims his system can genotype nearly 400,000 SNPs a day.