Tools for genomics

Edwin Southern

Genomics, in its early days, used a range of techniques that were developed to explore the composition and sequence organization of the nuclear DNA. High-throughput methods changed that, and most research in genomics is now done in factory-like laboratories, with robots doing much of the work. I have adapted to working in this way, but miss the immediacy of direct involvement with experiments. I want to see the data before the computer does. But solving problems at the bench has always given me as much enjoyment as the more cerebral aspects of science. This practical approach is, I think, deeply rooted in my childhood: I was born in the year before the outbreak of World War II, a period of austerity and 'make do and mend.' My father was a Royal Air Force aircraft mechanic, who took great pride in improvising solutions to the problems he faced daily, fixing damaged 'kites' as they limped back from their missions to his base in Reykjavik. I learned from him that if the tool for a job is not at hand, you make one. My theoretical contributions to genomics may endure longer, but I have to accept that I am better known for developing tools and it is for one of these that I am delighted to have been recognized with the Lasker Award.

The beginnings of genomics

As a postdoctoral research assistant in Cambridge, I studied the effects of ionizing radiation on polypeptides at the Low Temperature Research Station. The Medical Research Council Laboratory of Molecular Biology was not far away and its luminaries often gave seminars at the university that were open to all comers. I recall attending talks by Sydney Brenner, Frances Crick and, most influentially, Fred Sanger. The Laboratory of Molecular Biology also attracted star speakers and it was here that I met my future mentor, Peter Walker, who had come to talk about his work on mammalian satellite DNAs. These major fractions, which can make up as much as half of the genome[1], can be separated from the bulk of the DNA by density-gradient centrifugation, the technique used famously by Meselson and Stahl[2] to prove the semiconservative nature of DNA replication. The major satellites are absent from most genome sequences because their structure, which comprises very long stretches of tandem repeats, makes clones unstable, and sequence assembly from short stretches is nearly impossible. But in 1967, it was the simple nature of the sequences that made it seem worthwhile to give sequencing a try. The difficulty that stood in the way was the lack of a method for DNA sequencing. Here was an opening for someone with an enthusiasm for method development, so I packed my old car with my possessions and returned north to join Peter's group in Edinburgh.

The University of Edinburgh is very strong in science. In 1967, when the word 'genome' was understood by very few, Edinburgh was notable for the high concentration of scientists working in genomics. The sizes of eukaryotic genomes presented a puzzle and major challenge to the molecular biologist. Edinburgh was exceptional in the number of scientists developing molecular methods that could be applied to eukaryotic genomes. Conrad Waddington had built one of Europe's premier genetics laboratories. With great foresight, he founded a Department of Epigenetics, to which he recruited such pioneers in the molecular biology of eukaryotes as Max Birnstiel, John Bishop and Ken Jones. Max was the first to isolate a eukaryotic gene—the ribosomal genes of *Xenopus laevis*[3]; John devised ingenious methods for studying gene expression[4]; and Ken developed a method of *in situ* hybridization and used it to show that satellite DNAs clustered in the heterochromatic regions close to centromeres[5]. The Department of Molecular Biology, the first in any UK university, also had its share of scientists working on genes and genomes in the more genetically amenable prokaryotic systems. Ken and Noreen Murray were particularly important to my own development as a molecular biologist, but there were many others who contributed to the atmosphere of excitement and kept us abreast of developments beyond our own fields of specialization.

Michael Swann's Zoology Department also housed a remarkable number of scientists interested in molecules; there was a strong immunology group under James Howard; Murdoch Mitchison studied the cell life cycle of the fission yeast *Schizosaccharomyces pombe*; Peter Walker headed the Medical Research Council Mammalian Genome Unit. Peter was a professor of zoology, but also a qualified engineer, and for his PhD studies in the 1950s he had built an ultraviolet microspectrophotometer with a spatial resolution of 1 μm, which he used to measure the DNA and RNA contents of intact cells[6]: one of the most interesting puzzles presented by higher eukaryotes was the *c*-value paradox; the amount of DNA in the genome can vary by orders of magnitude in organisms that are closely related.

So when I joined Peter in Edinburgh to sequence satellite DNA, largely ignorant of nucleic acids, I turned for help to Ken Murray, who was adapting to DNA the methods Sanger had developed for RNA sequencing[7]. These methods, with some of my own variations, lacked the power of modern methods, but were good enough to give a picture of the structural organization of satellite DNA sequences[8]. We showed that the sequences did indeed comprise short repeats joined together in tandem, with variants introduced by point mutation.

Ken and Noreen had a keen interest in restriction and modification and when the type II restriction endonucleases were discovered by Ham Smith and his colleagues at Johns Hopkins University[9], they introduced these enzymes to the Edinburgh community. Ken founded a restriction endonuclease club. To join, you had to purify an enzyme and make it available to other members. I was given EcoRII as my entry ticket. This really was a piece of good luck: EcoRII is one of the few enzymes that finds a site in the sequence of mouse major satellite DNA. My colleague Gerard Roizes and I had shown

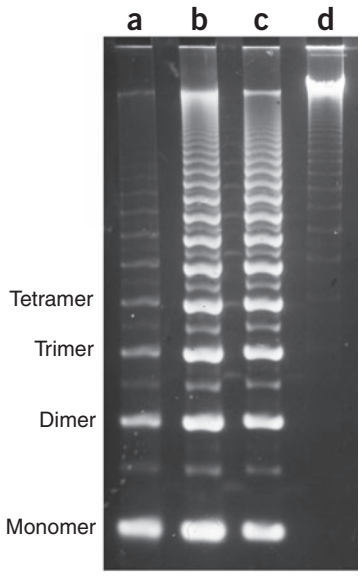

Figure 1 Restriction endonucleases and 'selfish DNA.' Digestion of mouse major satellite with EcoRII reveals the repetition of a short sequence motif, with variation introduced by mutation, which characterizes this type of DNA. Selfish DNAs require mechanisms for expanding the sequence. In this case, the mechanism is probably unequal crossover[11].

that gel electrophoresis of restriction fragments from complex genomes gave a smear[10], but we predicted that tandemly repeated sequences should show a distinct pattern. It was Gerard that took the gel to the light box and I could see we had an exciting result from his expression as he came out of the darkroom and exclaimed, "You have a ladder!" (**Fig. 1**). It was the kind of result that gives an instant buzz. At a glance, we could see in the ladder pattern all the main features of the long-range sequence organization of the satellite DNA—the tandem repetition, the unit length and even the effects of mutation and unequal crossing over were all evident in that simple pattern[11]. It is instructive, too, to realize that the information we gleaned from this simple experiment is beyond the reach of modern sequencing methods and well beyond the sequencing methods we were using at the time.

The structural features of the satellite DNAs suggested evolutionary mechanisms, but did not give any clues to possible functions. On the contrary, the sequence motifs were so short and the sequences so degraded by mutation as to suggest that they could not have a function that depended on their sequence. These observations were among the first to suggest that the bulk of eukaryotic DNA may carry no information, leading to the notion of 'parasitic,' 'selfish' or 'junk' DNA, and providing an explanation for the *c*-value paradox[12].

5S RNA genes and the hunt for their promoters

It may have been the possibility that the satellites had no functional importance that turned my interest to a DNA that certainly did have an important function. The sequencing methods available to us were developed for analyzing small RNA molecules and I used them in a collaboration with Peter Ford, which set me on the path to developing the gel-transfer or blotting method. Peter had found that the oocytes of *X. laevis* store large amounts of 5S ribosomal RNA in 40S particles[13], ready for the rapid burst of cell cleavage and protein synthesis that takes place after fertilization. We showed that the 5S RNA in this particle has a different sequence from the 5S RNA of somatic cells[14]. This indicated that there were two types of 5S RNA genes and that they were differentially regulated. It seemed a good system in which to study eukaryotic gene regulation. Our plan was to isolate the two types of 5S gene so that we could sequence the upstream regions, in the expectation that this would lead us to the promoter regions. At the time there was a great deal of excitement about the potential of cloning DNA sequences. But as there were fears that this powerful technology could somehow get out of hand and wreak genetic havoc, it was shelved while regulatory mechanisms were set in place.

The signs were that the 5S genes, like the satellite DNAs, were in blocks of tandem repeats. For these DNAs, it is usually possible to find a restriction endonuclease that finds no cutting site within the repeat unit, so the repeated genes would be left as large fragments (**Fig. 1**), whereas the bulk of the DNA is cut into small fragments that could be removed by gel electrophoresis. After isolating the gene as a large fragment, it could be cut by an enzyme that cleaved one site per repeat to give fragments corresponding in size to the repeat length. Calculations showed that just a few rounds of enrichment should provide genes essentially free of other DNA. My first attempts to find genes involved cutting the gel into thin slices and eluting the DNA, which I then bound to nitrocellulose discs. Each disc was hybridized with radioactive RNA in a vial, washed and counted in a scintillation counter using the hybridization method of Gillespie and Spiegelman[15]. This whole process was extremely tedious, but worse, the background was so high that it was difficult to discern the signal. So I set out to devise a method of '*in situ*' hybridization that could detect the bands in the gel and attempted, with Julia Thomson, to hybridize to DNA fragments trapped in dried gels. Though Oliver Smithies and his colleagues did get this method to work[16], in our attempts, the fragments diffused out of the gel. The story of how I stumbled on the realization that agarose gels are permeable enough to allow the DNA to be soaked out and trapped onto membranes has been written elsewhere[17]. Once I realized the gel was permeable, it was a simple matter to set up the transfer (**Fig. 2**). My first sight of bands on the autoradiograph was another great thrill. The full range of possibilities that was opened up by visualizing genes in this way was beyond my imagination, but I could think of enough to know that I was onto a good thing. And it was immediately clear that it was a simple and robust method; it worked well at the first attempt.

I did some experiments with bacterial 5S genes (**Fig. 3**) and eukaryotic ribosomal genes and sent a manuscript to the *Journal of Molecular Biology*. At first, it was rejected as a 'methods' paper, and it took me quite some months to do further experiments that were more biologically relevant[18]. In the meantime, word of the method was getting around the community and I was often surprised and delighted to hear speakers at meetings refer to results they had obtained with 'Southern's blotting method'—not my name for it, but one that must have come from Mike Matthews. Mike came to visit our group in Edinburgh, shortly after I had done the first transfers and he asked if he could use the method in his work on viruses. He returned to Cold Spring Harbor

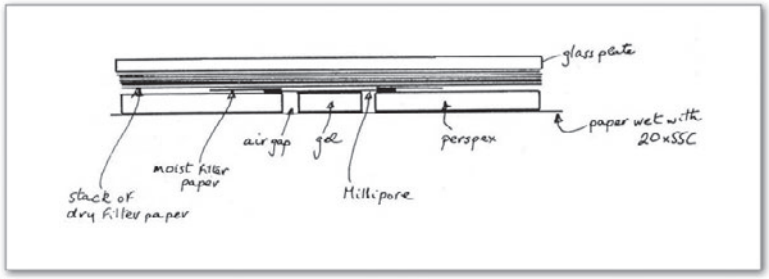

Figure 2 Early 'publication' of the gel transfer or blot. The method for transferring restriction fragments of DNA from a gel to a membrane is illustrated in this original sketch, which was sent to colleagues before publication of the method.

with a sketch of the method on a scrap of paper. He and his colleague, Mike Botchan, soon had the method working. Visitors to their laboratory asked to use it and Mike asked for permission to pass it on, acknowledging its origins, which I was happy to give. So the method was in widespread use well before it was published. Nonetheless, at one point the publication held the record for the *Journal of Molecular Biology*'s most cited paper.

In Edinburgh we used the method in studies of the genes for ribosomal RNAs and other repeated sequences[19]. We tried hard to increase the sensitivity so that we could detect single-copy genes, but we were beaten to this objective by the group in Cold Spring Harbor, who devised ways of using nick-translated DNA as probes which could be labeled to much higher specific activity than the RNAs we used[20]. Alec Jeffreys and Dick Flavell took a different route. By denaturing the DNA, they were able to load much more onto the gel. They used this method in their seminal work that showed the presence of intervening sequences in the rabbit gene encoding β-globin[21]; the first demonstration of introns in a nuclear gene. Alec went on to use the technique for his DNA fingerprinting method which has found so many important applications[22] (see Alec Jeffreys' paper in this issue).

Peter and I did not pursue the 5S DNA project, but, as Robert Roeder describes in his 'Lasker Commentary'[23], it was later discovered that 5S RNA genes are transcribed by RNA polymerase III. The transcription factor binding sites for this enzyme lie within the transcribed sequence—the region that Peter and I had already sequenced. In retrospect, it is perhaps a good thing that this possibility never crossed our minds.

RFLPs, disease genes and the Human Genome Project

In 1978, Y.W. Kan and A.M. Dozy used the transfer method to show that the mutation in the human gene encoding β-globin, which causes sickle-cell anemia in black Americans, is associated with a polymorphism in a nearby restriction site[24], making it possible to trace the disease-associated gene indirectly by a simple test on the DNA. The presence or absence of the restriction site in the region flanking the gene could be revealed as bands with different mobilities in a DNA transfer hybridized with a probe for the β-globin gene. Family members carrying the mutation but not manifesting the disease could be identified by this test. Polymorphisms of this type, which came to be called restriction fragment–length polymorphisms or RFLPs, were to have an enormous impact on human biology.

I got the first hint of this revolution when I met David Botstein in 1979, a meeting that changed my scientific direction. I had been invited to spend the summer in Woods Hole as the Lilley Fellow. David came on a visit, perched on a stool in the laboratory and explained at length to me his idea for constructing a genetic linkage map of the human genome, using random clones of DNA as hybridization probes to identify large numbers of RFLP markers. Genes for inherited traits segregating in families could then be placed in the map by finding markers that segregate with the trait. As Kan and Dozy had shown, such linked DNA marker loci could be used predictively for genetic counseling. Conceived as a way of locating genes for which there were no direct probes[25], the idea that David described to me that day at Woods Hole was the beginning of the program that led, some twenty years later, to sequencing the human genome.

Figure 3 Genomes on paper—making genes visible. An early 'blot,' the picture with the black backgrounds show all the fragments of DNA from the *E. coli* genome stained with a fluorescent dye. The picture with the white background shows the genes as bands that have been revealed by hybridizing with a probe for 5S DNA.

It took me some time to grasp the implications of what David told me, but when I realized its importance it occurred to me that a physical map of the genome would be a powerful complement to the genetic map. For example, a map based on contigs of clones would greatly simplify the identification and isolation of genes located by linkage analysis.

On my return to Edinburgh, I wrote a grant proposal to the Medical Research Council and set up a collaboration between a group in the Mammalian Genome Unit headed by Chris Bostock and one in the Clinical and Population Cytogenetics Unit headed by Nick Hastie. This was a project rich in possibilities for strategic planning and for the development of new techniques. Looking back on this period, it is clear that I spent too much effort on new techniques and too little on strategy—we were overtaken by groups who got on with the task using existing technology. But we did have some successes with new techniques. I mention one, briefly.

In 1984, David Schwartz and Charles Cantor described the pulsed field gradient gel electro-

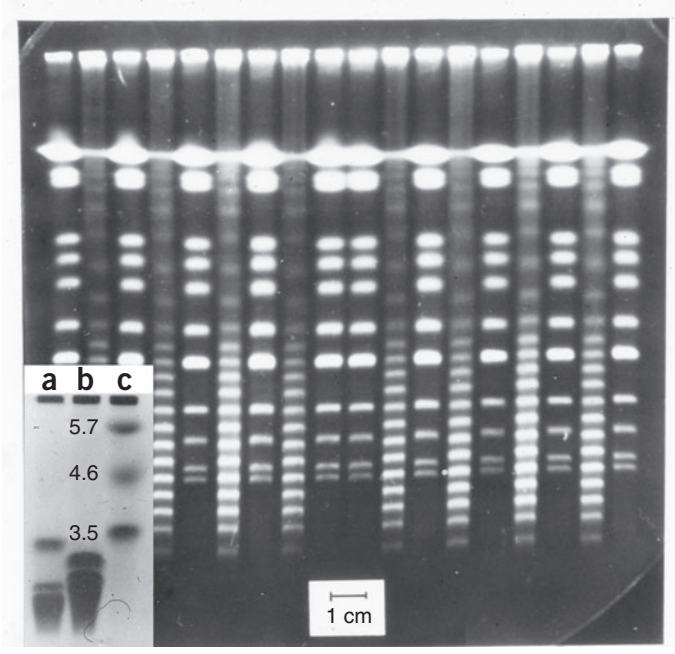

Figure 4 Genomes on gels. Pulsed field gel electrophoresis separates large DNAs. The inset shows separation of yeast chromosomes. The three *S. pombe* chromosomes (track c) are 3.5, 4.6 and 5.7 Mb corresponding to 1.1, 1.5 and 1.9 mm of DNA. The background shows separations of ladders of ligated lambda DNA interspersed with chromosomes of *S. cerevisiae* to illustrate the straight tracks obtained with the 'Waltzer'[30].

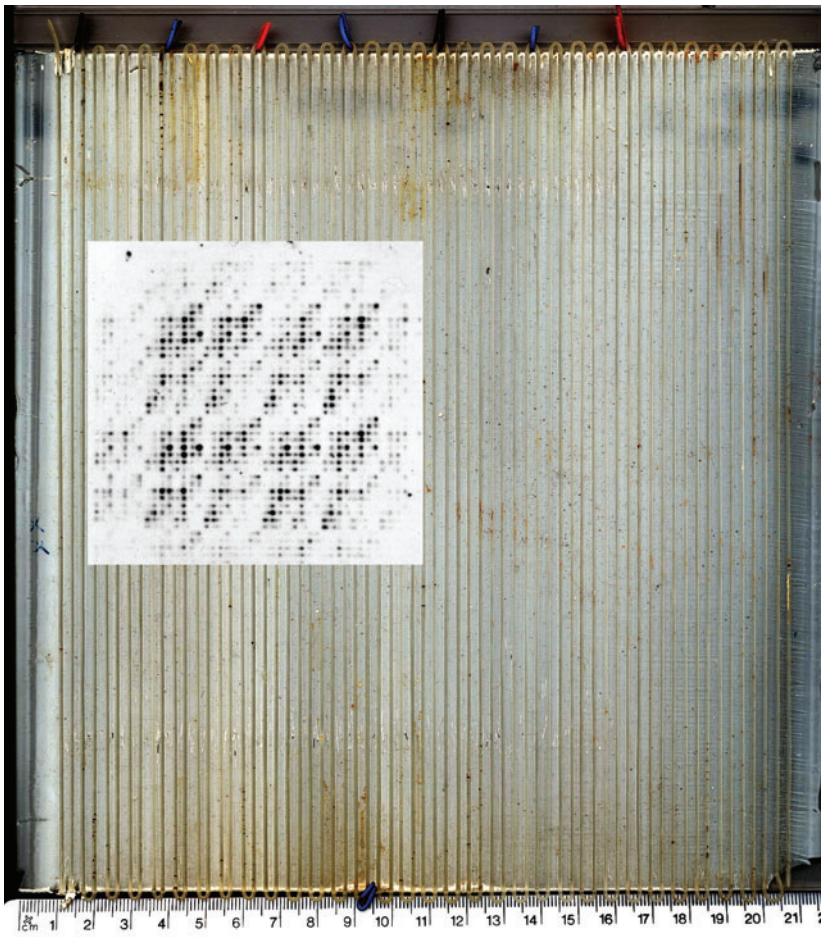

Figure 5 Low-technology tools for arrays. The background shows a channel plate made by gluing silicone rubber tubing to a sheet of glass, which was used to apply nucleotide precursors to another sheet of glass that had been activated for synthesis of oligonucleotides. An array of more than 2,000 oligonucleotides, including four copies of all 256 octapurines, was made by applying precursors for A and G on an ordered pattern of channels and by rotating the channel plate through 90° at appropriate points in the synthesis. The inset shows the pattern of hybridization from hybridizing a 20-mer oligopyrimidine to this array[36].

phoresis (PFGE) technique for separating very large fragments of DNA[26]. Conventional gel separations break down above about 50 kb, but this new method could separate fragments up to several Mb. This was a tremendous advantage for restriction-site mapping of human chromosomes, which in total comprise ~3,000 Mb of DNA. It is also one of those methods that grips the imagination: fully stretched, the molecules that can be separated by PFGE are up to 3 mm in length. The inventors devised the new separation method from an ingenious theory of how large DNA molecules behave in gels, which depends on its viscoelastic properties and which would only manifest itself in inhomogeneous electric fields. These fields were produced in an apparatus with complex electrode geometry that resulted in the DNA moving along curved tracks, making it difficult to compare mobilities in different tracks. In one of those 'light bulb' moments, it occurred to me that the separation could be explained by a much simpler mechanism, a ratcheting motion in which the DNA molecules reverse their direction at alternate pulses, and that this could operate in homogeneous fields. My colleague Duncan Fletcher quickly built a device that rotated the gel in a reciprocal movement in a stationary field, rather than moving the field relative to the gel[27]. The apparatus, which we called the Waltzer, gave beautiful separations in straight tracks (**Fig. 4**). My colleagues and others put the Waltzer to good use in their gene mapping[28], and in studies of chromosome structure[29].

Oligonucleotide arrays

In 1985, I was invited to the Whitley Chair of Biochemistry in Oxford. By that time, the human genome had become the subject of a large international program; I was ready for a change. I packed my family into our small car and we battled our way south against gale-force winds.

Few university departments have workshops nowadays, but it was my good fortune that Peter Walker had a well-equipped workshop in Edinburgh. My first task in Oxford was to set up a small machine shop in my laboratory where Martin Johnson could build specialized equipment. But much can still be done by improvisation, and the earliest experiments with oligonucleotide arrays were carried out using the modern equivalents of string and sealing wax—silicone rubber and silicone cement.

The idea behind oligonucleotide arrays came to me at a meeting in July 1987 in Okayama, organized by Akiyoshi Wada to discuss potential new methods for large-scale sequencing. Several well-known scientists, inspired by Sanger's successes in sequencing the increasingly complex genomes of bacteriophage ϕX174 (ref. 30), the human mitochondrion[31] and bacteriophage λ (ref. 32), proposed that sequencing the human genome would greatly advance human biology. But it was thought that the cost, using existing methods, would be too high. The Okayama meeting was one response to the challenge of bringing down the cost. At that meeting, a colleague of Professor Wada's demonstrated the exquisite discrimination that could be achieved by hybridizing RNA to oligonucleotides bound to a solid support in a chromatography column[33]. Single-base mismatches were enough to destabilize the duplex formed with a 20-mer. It occurred to me that this could be scaled up by printing large numbers of oligonucleotides as small spots on a planar surface. I thought that arrays of oligonucleotide probes would find a range of applications: mutation detection by resequencing PCR products, analysis of mRNA populations and even *de novo* sequencing[34].

I was joined by a young colleague, Uwe Maskos, who developed nice chemistry for derivatizing glass so that oligonucleotides could be synthesized on the surface and would remain covalently attached to take part in hybridization reactions[35]. We needed a simple way of applying the precursors to build up different sequences in confined regions. One solution was to create reaction stencils by smearing a glass plate with silicone cement and sticking thin silicone rubber tubing to the surface (**Fig. 5**). Pressing these devices against glass plates derivatized with Uwe's linker allowed us to direct the nucleotide precursors along channels or into small cavities. In this way, using our own version of combinatorial chemistry, we were able to build up quite complex sets of oligonucleotides in only a few steps. For example, we made four copies of an array of all 256 octapurines in a simple eight-step process[36] (**Fig. 5**).

Another device allowed us to make scanning arrays[37], also known as tiling path arrays, in a simple manual or automated process. These arrays (**Fig. 6**) were used in wide-ranging studies of the processes underpinning this new technology. For example, we found that adding long linkers between the oligonucleotides and the glass greatly increased hybridization yields[38]. We found that secondary structure in RNA molecules has a large effect on hybridization yield[39]; those regions that were open to hybridization on the arrays were good targets for antisense reagents[40].

Unlike the gel-transfer method, array technology requires complex equipment for most applications, and increasingly arrays are made by companies. But arrays, too, have found a number of applications. PubMed returns more than 13,000 hits for the search term 'oligonucleotide microarray.' Most of these are for gene-expression profiling, a technique with a very wide range of applications. Particularly important in the clinical field has been the characterization of genes expressed differently in normal cells and their corresponding tumors[41]. Arrays have also being used to discover transcribed regions in genomic DNA[42]; to detect polymorphism in copy number of regions of the genome[43], which may be a new and important class of mutation; to carry out large-scale analysis of single-nucleotide polymorphisms[44]; and to analyze amplifications and deletions that are associated with oncogenic transformation and some inherited conditions[45].

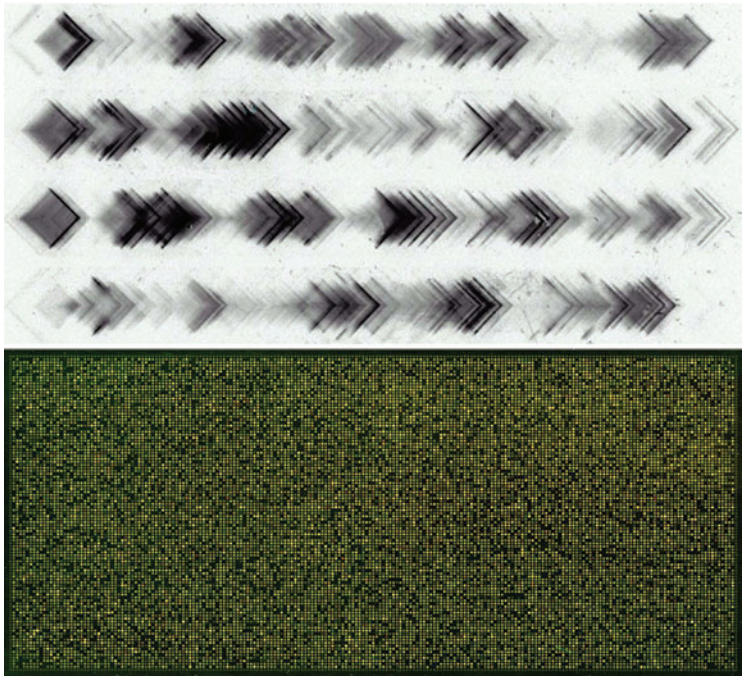

Figure 6 Genomes on arrays. An array of oligonucleotides representing complements of every position in the HIV genome was created by applying nucleotide precursors to the activated glass surface in a diamond-shaped reaction chamber, in the order in which they occur in the genome. Hybridization of labeled HIV transcripts reveals the regions in which the RNA is open to interaction with oligonucleotides (Fell, T.S. & Verhoef, K., unpublished data). The lower panel shows an array made by ink jet *in situ* synthesis at Oxford Gene Technology. The array, comprising 22,000 60-mers from chromosomes? and? was hybridized with mixed genomic DNAs from a normal male and a female with trisomy 21, labeled with red and green dyes? respectively. Intensity ratios of hybridization to probes from chromosomes Y and 21 confirmed the sensitivity of the method in comparative genomic hybridization.

From laboratory to clinic

Developed as a research tool, a number of diagnostic applications for arrays have been suggested, particularly in the cancer field. The first to be granted approval by the US Food and Drug Administration is the Roche AmpliChip for cytochrome P450. This test will help doctors determine an individual's genotype to help them determine appropriate drugs and doses to prescribe, minimizing harmful drug reactions (http://www.fda.gov/cdrh/mda/docs/k042259.html).

Retirement?

Science has brought me many rewards. It has sent me traveling the world, made me many friends and given me purpose, besides providing a decent income. I have no intention of giving all that up as I approach the official age for retirement. Oxford Gene Technology, the company I founded to take on the array technology, generously indulges my need to be involved in research and employs me in the role of Chief Scientific Officer. And, rather than spend all my time pruning fruit trees, I have taken up a new scientific interest. Some of the income from the microarray patent went into setting up a research charity, the Kirkhouse Trust. Our main focus is crop improvement in semiarid regions. This focus brings to light many important problems that can be tackled using the modern genetic methods that have been so successful in humans.

As I returned from a recent visit to Burkina Faso on Trust business, the G8 summit was about to begin in Scotland, with Africa high on the agenda. The outcomes of G8, debt relief, lifting of trade restrictions and increased aid, were hailed as successes. Regrettably, the role for science and technology in the advancement of African nations received little, if any, attention. In Africa itself, there are more encouraging signs.

Oagadougou is the capital of Burkina Faso, whose position as one of the world's poorest nations is confirmed by visitors' everyday observations. The Trust's meeting was hosted by Dr. Jeremy Ouedrago. He showed us his fine laboratory, recently refurbished by funds from the government of this poor country. The Trust spent two days in Oagadougou discussing cowpea, the 'poor man's meat,' which is an important source of high-quality protein in much of Africa. It is also used as a green vegetable and as fodder for animals. It improves soil by fixing nitrogen; its deep roots help it tolerate drought and allay soil erosion. Like many other crops of the semiarid regions, it is relatively neglected by crop scientists, particularly by the large seed companies; it is not a commodity in world markets. Can the genetic research that could improve this crop be done in Africa by African scientists? No and yes. In Oagadougou, we announced a plan to sequence the gene-rich regions of the cowpea genome, a task that is best carried out in a country with advanced facilities. Once that is done, it will provide sequences that can be used in marker-assisted breeding to hasten the introduction of desirable traits, such as resistance to pests. This is most certainly best done in Africa. Of course, African scientists work in difficult conditions. Part of the challenge is to adapt modern genetic methods to meet these conditions—plenty of scope for innovation there.

As I write, much of Europe is suffering a drought and French farmers are banned from irrigating their crops. Malaria is threatening a comeback. It is not too fanciful to speculate that in the future, as the climate changes, we in the more temperate regions may benefit from research by African scientists on tropical diseases and crops adapted to the semiarid regions.

ACKNOWLEDGMENTS

I have acknowledged in the text many of the colleagues who guided me, a chemist, into the complexities of biology and genetics and those with whom I enjoyed so many fruitful collaborations. One of the benefits of developing a method is that the inventor gets some of the credit for work done by others; it is a pleasure to acknowledge the generosity of the many scientists who have developed and shared new applications and improvements to the original method. Without their contributions, I would never have been in consideration for the prestigious Lasker Award. My research has been supported over the years by the Cancer Research Campaign (now Cancer Research UK), the Muscular Dystrophy Group and, especially, by the Medical Research Council. My family has indulged my obsession with science, and yet maintained a healthy and balanced life; I will always be grateful for that.

1. Hatch, F.T. & Mazrimas, J.A. Satellite DNAs in the kangaroo rat. *Biochim. Biophys. Acta* **12**, 291–294 (1970).
2. Meselson, M. & Stahl, F.W. The replication of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **23**, 9–12 (1958).
3. Birnstiel, M., Speirs, J., Purdom, I., Jones, K. & Loening, U.E. Properties and composition of the isolated ribosomal DNA satellite of *Xenopus laevis*. *Nature* **219**, 454–463 (1968).
4. Bishop, J.O. The effect of genetic complexity on the time-course of ribonucleic acid-deoxyribonucleic acid hybridization. *Biochem J.* **113**, 805–811 (1969).
5. Jones, K.W. Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature* **225**, 912–915 (1970).6. Walker, P.M. & Deeley, E.M. A recording microspectrophotometer. *Exp. Cell Res.***10**, 155–159 (1956).
7. Brownlee, G.G. & Sanger, F. Nucleotide sequences from the low molecular weight ribosomal RNA of *Escherichia coli*. *J. Mol. Biol.* **23**, 337–353 (1967).
8. Southern, E.M. Base sequence and evolution of guinea-pig alpha-satellite DNA. *Nature* **227**, 794–798 (1970).
9. Kelly, T.J. Jr. & Smith, H.O. A restriction enzyme from Hemophilus influenzae. II. *J. Mol. Biol.* **51**, 393–409 (1970).
10. Southern, E.M. & Roizes, G. The action of a restriction endonuclease on higher organism DNA. *Cold Spring Harb. Symp. Quant. Biol.* **38**, 429–33 (1974).
11. Southern, E.M. Long range periodicities in mouse satellite DNA. *J. Mol. Biol.* **94**, 51–69 (1975).
12. Orgel, L.E. & Crick, F.H. Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607 (1980).
13. Ford, P.J. Non-coordinated accumulation and synthesis of 5S ribonucleic acid by ovaries of Xenopus laevis. *Nature* **233**, 561–564 (1971).
14. Ford, P.J. & Southern, E.M. Different sequences for 5S RNA in kidney cells and ovaries of Xenopus laevis. *Nat. New Biol.* **241**, 7–12 (1973).
15. Gillespie, D. & Spiegelman, S. A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. *J. Mol. Biol.* **12**, 829–842 (1965).
16. Shinnick, T.M., Lund, E., Smithies, O. & Blattner, F.R. Hybridization of labeled RNA to DNA in agarose gels. *Nucleic Acids Res.* **2**, 1911–1929 (1975).
17. Southern, E.M. Blotting at 25. *Trends Biochem. Sci.* **25**, 585–588 (2000).
18. Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517 (1975).
19. Arnheim, N. & Southern, E.M. Heterogeneity of the ribosomal genes in mice and men. *Cell* **11**, 363–370 (1977).
20. Botchan, M., Topp, W. & Sambrook, J. The arrangement of simian virus 40 sequences in the DNA of transformed cells. *Cell* **9**, 269–287 (1976).
21. Jeffreys, A.J. & Flavell, R.A. The rabbit β-globin gene contains a large insert in the coding sequence. *Cell* **12**, 1097–1108 (1977).
22. Jeffreys, A.J., Wilson, V. & Thein, S.L. Individual-specific 'fingerprints' of human DNA. *Nature* **316**, 76–79 (1985).
23. Roeder, R.G. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen.*Nat. Med.* **9**, 1239–1244 (2003).
24. Kan, Y.W. & Dozy, A.M. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc. Natl. Acad. Sci. USA* **75**, 5631–5635 (1978).
25. Bostein, D., White, R.L., Skolnick, M. & Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331 (1980).
26. Schwartz, D.C. & Cantor, C.R. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75 (1984).
27. Southern, E.M., Anand, R., Brown, W. R. & Fletcher, D.S. A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Res.* **15**, 5925–5943 (1987).
28. Anand, R., Honeycombe, J., Whittaker, P.A., Elder, J.K. & Southern, E.M. Clones from an 840-kb fragment containing the 5′ region of the DMD locus enriched by pulsed field gel electrophoresis. *Genomics* **3**, 177–186 (1988).
29. Tyler-Smith, C. & Brown, W. R. Structure of the major block of alphoid satellite DNA on the human Y chromosome *J. Mol. Biol.* **195**, 457–470 (1987).
30. Sanger, F. *et al.* The nucleotide sequence of bacteriophage phiX174. *J. Mol. Biol.* **125**, 225–246 (1978).
31. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature.* **290,** 457–465 (1981).
32. Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. & Petersen, G.B. Nucleotide sequence of bacteriophage lambda DNA. *J. Mol. Biol.* **162**, 729–773 (1982).33.
33. Tsurui, H., Suyama, A. & Wada, A. Temperature gradient DNA-probe column chromatography: as tools for detection and purification of particular DNAs and RNAs. *Nucleic Acids Symp. Ser.* **19**, 49–52 (1988).
34. Southern, E.M. Apparatus and method for analyzing polynucleotide sequences and method of generating oligonucleotide arrays. US patent 5,700,637 (1997).
35. Maskos, U.. & Southern, E.M. Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. *Nucleic Acids Res.* **20**, 1679–1684 (1992).
36. Southern, E.M., Maskos, U. & Elder, J.K. Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. *Genomics* **13**, 1008–1017 (1992).
37. Southern, E.M. *et al.* Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. *Nucleic Acids Res.* **22,** 1368–1373 (1994).
38. Schepinov, M.S., Case-Green. S.C. & Southern, E.M. Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. *Nucleic Acids Res.* **25,** 1155–1161 (1997).
39. Sohail, M., Akhatn, S. & Southern, E.M. The folding of large RNAs studied by hybridization to arrays of complementary oligonucleotides. *RNA* **5**, 646–655 (1999).
40. Sohail, M. *et al.* Antisense oligonucleotides selected by hybridisation to scanning arrays are effective reagents *in vivo*. *Nucleic Acids Res.* **29**, 2041–2051 (2001).
41. Golub,T.R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
42. Hughes, T.R. *et al.* Experimental annotation of the human genome using microarray technology. *Nature.* **409**, 922–927 (2001).
43. Lucito, R. *et al.* Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.* **13**, 2291–2305 (2003)**.**
44. Kennedy, G.C. *et al.* Large-scale genotyping of complex DNA**.** *Nat. Biotechnol.* **21**, 1233–1237 (2003).
45. Anand, R. and Southern, E. M. Pulsed field gel electrophoresis. In *Gel Electrophoresis of Nucleic Acids* (eds Rickwood, D. & Hames, B.D.) 101–123 (IRL Press, Oxford, 1990).

Edwin Southern is Professor of Biochemistry at the Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.
e-mail: ed.southern@bioch.ox.ac.uk