# CHAPTER 4

THE FLOWERING OF THE AGE OF BIOTECHNOLOGY
1990–2000

The last decade of the 20th century was fittingly enough ushered in by another variation on the DNA race once more headed by the eponymous symbol of the age of DNA, James Watson. This decade of the genome began with the official launch of the Human Genome Project (HGP), the international effort to map all of the genes in the human body. The "Father of DNA" helped win funding from Congress by mollifying critics who deemed the project overcentralized Big Science of dubious practical value. The National Human Genome Research Institute (NHGRI) was given institute status at the National Institutes of Health as necessary for NHGRI's director to coordinate genome research with other projects at NIH. The actual term "Genomics" appeared for the first time in 1986 to describe the discipline of mapping, sequencing, and analyzing genes. The term was coined by Thomas
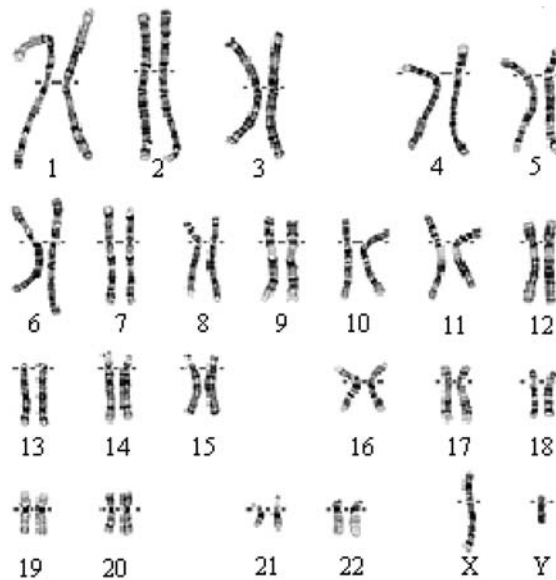
Roderick as a name for a new journal that he was developing. Following Chancellor Robert Sinsheimer's meeting in 1985 at UC Santa Cruz, to discuss the feasibility of sequencing the human genome, for much of the second half of the previous decade the merits of a human genome project were hotly debated (Leslie, 2001).

The jump-off event was at a meeting in 1986 at Watson's Cold Spring Harbor Laboratory appropriately titled "The Molecular Biology of *Homo sapiens*." In an influential editorial in Science in March 1986, Nobel laureate Renato Dulbecco discussed the potential of whole-genome sequencing for cancer research. About the same time, Charles DeLisi held a workshop to consider the plausibility of a concerted "crash program" to decode the human genome. As head of the office of Health and Environmental Research at the Department of Energy (DOE), DeLisi proposed and soon sought funding for the first stage of such a project. In the same year Sydney Brenner of the Medical Research Council (MRC) and father of the elegant "worm" *Caenorhabditis elegans, C. elegans* urged the European Union to undertake a concerted program to map and sequence the human genome and, being the entrepreneurial type started a small genome initiative at the MRC. When his former student and colleague John Sulston (along with Waterston, and Coulson) presented a genome map of *C. elegans* at a Cold Spring Harbor meeting in 1989 this result spurred efforts to sequence the genome as a model for the human project. In fact when Alan Coulson tacked up the worm genome map, Jim Watson after viewing it is reported to have said, 'You can't see this without wanting to sequence it'. According to Waterston the next day Watson agreed to consider the worm genome for the so-called 'security council' of the Human Genome Project. (Nature S1, 2006) The two were to share the 2002 Nobel Prize for disclosing the worm's secrets of reproduction and development (Sulston and Ferry, 2002).

Almost immediately it was decided to go forward with the decision sparking controversy stateside. A decade before Craig Venter became the bete noir of biotech by planning to go it alone without the public coffers, Walter Gilbert, perhaps still smarting from the insulin debacle of the preceding decade, resigned from the U.S. National Research Council (NRC) genome panel and announced plans to start Genome Corp., with the goal of sequencing and copyrighting the human genome and selling data for profit. A year later in a move indicative of the race having begun, Helen Donis-Keller and colleagues at Collaborative Research Inc. using restriction fragment length polymorphism published the "first" human genetic map with 403 markers, sparking a fight over credit and priority (Green et al., 1989).

That same year the race moved to a much more exalted platform when an advisory panel suggested that DOE should spend $1 billion on mapping and sequencing the human genome over the next 7 years and that DOE should lead the U.S. effort. And so it began, for the first half of the team. In 1987 a pivotal report was released by the National Research Council (NRC), of the National Academy of Sciences, from a committee that included former skeptics. Rather than a "crash program," the NRC suggested a phased approach with long-term, government-support and specific developmental milestones with a rapid scale-up to $200 million a year of new money. Although sequencing the genome remained the goal, the report underscored

the significance of developing genetic and physical maps of the genome, and the importance of comparing the human genome with those of other species. It also suggested a preliminary focus on improving current technology. At the request of the U.S. Congress, the Office of Technology Assessment (OTA) also studied the issue, and issued a document in 1987 – within days of the NRC report – that was similarly supportive. The OTA report discussed, in addition to scientific issues, social and ethical implications of a genome program together with problems of managing funding, negotiating policy and coordinating research efforts.

Prompted by advisers at a 1988 meeting in Reston, Virginia, James Wyngaarden, then director of the National Institutes of Health (NIH), decided that the agency should be a major player in the HGP, effectively seizing the lead from DOE. The start of the joint effort was in May 1990 (with an "official" start in October) when a 5-year plan detailing the goals of the U.S. Human Genome Project was presented to members of congressional appropriations committees in mid-February. This document co-authored by DOE and NIH and titled "Understanding Our Genetic Inheritance, the U.S. Human Genome Project: The First Five Years" examined the then current state of genome science. The plan also set forth complementary approaches of the two agencies for attaining scientific goals and presented plans for administering research agenda; it described collaboration between U.S. and international agencies and presented budget projections for the project.

According to the document, "a centrally coordinated project, focused on specific objectives, is believed to be the most efficient and least expensive way" to obtain the 3-billion base pair map of the human genome. In the course of the project, especially in the early years, the plan stated that "much new technology will be developed that will facilitate biomedical and a broad range of biological research, bring down the cost of many experiments (mapping and sequencing), and finding applications in numerous other fields." The plan built upon the 1988 reports of the Office of Technology Assessment and the National Research Council on mapping and sequencing the human genome. "In the intervening two years," the document said, "improvements in technology for almost every aspect of genomics research have taken place. As a result, more specific goals can now be set for the project."

The document describes objectives in the following areas mapping and sequencing the human genome and the genomes of model organisms; data collection and distribution; ethical, legal, and social considerations; research training; technology development; and technology transfer. These goals were to be reviewed each year and updated as further advances occured in the underlying technologies. They identified the overall budget needs to be the same as those identified by OTA and NRC, namely about $200 million per year for approximately 15 years. This came to $13 billion over the entire period of the project. Considering that in July 1990, the DNA databases contained only seven sequences greater than 0.1 Mb this was a major leap of faith.

This approach was a major departure from the single-investigator-based gene of interest focus that research took hitherto. This sparked much controversy both before and after its inception. Critics questioned the usefulness of genomic sequencing,

they objected to the high cost and suggested it might divert funds from other, more focused, basic research. The prime argument to support the latter position is that there appeared to be are far less genes than accounted for by the mass of DNA which would suggest that the major part of the sequencing effort would be of long stretches of base pairs with no known function, the so-called "junk DNA." And that was in the days when the number of genes was presumed to be 80–100,000. If, at that stage, the estimated number was guessed to be closer to the actual estimate of 35–40,000 (later reduced to 20–25,000) this would have made the task seem even more foolhardy and less worthwhile to some. However, the ever-powerful incentive of new diagnostics and treatments for human disease beyond what could be gleaned from the gene-by-gene approach and the rapidly evolving technologies, especially that of automated sequencing, made it both an attractive and plausible aim.

Charles Cantor (1990), a principal scientist for the Department of Energy's genome project contended that DOE and NIH were cooperating effectively to develop organizational structures and scientific priorities that would keep the project on schedule and within its budget. He noted that there would be small short-term costs to traditional biology, but that the long-term benefits would be immeasurable.

Genome projects were also discussed and developed in other countries and sequencing efforts began in Japan, France, Italy, the United Kingdom, and Canada. Even as the Soviet Union collapsed, a genome project survived as part of the Russian science program. The scale of the venture and the manageable prospect for pooling data via computer made sequencing the human genome a truly international initiative. In an effort to include developing countries in the project UNESCO assembled an advisory committee in 1988 to examine UNESCO's role in facilitating international dialogue and cooperation. A privately-funded Human Genome Organization (HUGO) had been founded in 1988 to coordinate international efforts and serve as a clearinghouse for data. In that same year the European Commission (EC) introduced a proposal entitled the "Predictive Medicine Programme." A few EC countries, notably Germany and Denmark, claimed the proposal lacked ethical sensitivity; objections to the possible eugenic implications of the program were especially strong in Germany (Dickson 1989). The initial proposal was dropped but later modified and adopted in 1990 as the "Human Genome Analysis Programme" (Dickman and Aldhous 1991). This program committed substantial resources to the study of ethical issues. The need for an organization to coordinate these multiple international efforts quickly became apparent. Thus the Human Genome Organization (HUGO), which has been called the "U.N. for the human genome," was born in the spring of 1988. Composed of a founding council of scientists from seventeen countries, HUGO's goal was to encourage international collaboration through coordination of research, exchange of data and research techniques, training, and debates on the implications of the projects (Bodmer 1991).

In August 1990 NIH began large-scale sequencing trials on four model organisms: the parasitic, cell-wall lacking pathogenic microbe *Mycoplasma capricolum*, the prokaryotic microbial lab rat *Escherichia coli*, the most simple animal *Caenorhabditis elegans*, and the eukaryotic microbial lab rat *Saccharomyces cerevisiae*. Each

research group agreed to sequence 3 megabases (Mb) at 75 cents a base within 3 years. A sub living organism was actually fully sequenced and the complete sequence of that genome, the human cytomegalovirus (HCMV) genome was 0.23 Mb.

That year also saw the casting of the first salvo in the protracted debate on "ownership" of genetic information beginning with the more tangible question of ownership of cells. And, as with the debates of the early eighties, which were to be revisited later in the nineties, the respondent was the University of California. Moore v. Regents of the University of California was the first case in the United States to address the issue of who owns the rights to an individual's cells. Diagnosed with leukemia, John Moore had blood and bone marrow withdrawn for medical tests. Suspicious of repeated requests to give samples because he had already been cured, Moore discovered that his doctors had patented a cell line derived from his cells and so he sued. The California Supreme Court found that Moore's doctor did not obtain proper informed consent, but, however, they also found that Moore cannot claim property rights over his body.

1. NASCENT TECHNOLOGIES

The quest for the holy grail of the human genome was both inspired by the rapidly evolving technologies for mapping and sequencing and subsequently spurred on the development of ever more efficient tools and techniques. Advances in analytical tools, automation, and chemistries as well as computational power and algorithms revolutionized the ability to generate and analyze immense amounts of DNA sequence and genotype information. In addition to leading to the determination of the complete sequences of a variety of microorganisms and a rapidly increasing number of model organisms, these technologies have provided insights into the repertoire of genes that are required for life, and their allelic diversity as well as

their organization in the genome. But back in 1990 many of these were still nascent technologies.

The technologies required to achieve this end could be broadly divided into three categories: equipment, techniques, and computational analysis. These are not truly discrete divisions and there was much overlap in their influence on each other.

2. EQUIPMENT

As noted, Lloyd Smith, Michael and Tim Hunkapiller, and Leroy Hood conceived the automated sequencer and Applied Biosystems Inc. brought it to market in June 1986. There is no much doubt that when Applied Biosystems Inc. put it on the market that which had been a dream became decidedly closer to an achievable reality. In automating Sangers chain termination sequencing system, Hood modified both the chemistry and the data-gathering processes. In the sequencing reaction itself, he replaced radioactive labels, which were unstable, posed a health hazard, and required separate gels for each of the four bases. Hood developed chemistry that used fluorescent dyes of different colors for each of the four DNA bases. This system of "color-coding" eliminated the need to run several reactions in overlapping gels. The fluorescent labels addressed another issue which contributed to one of the major concerns of sequencing – data gathering. Hood integrated laser and computer technology, eliminating the tedious process of information-gathering by hand. As the fragments of DNA passed a laser beam on their way through the gel the fluorescent labels were stimulated to emit light. The emitted light was transmitted by a lens and the intensity and spectral characteristics of the fluorescence are measured by a photomultiplier tube and converted to a digital format that could be read directly into a computer. During the next thirteen years, the machine was constantly improved, and by 1999 a fully automated instrument could sequence up to 150,000,000 base pairs per year.

In 1990 three groups came up with a variation on this approach. They developed what is termed capillary electrophoresis, one team was led by Lloyd Smith (Luckey, 1990), the second by Barry Karger (Cohen, 1990), and the third by Norman Dovichi. In 1997 Molecular Dynamics introduced the MegaBACE, a capillary sequencing machine. And not to be outdone the following year in 1998, the original of the species came up with the ABI Prism 3700 sequencing machine. The 3700 is also a capillary-based machine designed to run about eight sets of 96 sequence reactions per day.

3. TECHNIQUES

On the biology side, one of the biggest challenges was the construction of a physical map to be compiled from many diverse sources and approaches in such a way as to insure continuity of physical mapping data over long stretches of DNA. The development of DNA Sequence Tagged Sites (STSs) to correlate diverse types of DNA clones aided this standardization of the mapping component by providing

Chromosomal   Linkage   Physical

Long Range   Short Range
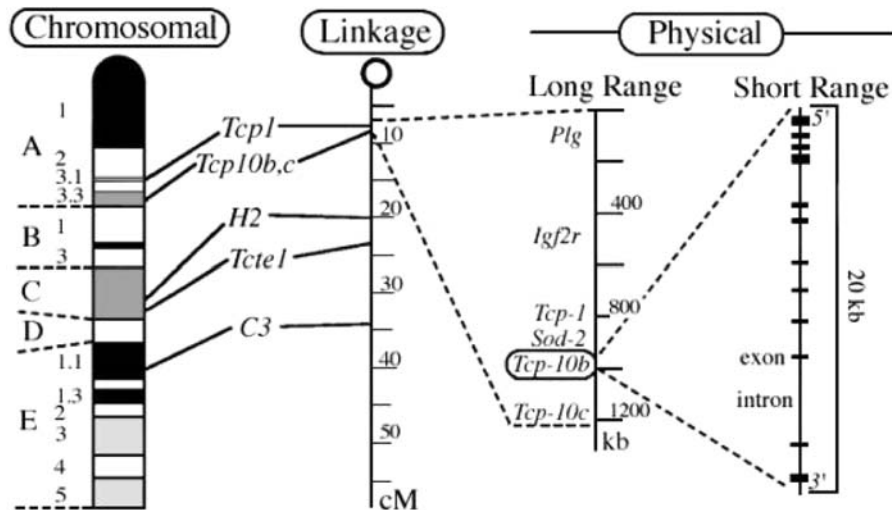
mappers with a common language and a system of landmarks for all the libraries from such varied sources as cosmids, yeast artificial chromosomes (YACs) and other rDNAs clones. This way each mapped element (individual clone, contig, or sequenced region) would be defined by a unique STS. A crude map of the entire genome, showing the order and spacing of STSs, could then be constructed. The order and spacing of these unique identifier sequences composing an STS map was made possible by development of Mullis' polymerase chain reaction (PCR), which allows rapid production of multiple copies of a specific DNA fragment, for example, an STS fragment.

Sequence information generated in this way could be recalled easily and, once reported to a database, would be available to other investigators. With the STS sequence stored electronically, there would be no need to obtain a probe or any other reagents from the original investigator. No longer would it be necessary to exchange and store hundreds of thousands of clones for full-scale sequencing of the human genome-a significant saving of money, effort, and time. By providing a common language and landmarks for mapping, STS's allowed genetic and physical maps to be cross-referenced.

With a refinement on this technique to go after actual genes, Sydney Brenner proposed sequencing human cDNAs to provide rapid access to the genes stating that 'One obvious way of finding at least a large part of the important [fraction] of the human genome is to look at the sequences of the messenger RNA's of expressed genes' (Brenner, 1990). The following year the man who was to play a pivotal role on the world stage that became the human genome project suggested a way to implement Sydney's approach. That player, NIH biologist J. Craig Venter announced a strategy to find expressed genes, using ESTs (Expressed Sequence Tag) (Adams, 1991).

These so called ESTs represent a unique stretch of DNA within a coding region of a gene, which as Sydney suggested would be useful for identifying full-length genes and as a landmark for mapping. So using this approach projects were begun to mark gene sites on chromosome maps as sites of mRNA expression. To help with this a more efficient method of handling large chunks of sequences was needed and two approaches were developed. Yeast artificial chromosomes, which were developed by David Burke, Maynard Olson, and George Carle, increased insert size 10-fold (David T. Burke et al., 1987). Caltech's second major contribution to the genome project was developed by Melvin Simon, and Hiroaki Shizuya. Their approach to handling large DNA segments was to develop "bacterial artificial chromosomes" (BACs), which basically allow bacteria to replicate chunks greater than 100,000 base pairs in length. This efficient production of more stable, large-insert BACs made the latter an even more attractive option, as they had greater flexibility than YACs. In 1994 in a collaboration that presages the SNP Consortium, Washington University, St Louis MO, was funded by the pharmaceutical company Merck and the National Cancer Institute to provide sequence from those ESTs. More than half a million ESTs were submitted during the project (Murr L et al., 1996).

4. ANALYTICAL TOOLS

On the analysis side was the major challenge to manage and mine the vast amount of DNA sequence data being generated. A rate-limiting step was the need to develop semi-intelligent algorithms to achieve this Herculean task. This is where the discipline of bioinformatics came into play. It had been evolving as a discipline since Margaret Oakley Dayhoff used her knowledge of chemistry, mathematics, biology and computer science to develop this entirely new field in the early sixties. She is in fact credited today as a founder of the field of bioinformatics in which biology, computer science, and information technology merge into a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. There are three important sub-disciplines within bioinformatics: the development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information.

Paralleling the rapid and very public ascent of recombinant DNA technology during the previous two decades, the analytic and management tools of the discipline that was to become bioinformatics evolved at a more subdued but equally impressive pace. Some of the key developments included tools such as the Needleman-Wunsch algorithm for sequence comparison which appeared even before recombinant DNA technology had been demonstrated as early as 1970; the Smith-Waterman algorithm for sequence alignment (1974); the FASTP algorithm (1985) and the FASTA algorithm for sequence comparison by Pearson and Lupman in 1988 and Perl

(Practical Extraction Report Language) released by Larry Wall in 1987. On the data management side several databases with ever more effective storage and mining capabilities were developed over the same period. The first bioinformatic/biological databases were constructed a few years after the first protein sequences began to become available. The first protein sequence reported was that of bovine insulin in 1956, consisting of 51 residues. Nearly a decade later, the first nucleic acid sequence was reported, that of yeast alanine tRNA with 77 bases. Just one year later, Dayhoff gathered all the available sequence data to create the first bioinformatic database. One of the first dedicated databases was the Brookhaven Protein DataBank whose collection consisted of ten X-ray crystallographic protein structures (Acta. Cryst. B, 1973). The year 1982 saw the creation of the Genetics Computer Group (GCG) as a part of the University of Wisconsin Biotechnology Center. The group's primary and much used product was the Wisconsin Suite of molecular biology tools. It was spun off as a private company in 1989. The SWISS-PROT database made its debut in 1986 in Europe at the Department of Medical Biochemistry of the University of Geneva and the European Molecular Biology Laboratory (EMBL).

The first dedicated "bioinformatics" company IntelliGenetics, Inc. was founded in California in 1980. Their primary product was the IntelliGenetics Suite of programs for DNA and protein sequence analysis. The first unified federal effort, the National Center for Biotechnology Information (NCBI) was created at NIH/NLM in 1988 and it was to play a crucial part in coordinating public databases, developing software tools for analyzing genome data, and disseminating information. And on the other side of the Atlantic, Oxford Molecular Group, Ltd. (OMG) was founded in Oxford, UK by Anthony Marchington, David Ricketts, James Hiddleston, Anthony Rees, and W. Graham Richards. Their primary focus was on rational drug design and their products such as Anaconda, Asp, and Chameleon obviously reflected this as they were applied in molecular modeling, and protein design engineering.

Within two years NCBI were making their mark when David Lipman, Eugene Myers, and colleagues at the NCBI published the Basic Local Alignment Search Tool BLAST algorithm for aligning sequences (Altschul et al., 1990). It is used to compare a novel sequence with those contained in nucleotide and protein databases by aligning the novel sequence with previously characterized genes. The emphasis of this tool is to find regions of sequence similarity, which will yield functional and evolutionary clues about the structure and function of this novel sequence. Regions of similarity detected via this type of alignment tool can be either local, where the region of similarity is based in one location, or global, where regions of similarity can be detected across otherwise unrelated genetic code. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP). An HSP consists of two sequence fragments of arbitrary but equal length whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score. This system has been refined and modified over the years the two principal variants presently in use being the NCBI BLAST and WU-BLAST (WU signifying Washington University).

The same year that BLAST was launched two other bioinformatics companies were launched. One was InforMax in Bethesda, MD whose products addressed sequence analysis, database and data management, searching, publication graphics, clone construction, mapping and primer design. The second, Molecular Applications Group in California, was to play a bigger part on the proteomics end (Michael Levitt and Chris Lee). Their primary products were Look and SegMod which are used for molecular modeling and protein design. The following year in 1991 the Human chromosome mapping data repository, Genome Data Base (GDB) was established. On a more global level, the development of computational capabilities in general and the Internet in specific was also to play a considerable part in the sharing of data and access to databases that rendered the rapidity of the forward momentum of the HGP possible. Also in 1991 Edward Uberbacher of Oak Ridge National Laboratory in Tennessee developed GRAIL, the first of many gene-finding programs.

In 1992 the first two "genomics" companies made their appearance. Incyte Pharmaceuticals, a genomics company headquartered in Palo Alto, California, was formed and Myriad Genetics, Inc. was founded in Utah. Incyte's stated goal was to lead in the discovery of major common human disease genes and their related pathways. The company discovered and sequenced, with its academic collaborators (originally Synteni from Pat Brown's lab at Stanford), a number of important genes including BRCA1 and BRCA2, with Mary Claire King, epidemiologist at UC-Berkeley, the genes linked to breast cancer in families with a high degree of incidence before age 45. By 1992 a low-resolution genetic linkage map of the entire human genome was published and U.S. and French teams completed genetic maps of both mouse and man. The mouse with an average marker spacing of 4.3 cM as determined by Eric Lander and colleagues at Whitehead and the human, with an average marker spacing of 5 cM by Jean Weissenbach and colleagues at CEPH (Centre d'Etude du Polymorphisme Humaine). The latter institute was the subject of a rather scathing book by Paul Rabinow (1999) based on what they did with this genome map. In 1993, an American biotechnology company, Millennium Pharma-ceuticals, and the CEPH, developed plans for a collaborative effort to discover diabetes genes. The results of this collaboration could have been medically signif-icant and financially lucrative. The two parties had agreed that CEPH would supply Millennium with germplasm collected from a large coterie of French families, and Millennium would supply funding and expertise in new technologies to accelerate the identification of the genes, terms to which the French government had agreed. But in early 1994, just as the collaboration was to begin, the French government cried halt! The government explained that the CEPH could not be permitted to give the Americans that most precious of substances for which there was no precedent in law – French DNA. Rabinow's book discusses the tangled relations and conceptions such as, can a country be said to have its own genetic material, the first but hardly the last Franco-American disavowal of détente (Paul Rabinow, 1999).

The latest facilities such as the Joint Genome Institute (JGI), Walnut Creek, CA are now able to sequence up to 10Mb per day which makes it possible to sequence whole microbial genomes within a day. Technologies currently under

development will probably increase this capacity yet further through massively parallel sequencing and/or microfluidic processing making it possible to sequence multiple genotypes from several species.

5. AND THE BEAT GOES ON

Nineteen ninety-two saw one of the first shakeups in the progress of the HGP. That was the year that the first major outsider entered the race when Britain's Wellcome Trust plunked down $95 million to join the HGP. This caused a mere ripple while the principal shake-ups occurred stateside. Much of the debate and subsequently the direction all the way through the HGP process was shaped by the personalities involved. As noted the application of one of the innovative techniques, namely ESTs, to do an end run on patenting introduced one of those major players to the fray, Craig Venter. Venter, the high school drop out who reached the age of majority in the killing fields of Vietnam was to play a pivotal role in a more "civilized" but no less combative field of human endeavor.

He came onto the world stage through his initial work on ESTs while at the National Institute of Neurological Disorders and Stroke (NINDS) from 1984 to 1992. He noted in an interview with The Scientist magazine in 1995, that there was a degree of ambiguity at NINDS about his venturing into the field of genomics, while they liked the prestige of hosting one of the leaders and innovators in his newly emerging field, they were concerned about him moving outside the NIND purview of the human brain and nervous system. Ultimately, while he proclaimed to like the security and service infrastructure this institute afforded him, that same system became too restrictive for his interests and talent. He wanted the whole canvas of human-gene expression to be his universe, not just what was confined to the central nervous system. He was becoming more interested in taking a whole genome approach to understanding the overall structure of genomes and genome evolution, which was much broader than the mission of NINDS. He noted, with some irony, in later years that the then current NIH director Harold Varmus had wished in hindsight that NIH had pushed to do a similar database in the public domain, clearly in Venter's opinion Varmus was in need of a refresher course in history!

Bernadine Healy, NIH director in 1994, was one of the few in a leadership role who saw the technical and fiscal promise of Venter's work and, like all good administrators, it also presented an opportunity to resolve a thorny "personnel" issue. She appointed him head of the ad hoc committee to have an intramural genome program at NIH to give the head of the HGP (that other larger than life personality Jim Watson) notice that he was not the sole arbitrator of the direction for the Human Genome Project. However Venter very soon established himself as an equally non-conformist character and with the tacit consent of his erstwhile benefactor.

He initially assumed the mantle of a non-conformist through guilt by association rather than direct actions when it was revealed that NIH was filing patent applications on thousands of these partial genes based on his ESTs catalyzing the first

HGP fight at a congressional hearing. NIH's move was widely criticized by the scientific community because, at the time, the function of genes associated with the partial sequences was unknown. Critics charged that patent protection for the gene segments would forestall future research on them. The Patent Office eventually rejected the patents, but the applications sparked an international controversy over patenting genes whose functions were still unknown.

Interestingly enough despite NIH's reliance on the EST/cDNA technique, Venter, who was now clearly venturing outside the NINDS mandated rubric, could not obtain government funding to expand his research, prompting him to leave NIH in 1992. He moved on to become president and director of The Institute for Genomic Research (TIGR), a nonprofit research center based in Gaithersburg, Md. At the same time William Haseltine formed a sister company, Human Genome Sciences (HGS), to commercialize TIGR products. Venter continued EST work at TIGR, but also began thinking about sequencing entire genomes. Again, he came up with a quicker and faster method: whole genome shotgun sequencing. He applied for an NIH grant to use the method on *Hemophilus influenzae*, but started the project before the funding decision was returned. When the genome was nearly complete, NIH rejected his proposal saying the method would not work. In a triumphal flurry in late May 1995 and with a metaphorical nose-thumbing at his recently rejected "unworkable" grant Venter announced that TIGR and collaborators had fully sequenced the first free-living organism – *Haemophilus influenzae*. In November 1994, controversy surrounding Venter's research escalated. Access restrictions associated with a cDNA database developed by TIGR and its Rockville, Md.–based biotech associate, Human Genome Sciences (HGS) Inc. – including HGS's right to preview papers on resulting discoveries and for first options to license products – prompted Merck and Co. Inc. to fund a rival database project. In that year also Britain "officially" entered the HGP race when the Wellcome Trust trumped down $95 million (as mentioned earlier).

The following year HGS was involved in yet another patenting debacle forced by the rapid march of technology into uncharted patent law territory. On June 5, 1995 HGS applied for a patent on a gene that produces a "receptor" protein that is later called CCR5. At that time HGS has no idea that CCR5 is an HIV receptor. In December 1995, U.S. researcher Robert Gallo, the co-discoverer of HIV, and colleagues found three chemicals that inhibit the AIDS virus but they did not know how the chemicals work. In February 1996, Edward Berger at the NIH discovered that Gallo's inhibitors work in late-stage AIDS by blocking a receptor on the surface of T-cells. In June of that year in a period of just 10 days, five groups of scientists published papers saying CCR5 is the receptor for virtually all strains of HIV. In January 2000, Schering-Plough researchers told a San Francisco AIDS conference that they have discovered new inhibitors. They knew that Merck researchers had made similar discoveries. As a significant Valentine in 2000 the U.S. Patent and Trademark Office (USPTO) grants HGS a patent on the gene that makes CCR5 and on techniques for producing CCR5 artificially. The decision sent HGS stock

flying and dismayed researchers. It also caused the USPTO to revise its definition of a "patentable" drug target.

In the meantime Haseltine's partner in rewriting patenting history, Venter turned his focus to the human genome. He left TIGR and started the for-profit company Celera, a division of PE Biosystems, the company that at times, thanks to Hood and Hunkapillar, led the world in the production of sequencing machines. Using these machines, and the world's largest civilian supercomputer, Venter finished assembling the human genome in just three years.

Following the debacle with the then NIH director Bernine Healy over patenting the partial genes that resulted from EST analysis, another major personality-driven event in that same year occurred. Watson strongly opposed the idea of patenting gene fragments fearing that it would discourage research, and commented that "the automated sequencing machines 'could be run by monkeys.'" (Nature June 29, 2000) with this dismissal Watson resigned his NIH NCHGR post in 1992 to devote his full-time effort to directing Cold Spring Harbor Laboratory. His replacement was of a rather more pragmatic, less flamboyant nature.

While Venter maybe was described as an idiosyncratic Shogun of the shotgun, Francis Collins was once described as the King Arthur of the Holy Grail that is the Human Genome Project. Collins became the Director of the National Human Genome Research Institute in 1993. He was considered the right man for the job following his 1989 success (along with Lap-Chee Tsui) in identifying the gene for the cystic fibrosis transmembrane (CFTR) chloride channel receptor that, when mutated, can lead to the onset of cystic fibrosis. Although now indelibly connected with the topic non-plus tout in biology, like many great innovators in this field before him, Francis Collins had little interest in biology as he grew up on a farm in the Shenandoah Valley of Virginia. From his childhood he seemed destined to be at the center of drama, his father was professor of dramatic arts at Mary Baldwin College and the early stage management of career was performed on a stage he built on the farm. While the physical and mathematical sciences held appeal for him, being possessed of a highly logical mind, Collins found the format in which biology was taught in the high school of his day mind-numbingly boring, filled with dissections and rote memorization. He found the contemplation of the infinite outcomes of dividing by zero (done deliberately rather than by accident as in Einstein's case) far more appealing than contemplating the innards of a frog.

That biology could be gloriously logical only became clear to Collins when, in 1970, he entered Yale with a degree in chemistry from the University of Virginia and was first exposed to the nascent field of molecular biology. Anecdotally it was the tome, the Book of Life, penned by the theoretical physicist father of molecular biology, Edwin Schrodinger, while exiled in Trinity College Dublin in 1942 that was the catalyst for his conversion. Like Schrodinger he wanted to do something more obviously meaningful (for less than hardcore physicists at least!) than theoretical physics, so he went to medical school at UNC-Chapel Hill after completing his chemistry doctorate in Yale, and returned to the site of his road to Damascus for post-doctoral study in the application of his newfound interest in human genetics.

During this sojourn at Yale, Collins began working on developing novel tools
to search the genome for genes that cause human disease. He continued this work,
which he dubbed "positional cloning," after moving to the University of Michigan
as a professor in 1984. He placed himself on the genetic map when he succeeded
in using this method to put the gene that causes cystic fibrosis on the physical map.
While a less colorful-in-your-face character than Venter he has his own personality
quirks, for example, he pastes a new sticker onto the back of his motorcycle helmet
every time he finds a new disease gene. One imagines that particular piece of really
estate is getting rather crowded.

Interestingly it was not these four hundred pound US gorillas who proposed the
eventually prescient timeline for a working draft but two from the old power base.
In meetings in the US in 1994, John Sulston and Bob Waterston proposed to produce
a 'draft' sequence of the human genome by 2000, a full five years ahead of schedule.
While agreed by most to be feasible it meant a rethinking of strategy and involved
focusing resources on larger centers and emphasizing sequence acquisition. Just
as important, it asserts the value of draft quality sequence to biomedical research.
Discussion started with the British based Wellcome Trust as possible sponsors
(Marshall E. 1995).

By 1995 a rough draft of the human genome map was produced showing the
locations of more than 30,000 genes. The map was produced using yeast artificial
chromosomes and some chromosomes – notably the littlest 22 – were mapped in
finer detail. These maps marked an important step toward clone-based sequencing.
The importance was illustrated in the devotion of an entire edition of the journal
Nature to the subject. (Nature 377: 175–379 1995)

The duel between the public and private face of the HGP progressed at a pace
over the next five years. Following release of the mapping data some level of
international agreement was decided on sequence data release and databases. They
agreed on the release of sequence data, specifically, that Primary Genomic Sequence
should be in the Public Domain to encourage research and development to maximize
its benefit to society. Also that it be rapidly released on a daily basis with assemblies
of greater than 1 Kb and that the finished annotated sequence should be submitted
immediately to the public databases.

In 1996 an international consortium completed the sequence of the genome of the
workhorse yeast *Saccharomyces cerevisiae.* Data had been released as the individual
chromosomes were completed. The Saccharomyces Genome Database (SGD) was
created to curate this information. The project collects information and maintains a
database of the molecular biology of *S. cerevisiae*. This database includes a variety
of genomic and biological information and is maintained and updated by SGD
curators. The SGD also maintains the S. cerevisiae Gene Name Registry, a complete
list of all gene names used in *S. cerevisiae.*

In 1997 a new more powerful diagnostic tool termed SNPs (Single Nucleotide
Polymorphisms) was developed. SNPs are changes in single letters in our DNA
code that can act as markers in the DNA landscape. Some SNPs are associated
closely with susceptibility to genetic disease, our response to drugs or our ability to

remove toxins. The SNP Consortium although designated a limited company is a non-profit foundation organized for the purpose of providing public genomic data. It is a collaborative effort between pharmaceutical companies and the Wellcome Trust with the idea of making available widely accepted, high-quality, extensive, and publicly accessible SNP map. Its mission was to develop up to 300,000 SNPs distributed evenly throughout the human genome and to make the information related to these SNPs available to the public without intellectual property restrictions. The project started in April 1999 and was anticipated to continue until the end of 2001. In the end, many more SNPs, about 1.5 million total, were discovered than was originally planned.

By 1998 the complete genome sequence of *Mycobacterium tuberculosis* was published by teams from the UK, France, US and Denmark in June 1998. The ABI Prism 3700 sequencing machine, a capillary-based machine designed to run about eight sets of 96 sequence reactions per day also reached the market that year. That same year the genome sequence of the first multicellular organism, *C. elegans* was completed. *C. elegans* has a genome of about 100 Mb and, as noted, is a primitive animal model organism used in a range of biological disciplines.

By November 1999 the human genome draft sequence reached 1000 Mb and the first complete human chromosome was sequenced – this first was reached on the East side of the Atlantic by the HGP team led by the Sanger Centre, producing a finished sequence for chromosome 22, which is about 34 million base-pairs and includes at least 550 genes. According to anecdotal evidence when visiting his namesake centre, Sanger asked: "What does this machine do then?" "Dideoxy sequencing" came the reply, to which Fred retorted: "Haven't they come up with anything better yet?"

As will be elaborated in the final chapter the real highlight of 2000 was production of a 'working draft' sequence of the human genome, which was announced simultaneously in the US and the UK. In a joint event, Celera Genomics announced completion of their 'first assembly' of the genome. In a remarkable special issue, Nature included a 60-page article by the Human Genome Project partners, studies of mapping and variation, as well as analysis of the sequence by experts in different areas of biology. Science published the article by Celera on their assembly of HGP and Celera data as well as analyses of the use of the sequence. However to demonstrate the sensitivity of the market place to presidential utterances the joint appearances by Bill Clinton and Tony Blair touting this major milestone turned into a major cold shower when Clinton's reassurance of access of the people to their genetic information caused a precipitous drop in Celera's share value overnight. Clinton's assurance that, "The effort to decipher the human genome . . . will be the scientific breakthrough of the century – perhaps of all time. We have a profound responsibility to ensure that the life-saving benefits of any cutting-edge research are available to all human beings." (President Bill Clinton, Wednesday, March 14, 2000) stands in sharp contrast to the statement from Venter's Colleague that " Any company that wants to be in the business of using genes, proteins, or antibodies as drugs has a very high probability of running afoul of our patents. From a commercial point of view, they are severely constrained – and far more than they

realize." (William A. Haseltine, Chairman and CEO, Human Genome Sciences). The huge sell-off in stocks ended weeks of biotech buying in which those same stocks soared to unprecedented highs. By the next day, however, the genomic company spin doctors began to recover ground in a brilliant move which turned the Clinton announcement into a public relations coup.

All major genomics companies issued press releases applauding President Clinton's announcement. The real news they argued, was that "for the first time a President strongly affirmed the importance of gene based patents." And the same Bill Haseltine of Human Genome Sciences positively gushed as he happily pointed out that he "could begin his next annual report with the [President's] monumental statement, and quote today as a monumental day."

As distinguished Harvard biologist Richard Lewontin notes: "No prominent molecular biologist of my acquaintance is without a financial stake in the biotechnology business. As a result, serious conflicts of interest have emerged in universities and in government service (Lewontin, 2000).

Away from the spin doctors perhaps Eric Lander may have best summed up the Herculean effort when he opined that for him "the Human Genome Project has been the ultimate fulfilment: the chance to share common purpose with hundreds of wonderful colleagues towards a goal larger than ourselves. In the long run, the Human Genome Project's greatest impact might not be the three billion nucleotides of the human chromosomes, but its model of scientific community." (Ridley, 2000)

6. GENE THERAPY

The year 1990 also marked the passing of another milestone that was intimately connected to one of the fundamental drivers of the HGP. The California Hereditary Disorders Act came into force and with it one of the potential solutions for human hereditary disorders. W. French Anderson in the USA reported the first successful application of gene therapy in humans. The first successful gene therapy for a human disease was successfully achieved for Severe Combined Immune Deficiency (SCID) by introducing the missing gene, adenosine deaminase deficiency (ADA) into the peripheral lymphocytes of a 4-year-old girl and returning modified lymphocytes to her. Although the results are difficult to interpret because of the concurrent use of polyethylene glycol-conjugated ADA commonly referred to as pegylated ADA (PGLA) in all patients, strong evidence for *in vivo* efficacy was demonstrated. ADA-modified T cells persisted *in vivo* for up to three years and were associated with increases in T-cell number and ADA enzyme levels, T cells derived from transduced PGLA were progressively replaced by marrow-derived T cells, confirming successful gene transfer into long-lived progenitor cells. Ashanthi DeSilva, the girl who received the first credible gene therapy, continues to do well more than a decade later. Cynthia Cutshall, the second child to receive gene therapy for the same disorder as DeSilva, also continues to do well. Within 10 years (by January 2000), more than 350 gene therapy protocols had been approved in the US and

worldwide, researchers launched more than 400 clinical trials to test gene therapy against a wide array of illnesses. Surprisingly, a disease not typically heading the charts of heritable disorders, cancer has dominated the research. In 1994 cancer patients were treated with the tumor necrosis factor gene, a natural tumor fighting protein which worked to a limited extent. Even more surprisingly, after the initial flurry of success little has worked. Gene therapy, the promising miracle of 1990 failed to deliver on its early promise over the decade.

Apart from those examples, there are many diseases whose molecular pathology is, or soon will be, well understood, but for which no satisfactory treatments have yet been developed. At the beginning of the nineties it appeared that gene therapy did offer new opportunities to treat these disorders both by restoring gene functions that have been lost through mutation and by introducing genes that can inhibit the replication of infectious agents, render cells resistant to cytotoxic drugs, or cause the elimination of aberrant cells. From this "genomic" viewpoint genes could be said to be viewed as medicines, and their development as therapeutics should embrace the issues facing the development of small-molecule and protein therapeutics such as bioavailability, specificity, toxicity, potency, and the ability to be manufactured at large scale in a cost-effective manner.

Of course for such a radical approach certain basal level criteria needed to be established for selecting disease candidates for human gene therapy. These include, such factors as the disease is an incurable, life-threatening disease; organ, tissue, and cell types affected by the disease have been identified; the normal counterpart of the defective gene has been isolated and cloned; either the normal gene can be introduced into a substantial subfraction of the cells from the affected tissue, or the introduction of the gene into the available target tissue, such as bone marrow, will somehow alter the disease process in the tissue affected by the disease; the gene can be expressed adequately (it will direct the production of enough normal protein to make a difference); and techniques are available to verify the safety of the procedure.

An ideal gene therapeutic should, therefore, be stably formulated at room temperature and amenable to administration either as an injectable or aerosol or by oral delivery in liquid or capsule form. The therapeutic should also be suitable for repeat therapy, and when delivered, it should neither generate an immune response nor be destroyed by tissue-scavenging mechanisms. When delivered to the target cell, the therapeutic gene should then be transported to the nucleus, where it should be maintained as a stable plasmid or chromosomal integrant, and be expressed in a predictable, controlled fashion at the desired potency in a cell-specific or tissue-specific manner.

In addition to the ADA gene transfer in children with severe combined immunodeficiency syndrome, a gene-marking study of Epstein–Barr virus-specific cytotoxic T cells, and trials of gene-modified T cells expressing suicide or viral resistance genes in patients infected with HIV were studied in the early nineties. Additional strategies for T-cell gene therapy which were pursued later in the decade involve the engineering of novel T-cell receptors that impart antigen specificity for virally

infected or malignant cells. Issues which still are not resolved include nuclear transport, integration, regulated gene expression and immune surveillance. This knowledge, when finally understood and applied to the design of delivery vehicles of either viral or non-viral origin, will assist in the realization of gene therapeutics as safe and beneficial medicines that are suited to the routine management of human health.

Scientists are also working on using gene therapy to generate antibodies directly inside cells to block the production of harmful viruses such as HIV or even cancer-inducing proteins. There is a specific connection with Francis Collins, as his motivation for pursuing the HGP was his pursuit of defective genes beginning with the cystic fibrosis gene. This gene, called the CF transmembrane conductance regulator, codes for an ion channel protein that regulates salts in the lung tissue. The faulty gene prevents cells from excreting salt properly causing a thick sticky mucus to build up and destroy lung tissue. Scientists have spliced copies of the normal genes into disabled adeno viruses that target lung tissues and have used bronchioscopes to deliver them to the lungs. The procedure worked well in animal studies however clinical trials in humans were not an unmitigated success. Because the cells lining the lungs are continuously being replaced the effect is not permanent and must be repeated. Studies are underway to develop gene therapy techniques to replace other faulty genes. For example, to replace the genes responsible for factor VIII and factor IX production whose malfunctioning causes hemophilia A and B respectively; and to alleviate the effects of the faulty gene in dopamine production that results in Parkinson's disease.

Apart from technical challenges such a radical therapy also engenders ethical debate. Many persons who voice concerns about somatic-cell gene therapy use a "slippery slope" argument. It sounds good in theory but where does one draw the line. There are many issues yet to be resolved in this field of thorny ethics "good" and "bad" uses of the gene modification, difficulty of following patients in long-term clinical research and such. Many gene therapy candidates are children who are too young to understand the ramifications of this treatment: Conflict of interest – pits individuals' reproductive liberties and privacy interests against the interests of insurance companies or society. One issue that is unlikely to ever gain acceptance is germline therapy, the removal of deleterious genes from the population. Issues of justice and resource allocation also have been raised: In a time of strain on our health care system, can we afford such expensive therapy? Who should receive gene therapy? If it is made available only to those who can afford it, then a number of civil rights groups claim that the distribution of desirable biological traits among different socioeconomic and ethnic groups would become badly skewed adding a new and disturbing layer of discriminatory behavior.

Indeed a major setback occurred before the end of the decade in 1999. Jesse Gelsinger was the first person to die from gene therapy, on September 17, 1999, and his death created another unprecedented situation when his family sued not only the research team involved in the experiment (U Penn), the company Genovo

Inc., but also the ethicist who offered moral advice on the controversial project. This inclusion of the ethicist as a defendant alongside the scientists and school was a surprising legal move that puts this specialty on notice, as will no doubt be the case with other evolving technologies such as stem cells and therapeutic cloning, that its members could be vulnerable to litigation over the philosophical guidance they provide to researchers.

The Penn group principal investigator James Wilson approached ethicist Arthur Caplan about their plans to test the safety of a genetically engineered virus on babies with a deadly form of the liver disorder, ornithine transcarbamylase deficiency. The disorder allows poisonous levels of ammonia to build up in the blood system. Caplan steered the researchers away from sick infants, arguing that desperate parents could not provide true informed consent. He said it would be better to experiment on adults with a less lethal form of the disease who were relatively healthy. Gelsinger fell into that category. Although he had suffered serious bouts of ammonia buildup, he was doing well on a special drug and diet regimen. The decision to use relatively healthy adults was controversial because risky, unproven experimental protocols generally use very ill people who have exhausted more traditional treatments, so have little to lose. In this case, the virus used to deliver the genes was known to cause liver damage, so some scientists were concerned it might trigger an ammonia crisis in the adults.

Wilson underestimated the risk of the experiment, omitted the disclosure about possible liver damage in earlier volunteers in the experiment and failed to mention the deaths of monkeys given a similar treatment during pre-clinical studies. A Food and Drug Administration investigation after Gelsinger's death found numerous regulatory violations by Wilson's team, including the failure to stop the experiment and inform the FDA after four successive volunteers suffered serious liver damage prior to the teen's treatment. In addition, the FDA said Gelsinger did not qualify for the experiment, because his blood ammonia levels were too high just before he underwent the infusion of genetic material. The FDA suspended all human gene experiments by Wilson and the University of Penn subsequently restricting him solely to animal studies. A follow-up FDA investigation subsequently alleged he improperly tested the experimental treatment on animals. Financial conflicts of interest also surrounded James Wilson, who stood to personally profit from the experiment through Genovo his biotechnology company. The lawsuit was settled out of court for undisclosed terms in November 2000.

The FDA also suspended gene therapy trials at St. Elizabeth's Medical Center in Boston, a major teaching affiliate of Tufts University School of Medicine, which sought to use gene therapy to reverse heart disease, because scientists there failed to follow protocols and may have contributed to at least one patient death. In addition, the FDA temporarily suspended two liver cancer studies sponsored by the Schering-Plough Corporation because of technical similarities to the University of Pennsylvania study.

Some research groups voluntarily suspended gene therapy studies, including two experiments sponsored by the Cystic Fibrosis Foundation and studies at Beth Israel

Deaconess Medical Center in Boston aimed at hemophilia. The scientists paused to make sure they learned from the mistakes.

7. LAYING DOWN THE CHIPS

The nineties also saw the development of another "high-thoughput" breakthrough, a derivative of the other high tech revolution namely DNA chips. In 1991 Biochips were developed for commercial use under the guidance of Affymetrix. DNA chips or microarrays represent a "massively parallel" genomic technology. They facilitate high throughput analysis of thousands of genes simultaneously, and are thus potentially very powerful tools for gaining insight into the complexities of higher organisms including analysis of gene expression, detecting genetic variation, making new gene discoveries, fingerprinting strains and developing new diagnostic tools. These technologies permit scientists to conduct large scale surveys of gene expression in organisms, thus adding to our knowledge of how they develop over time or respond to various environmental stimuli. These techniques are especially useful in gaining an integrated view of how multiple genes are expressed in a coordinated manner. These DNA chips have broad commercial applications and are now used in many areas of basic and clinical research including the detection of drug resistance mutations in infectious organisms, direct DNA sequence comparison of large segments of the human genome, the monitoring of multiple human genes for disease associated mutations, the quantitative and parallel measurement of mRNA expression for thousands of human genes, and the physical and genetic mapping of genomes.

However the initial technologies, or more accurately the algorithms used to extract information, were far from robust and reproducible. The erstwhile serial entrepreneur, Al Zaffaroni (the rebel who in 1968 founded Alza when Syntex ignored his interest in developing new ways to deliver drugs) founded yet another company, Affymetrix, under the stewardship of Stephen Fodor, which was subject to much abuse for providing final extracted data and not allowing access to raw data. As with other personalities of this high through put era, Seattle-bred Steve Fodor was also somewhat of a polymath having contributed to two major technologies, microarrays and combinatorial chemistry, the former has delivered on it's, promise while the latter, like gene therapy, is still in a somewhat extended gestation. And despite the limitations of being an industrial scientist he has had a rather prolific portfolio of publications. His seminal manuscripts describing this work have been published in all the journals of note, Science, Nature and PNAS and was recognized in 1992 by the AAAS by receiving the Newcomb-Cleveland Award for an outstanding paper published in Science. Fodor began his industrial career in yet another Zaffaroni firm. In 1989 he was recruited to the Affymax Research Institute in Palo Alto where he spearheaded the effort to develop high-density arrays of biological compounds. His initial interest was in the broad area of what came to be called combinatorial chemistry. Of the techniques developed, one approach permitted high resolution chemical synthesis in a light-directed, spatially-defined format.

In the days before positive selection vectors, a researcher might have screened thousands of clones by hand with an oligonucleotide probe just to find one elusive insert. Fodor's (and his successors) DNA array technology reverses that approach. Instead of screening an array of unknowns with a defined probe – a cloned gene, PCR product, or synthetic oligonucleotide – each position or "probe cell" in the array is occupied by a defined DNA fragment, and the array is probed with the unknown sample.

Fodor used his chemistry and biophysics background to develop very dense arrays of these biomolecules by combining photolithographic methods with traditional chemical techniques. The typical array may contain all possible combinations of all possible oligonucleotides (8-mers, for example) that occur as a "window" which is tracked along a DNA sequence. It might contain longer oligonucleotides designed from all the open reading frames identified from a complete genome sequence. Or it might contain cDNAs – of known or unknown sequence – or PCR products.

Of course it is one thing to produce data it is quite another to extract it in a meaningful manner. Fodor's group also developed techniques to read these arrays, employing fluorescent labeling methods and confocal laser scanning to measure each individual binding event on the surface of the chip with extraordinary sensitivity and precision. This general platform of microarray based analysis coupled to confocal laser scanning has become the standard in industry and academia for large-scale genomics studies. In 1993, Fodor co-founded Affymetrix where the chip technology has been used to synthesize many varieties of high density oligonucleotide arrays containing hundreds of thousands of DNA probes. In 2001, Steve Fodor founded Perlegen, Inc., a new venture that applied the chip technology towards uncovering the basic patterns of human diversity. His company's stated goals are to analyze more than one million genetic variations in clinical trial participants to explain and predict the efficacy and adverse effect profiles of prescription drugs. In addition, Perlegen also applies this expertise to discovering genetic variants associated with disease in order to pave the way for new therapeutics and diagnostics.

Fodor's former company diversified into plant applications by developing a chip of the archetypal model of plant systems Arabidopsis and supplied Pioneer Hi Bred with custom DNA chips for monitoring maize gene expression. They (Affymetrix) have established programs where academic scientists can use company facilities at a reduced price and set up 'user centers' at selected universities.

A related but less complex technology called 'spotted' DNA chips involves precisely spotting very small droplets of genomic or cDNA clones or PCR samples on a microscope slide. The process uses a robotic device with a print head bearing fine "repeatograph" tips that work like fountain pens to draw up DNA samples from a 96-well plate and spot tiny amounts on a slide. Up to 10,000 individual clones can be spotted in a dense array within one square centimeter on a glass slide. After hybridization with a fluorescent target mRNA, signals are detected by a custom scanner. This is the basis of the systems used by Molecular Dynamics

and Incyte (who acquired this technology when it took over Synteni). In 1997, Incyte was looking to gather more data for its library and perform experiments for corporate subscribers. The company considered buying Affymetrix GeneChips but opted instead to purchase the smaller Synteni, which had sprung out of Pat Brown's Stanford array effort. Synteni's contact printing technology resulted in dense – and cheaper – arrays. Though Incyte used the chips only internally, Affymetrix sued, claiming Synteni/Incyte was infringing on its chip density patents. The suit argued that dense biochips – regardless of whether they use photolithography – cannot be made without a license from Affymetrix! And in a litigious Congo line endemic of this hi-tech era Incyte countersued and for good measure also filed against genetic database competitor Gene Logic for infringing Incyte's patents on database building. Meanwhile, Hyseq sued Affymetrix, claiming infringement of nucleotide hybridization patents obtained by its CSO. Affymetrix, in turn, filed a countersuit, claiming Hyseq infringed the spotted array patents. Hyseq then reached back and found an additional hybridization patent it claimed that Affymetrix had infringed. And so on into the next millennium!

In part to avoid all of this another California company Nanogen, Inc. took a different approach to single nucleotide polymorphism discrimination technology. In an article in the April 2000 edition of Nature Biotechnology, entitled "Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips," Nanogen describes the use of microchips to identify variants of the mannose binding protein gene that differ from one another by only a single DNA base. The mannose binding protein (MBP) is a key component of the innate immune system in children who have not yet developed immunity to a variety of pathogens. To date, four distinct variants (alleles) of this gene have been identified, all differing by only a single nucleotide of DNA. MBP was selected for this study because of its potential clinical relevance and its genetic complexity. The samples were assembled at the NCI laboratory in conjunction with the National Institutes of Health and transferred to Nanogen for analysis.

However, from a high throughput perspective there is a question mark over microarrays. Mark Benjamin, senior director of business development at Rosetta Inpharmatics (Kirkland, WA), is skeptical about the long-term prospects for standard DNA arrays in high-throughput screening as the first steps require exposing cells and then isolating RNA, which is something that is very hard to do in a high-throughput format.

Another drawback is that most of the useful targets are likely to be unknown (particularly in the agricultural sciences where genome sequencing is still in its infancy), and DNA arrays that are currently available test only for previously sequenced genes. Indeed, some argue that current DNA arrays may not be sufficiently sensitive to detect the low expression levels of genes encoding targets of particular interest. And the added complication of the companies' reluctance to provide "raw data" means that derived data sets may be created with less than optimum algorithims thereby irretrievably losing potentially valuable information from the starting material. Reverse engineering is a possible approach but this is

laborious and time consuming and being prohibited by many contracts may arouse the interest of the ever-vigilant corporate lawyers.

8. RISE OF THE " -OMICS"

Over the course of the nineties, outgrowths of functional genomics have been termed proteomics and metabolomics, which are the global studies of gene expression at the protein and metabolite levels respectively. The study of the integration of information flow within an organism is emerging as the field of systems biology. In the area of proteomics, the methods for global analysis of protein profiles and cataloging protein-protein interactions on a genome-wide scale are technically more difficult but improving rapidly, especially for microbes. These approaches generate vast amounts of quantitative data. The amount of expression data becoming available in the public and private sectors is already increasing exponentially. Gene and protein expression data rapidly dwarfed the DNA sequence data and is considerably more difficult to manage and exploit.

In microbes, the small sizes of the genomes and the ease of handling microbial cultures, will enable high throughput, targeted deletion of every gene in a genome, individually and in combinations. This is already available on a moderate throughput scale in model microbes such as *E. coli* and yeast. Combining targeted gene deletions and modifications with genome-wide assay of mRNA and protein levels will enable intricate inter-dependencies among genes to be unraveled. Simultaneous measurement of many metabolites, particularly in microbes, is beginning to allow the comprehensive modeling and regulation of fluxes through interdependent pathways. Metabolomics can be defined as the quantitative measurement of all low molecular weight metabolites in an organism's cells at a specified time under specific environmental conditions. Combining information from metabolomics, proteomics and genomics will help us to obtain an integrated understanding of cell biology.

The next hierarchical level of phenotype considers how the proteome within and among cells cooperates to produce the biochemistry and physiology of individual cells and organisms. Several authors have tentatively offered "physiomics" as a descriptor for this approach. The final hierarchical levels of phenotype include anatomy and function for cells and whole organisms. The term "phenomics" has been applied to this level of study and unquestionably the more well known omics namely economics, has application across all those fields.

And, coming slightly out of left field this time, the spectre of eugenics needless to say was raised in the omics era. In the year 1992 American and British scientists unveiled a technique which has come to be known as pre-implantation genetic diagnosis (PID) for testing embryos *in vitro* for genetic abnormalities such as cystic fibrosis, hemophilia, and Down's Syndrome (Wald, 1992). This might be seen by most as a step forward, but it led ethicist David S. King (1999) to decry PID as a technology that could exacerbate the eugenic features of prenatal testing and make possible an expanded form of free-market eugenics. He further argues that due to

social pressures and eugenic attitudes held by clinical geneticists in most countries, it results in eugenic outcomes even though no state coercion is involved and that, as abortion is not involved, and multiple embryos are available, PID is radically more effective as a tool of genetic selection.

9. AGRICULTURAL/INDUSTRIAL BIOTECH IN THE 1990S

The first regulatory approval of a recombinant DNA technology in the U.S. food supply was not a plant but an industrial enzyme that has become the hallmark of food biotechnology success. Enzymes were important agents in food production long before modern biotechnology was developed. They were used, for instance, in the clotting of milk to prepare cheese, the production of bread and the production of alcoholic beverages. Nowadays, enzymes are indispensable to modern food processing technology and have a great variety of functions. They are used in almost all areas of food production including grain processing, milk products, beer, juices, wine, sugar and meat. Chymosin, known also as rennin, is a proteolytic enzyme whose role in digestion is to curdle or coagulate milk in the stomach, efficiently converting liquid milk to a semisolid like cottage cheese, allowing it to be retained for longer periods in a neonate's stomach. The dairy industry takes advantage of this property to conduct the first step in cheese production. Chy-Max™, an artificially produced form of the chymosin enzyme for cheese-making, was approved in 1990.

In some instances they replace less acceptable "older" technology, for example the enzyme chymosin. Unlike crops industrial enzymes have had relatively easy passage to acceptance for a number of reasons. As noted they are part of the processing system and theoretically do not appear in the final product. Today about 90% of the hard cheese in the US and UK is made using chymosin from genetically-modified microbes. It is easier to purify, more active (95% as compared to 5%) and less expensive to produce (Microbes are more prolific, more productive and cheaper to keep than calves). Like all enzymes it is required only in very small quantities and because it is a relatively unstable protein it breaks down as the cheese matures. Indeed, if the enzyme remained active for too long it would adversely affect the development of the cheese, as it would degrade the milk proteins to too great a degree. Such enzymes have gained the support of vegetarian organizations and of some religious authorities.

For plants the nineties was the era of the first widespread commercialization of what came to be known in often deprecating and literally inaccurate terms as GMOs (Genetically Modified Organisms). When the nineties dawned dicotyledonous plants were relatively easily transformed with *Agrobacterium tumefaciens* but many economically important plants, including the cereals, remained inaccessible for genetic manipulation because of lack of effective transformation techniques. In 1990 this changed with the technology that overcame this limitation. Michael Fromm, a molecular biologist at the Plant Gene Expression Center, reported the stable transformation of corn using a high-speed gene gun. The method known as biolistics uses a "particle gun" to shoot metal particles coated with DNA into

cells. Initially a gunpowder charge subsequently replaced by helium gas was used to accelerate the particles in the gun. There is a minimal disruption of tissue and the success rate has been extremely high for applications in several plant species. The technology rights are now owned by DuPont. In 1990 some of the first of the field trials of the crops that would dominate the second half of the nineties began, including Bt corn (with the *Bacillus thuriengenesis* Cry protein discussed in chapter three).

In 1992 the FDA declared that genetically engineered foods are "not inherently dangerous" and do not require special regulation. Since 1992, researchers have pinpointed and cloned several of the genes that make selected plants resistant to certain bacterial and fungal infections; some of these genes have been successfully inserted into crop plants that lack them. Many more infection-resistant crops are expected in the near future, as scientists find more plant genes in nature that make plants resistant to pests. Plant genes, however, are just a portion of the arsenal; microorganisms other than Bt also are being mined for genes that could help plants fend off invaders that cause crop damage.

The major milestone of the decade in crop biotechnology was approval of the first bioengineered crop plant in 1994. It represented a double first not just of the first approved food crop but also of the first commercial validation of a technology which was to be surpassed later in the decade. That technology, antisense technology works because nucleic acids have a natural affinity for each other. When a gene coding for the target in the genome is introduced in the opposite orientation, the reverse RNA strand anneals and effectively blocks expression of the enzyme. This technology was patented by Calgene for plant applications and was the technology behind the famous FLAVR SAVR tomatoes. The first success for antisense in medicine was in 1998 when the U.S. Food and Drug Administration gave the go-ahead to the cytomegalovirus (CMV) inhibitor fomivirsen, a phosphorothionate antiviral for the AIDS-related condition CMV retinitis making it the first drug belonging to Isis, and the first antisense drug ever, to be approved.

Another technology, although not apparent at the time was behind the second approval and also the first and only successful to date in a commercial tree fruit biotech application. The former was a virus resistant squash the second the papaya ringspot resistant papaya. Both owed their existence as much to historic experience as modern technology. Genetically engineered virus-resistant strains of squash and cantaloupe, for example, would never have made it to farmers' fields if plant breeders in the 1930's had not noticed that plants infected with a mild strain of a virus do not succumb to more destructive strains of the same virus. That finding led plant pathologist Roger Beachy, then at Washington University in Saint Louis, to wonder exactly how such "cross-protection" worked – did part of the virus prompt it?

In collaboration with researchers at Monsanto, Beachy used an *A. tumefaciens* vector to insert into tomato plants a gene that produces one of the proteins that makes up the protein coat of the tobacco mosaic virus. He then inoculated these plants with the virus and was pleased to discover, as reported in 1986, that the vast majority of plants did not succumb to the virus.

Eight years later, in 1994, virus-resistant squash seeds created with Beachy's method reached the market, to be followed soon by bioengineered virus-resistant seeds for cantaloupes, potatoes, and papayas. (Breeders had already created virus-resistant tomato seeds by using traditional techniques.) And the method of protection still remained a mystery when the first approvals were given in 1994 and 1996. Gene silencing was perceived initially as an unpredictable and inconvenient side effect of introducing transgenes into plants. It now seems that it is the consequence of accidentally triggering the plant's adaptive defense mechanism against viruses and transposable elements. This recently discovered mechanism, although mechanistically different, has a number of parallels with the immune system of mammals. How this system worked was not elucidated until later in the decade by a researcher who was seeking a very different holy grail – the black rose! Rick Jorgensen, at that time at DNA Plant Technologies in Oakland, CA and subsequently of, of the University of California Davis attempted to overexpress the chalcone synthase gene by introducing a modified copy under a strong promoter.Surprisingly he obtained white flowers, and many strange variegated purple and white variations in between. This was the first demonstration of what has come to be known as post-transcriptional gene silencing (PTGS). While initially it was considered a strange phenomenon limited to petunias and a few other plant species, it is now one of the hottest topics in molecular biology. RNA interference (RNAi) in animals and basal eukaryotes, quelling in fungi, and PTGS in plants are examples of a broad family of phenomena collectively called RNA silencing (Hannon 2002; Plasterk 2002). In addition to its occurrence in these species it has roles in viral defense (as demonstrated by Beachy) and transposon silencing mechanisms among other things. Perhaps most exciting, however, is the emerging use of PTGS and, in particular, RNAi – PTGS initiated by the introduction of double-stranded RNA (dsRNA) – as a tool to knock out expression of specific genes in a variety of organisms.

Nineteen ninety one also heralded yet another first. The February 1, 1991 issue of Science reported the patenting of "molecular scissors": the Nobel-prize winning discovery of enzymatic RNA, or "ribozymes," by Thomas Czech of the University of Colorado. It was noted that the U.S. Patent and Trademark Office had awarded an "unusually broad" patent for ribozymes. The patent is U.S. Patent No. 4,987,071, claim 1 of which reads as follows: "An enzymatic RNA molecule not naturally occurring in nature having an endonuclease activity independent of any protein, said endonuclease activity being specific for a nucleotide sequence defining a cleavage site comprising single-stranded RNA in a separate RNA molecule, and causing cleavage at said cleavage site by a transesterification reaction." Although enzymes made of protein are the dominant form of biocatalyst in modern cells, there are at least eight natural RNA enzymes, or ribozymes, that catalyze fundamental biological processes. One of which was yet another discovery by plant virologists, in this instance the hairpin ribozyme was discovered by George Bruening at UC Davis. The self-cleavage structure was originally called a paperclip, by the Bruening laboratory which discovered the reactions.

As mentioned in chapter 3, it is believed that these ribozymes might be the remnants of an ancient form of life that was guided entirely by RNA. Since a

ribozyme is a catalytic RNA molecule capable of cleaving itself and other target RNAs it therefore can be useful as a control system for turning off genes or targeting viruses. The possibility of designing ribozymes to cleave any specific target RNA has rendered them valuable tools in both basic research and therapeutic applications. In the therapeutics area, they have been exploited to target viral RNAs in infectious diseases, dominant oncogenes in cancers and specific somatic mutations in genetic disorders. Most notably, several ribozyme gene therapy protocols for HIV patients are already in Phase 1 trials. More recently, ribozymes have been used for transgenic animal research, gene target validation and pathway elucidation. However, targeting ribozymes to the cellular compartment containing their target RNAs has proved a challenge. At the other bookend of the decade in 2000, Samarsky et al. reported that a family of small RNAs in the nucleolus (snoRNAs) can readily transport ribozymes into this subcellular organelle.

In addition to the already extensive panoply of RNA entities yet another has potential for mischief. Viroids are small, single-stranded, circular RNAs containing 246–463 nucleotides arranged in a rod-like secondary structure and are the smallest pathogenic agents yet described. The smallest viroid characterized to date is rice yellow mottle sobemovirus (RYMV), at 220 nucleotides. In comparison, the genome of the smallest known viruses capable of causing an infection by themselves, the single-stranded circular DNA of circoviruses, is around 2 kilobases in size. The first viroid to be identified was the Potato spindle tuber viroid (PSTVd). Some 33 species have been identified to date. Unlike the many satellite or defective interfering RNAs associated with plant viruses, viroids replicate autonomously on inoculation of a susceptible host. The absence of a protein capsid and of detectable messenger RNA activity implies that the information necessary for replication and pathogenesis resides within the unusual structure of the viroid genome. The replication mechanism actually involves interaction with RNA polymerase II, an enzyme normally associated with synthesis of messenger RNA, and "rolling circle" synthesis of new RNA. Some viroids have ribozyme activity which allow self-cleavage and ligation of unit-size genomes from larger replication intermediates. It has been proposed that viroids are "escaped introns".

Viroids are usually transmitted by seed or pollen. Infected plants can show distorted growth.

10. AND THE FLIP SIDE

From its earliest years, biotechnology attracted interest outside scientific circles. Initially the main focus of public interest was on the safety of recombinant DNA technology, and of the possible risks of creating uncontrollable and harmful novel organisms (Berg , 1975). The debate on the deliberate release of genetically modified organisms, and on consumer products containing or comprising them, followed some years later (NAS, 1987). It is interesting to note that within the broad ethical tableau of potential issues within the science and products of biotechnology, the seemingly innocuous field of plant modification has been one of the major players

of the 1990's. The success of agricultural biotechnology is heavily dependent on its acceptance by the public, and the regulatory framework in which the industry operates is also influenced by public opinion. As the focus for molecular biology research shifted from the basic pursuit of knowledge to the pursuit of lucrative applications, once again as in the previous two decades the specter of risk arose as the potential of new products and applications had to be evaluated outside the confines of a laboratory.

However, the specter now became far more global as the implications of commercial applications brought not just worker safety into the loop but also, the environment, agricultural and industrial products and the safety and well being of all living things. Beyond "deliberate" release, the RAC guidelines were not designed to address these issues, so the matter moved into the realm of the federal agencies who had regulatory authority which could be interpreted to oversee biotechnology issues. This adaptation of oversight is very much a dynamic process as the various agencies wrestle with the task of applying existing regulations and developing new ones for oversight of this technology in transition.

As the decade progressed focus shifted from basic biotic stress resistance to more complex modifications The next generation of plants will focus on value added traits in which valuable genes and metabolites will be identified and isolated, with some of the later compounds being produced in mass quantities for niche markets. Two of the more promising markets are nutraceuticals or so-called "Functional Foods" and plants developed as bioreactors for the production of valuable proteins and compounds, a field known as Plant Molecular Farming.

Developing plants with improved quality traits involves overcoming a variety of technical challenges inherent to metabolic engineering programs. Both traditional plant breeding and biotechnology techniques are needed to produce plants carrying the desired quality traits. Continuing improvements in molecular and genomic technologies are contributing to the acceleration of product development in this space.

By the end of the decade in 1999, applying nutritional genomics, Della Penna (1999) isolated a gene, which converts the lower activity precursors to the highest activity vitamin E compound, alpha-tocopherol. With this technology, the vitamin E content of *Arabidopsis* seed oil has been increased nearly 10-fold and progress has been made to move the technology to crops such as soybean, maize, and canola. This has also been done for folates in rice. Omega three fatty acids play a significant role in human health, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), which are present in the retina of the eye and cerebral cortex of the brain, respectively, are some of the most well documented from a clinical perspective. It is believed that EPA and DHA play an important role in the regulation of inflammatory immune reactions and blood pressure, treatment of conditions such as cardiovascular disease and cystic fibrosis, brain development *in utero*, and, in early postnatal life, the development of cognitive function. They are mainly found in fish oil and the supply is limited. By the end of the decade Ursin (2000) had succeeded in engineering canola to produce these fatty acids.

From a global perspective another value-added development had far greater impact both technologically and socio-economically. A team led by Ingo Potrykus (1999) engineered rice to produce pro-Vitamin A, which is an essential micronutrient. Widespread dietary deficiency of this vitamin in rice-eating Asian countries, which predisposes children to diseases such as blindness and measles, has tragic consequences. Improved vitamin A nutrition would alleviate serious health problems and, according to UNICEF, could also prevent up to two million infant deaths due to vitamin A deficiency.

Adoption of the next stage of GM crops may proceed more slowly, as the market confronts issues of how to determine price, share value, and adjust marketing and handling to accommodate specialized end-use characteristics. Furthermore, competition from existing products will not evaporate. Challenges that have accompanied GM crops with improved agronomic traits, such as the stalled regulatory processes in Europe, will also affect adoption of nutritionally improved GM products. Beyond all of this, credible scientific research is still needed to confirm the benefits of any particular food or component. For functional foods to deliver their potential public health benefits, consumers must have a clear understanding of, and a strong confidence level in, the scientific criteria that are used to document health effects and claims. Because these decisions will require an understanding of plant biochemistry, mammalian physiology, and food chemistry, strong interdisciplinary collaborations will be needed among plant scientists, nutritionists, and food scientists to ensure a safe and healthful food supply.

In addition to being a source of nutrition, plants have been a valuable wellspring of therapeutics for centuries. During the nineties, however, intensive research has focused on expanding this source through rDNA biotechnology and essentially using plants and animals as living factories for the commercial production of vaccines, therapeutics and other valuable products such as industrial enzymes and biosynthetic feedstocks.

Possibilities in the medical field include a wide variety of compounds, ranging from edible vaccine antigens against hepatitis B and Norwalk viruses (Arntzen, 1997) and *Pseudomonas aeruginosa* and *Staphylococcus aureus* to vaccines against cancer and diabetes, enzymes, hormones, cytokines, interleukins, plasma proteins, and human alpha-1-antitrypsin. Thus, plant cells are capable of expressing a large variety of recombinant proteins and protein complexes. Therapeutics produced in this way are termed plant made pharmaceuticals (PMPs). And non-therapeutics are termed plant made industrial products (PMIPs) (Newell-McGloughlin, 2006).

The first reported results of successful human clinical trials with their transgenic plant-derived pharmaceuticals were published in 1998. They were an edible vaccine against *E. coli*-induced diarrhea and a secretory monoclonal antibody directed against *Streptococcus mutans*, for preventative immunotherapy to reduce incidence of dental caries. Haq et al. (1995) reported the expression in potato plants of a vaccine against *E. coli* enterotoxin (ETEC) that provided an immune response against the toxin in mice. Human clinical trials suggest that oral vaccination against either of the closely related enterotoxins of *Vibrio cholerae* and *E. coli* induces production of antibodies that can neutralize the respective toxins by preventing

them from binding to gut cells. Similar results were found for Norwalk virus oral vaccines in potatoes. For developing countries, the intention is to deliver them in bananas or tomatoes (Newell-McGloughlin, 2006).

Plants are also faster, cheaper, more convenient and more efficient than the principal eukaryotic production system, namely Chinese Hamster Ovary (CHO) cells for the production of pharmaceuticals. Hundreds of acres of protein-containing seeds could inexpensively double the production of a CHO bioreactor factory. In addition, proteins can be expressed at the highest levels in the harvestable seed and plant-made proteins and enzymes formulated in seeds have been found to be extremely stable, reducing storage and shipping costs. Pharming may also enable research on drugs that cannot currently be produced. For example, CropTech in Blacksburg, Va., is investigating a protein that seems to be a very effective anticancer agent. The problem is that this protein is difficult to produce in mammalian cell culture systems as it inhibits cell growth. This should not be a problem in plants.

Furthermore, production size is flexible and easily adjustable to the needs of changing markets. Making pharmaceuticals from plants is also a sustainable process, because the plants and crops used as raw materials are renewable. The system also has the potential to address problems associated with provision of vaccines to people in developing countries. Products from these alternative sources do not require a so-called "cold chain" for refrigerated transport and storage. Those being developed for oral delivery obviates the need for needles and aspectic conditions which often are a problem in those areas. Apart from those specific applications where the plant system is optimum there are many other advantages to using plant production. Many new pharmaceuticals based on recombinant proteins will receive regulatory approval from the United States Food and Drug Administration (FDA) in the next few years. As these therapeutics make their way through clinical trials and evaluation, the pharmaceutical industry faces a production capacity challenge. Pharmaceutical discovery companies are exploring plant-based production to overcome capacity limitations, enable production of complex therapeutic proteins, and fully realize the commercial potential of their biopharmaceuticals (Newell-McGloughlin, 2006).

11. ANIMAL BIOTECH

Nineteen ninety also marked a major milestone in the animal biotech world when Herman made his appearance on the world's stage. Since the Palmiter's mouse, transgenic technology has been applied to several species including agricultural species such as sheep, cattle, goats, pigs, rabbits, poultry, and fish. Herman was the first transgenic bovine created by GenPharm International, Inc., in a laboratory in the Netherlands at the early embryo stage. Scientist's microinjected recently fertilized eggs with the gene coding for human lactoferrin. The scientists then cultured the cells *in vitro* to the embryo stage and transferred them to recipient cattle. Lactoferrin, an iron-containing anti-bacterial protein is essential for infant

growth. Since cow's milk doesn't contain lactoferrin, infants must be fed from other sources that are rich in iron – formula or mother's milk (Newell-McGloughlin, 2001).

As Herman was a boy he would be unable to provide the source, that would require the production of daughters which was not necessarily a straightforward process. The Dutch parliments permission was required. In 1992 they finally approved a measure that permitted the world's first genetically engineered bull to reproduce. The Leiden-based Gene Pharming proceeded to artificially inseminate 60 cows with Herman's sperm. With a promise that the protein, lactoferrin, would be the first in a new generation of inexpensive, high-tech drugs derived from cows' milk to treat complex diseases like AIDS and cancer. Herman, became the father of at least eight female calves in 1994, and each one inherited the gene for lactoferrin production. While their birth was initially greeted as a scientific advancement that could have far-reaching effects for children in developing nations, the levels of expression were too low to be commercially viable.

By 2002, Herman, who likes to listen to rap music to relax, had sired 55 calves and outlived them all. His offspring were all killed and destroyed after the end of the experiment, in line with Dutch health legislation. Herman was also slated for the abattoir, but the Dutch public – proud of making history with Herman – rose up in protest, especially after a television program screened footage showing the amiable bull licking a kitten. Herman won a bill of clemency from parliament. However, instead of retirement on a comfortable bed of straw, listening to rap music, Herman was pressed into service again. He now stars at a permanent biotech exhibit in Naturalis, a natural history museum in the Dutch city of Leiden. After his death, he will be stuffed and remain in the museum in perpetuity (A fate similar to what awaited an even more famous mammalian first born later in the decade).

The applications for transgenic animal research fall broadly into two distinct areas, namely medical and agricultural applications. The recent focus on developing animals as bioreactors to produce valuable proteins in their milk can be catalogued under both areas. Underlying each of these, of course, is a more fundamental application, that is the use of those techniques as tools to ascertain the molecular and physiological bases of gene expression and animal development. This understanding can then lead to the creation of techniques to modify development pathways.

In 1992 a European decision with rather more far-reaching implications than Hermans sex life was made. The first European patent on a transgenic animal was issued for a transgenic mouse sensitive to carcinogens – Harvard's "Oncomouse". The oncomouse patent application was refused in Europe in 1989 due primarily to an established ban on animal patenting. The application was revised to make narrower claims, and the patent was granted in 1992. This has since been repeatedly challenged, primarily by groups objecting to the judgement that benefits to humans outweigh the suffering of the animal. Currently, the patent applicant is awaiting protestors' responses to a series of possible modifications to the application. Predictions are that agreement will not likely be forthcoming and that the legal wrangling will continue into the future.

Bringing animals into the field of controversy starting to swirl around GMOs and preceding the latter's commercialization, was the approval by the FDA of bovine somatotropin (BST) for increased milk production in dairy cows. The FDA's Center for Veterinary Medicine (CVM) regulates the manufacture and distribution of food additives and drugs that will be given to animals. Biotechnology products are a growing proportion of the animal health products and feed components regulated by the CVM. The Center requires that food products from treated animals must be shown to be safe for human consumption. Applicants must show that the drug is effective and safe for the animal and that its manufacture will not affect the environment. They must also conduct geographically dispersed clinical trials under an Investigational New Animal Drug application with the FDA through which the agency controls the use of the unapproved compound in food animals. Unlike within the EU, possible economic and social issues cannot be taken into consideration by the FDA in the premarket drug approval process. Under these considerations the safety and efficacy of rBST was determined. It was also determined that special labeling for milk derived from cows that had been treated with rBST is not required under FDA food labeling laws because the use of rBST does not effect the quality or the composition of the milk.

Work with fish proceeded a pace throughout the decade. Gene transfer techniques have been applied to a large number of aquatic organisms, both vertebrates and invertebrates. Gene transfer experiments have targeted a wide variety of applications, including the study of gene structure and function, aquaculture production, and use in fisheries management programs.

Because fish have high fecundity, large eggs, and do not require reimplantation of embryos, transgenic fish prove attractive model systems in which to study gene expression. Transgenic zebrafish have found utility in studies of embryogenesis, with expression of transgenes marking cell lineages or providing the basis for study of promoter or structural gene function. Although not as widely used as zebrafish, transgenic medaka and goldfish have been used for studies of promoter function. This body of research indicates that transgenic fish provide useful models of gene expression, reliably modeling that in "higher" vertebrates.

Perhaps the largest number of gene transfer experiments address the goal of genetic improvement for aquaculture production purposes. The principal area of research has focused on growth performance, and initial transgenic growth hormone (GH) fish models have demonstrated accelerated and beneficial phenotypes. DNA microinjection methods have propelled the many studies reported and have been most effective due to the relative ease of working with fish embryos. Bob Devlins' group in Vancouver has demonstrated extraordinary growth rate in coho salmon which were transformed with a growth hormone from sockeye salmon. The transgenics achieve up to eleven times the size of their littermates within six months, reaching maturity in about half the time. Interestingly this dramatic effect is only observed in feeding pins where the transgenics' ferocious appetites demands constant feeding. If the fish are left to their own devices and must forage for themselves, they appear to be out-competed by their smarter siblings.

However most studies, such as those involving transgenic Atlantic salmon and channel catfish, report growth rate enhancement on the order of 30–60%. In addition to the species mentioned, GH genes also have been transferred into striped bass, tilapia, rainbow trout, gilthead sea bream, common carp, bluntnose bream, loach, and other fishes.

Shellfish also are subject to gene transfer toward the goal of intensifying aquaculture production. Growth of abalone expressing an introduced GH gene is being evaluated; accelerated growth would prove a boon for culture of the slow-growing mollusk. A marker gene was introduced successfully into giant prawn, demonstrating feasibility of gene transfer in crustaceans, and opening the possibility of work involving genes affecting economically important traits. In the ornamental fish sector of aquaculture, ongoing work addresses the development of fish with unique coloring or patterning. A number of companies have been founded to pursue commercialization of transgenics for aquaculture. As most aquaculture species mature at 2–3 years of age, most transgenic lines are still in development and have yet to be tested for performance under culture conditions.

Extending earlier research that identified methylfarnesoate (MF) as a juvenile hormone in crustaceans and determined its role in reproduction, researchers at the University of Connecticut have developed technology to synchronize shrimp egg production and to increase the number and quality of eggs produced. Females injected with MF are stimulated to produce eggs ready for fertilization. The procedure produces 180 percent more eggs than the traditional crude method of removing the eyestalk gland. This will increase aquaculture efficiency.

A number of experiments utilize gene transfer to develop genetic lines of potential utility in fisheries management. Transfer of GH genes into northern pike, walleye, and largemouth bass are aimed at improving the growth rate of sport fishes. Gene transfer has been posed as an option for reducing losses of rainbow trout to whirling disease, although suitable candidate genes have yet to be identified. Richard Winn of the University of Georgia is developing transgenic killifish and medaka as biomonitors for environmental mutagens, which carry the bacteriophage phi X 174 as a target for mutation detection. Development of transgenic lines for fisheries management applications generally is at an early stage, often at the founder or F1 generation.

Broad application of transgenic aquatic organisms in aquaculture and fisheries management will depend on showing that particular GMOs can be used in the environment both effectively and safely. Although our base of knowledge for assessing ecological and genetic safety of aquatic GMOs currently is limited, some early studies supported by the USDA biotechnology risk assessment program have yielded results. Data from outdoor pond-based studies on transgenic catfish reported by Rex Dunham of Auburn University show that transgenic and non-transgenic individuals interbreed freely, that survival and growth of transgenics in unfed ponds was equal to or less than that of non-transgenics, and that predator avoidance is not affected by expression of the transgene.

However, unquestionably the seminal event for animal biotech in the nineties was Ian Wilmut's landmark work using nuclear transfer technology to generate the lambs Morag and Megan reported in 1996 (from an embryonic cell nuclei) and the truly ground-breaking work of creating Dolly from an adult somatic cell nucleus, reported in February, 1997 (Wilmut, 1997). Wilmut and his colleagues at the Roslin Institute demonstrated for the first time with the birth of Dolly the sheep that the nucleus of an adult somatic cell can be transferred to an enucleated egg to create cloned offspring. It had been assumed for some time that only embryonic cells could be used as the cellular source for nuclear transfer. This assumption was shattered with the birth of Dolly. This example of cloning an animal using the nucleus of an adult cell was significant because it demonstrated the ability of egg cell cytoplasm to "reprogram" an adult nucleus. When cells differentiate, that is, evolve from primitive embryonic cells to functionally defined adult cells, they lose the ability to express most genes and can only express those genes necessary for the cell's differentiated function. For example, skin cells only express genes necessary for skin function, and brain cells only express genes necessary for brain function. The procedure that produced Dolly demonstrated that egg cytoplasm is capable of reprogramming an adult differentiated cell (which is only expressing genes related to the function of that cell type). This reprogramming enables the differentiated cell nucleus to once again express all the genes required for the full embryonic development of the adult animal. Since Dolly was cloned, similar techniques have been used to clone a veritable zoo of vertebrates including mice, cattle, rabbitts, mules, horses, fish, cats and dogs from donor cells obtained from adult animals. These spectacular examples of cloning normal animals from fully differentiated adult cells demonstrate the universality of nuclear reprogramming although the next decade called some of these assumptions into question.

This technology supports the production of genetically identical and genetically modified animals. Thus, the successful "cloning" of Dolly has captured the imagination of researchers around the world. This technological breakthrough should play a significant role in the development of new procedures for genetic engineering in a number of mammalian species. It should be noted that nuclear cloning, with nuclei obtained from either mammalian stem cells or differentiated "adult" cells, is an especially important development for transgenic animal research. As the decade reached its end the clones began arriving rapidly with specific advances made by a Japanese group who used cumulus cells rather than fibroblasts to clone calves. They found that the percentage of cultured, reconstructed eggs that developed into blastocysts was 49% for cumulus cells and 23% for oviductal cells. These rates are higher than the 12% previously reported for transfer of nuclei from bovine fetal fibroblasts. Following on the heels of Dolly, Polly and Molly became the first genetically engineered transgenic sheep produced through nuclear transfer technology. Polly and Molly were engineered to produce human factor IX (for hemophiliacs) by transfer of nuclei from transfected fetal fibroblasts. Until then germline competent transgenics had only been produced in mammalian species, other than mice, using DNA microinjection.

Researchers at the University of Massachusetts and Advanced Cell Technology (Worcester, MA) teamed up to produce genetically identical calves utilizing a strategy similar to that used to produce transgenic sheep. In contrast to the sheep cloning experiment, the bovine experiment involved the transfer of nuclei from an actively dividing population of cells. Previous results from the sheep experiments suggested that induction of quiescence by serum starvation was required to reprogram the donor nuclei for successful nuclear transfer. The current bovine experiments indicate that this step may not be necessary.

Typically about 500 embryos needed to be microinjected to obtain one transgenic cow, whereas nuclear transfer produced three transgenic calves from 276 reconstructed embryos. This efficiency is comparable to the previous sheep research where six transgenic lambs were produced from 425 reconstructed embryos. The ability to select for genetically modified cells in culture prior to nuclear transfer opens up the possibility of applying the powerful gene targeting techniques that have been developed for mice. One of the limitations of using primary cells, however, is their limited lifespan in culture. Primary cell cultures such as the fetal fibroblasts can only undergo about 30 population doublings before they senesce. This limited lifespan would preclude the ability to perform multiple rounds of selection. To overcome this problem of cell senescence, these researchers showed that fibroblast lifespan could be prolonged by nuclear transfer. A fetus, which was developed by nuclear transfer from genetically modified cells, could in turn be used to establish a second generation of fetal fibroblasts. These fetal cells would then be capable of undergoing another 30 population doublings, which would provide sufficient time for selection of a second genetic modification.

As noted, there is still some uncertainty over whether quiescent cells are required for successful nuclear transfer. Induction into quiescence was originally thought to be necessary for successful nuclear reprogramming of the donor nucleus. However, cloned calves have been previously produced using non-quiescent fetal cells. Furthermore, transfer of nuclei from Sertoli and neuronal cells, which do not normally divide in adults, did not produce a liveborn mouse; whereas nuclei transferred from actively dividing cumulus cells did produce cloned mice.

The fetuses used for establishing fetal cell lines in a Tufts goat study were generated by mating nontransgenic females to a transgenic male containing a human antithrombin (AT) III transgene. This AT transgene directs high level expression of human AT into milk of lactating transgenic females. As expected, all three offspring derived from female fetal cells were females. One of these cloned goats was hormonally induced to lactate. This goat secreted 3.7–5.8 grams per liter of AT in her milk. This level of AT expression was comparable to that detected in the milk of transgenic goats from the same line obtained by natural breeding.

The successful secretion of AT in milk was a key result because it showed that a cloned animal could still synthesize and secrete a foreign protein at the expected level. It will be interesting to see if all three cloned goats secrete human AT at the identical level. If so, then the goal of creating a herd identical transgenic animals, which secrete identical levels of an important pharmaceutical, would become a

reality. No longer would variable production levels exist in subsequent generations due to genetically similar but not identical animals. This homogeneity would greatly aid in the production and processing of a uniform product. As nuclear transfer technology continues to be refined and applied to other species, it may eventually replace microinjection as the method of choice for generating transgenic livestock.

Nuclear transfer has a number of advantages: 1) nuclear transfer is more efficient than microinjection at producing a transgenic animal, 2) the fate of the integrated foreign DNA can be examined prior to production of the transgenic animal, 3) the sex of the transgenic animal can be predetermined, and 4) the problem of mosaicism in first generation transgenic animals can be eliminated.

DNA microinjection has not been a very efficient mechanism to produce transgenic mammals. However, in November, 1998, a team of Wisconsin researchers reported a nearly 100% efficient method for generating transgenic cattle. The established method of cattle transgenes involves injecting DNA into the pronuclei of a fertilized egg or zygote. In contrast, the Wisconsin team injected a replication-defective retroviral vector into the perivitelline space of an unfertilized oocyte. The perivitelline space is the region between the oocyte membrane and the protective coating surrounding the oocyte known as the zona pellucida.

In addition to ES (embryonic stem) cells other sources of donor nuclei for nuclear transfer might be used such as embryonic cell lines, primordial germ cells, or spermatogonia to produce offspring. The utility of ES cells or related methodologies to provide efficient and targeted *in vivo* genetic manipulations offer the prospects of profoundly useful animal models for biomedical, biological and agricultural applications. The road to such success has been most challenging, but recent developments in this field are extremely encouraging.

12. REPLACEMENT PARTS

With the May 1999 announcement of Geron buying out Ian Wilmuts company Roslin BioMed, they declared it the dawn of an new era in biomedical research. Geron's technologies for deriving transplantable cells from human pluripotent stem cells (hPSCs) and extending their replicative capacity with telomerase was combined with the Roslin Institute nuclear transfer technology, the technology that produced Dolly the cloned sheep. The goal was to produce transplantable, tissue-matched cells that provide extended therapeutic benefits without triggering immune rejection. Such cells could be used to treat numerous major chronic degenerative diseases and conditions such as heart disease, stroke, Parkinson's disease, Alzheimer's disease, spinal cord injury, diabetes, osteoarthritis, bone marrow failure and burns.

The stem cell is a unique and essential cell type found in animals. Many kinds of stem cells are found in the body, with some more differentiated, or committed, to a particular function than others. In other words, when stem cells divide, some of the progeny mature into cells of a specific type (heart, muscle, blood, or brain cells), while others remain stem cells, ready to repair some of the everyday wear and tear

undergone by our bodies. These stem cells are capable of continually reproducing themselves and serve to renew tissue throughout an individual's life. For example, they continually regenerate the lining of the gut, revitalize skin, and produce a whole range of blood cells. Although the term "stem cell" commonly is used to refer to the cells within the adult organism that renew tissue (e.g., hematopoietic stem cells, a type of cell found in the blood), the most fundamental and extraordinary of the stem cells are found in the early-stage embryo. These embryonic stem (ES) cells, unlike the more differentiated adult stem cells or other cell types, retain the special ability to develop into nearly any cell type. Embryonic germ (EG) cells, which originate from the primordial reproductive cells of the developing fetus, have properties similar to ES cells.

It is the potentially unique versatility of the ES and EG cells derived, respectively, from the early-stage embryo and cadaveric fetal tissue that presents such unusual scientific and therapeutic promise. Indeed, scientists have long recognized the possibility of using such cells to generate more specialized cells or tissue, which could allow the generation of new cells to be used to treat injuries or diseases, such as Alzheimer's disease, Parkinson's disease, heart disease, and kidney failure. Likewise, scientists regard these cells as an important – perhaps essential – means for understanding the earliest stages of human development and as an important tool in the development of life-saving drugs and cell-replacement therapies to treat disorders caused by early cell death or impairment.

Geron Corporation and its collaborators at the University of Wisconsin – Madison (Dr. James A. Thomson) and Johns Hopkins University (Dr. John D. Gearhart) announced in November 1998 the first successful derivation of hPSCs from two sources: (i) human embryonic stem (hES) cells derived from *in vitro* fertilized blastocysts (Thomson 1998) and (ii) human embryonic germ (hEG) cells derived from fetal material obtained from medically terminated pregnancies (Shamblott et al. 1998). Although derived from different sources by different laboratory processes, these two cell types share certain characteristics but are referred to collectively as human pluripotent stem cells (hPSCs). Because hES cells have been more thoroughly studied, the characteristics of hPSCs most closely describe the known properties of hES cells.

Stem cells represent a tremendous scientific advancement in two ways: first, as a tool to study developmental and cell biology; and second, as the starting point for therapies to develop medications to treat some of the most deadly diseases. The derivation of stem cells is fundamental to scientific research in understanding basic cellular and embryonic development. Observing the development of stem cells as they differentiate into a number of cell types will enable scientists to better understand cellular processes and ways to repair cells when they malfunction. It also holds great potential to yield revolutionary treatments by transplanting new tissue to treat heart disease, atherosclerosis, blood disorders, diabetes, Parkinson's, Alzheimer's, stroke, spinal cord injuries, rheumatoid arthritis, and many other diseases. By using stem cells, scientists may be able to grow human skin cells to treat wounds and burns. And, it will aid the understanding of fertility disorders. Many patient and scientific organizations recognize the vast potential of stem cell research.

Another possible therapeutic technique is the generation of "customized" stem cells. A researcher or doctor might need to develop a special cell line that contains the DNA of a person living with a disease. By using a technique called "somatic cell nuclear transfer" the researcher can transfer a nucleus from the patient into an enucleated human egg cell. This reformed cell can then be activated to form a blastocyst from which customized stem cell lines can be derived to treat the individual from whom the nucleus was extracted. By using the individual's own DNA, the stem cell line would be fully compatible and not be rejected by the person when the stem cells are transferred back to that person for the treatment.

Preliminary research is occurring on other approaches to produce pluripotent human ES cells without the need to use human oocytes. Human oocytes may not be available in quantities that would meet the needs of millions of potential patients. However, no peer-reviewed papers have yet appeared from which to judge whether animal oocytes could be used to manufacture "customized" human ES cells and whether they can be developed on a realistic timescale. Additional approaches under consideration include early experimental studies on the use of cytoplasmic-like media that might allow a viable approach in laboratory cultures.

On a much longer timeline, it may be possible to use sophisticated genetic modification techniques to eliminate the major histocompatibility complexes and other cell-surface antigens from foreign cells to prepare master stem cell lines with less likelihood of rejection. This could lead to the development of a bank of universal donor cells or multiple types of compatible donor cells of invaluable benefit to treat all patients. However, the human immune system is sensitive to many minor histocompatibility complexes and immunosuppressive therapy carries life-threatening complications.

Stem cells also show great potential to aid research and development of new drugs and biologics. Now, stem cells can serve as a source for normal human differentiated cells to be used for drug screening and testing, drug toxicology studies and to identify new drug targets. The ability to evaluate drug toxicity in human cell lines grown from stem cells could significantly reduce the need to test a drug's safety in animal models.

There are other sources of stem cells, including stem cells that are found in blood. Recent reports note the possible isolation of stem cells for the brain from the lining of the spinal cord. Other reports indicate that some stem cells that were thought to have differentiated into one type of cell can also become other types of cells, in particular brain stem cells with the potential to become blood cells. However, since these reports reflect very early cellular research about which little is known, we should continue to pursue basic research on all types of stem cells. Some religious leaders will advocate that researchers should only use certain types of stem cells. However, because human embryonic stem cells hold the potential to differentiate into any type of cell in the human body, no avenue of research should be foreclosed. Rather, we must find ways to facilitate the pursuit of all research using stem cells while addressing the ethical concerns that may be raised.

Another seminal and intimately related event at the end of the nineties occurred in Madison Wisconsin. Up until November of 1998, isolating ES cells in mammals other than mice proved elusive, but in a milestone paper in the November 5, 1998 issue of Science, James A. Thomson, (1998) a developmental biologist at UW-Madison reported the first successful isolation, derivation and maintenance of a culture of human embryonic stem cells (hES cells). It is interesting to note that this leap was made from mouse to man. As Thomson himself put it, these cells are different from all other human stem cells isolated to date and as the source of all cell types, they hold great promise for use in transplantation medicine, drug discovery and development, and the study of human developmental biology. The new century is rapidly exploiting this vision.
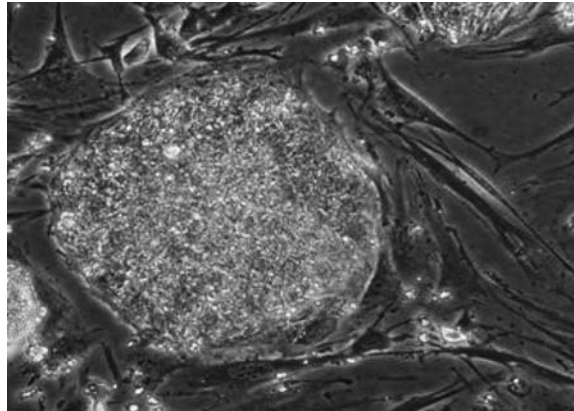

13. CHIPS AHOY

When Steve Fodor was asked in 2003 "How do you really take the Human genome sequence and transform it into knowledge?" he answered from Affymetrix's perspective, it is a technology development task. He sees the colloquially named affychips being the equivalent of a CD-ROM of the genome. They take information from the genome and write it down. The company has come a long way from the early days of Venter's ESTs and less than robust algorithms as described earlier.

One surprising fact unearthed by the newer more sophisticated generation of chips is that 30 to 35 percent of the non-repetitive DNA is being expressed as accepted knowledge was that only 1.5 to 2 percent of the genome would be expressed. Since much of that sequence has no protein-coding capacity it is most likely coding for regulatory functions. In a parallel to astrophysics this is often referred to in common parlance as the "dark matter of the genome" and like dark matter for many it is the most exciting and challenging aspect of uncovering the occult genome.

It could be, and most probably is, involved in regulatory functions, networks, or development. And like physical dark matter it may change our whole concept of what exactly a gene is or is not! Since Beadle and Tatum's circumspect view of the protein world no longer holds true it adds a layer of complexity to organizing chip design. Depending on which sequences are present in a particular transcript, you can, theoretically, design a set of probes to uniquely distinguish that variant. At the DNA level itself there is much potential for looking at variants either expressed or not at a very basic level as a diagnostic system, but ultimately the real paydirt is the information that can be gained from looking at the consequence of non-coding sequence variation on the transcriptome itself.

And fine tuning when this matters and when it is irrelevant as a predicative model is the auspices of the Affymetrix spin-off Perlegen. Perlegen came into being in

Eukaryotic Target Labeling for GeneChip® Probe Arrays
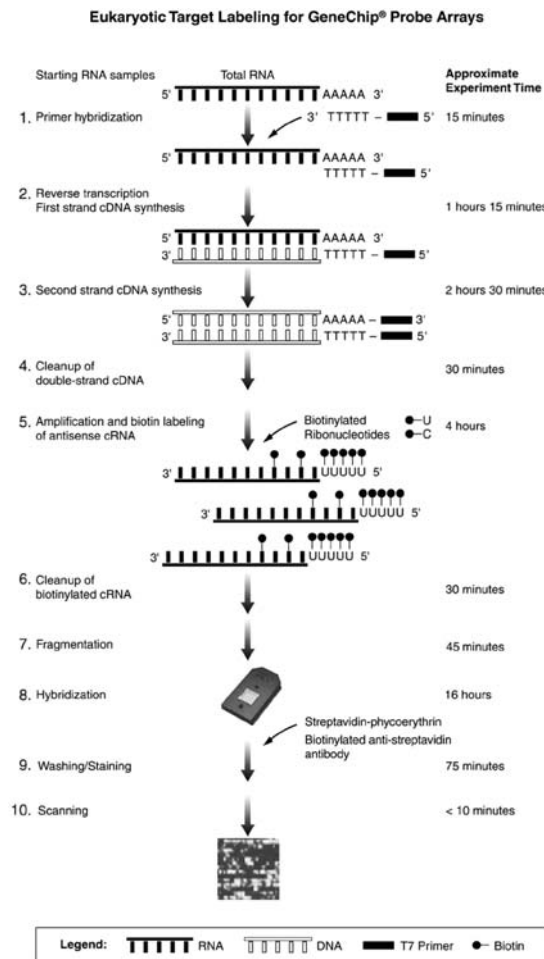
late 2000 to accelerate the development of high-resolution, whole genome scanning. And they have stuck to that purity of purpose. To paraphrase Dragnet's Sergeant Joe Friday, they focus on the facts of DNA just the DNA. Perlegen owes its true genesis to the desire of one of its cofounders to use DNA chips to help understand the dynamics underlying genetic diseases. Brad Margus' two sons have the rare disease "ataxia telangiectasia" (A-T). A-T is a progressive, neurodegenerative childhood disease that affects the brain and other body systems. The first signs of the disease, which include delayed development of motor skills, poor balance, and slurred speech, usually occur during the first decade of life. Telangiectasias (tiny, red "spider" veins), which appear in the corners of the eyes or on the surface of the ears and cheeks, are characteristic of the disease, but are not always present. Many individuals with A-T have a weakened immune system, making them susceptible to recurrent respiratory infections. About 20% of those with A-T develop cancer, most frequently acute lymphocytic leukemia or lymphoma suggesting that the sentinel competence of the immune system is compromised.

Having a focus so close to home is a powerful driver for any scientist. His co-founder David Cox is a polymath pediatrician whose training in the latter informs his application of the former in the development of patient-centered tools. From that perspective, Perlegen's stated mission is to collaborate with partners to rescue or improve drugs and to uncover the genetic bases of diseases. They have created a whole genome association approach that enables them to genotype millions of unique SNPs in thousands of cases and controls in a timeframe of months rather than years. As mentioned previously, SNP (single nucleotide polymorphism) markers are preferred over microsatellite markers for association studies because of their abundance along the human genome, the low mutation rate, and accessibilities to high-throughput genotyping. Since most diseases, and indeed responses to drug
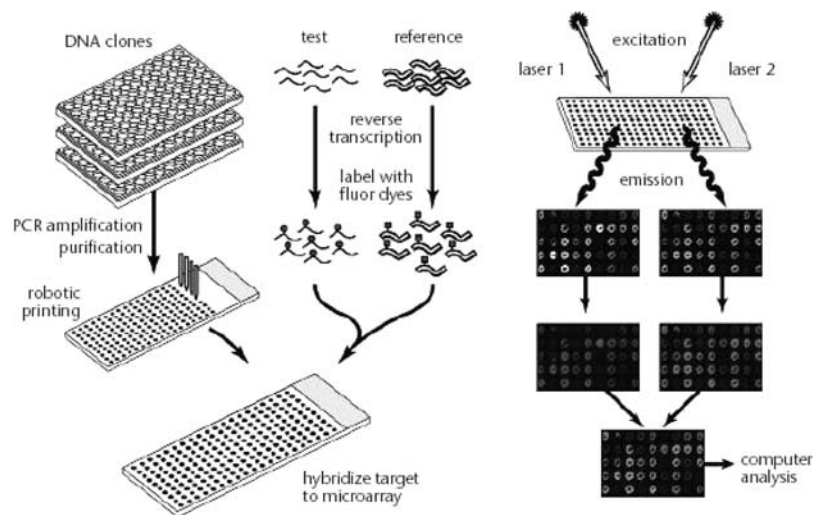
interventions, are the products of multiple genetic and environmental factors it is a challenge to develop discriminating diagnostics and, even more so, targeted-therapeutics. Because mutations involved in complex diseases act probabilistically – that is, the clinical outcome depends on many factors in addition to variation in the sequence of a single gene – the effect of any specific mutation is smaller. Thus, such effects can only be revealed by searching for variants that differ in frequency among large numbers of patients and controls drawn from the general population. Analysis of these SNP patterns provides a powerful tool to help achieve this goal.

Although most bi-alleic SNPs are rare, it has been estimated that just over 5 million common SNPs, each with a frequency of between 10 and 50%, account for the bulk of the DNA sequence difference between humans. Such SNPs are present in the human genome once every 600 base pairs or so. As is to be expected from linkage disequilibrium studies, alleles making up blocks of such SNPs in close physical proximity are often correlated, resulting in reduced genetic variability and defining a limited number of "SNP haplotypes," each of which reflects descent from a single, ancient ancestral chromosome. In 2001 Cox's group, using high level scanning with some old-fashioned somatic cell genetics, constructed the SNP map of Chromosome 21.The surprising findings were blocks of limited haplotype diversity in which more than 80% of a global human sample can typically be characterized by only three common haplotypes (interestingly enough the prevalence of each hapolytype in the examined population was in the ratio 50:25:12.5).From this the conclusion could be drawn that by comparing the frequency of genetic variants in unrelated cases and controls, genetic association studies could potentially identify specific haplotypes in the human genome that play important roles in disease, without need of knowledge of the history or source of the underlying sequence, which hypothesis they subsequently went on to prove.

Following Cox et al. pioneering work on "blocking" Chromosome 21 into characteristic haplotypes, Tien Chen came to visit him from University of Southern California and following the visit his group developed discriminating algorithms which took advantage of the fact that the haplotype block structure can be decomposed into large blocks with high linkage disequilibrium and relatively limited haplotype diversity, separated by short regions of low disequilibrium. One of the practical implications of this observation is as suggested by Cox that only a small fraction of all the SNPs they refer to as "tag" SNPs can be chosen for mapping genes responsible for complex human diseases, which can significantly reduce genotyping effort, without much loss of power. They developed algorithms to partition haplotypes into blocks with the minimum number of tag SNPs for an entire chromosome. In 2005 they reported that they had developed an optimized suite of programs to analyze these block linkage disequilibrium patterns and to select the corresponding tag SNPs that will pick the minimum number of tags for the given criteria. In addition the updated suite allows haplotype data and genotype data from unrelated individuals and from general pedigrees to be analyzed.

Using an approach similar to Richard Michelmore's bulk segregant analysis in plants of more than a decade previously, Perlegen subsequently made use of these

SNP haplotype and statistical probability tools to estimate total genetic variability of a particular complex trait coded for by many genes, with any single gene accounting for no more than a few percent of the overall variability of the trait. Cox's group have determined that fewer than 1000 total individuals provide adequate power to identify genes accounting for only a few percent of the overall genetic variability of a complex trait, even using the very stringent significance levels required when testing large numbers of DNA variants. From this it is possible to identify the set of major genetic risk factors contributing to the variability of a complex disease and/or treatment response. So, while a single genetic risk factor is not a good predictor of treatment outcome, the sum of a large fraction of risk factors contributing to a treatment response or common disease can be used to optimize personalized treatments without requiring knowledge of the underlying mechanisms of the disease.They feel that a saturating level of coverage is required to produce repeatable prediction of response to medication or predisposition to disease and that taking shortcuts will for the most part lead to incomplete, clinically-irrelevant results.

In 2005 Hinds et al. in Science describe even more dramatic progresss. They describe a publicly available, genome-wide data set of 1.58 million common single-nucleotide polymorphisms (SNPs) that have been accurately genotyped in each of 71 people from three population samples. A second public data set of more than 1 million SNPs typed in each of 270 people has been generated by the International Haplotype Map (HapMap) Project. These two public data sets, combined with multiple new technologies for rapid and inexpensive SNP genotyping, are paving the way for comprehensive association studies involving common human genetic variations.

Perlegen basically is taking to the next level Fodor's stated reason for the creation of Affymetrix, the belief that understanding the correlation between genetic variability and its role in health and disease would be the next step in the genomics revolution. The other interesting aspect of this level of coverage is, of course, the notion of discrete identifiable groups based on ethnicity, centers of origin and such breaks down and a spectrum of variation arises across all populations which makes the Perlegen chip, at one level, a true unifier of humanity but at another adds a whole layer of complexity for HMOs!

At the turn of the century, this personalized chip approach to medicine received some validation at a simpler level in a closely related disease area to the one to which one fifth of A-T patients ultimately succumb when researchers at the Whitehead Institute used DNA chips to distinguish different forms of leukemia based on patterns of gene expression in different populations of cells. Moving cancer diagnosis away from visually based systems to such molecular based systems is a major goal of the National Cancer Institute. In the study, scientists used a DNA chip to examine gene activity in bone marrow samples from patients with two different types of acute leukemia – acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). Then, using an algorithm, developed at the Whitehead, they identified signature patterns that could distinguish the two types.

When they cross-checked the diagnoses made by the chip against known differences in the two types of leukemia, they found that the chip method could automatically make the distinction between AML and ALL without previous knowledge of these classes. Taking it to a level beyond where Perlegen are initially aiming, Eric Lander, leader of the study said, mapping not only what is in the genome, but also what the things in the genome do, is the real secret to comprehending and ultimately curing cancer and other diseases.

Chips gained recognition on the world stage in 2003 when they played a key role in the search for the cause of Severe Acute Respiratory Syndrome (SARS) and probably won a McArthur genius award for their creator. UCSF Assistant Professor Joseph DeRisi, already famous in the scientific community as the wunderkind originator of the online DIY chip maker in Pat Brown's lab at Stanford, built a gene microarray containing all known completely sequenced viruses (12,000 of them) and, using a robot arm that he also customized, in a three day period used it to classify a pathogen isolated from SARS patients as a novel coronavirus. When a whole galaxy of dots lit up across the spectrum of known vertebrate cornoviruses DeRisis knew this was a new variant. Interestingly the sequence had the hottest signal with Avian Infectious Bronchitis Virus. His work subsequently led epidemiologists to target the masked palm civet, a tree-dwelling animal with a weasel-like face and a catlike body as the probable primary host. The role that DeRisi's team at UCSF played in identifying a coronavirus as a suspected cause of SARS came to the attention of the national media when CDC Director Dr. Julie Gerberding recognized Joe in March 24, 2003 press conference and in 2004 when Joe was honored with one of the coveted McArthur genius awards.

This and other tools arising from information gathered from the human genome sequence and complementary discoveries in cell and molecular biology, new tools such as gene-expression profiling, and proteomics analysis are converging to finally show that rapid robust diagnostics and "rational" drug design has a future in disease research.

Another virus that puts SARS deaths in perspective benefitted from rational drug design at the turn of the century. Influenza, or flu, is an acute respiratory infection caused by a variety of influenza viruses. Each year, up to 40 million Americans develop the flu, with an average of about 150,000 being hospitalized and 10,000 to 40,000 people dying from influenza and its complications. The use of current influenza treatments has been limited due to a lack of activity against all influenza strains, adverse side effects, and rapid development of viral resistance. Influenza costs the United States an annual $14.6 billion in physician visits, lost productivity and lost wages. And least we still dismiss it as a nuisance we are well to remember that the "Spanish" influenza pandemic killed over 20 million people in 1918 and 1919, making it the worst infectious pandemic in history beating out even the more notorious black death of the Middle Ages. This fear has been rekindled as the dreaded H5N1 (H for haemaglutenin and N for neuraminidase as described below) strain of bird flu has the potential to mutate and recognise homo sapiens as a desirable host. Since RNA viruses are notoriously faulty in their replication this

accelerated evolutionary process gives then a distinct advantage when adapting to new environments and therefore finding more amenable hosts.

Although inactivated influenza vaccines are available, their efficacy is suboptimal partly because of their limited ability to elicit local IgA and cytotoxic T cell responses. The choices of treatments and preventions for influenza hold much more promise in this millennium. Clinical trials of cold-adapted live influenza vaccines now under way suggest that such vaccines are optimally attenuated, so that they will not cause influenza symptoms but will still induce protective immunity. Aviron (Mountain View, CA), BioChem Pharma (Laval, Quebec, Canada), Merck (Whitehouse Station, NJ), Chiron (Emeryville, CA), and Cortecs (London), all had influenza vaccines in the clinic at the turn of the century, with some of them given intra-nasally or orally. Meanwhile, the team of Gilead Sciences (Foster City, CA) and Hoffmann-La Roche (Basel, Switzerland) and also GlaxoWellcome (London) in 2000 put on the market neuraminidase inhibitors that block the replication of the influenza virus.

Gilead was one of the first biotechnology companies to come out with an anti-flu therapeutic. Tamiflu™ (oseltamivir phosphate) was the first flu pill from this new class of drugs called neuraminidase inhibitors (NI) that are designed to be active against all common strains of the influenza virus. Neuraminidase inhibitors block viral replication by targeting a site on one of the two main surface structures of the influenza virus, preventing the virus from infecting new cells. Neuraminidase is found protruding from the surface of the two main types of influenza virus, type A and type B. It enables newly formed viral particles to travel from one cell to another in the body. Tamiflu is designed to prevent all common strains of the influenza virus from replicating. The replication process is what contributes to the worsening of symptoms in a person infected with the influenza virus. By inactivating neuraminidase, viral replication is stopped, halting the influenza virus in its tracks.

In marked contrast to the usual protracted process of clinical trials for new therapeutics, the road from conception to application for Tamiflu was remarkably expeditious. In 1996, Gilead and Hoffmann-La Roche entered into a collaborative agreement to develop and market therapies that treat and prevent viral influenza. In 1999, as Gilead's worldwide development and marketing partner, Roche led the final development of Tamiflu, 26 months after the first patient was dosed in clinical trials in April 1999, Roche and Gilead announced the submission of a New Drug Application to the U.S. Food and Drug Administration (FDA) for the treatment of influenza. Additionally, Roche filed a Marketing Authorisation Application (MAA) in the European Union under the centralized procedure in early May 1999. Six months later in October 1999, Gilead and Roche announced that the FDA approved Tamiflu for the treatment of influenza A and B in adults. These accelerated efforts allowed Tamiflu to reach the U.S. market in time for the 1999–2000 flu season. One of Gilead's studies showed an increase in efficacy from 60% when the vaccine was used alone to 92% when the vaccine was used in conjunction with a neuraminidase inhibitor. Outside of the U.S., Tamiflu also has been approved for the treatment

of influenza A and B in Argentina, Brazil, Canada, Mexico, Peru and Switzerland. Regulatory review of the Tamiflu MAA by European authorities is ongoing. With the H5N1 birdflu strain's relentless march (or rather flight) across Asia, in 2006 through Eastern Europe to a French farmyard, an unwelcome stowaway on a winged migration, and no vaccine in sight, Tamiflu, although untested for this species, seen as the last line of defense is now being horded and its patented production right's fought over like an alchemist's formula.

Tamiflu's main competitor, Zanamivir marketed as Relenza™ was one of a group of molecules developed by GlaxoWellcome and academic collaborators using structure-based drug design methods targeted, like Tamiflu, at a region of the neuraminidase surface glycoprotein of influenza viruses that is highly conserved from strain to strain. Glaxo filed for marketing approval for Relenza in Europe and Canada.

The Food and Drug Administration's accelerated drug-approval timetable began to show results by 2001, its evaluation of Novartis's Gleevec took just three months compared with the standard 10–12 months. Another factor in improving biotherapeutic fortunes in the new century was the staggering profits of early successes. In 2003, $1.9 billion of the $3.3 billion in revenue collected by Genentech in South San Francisco came from oncology products, mostly the monoclonal antibody-based drugs Rituxan, used to treat non-Hodgkin's lymphoma, and Herceptin for breast cancer. In fact two of the first cancer drugs to use the new tools for 'rational' design Herceptin and Gleevec, a small-molecule chemotherapeutic for some forms of leukemia are proving successful, and others such as Avastin (an anti-vascular endothelial growth factor) for colon cancer and Erbitux are already following in their footsteps. Gleevec led the way in exploiting signal-transduction pathways to treat cancer as it blocks a mutant form of tyrosine kinase (termed the Philadelphia translocation recognized in 1960's) that can help to trigger out-of-control cell division.

About 25% of biotech companies raising venture capital during the third quarter of 2003 listed cancer as their primary focus, according to online newsletter VentureReporter. By 2002 according to the Pharmaceutical Research and Manufacturers of America, 402 medicines were in development for cancer up from 215 in 1996. Another new avenue in cancer research is to combine drugs. Wyeth's Mylotarg, for instance, links an antibody to a chemotherapeutic, and homes in on CD33 receptors on acute myeloid leukemia cells. Expertise in biochemistry, cell biology and immunology is required to develop such a drug. This trend has created some bright spots in cancer research and development, even though drug discovery in general has been adversely affected by mergers, a few high-profile failures and a shaky US economy in the early 2000's.

As the millennium approached observers as diverse as Microsoft's Bill Gates and President Bill Clinton predicted the 21st century wiould be the "biology century". By 1999 the many programs and initiatives underway at major research institutions and leading companies were already giving shape to this assertion. These initiatives have ushered in a new era of biological research anticipated to generate technological changes of the magnitude associated with the industrial revolution and the computer-based information revolution.

REFERENCES

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 21:1651–1656

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arntzen CJ (1997) High-tech herbal medicine: Plant-Based Vaccines. Nat Biotechnol 15:221–222

Berg P, et al. (1975) Asilomar conference on recombinant DNA molecules. Science 188:991–994.

Berg P, Baltimore D, Boyer HW (1974) Potential biohazards of recombinant DNA molecules. Science 185:303.

Bodmer WF (1991) HUGO: The Human Genome Organization. FASEB J 5(1):73–74

Brennan FR, Bellaby T, Helliwell SM, Jones TD, Ka-mstrup S, Dalsgaard K, Flock JI, Hamilton WDO (1999) Chimeric plant virus particles administered nasally or orally induce systemic and mucosal immune responses in mice. J Virol 73:930–935

Brenner S (1990) The human genome: the nature of the enterprise. In Human Genetic Information: Science, Law and Ethics. Ciba Found Symp 149:6–12

Charles C (1990) Orchestrating the Human Genome Project. Science 248(4951):49–51

Cohen AS, Najarian DR, Karger BL (1990) Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis. J Chromatogr 516(1):49–60

Della Penna D (1999) Nutritional genomics: Manipulating plant micronutrients to improve human health. Science 285:375–379

Dickman S, Aldhous P (1991) Helping Europe compete in human genome research. Nature 350(6316):261

Dickson D (1989) Genome project gets rough ride in Europe. Science 243:599 Down Memory Lane, Nature S1, 2006

Green P, Helms C, Weiffenbach B, Stephens K, Keith T, Bowden D, Smith D, Donis-Keller H (1989) Construction of a linkage map of the human genome, and its application to mapping genetic diseases. Clin Chem 35(7 Suppl.):B33–B37

Heiger DN, Cohen AS, Karger BL (1990) Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. J Chromatogr 516(1):33–48

King DS (1999) Preimplantation and the 'new' genetics. J Med Ethics 25:176–182

Leslie R, Davenport RJ, Pennisi E, Marshall E (2001) A history human genome project. Science 291(5507):1195 (in News Focus)

Lewontin R (2000) It aint necessarily so: The dream of the human genome and other illusions. New York Review of Books

Luckey JA, Drossman H, Kostichka AJ, Mead DA, D'Cunha J, Norris TB, Smith LM (1990) High speed DNA sequencing by capillary electrophoresis. Nucleic Acids Res 18(15):4417–4421

Marshall E (1995) A strategy for sequencing the genome 5 years early. Science 267:783–784

Mason H et al. (1996) Expression of norwalk virus capsid protein in transgenic tobacco and potato and its oral immunogenicity in mice. Proc Natl Acad Sci USA 93:5335–5340

McCormick AA, Kumagai MH, Hanley K, Turpen TH, Hakim I, Grill LK, Tusé D, Levy S, Levy R (1999) Rapid production of specific vaccines for lymphoma by expression of the tumor-derived single-chain Fv epitopes in tobacco plants. Proc Natl Acad Sci USA 96:703–708

Murr L et al. (1996) Generation and analysis of 280,000 human expressed sequence tags. Genome Res 6:807–828

NAS (1987) National Academy of Sciences. Introduction of recombinant DNA-engineered organisms into the environment: key issues. National Academy Press, Washington, D.C.

Newell-McGloughlin M (2006) Functional Foods and Biopharmaceuticals: The Next Generation of the GM Revolution in Let Them Eat Precaution, Jon Entine, Ed. published by AEI Press, pp 163–178.

Newell-McGloughlin M and Burke J (2001) Biotechnology: A Review of Technological Developments, Publishers Forfas, Dublin Ireland

Potrykus I (1999) Vitamin-A and iron-enriched rices may hold key to combating blindness and malnutrition: a biotechnology advance. Nat Biotechnol 17:37

Rabinow P (1999) French DNA: Trouble in Purgatory, University of Chicago Press, Chicago, viii + 201 pp

Ridley M (2000) Genome: The Autobiography of a Species in 23 Chapters Harper Collins, New York.

Shamblott, Michael Shamblott, Joyce Axelman, Shunping Wang, Elizabeth M. Bugg, John W. Littlefield, Peter J. Donovan, Paul D. Blumenthal, George R. Huggins, and John D. Gearhart (1998) "Derivation of pluripotent stem cells from cultured human primordial germ cells," *PNAS* 95:13726–13731

Sijmons PC, Dekker BMM, Schrammeijer B, Verwoerd TC, van den Elzen PJM, Hoekema A (1990) Production of correctly processed human serum albumin in transgenic plants. Bio/Technology 8:217–221

Staub J, Garcia B, Graves J, Hajdukiewicz P, Hunter P, Nehra N, Paradkar V, Schlittler M, Carroll J, Spatola L, Ward D, Ye G, Russell D (2000) High-yield production of a human therapeutic protein in tobacco chloroplasts. Nat Biotechnol 18:333–338

Sulston J, Ferry G (2002) The common thread: A story of science, politics, ethics and the human genome, 1st edn. National Academies Press

Swerdlow H, Wu SL, Harke H, Dovichi NJ (1990) Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. J Chromatogr 516(1):61–67

Terashima M, Murai Y, Kawamura M, Nakanishi S, Stoltz T, Chen L, Drohan W, Rodriguez RL, Katoh S (1999) Production of functional human alpha 1-antitrypsin by plant cell culture. Appl Microbiol Biotechnol 52:516–523

Thomson, James A, Joseph Itskovitz-Eldor, Sander S, Shapiro, Michelle A, Waknitz, Jennifer J, Swiergiel, Vivienne S, Marshall, and Jeffrey M, (1998) Jones Science 6 November, 282:1145–1147

Ursin V (2000) Genetic modification of oils for improved health benefits, Presentation at conference, Dietary Fatty Acids and Cardiovascular Health: Dietary Recommendations for Fatty Acids: Is There Ample Evidence? American Heart Association, Reston VA, June 5–6

Verwoerd TC, van Paridon PA, van Ooyen AJJ, van Lent JWM, Hoekema A, Pen J (1995) Stable accumulation of *Aspergillus niger* phytase in transgenic tobacco leaves. Plant Physiol 109:1199–1205

Wald NJ, Kennard A, Densem JW, Cuckle HS, Chard T, Butler L (1992) Antenatal maternal serum screening for Down's syndrome: results of a demonstration project. Br Med J 305:391–394

Wilmut I, Schnieke AE, McWhir J, Kind WAJ, Campbell KHS, (1997) Viable offspring derived from fetal and adult mammalian cells. Nature 385:810