1 CHAPTER 1

The gene in the twenty-first century

Choong-Chin Liew, PhD*, & Victor J. Dzau,* MD

Introduction

When the word was first used in 1909, "gene" was a hypothesis necessary to explain puzzling observations about heredity. As the century progressed, the hypothesis began to acquire reality as the structure and functions of the gene were gradually elucidated. Earlier and simpler concepts became superseded as evidence led to better understanding of the gene. Today the gene is recognized to be a highly complex entity. The genomics revolution is well underway but there is much that remains for twenty-first century science to learn before the potential of molecular biology and technology can be fully realized.

The search for the gene

Much of the science that laid the foundation for the genetics and genomics revolution took place in the very near past; 1900 is the date often considered to be the beginning of modern genetics. In that year, three botanists working on plant hybridization, independently, in three different countries, published their rediscovery of Gregor Mendel's (1822–1884) rules of inheritance, first presented in 1865 and then largely forgotten [1]. Carl Correns (1864–1933) in Germany, Hugo de Vries (1848–1935) in the Netherlands and Erich von Tschermak (1871–1962) in Austria each published their findings in the *Berichte der Deutsche Botanischen Gesellschaft* (Proceedings of the German Botanical Society) [2–4]. The botanists recognized that Mendel's concept of dominant and recessive traits could be used to explain how traits can skip generations, appearing and disappearing through the years. Hugo de Vries named the transmitted substances "pangens"; he later coined the term "mutation" to signify the appearance of a new pangen [5].

Cambridge evolutionist William Bateson (1861–1926) translated Mendel into English and worked vigourously to promote Mendel's ideas in the English-speaking scientific world. Bateson himself coined the term "genetics" in 1906 [6]. The word "gene" was not introduced until 1909, when Wilhelm Johannsen (1857–1927), a Danish botanist, offered this term in preference to earlier terms [7].

A next major step towards an elucidation of the gene came with the discovery that genes have physical locations on chromosomes in studies on *Drosophila* carried out by Thomas Hunt Morgan (1866–1945) and his colleagues at the zoology department of Columbia University [8,9]. Morgan's student, Alfred Sturtevant (1891–1970) was able to show that the gene for a trait was localized in a fixed location or locus arranged "like beads on a string" in his often quoted metaphor [10]. Later, Calvin Bridges was able to visualize this arrangement using light microscopy to show in detail the parallel bands on the chromosomes of the salivary gland cells of larval fruit flies [11].

In 1927, another of Morgan's students, Herman J. Muller (1890–1967) proved in studies at the University of Texas, Austin, that ionizing radiation from X-rays and other mutagens could be used to create genetic mutations in fruit flies, and that some of these mutations were able to pass to offspring [12].

Muller believed as early as the 1920s that genes were "the basis of life" [13]. However, it was not until the 1940s that researchers began to work out

the physical and material properties of genes. In 1944, Rockefeller University researchers, Oswald Avery (1877–1955), Colin MacLeod (1909–1972) and Maclyn McCarty (1911–2005) demonstrated that it was DNA that was the carrier of genetic information [14]. In 1952, Alfred Hershey (1908–1997) and his laboratory assistant Martha Chase (1928–2003) confirmed Avery's findings [15].

Against this background can be understood the importance mid-century of James Watson (b. 1928) and Francis Crick's (1916–2004) double helix. In their landmark paper, published in *Nature* in 1953, Watson and Crick presented for the first time a comprehensible model of a unit of heredity [16]. Briefly, their double helix is composed of two long polymers of alternating sugar-phosphate deoxyribose molecules, like the sides of a twisted spiral ladder. To these molecules Watson and Crick attached the ladder's rungs, four nucleotide bases: adenine and guanine (A and G) and cytosine and thymine (C and T). The property of each base is such that it attracts and bonds to its complementary base forming arrangements known as base pairs: "A" can only pair with "T" and "C" can only pair with "G." The DNA bases are loosely attached to each other by weak bonds; they are released from each other by disrupting the bonds. Thus, every time a cell divides it copies its DNA program, in the human cell, it copies its entire three billion base pair human genome.

Once the structure of the gene was described the molecule could take its place in the scientific ontology of the twentieth century. With the double helix, classic genetics began to shift to molecular genetics [8].

Gene function

Studies of gene function proceeded largely independently of investigations into gene structure. The first clue to the biologic behaviour of the gene in the organism came in 1902, with the work of London physician Archibald Garrod (1857–1936) [17]. In his famous paper published in the *Lancet* in 1902, Garrod hypothesized that alkaptonuria was a consequence of some flaw in body chemistry that disrupts one of the chemical steps in the metabolism of tyrosine [18]. He explained alkaptonuria as a recessive disorder, using the terms of the new Mendelian genetics and conjectured that it is an absence of the enzyme involved that leads to alkaptonuria and other "inborn errors of metabolism" [19].

Garrod's hypothesis was given experimental support in an important series of studies on *Neurospora crassa* carried out at Stanford University by George Beadle (1903–1989) and Edward Tatum (1909–1975) between 1937 and 1941 [20]. Because biochemical processes are catalyzed by enzymes and because mutations affect genes, reasoned Beadle and Tatum, then genes must make enzymes: the "one-gene-one-enzyme" hypothesis, later made famous by Beadle.

The hypothesis was further developed in studies on sickle cell anemia. In 1949, the hereditary basis of the disorder was shown by James Neel (1915–2000) [21]. Also in 1949, Linus Pauling (1901–1994) and Harvey Itano showed that the disease was linked to a modification in hemoglobin, such that the hemoglobin in sickled cells carries a charge different to the charge of the molecule in normal cells [22].

Eight years later, Vernon Ingram (1924–2006) and Francis Crick demonstrated that this difference was caused by the replacement of a single amino acid, glutamic acid, by another, valine, at a specific position in the long hemoglobin protein [23]. Sickle cell anemia was the first disease explicitly identified as a disorder flowing from a derangement at the molecular level, or as Pauling himself put it, "a molecular disease" [22].

Understanding of gene function sped up once Watson and Crick had elucidated the structure of the double helix. As summarized in Crick's famous central dogma of 1958, information flows from DNA to RNA to protein [24]. The central dogma captured the imagination of biologists, the public and the media in the 1960s and 1970s [25]. "Stunning in its simplicity," Evelyn Fox Keller writes, the central dogma allows us to think of the cell's DNA as "the genetic program, the lingua prima, or perhaps, best of all, the book of life" [25].

The gene since 1960

By 1960, the definition of a gene was that implied by the central dogma: a gene is a segment of DNA that codes for a protein [26]. The first significant challenge to that definition arose with the work of

microbiologists Francois Jacob (b. 1920) and Jacques Monod (1910–1976) of the Institut Pasteur in Paris [27].

Monod and Jacob's operon model explained gene function in terms of gene cluster. However, such a model adds levels of complexity to the gene and makes it more difficult to determine precisely what is a gene. What should be included in one gene? Its regulatory elements? Its coding elements? What are the boundaries of the gene? [25]. Furthermore, the Monod/Jacob gene loses some of its capacity for self-regulation: on the operon model the gene acts, not autonomously, but in response to proteins within the cell and between the cell and its environment [28].

Although Monod famously asserted that what was true for *Escherichia coli* would be true for the elephant, in fact the operon model of gene regulation characterizes prokaryotes (simple unicellular organisms without nuclei). In eukaryotes (animals and plants whose cells contain nuclei), gene regulation is far more complicated. Later research showed that in some cases, regulatory elements were scattered at sites far away from the coding regions of the gene; in other cases, regulator genes were found to be shared by several genes; gene regulation included further levels of control including positive control mechanisms, attenuation mechanisms, complex promoters, enhancers and multiple polyadenylation sites, making it even more difficult to clarify the boundaries of the gene (for discussion on difficulties in defining the gene see [25,29–31]).

Later, in 1970, Howard Temin (1934–1994) and David Baltimore (b. 1938) also posed challenges to the one way DNA-to-protein pathway implicit in the central dogma. In their work on viruses that can cause cancer they discovered an enzyme, reverse transcriptase, which uses RNA as the template to synthesize DNA [32,33].

Another unexpected finding to shake the central dogma occurred with the discovery of the split gene. In 1977, Phillip Sharp (b. 1944) at the Massachusetts Institute of Technology and Richard Rogers (b. 1943) at Cold Spring Harbor Laboratory showed that not all genes are made of one continuous series of nucleotides. Their electron microscopy comparisons of adenovirus DNA and mRNA showed that some genes are split, or fragmented into regions of coding pieces of DNA interrupted by stretches of non-coding DNA [34,35]. Walter Gilbert (b. 1932) of Harvard University later coined the terms "exon" and "intron" to describe these regions [36].

Split genes can be spliced, or alternatively spliced, in different ways: exons can be excised out, some introns can be left in, or the primary transcript can be otherwise recombined (for review see [37]). The proteins thereby produced are similar although slightly different isoforms. Split genes play havoc with the straightforward one-gene-one-enzyme hypothesis. As Keller has pointed out "one gene – many proteins" is an expression common in the literature of molecular biology today [25].

Other nontraditional genes discovered (or become accepted by the research community) since the 1960s include transposons, moveable genes that travel from place to place in the genome of a cell where they affect the expression of other genes discovered by Barbara McClintock [38]; nested genes, whose exon sequences are contained within other genes; and pseudogenes which are "dead" or nonfunctional gene remnants, overlapping genes, repeated genes and other gene types (these and other "nonclassical" genes are reviewed in [30]).

Most recently, with the discovery of nonprotein coding RNAs, the idea that genes necessarily make proteins at all has been called into question. As far back as 1968, Roy Britten and David Kohne published a paper in *Science* reporting that long stretches of DNA do not seem to code for proteins at all [39]. Large areas of genome – hundreds to thousands of base pairs – seemed to consist of monotonous nucleotide sequence repetition of DNA. Such noncoding DNA, which includes introns within genes and areas between coding genes, represents a surprising fraction of the genomes, at least genomes of higher organisms. In humans, 98% of human DNA appears not to code for anything. Only a tiny percentage – about 2% – of the three billion base pairs of the human genome corresponds to the 20,000–25,000 protein coding genes tallied by the International Human Genome Sequencing Consortium [40]. Much of it consists of repeated DNA; some elements are repeated over 100,000 times in the genome with no apparent purpose. Such noncoding elements were long dismissed as parasitic or "junk" DNA: a chance by-product of evolution with no discernible function.

Table 1.1 Recently discovered noncoding RNA families and their functions.

Family		*Processes affected*
miRNAs	microRNAs	translation/regulation
siRNAs	small interfering RNAs	RNA interference/gene silencing
snRNAs	small nuclear RNAs	RNA processing/spliceosome components
st RNAs	small temporal RNAs	temporal regulation/translation
snoRNAs	small nucleolar RNAs	ribosomal RNA processing/modification
cis-antisense RNAs		transcription elongation/RNA processing/stability/mRNA translation

Adapted from Storz *et al.* [44].

Since the sequencing of the human and other genomes, however, and with the availability of transcriptomes and novel genomic technologies such as cDNA cloning approaches and genome tiling microarrays, researchers have begun to explore intronic and intragenic space. Increasingly since 2001 it appears that far from being junk, these stretches of DNA are rich in "gems" [41]: small genes that produce RNAs, called noncoding RNAs (reviewed in [42–44]) (Table 1.1).

Noncoding RNAs are not messenger RNAs, transfer RNAs or ribosomal RNAs, RNA species whose functions have long been known. They vary in size from tiny 20–30 nucleotide-long microRNAs to 100–200 nucleotide-long nonprotein coding RNAs (ncRNAs) in bacteria to more than 10,000 nucleotide-long RNAs involved in gene silencing. Many of these intriguing ncRNAs are highly conserved through evolution, and many seem to have important structural, catalytic and regulatory properties [45].

Noncoding RNAs were thought at first to be unusual; however, over the past 5 years increasing numbers of these intriguing elements have been emerging. The number of ncRNAs in mammalian transcriptomes is unknown, but there may be tens of thousands; it has been estimated that some 50% of the human genome transcriptome is made up of ncRNAs [46].

The function of these elements is only beginning to be explored; and their structural features are beginning to be modeled [47]. Nonprotein coding RNAs seem to be fundamental agents in primary molecular biologic processes, affecting complex regulatory networks, RNA signaling, transcription initiation, alternative splicing, developmental timing, gene silencing and epigenetic pathways [44].

One class of ncRNAs has been the focus of much research attention. MicroRNAs are hairpin-shaped RNAs first discovered in *Caenorhabditis elegans* [48,49]. These tiny, approximately 22 nucleotide elements seem to control aspects of gene expression in higher eukaryote plants and animals. Many microRNAs are highly conserved through evolution; others are later evolutionary elements. For example, of some 1500 microRNAs in the human genome [46], 53 are unique to primates [50]. Each microRNA may regulate as many as 200 target genes in a cell, or one-third of the genes in the human genome [51].

In animals, microRNAs appear to repress translation initiation or destabilize messenger RNA. In animals, microaRNAs so far characterized seem to be involved in developmental timing, neuronal cell fate, cell death, fat storage and hematopoietic cell fate [49]. The potential effects of these RNA elements on gene expression have led to the hypothesis that these elements may be involved in disease processes. For example, microRNAs have been suggested to be involved in cancer pathogenesis, acting as oncogenes and tumor suppressors [52]. Calin *et al.* [53] recently reported a unique microRNA microarray signature, predicting factors associated with the clinical course of human chronic lymphocytic leukemia.

In 2006 Andrew Fine of Stanford University and Craig Mello of the University of Massachusetts Medical School shared the Nobel Prize in Physiology or Medicine for their work in RNA interference gene silencing by double-stranded RNA.

The Human Genome Project

By this first decade of the twenty-first century the simple "bead on a string," "one-gene-one-enzyme" concept of the gene has given way to a far more detailed understanding of genes and gene function. In addition, there has been a fundamental shift in scientific emphasis since 2000 from gene to the genome: the whole complement of DNA of an organism and includes genes as well as intergenic and intronic space [25,30,54].

In February 2001, two independent drafts of the genome sequence were published simultaneously in the journals *Science* [55] and *Nature* [56]. The work highlighted in *Science* had been carried out by Celera, Rockville, Maryland, a company founded by Craig Venter; that in *Nature* was work by the International Human Genome Sequencing Consortium. The Human Genome Project, the culmination of decades of discussion, had officially begun in 1990 by the US National Institutes of Health and US Department of Energy. Completing the sequencing project and determining the location of the protein encoding genes opened a new "era of the genome" in the biologic sciences. Hopes have continued high that the project would provide the tools for a better and more fundamental level of understanding of human genetic diseases, of which there are some 4000 known, as well as providing new insights into complex multifactorial polygenic diseases.

Also in 2001, our team working at the University of Toronto was the first to describe the total number of genes expressed in a single organ system, the cardiovascular system [57]. This work had developed out of our 1990s research project using the expressed sequence tag (EST) strategy to identify genes in human heart and artery. We sequenced more than 57,000 ESTs from 13 different cardiovascular tissue cDNA libraries and in 1997 published a comprehensive analysis of cardiovascular gene expression, the largest existing database for a single human organ [58].

Even when the first draft of the genome was published in 2001 – still incomplete and with many gaps – researchers were surprised at the small number of genes in the human genome: approximately 30,000–40,000 genes and far fewer than the original (and often quoted) 100,000 genes that had been informally calculated by Walter Gilbert in the 1980s [59]. When in 2004 the almost completed final sequence of the genome appeared in *Nature*, our species' total gene count was further reduced to 20,000–25,000 [40]. Furthermore, when compared with the genomes of other organisms, humans seem to have surprisingly few genes: only about twice as many genes as fruit flies; and only half as many genes as the corn plant (Table 1.2).

Table 1.2 Genes in the genome.

Organism	*Number of genes*
Maize (*Zea mays*)	50,000
Mustard (*Arabidopsis thaliana*)	26,000
Human (*Homo sapiens*)	20,000–25,000
Nematode worm *(Caenorhabditis elegans)*	19,000
Fruit fly (*Drosophila melanogaster)*	14,000
Baker's yeast (*Saccharomyces cerevisiae*)	6000
Bacterium (*Escherichia coli*)	3000
Human immunodeficiency virus	9

Adapted from Functional and Comparative Genomics Factsheet. Human Genome Project. http://www.ornl.gov/sci/techresources/Human_Genome/faq/compgen.shtml#compgen and Human Genome Program, US Department of Energy, Genomics and Its Impact on Science and Society: A Primer, 2003. http://www.ornl.gov/sci/techresources/Human_Genome/publicat/primer2001/index.shtml

Certainly a challenge for the Human Genome Project and a major challenge in the transition from structural to functional genomics was to identify the entire set of human genes in the genome. About 98% of the DNA in the genome does not code for any known functional gene product and only 2% encodes protein producing genes. In 1991, Mark Adams and J. Craig Venter and colleagues at the National Institutes of Health [60] had proposed the EST approach to gene identification. In this approach individual clones are randomly selected from cDNA libraries representing the genes expressed in a cell type, tissue or organ of interest. Selected clones are amplified and sequenced in a single pass from one or both ends, yielding partial gene sequences known as ESTs. These are then compared with gene sequences in existing nucleotide databases to determine whether they match known genes, or whether they represent uncharacterized genes.

Venter and his colleagues used automated fluorescent DNA sequencing technology to increase the efficiency and scale of EST generation; they were able to rapidly generate ESTs representing over 600 cDNA clones randomly selected from a human brain cDNA library [60]. More than half of these were human genes that had previously been unknown. Venter argued that this strategy could lead to the identification and tagging of 80–90% of human genes in a short period of time and at dramatically less cost than complete genome sequencing, a full decade before the proposed date of completion of the human genomic nucleotide sequence [61]. At about the same time at our laboratory at the University of Toronto we launched the first human heart EST project as we began our catalog of the complete set of genes expressed in the cardiovascular system [58,62,63].

The EST approach ultimately overcame skepticism [64,65] and became recognized as an important and powerful strategy complementing complete genome sequencing. It has been found that ESTs are an efficient vehicle for new gene discovery; ESTs provide information on gene expression levels in different cells/tissues and EST sequences can be used to design PCR primers for physical mapping of the genome. ESTs may also be useful in the search for new genes involved in genetic disease. Chromosomal localization of ESTs increases the ability to identify novel disease genes. Such positional candidate strategies were used, for example, to identify novel candidates for a familial Alzheimer's disease gene [66].

Early EST-based strategies for gene expression investigation were expensive and labor-intensive. Another important technology to emerge from the Human Genome Project, microarray technology enables data similar to EST data to be produced for thousands of genes, simultaneously, in a single experiment. Indeed, while ESTs have been useful for monitoring gene expression in different tissues or cells, their primary utility is now to provide materials for cDNA microarrays [67]. By tagging and identifying thousands of genes, EST repositories presently serve as the primary source of cDNA clones for microarrays.

The two types of microarray systems in widespread use are the photolithographic synthesis of oligodeoxynucleotides directly on to silicon chips and an X-Y-Z robotic system, which spots DNA onto coated standard glass microscope slides or nylon membranes [67–70]. Microarray will be discussed more fully in Chapter 2. DNA microarray technology can profile and compare thousands of genes between mRNA populations simultaneously.

The DNA microarray is also a novel tool to pinpoint differences in expression between single genes on a large scale. A series of transcript profiling experiments can be analyzed to determine relationships between genes or samples in multiple dimensions. A set of expression fingerprints, or profiles, similarities and differences in gene expression are used in order to group different mRNA populations or genes into discrete related sets or clusters. Clusters of co-regulated genes often belong to the same biologic pathways, or the same protein complex, whereas the clusters of mRNA populations are defined by their "expression fingerprint" providing a means to define differences between samples. Thus, the microarray is a powerful technique.

For example, a molecular profile of cancer has been a subject for cDNA microarray analysis. Perou *et al.* [71] compared transcript profiles between cultured human mammary epithelial cells subjected to a variety of growth factors or cytokines and primary breast tumors. Interestingly, a correlation between two subsets of genes with similar expression patterns *in vitro* and in the primary tumors was found, suggesting that these genes could be used for tumor classification. Other transcriptomal cancer studies have also yielded findings, such as new candidate genes that may now be further investigated in population based studies [72–74].

Microarrays are also increasingly being used to investigate gene expression in heart failure – a condition that has complex etiologies and secondary adaptations that make it difficult to study at the level of cellular and molecular mechanisms [75]. A few cardiovascular-based microarray studies have been published. For example, Friddle *et al.* [76] used microarray technology to identify gene expression patterns altered during induction and regression of cardiac hypertrophy induced by administration of angiotensin II and isoproterenol in a mouse model. The group identified 55 genes during induction or regression of cardiac hypertrophy. They confirmed 25 genes or pathways previously shown to be altered by hypertrophy and further identified 30

genes whose expression had not previously been associated with cardiac hypertrophy or regression. Among the 55 genes, 32 genes were altered only during induction, and eight were altered only during regression. This study used a genome-wide approach to show that a set of known and novel genes was involved in cardiac remodeling during regression and that these genes were distinct from those expressed during induction of hypertrophy.

In the first reported human microarray study in end stage heart failure, Yang *et al.* [77] used high density oligonucleotide arrays to investigate failing and nonfailing human hearts (end stage ischemic and dilated cardiomyopathy). Similar changes were identified in 12 genes in both types of heart failure, which, the authors maintain, indicate that these changes may be intrinsic to heart failure. They found altered expression in cytoskeletal and myofibrillar genes, in genes involved in degradation and disassembly of myocardial proteins, in metabolism, in protein synthesis and genes encoding stress proteins.

Our "CardioChip" microarray, an in-house 10,848-element human cardiovascular-based expressed sequence tag glass slide cDNA microarray, has also proved highly useful in helping elucidate molecular and genetic events leading to end stage heart failure. Our group used the CardioChip to explore expression analysis in heart failure [78,79]. We compared left ventricle heart transplant tissue with nonfailing heart controls and found some 100 transcripts that were consistently differentially expressed in dilated cardiomyopathy samples by more than one and a half times.

Microarrays have revolutionized our approach to studying the molecular aspects of disease. The whole genome scan opens a window through which we can monitor molecular pathways of interest and determine how gene expression changes in response to various stimuli (such as drug therapy). These comparisons offer the ability to study disease as it evolves over different time points and to compare patients with different epigenetic risk factor profiles and under different environmental influences. By examining tissue biopsies or cell samples, researchers can identify a whole-genome "portrait" of gene expression, extract candidate genes and conduct targeted follow-up studies that directly relate to specific cellular functions. Current microarray studies typically utilize tissue samples, and of necessity rely on tissue biopsy. In many cases, however, such as in the cardiovascular studies above, tissue samples can only be obtained in very late stage disease, at transplant or after death. The need for a simple noninvasive cost-effective method to replace tissue biopsy to identify early stage disease is clear. Hence, research interest has begun to turn to investigating the use of blood based gene expression profiling. Blood samples have a number of advantages over tissue samples, in particular that blood can be obtained early during disease development and causes little discomfort to patients.

There is a growing body of evidence that the blood contains substantial bioinformation and that biomarkers derived from blood RNA may provide an alternative to tissue biopsy for the diagnosis and prognosis of disease [80]. Recent studies have shown that blood cell gene expression profiles reflect individual characteristics [81,82], and alterations in blood cell transcriptomes have been found to characterize a wide range of diseases and disorders occurring in different tissues and organs, including juvenile arthritis [83], hypertension [84–86], colorectal cancer [87], chronic fatigue syndrome [88] and neuronal injuries [89,90]. Circulating blood cells also show distinctive expression patterns under various environmental pressures and stimuli, such as exercise [91], hexachlorobenzene exposure [92], arsenic exposure [93] and smoking [94].

Such research findings provide convincing support to the hypothesis that circulating blood cells act as a "sentinels" which detect and respond to microenvironmental changes in the body. Our laboratory, Gene News Corp., in Toronto has developed a methodology to establish the *Sentinel Principle*™. We have profiled gene expression from peripheral blood and we have identified mRNA biomarkers for different diseases. In an initial study, blood samples were drawn from patients with coronary artery disease and gene expression compared with healthy control samples [95]. Differentially expressed genes identified in the circulating blood successfully discriminated the coronary patients from healthy control subjects [95]. We have also used the principle to discriminate successfully between patients with schizophrenia and those with bipolar disorder and between patients and controls [96], which findings have been verified in later studies

[97]. Our group has also identified biomarkers in blood that have utility in identification of early osteoarthritis [98] and bladder cancer [99].

The new technologies of the Human Genome Project allow us to view the entire genome of an organism and permit better characterization of disease as a dynamic process. Although at an early stage as yet, the possibility of using blood samples as the basis for microarray studies of biology and disease opens up new vistas of research for the future.

Conclusions

The twentieth century opened with the start of the search for the gene. The concept grew in stature and importance with the double helix and the central dogma. However, research since 1960 has led to changes in traditional ideas about the gene. No longer is the gene the autonomous self-replicating unit of inheritance of 1953; rather it requires the assistance of a host of accessory regulatory proteins [25]. Indeed, when in 1986 Walter Gilbert proposed the "RNA world hypothesis": that RNA, which can self-replicate, must be the primary molecule in evolution, the traditional gene even lost its ascendancy over other molecules as "the basis of life" [100].

Since 2001, the date of the first draft of the Human Genome Project, and since the release of the genome sequencing projects of other organisms, floods of new genome data have been generated and novel technologies have been developed to attempt to make sense of that data. High throughput microarray technology has provided a "new kind of microscope" [101] for post genomic analysis. It is now possible to look at thousands of base pair sequences simultaneously. The one gene at a time paradigm has been replaced, or at least supplemented, with a more holistic model of the gene in its surrounding molecular landscape.

For example, the central dogma presupposes a correspondence between genes and complexity and one of the big surprises of the Human Genome Project has been the scarcity of genes in the genome. The human genome contains in fact very few protein coding genes and fewer than many "simpler" organisms, a mere one-quarter to one-fifth of the original estimates [40]. To begin to explain the paradoxical genome data, researchers have had to shift their emphasis away from genes and proteins and towards gene regulation. Why do humans have so few genes [102] has been replaced by the question: How do so few genes create such complexity?

Clearly, it is not genes themselves, per se, that confer complexity. Rather complexity occurs as a result of gene–gene interactions and programs – molecular pathways that modulate development. Alternative splicing is one possible mechanism that might allow the cell to produce numerous proteins from one basic gene, and the mechanisms, pathways and regulators governing alternative splicing and spliceosomes are the subject of intensive research investigation. In addition, the large amount of noncoding DNA in genomes suggests that noncoding DNA may have functional biological activity [103]. In particular, ncRNAs may prove to be the programmers controlling complexity [42].

Science in the post genome era recognizes that gene activity does not occur in isolation. Rather, a full understanding of the development, the disease and decay of organisms will be found when the "genes," including the protein gene, the RNA gene or any other genes that might be discovered, are considered together with gene regulatory factors, gene–gene interactions, gene–cell interactions, epigenetic factors and signaling pathways in gene expression. Understanding signaling pathways in gene expression is a major research focus.

Gene function is beginning to be understood in different ways, with different ways to pose the problems. For example, rarely today do we speak of a gene as causing a particular disease or giving rise to a specific trait; diseases, even the so-called single gene diseases, and traits are, rather, understood to be the results of hundreds and even thousands of genes operating in complex regulatory networks. This is especially true in cardiology, where such complex multifactorial diseases as coronary artery disease, heart failure, hypertension and atherosclerosis are caused by genetic factors together with a host of environmental and other factors. Even in the case of the "single" gene diseases, such as hypertrophic cardiomyopathy, dilated cardiomyopathy and other disorders considered to be the result of mutations of a single gene, it is becoming increasingly clear that such disorders are actually far more complex than previously thought [104–107].

Already with microarray and other novel technologies, holistic approaches to investigating the health and disease of organisms are becoming possible. As Evelyn Fox Keller put it, the twenty-first century will be "the century of the genome" [25].

A closer look at some genes of importance in cardiology

Cardiac myosin heavy chain genes

A family of genes of major importance in cardiology are the myosin heavy chain genes [108]. Myosin, the contractile protein of muscle, makes up the thick filaments of cardiac and skeletal muscle. Conventional myosin contains two heavy chains (220,000 kDa) which form the helical coiled rod region of the molecule and four light chains (26,000 and 18,000 kDA) which form the pair-shaped head regions. Striated muscle contraction is generated by interaction between myosin and thin filament actin. Upon fibre activation the myosin head binds to actin, which slides a short distance along the thick filament. Linkage is broken by adenosine triphosphate (ATP) hydrolysis whereupon actin and myosin dissociate. By this process the filaments are pulled along each other, rachet-like, in the classic sliding filament motion.

Myosin heavy chain genes are highly conserved and structurally similar [109–111]. Mammalian myocardial genes are large and complex, spanning approximately 24 kb and split into 40–41 exons and approximately the same number of introns, of various sizes [112]. Two isoforms of myosin heavy chain gene are expressed in myocardial cells, α-*MYH* and β-*MYH*, extending over 51 kb on chromosome 14 in humans; α-*MYH* and β-*MYH* are separated intragenically by about 4.5 kb; similar in overall structure, their sequences in the 5′ flanking regions are quite different, suggesting independent regulation of these genes [113].

The α and β cardiac heavy chain genes are tandemly linked, and are arranged in order of their expression during fetal development. The β-*MHC* is located 5′ upstream of the α-MHC sequence and is expressed first during heart development, followed by α-MHC gene expression. Despite the fact that there is almost 93% sequence identity between α-*MYH* and β-*MYH*, their ATPase activity differs by twofold suggesting functional differences.

Table 1.3 Response to stimuli of cardiac myosin heavy chain genes.

	α-MYH	*β*-MYH
+ Thyroid (T3)	Upregulated	Downregulated
– Thyroid (T3)	Downregulated	Upregulated
Exercise	Upregulated	Downregulated
Pressure	Downregulated	Upregulated
Aging	Downregulated	Upregulated

Adapted from Weiss & Leinwald [108].

α-*MYH* and β-*MYH* isoforms are tissue specific and differentially developmentally regulated (reviewed in [114]). Thus, α-*MYH* and β-*MYH* are both expressed at high levels throughout the cells of the developing fetal heart tube at about 7.5–8 days post coitum [115]. As ventricular and atrial chambers begin to form, isoform expression patterns begin to diverge: β-*MYH* begins to be restricted to ventricular myocytes in humans, and α-*MHC* levels diminish in ventricular cells, but continue to be expressed in adult human atrial cells [116]. Cardiac myosin heavy chain gene expression and proportion of α-*MYH* and β-*MYH* expressed is regulated by a number of factors, including thyroid hormone during development, pressure or volume overload, diabetes, catecholamine levels and aging (Table 1.3) [108,114]. Regulatory elements in cardiac myosin heavy chain genes have been studied extensively (reviewed in [108]).

Disease mutations associated with *MYH* genes include, most notably, hypertrophic cardiomyopathy. Hypertrophic cardiomyopathy, a primary disorder of the myocardium and an important cause of heart failure, was first associated with mutations in the β myosin heavy chain gene in 1990 when a missense mutation in R403Q was identified [117]. Subsequently, more than 80 mutations linked with hypertrophic cardiomyopathy have been identified in the β myosin heavy chain gene, and the list continues to grow [118].

In addition to mutations in the β myosin heavy chain gene, researchers have identified hundreds of mutations in at least 10 other genes, all encoding for proteins involved in the cardiac contractile apparatus including α-myosin heavy chain gene, cardiac myosin binding protein C, cardiac troponin T2, C and I, α-tropomyosin, myosin regulatory and

essential light chains, actin and titin [119]. Because all of the genes identified as being causal in primary hypertrophic cardiomyopathy encode for the sarcomeric proteins, primary hypertrophic cardiomyopathy is now widely recognized as a disorder of the sarcomere [105].

Primer of genes and genomics

DNA

The deoxyribonucleic acid (DNA) of a living organism contains all of the genetic information necessary to construct a specific organism and to direct the activity of the organism's cells.

DNA is a very long, twisted, double stranded molecule made up of two chains of nucleotides. Each DNA nucleotide contains one of the four DNA bases: *guanine* (G), *adenine* (A), *thymine* (T) and *cytosine* (C). These bases are arranged side by side (for example, AAGTTAAG) and it is their sequence arrangement that will determine the protein constructed by the gene.

Gene

The basic unit of heredity, a gene is an ordered sequence of DNA nucleotides that can be decoded to produce a gene product. The overwhelming majority of genes of the human genome are protein-coding genes; noncoding genes produce RNA molecules, mainly involved in gene expression.

Gene expression

Gene expression is the complex process by which information in the gene is transcribed into RNA and translated into proteins. Gene expression is carried out in two stages: transcription and translation. During transcription genetic information is transcribed into an mRNA copy of a gene, which must then be translated into a protein.

Although each cell of the human body contains a complete genome and set of 20,000–25,000 genes, only a subset of these genes are expressed or turned on, depending on cell type. Such cell-specific gene expression determines whether a cell will be a brain cell, a heart cell or a liver cell, for example. Some genes that carry out basic cellular functions are expressed all the time in all the cells – they are called housekeeping genes. Others are expressed only under certain conditions, such as when activated by signals such as hormones. Researchers study changes in gene expression to gain understanding as to how cells behave in response to changes in stimuli.

Gene structure

The gene is a structured molecule comprising exons, introns and regulatory sequences. The region of the gene that codes for a gene product (usually a protein) is called the exon; between the exons are sequences of noncoding DNA, called introns. Introns must be edited out of the gene during transcription and before translation of the protein.

Stretches of DNA indicate the beginning and end of genes. Coding begins with the initiation codon or start codon "ATG" and ends with termination or stop codons: TAA, TAG or TGA.

Genome

Genome is a word compound of "gene" and "chromosome." A genome is the complete DNA required to build a living organism, and an organism's genome is contained in each of its cells. Some genomes are small, such as bacterial genomes which may contain less than a million base pairs and some are very large: the human genome comprises about three billion base pairs.

Human Genome Project

The Human Genome Project is an international consortium to sequence all of the three billion base pairs of the human genome. The Human Genome Project formally commenced in 1990, led by the US Department of Energy and the National Institutes of Health. The project was completed in April 2003 with the announcement that the human genome contains some 20,000–25,000 genes.

The benefits of the Human Genome Project are beginning to make themselves felt. As a result of the research project, powerful and novel technologies and resources have been developed which will lead eventually to an understanding of biology at the deepest levels. Major advances in diagnosis and treatment of many diseases, and disease prevention is expected as a result of Human Genome Project efforts.

How many genes in the human genome?

As of October 2004, the latest estimate from the Human Genome Project is that the human gen-

ome contains some 20,000–25,000 protein-coding genes.

Genomics

Genome is a word combining "gene" and "chromosome", and the genome includes the entire set of an organism's protein coding genes and all of the DNA sequences between the genes. Genomics uses the techniques of molecular biology and bioinformatics to study not just the individual genes of an organism but of the whole genome.

Metabolomics

By analogy with genomics and proteomics, metabolomics is the large-scale study of the all the metabolites of an organism. Understanding the metabolome offers an opportunity to understand genotype–phenotype and genotype–environment interactions.

Microarray

Microarray is an enabling technology that allows researchers to compare gene portraits of tissue samples at a snapshot in time. A microarray is a slide or membrane to which is attached an orderly array of DNA sequences of known genes. The researcher pipettes samples of mRNA onto the slide, containing unknown transcripts obtained from a tissue of interest. mRNA has the property that it is complementary to the DNA template of origin. Thus, mRNA binds or hybridizes to the slide DNA and can be calculated by computer to provide a portrait or snapshot of which genes are active in the sample.

By monitoring and comparing thousands of genes at a time – instead of one by one – a microarray gene chip data can be used to see which genes in a tissue are turned on or expressed and which are turned off.

Microarray gene expression profiling

Understanding gene function is crucially important to understanding health and disease. Most of the common and serious diseases afflicting humans are polygenic: that is, it takes hundreds if not thousands of genes interacting with each other and with the environment to cause such diseases as cancer and heart disease. By monitoring and comparing thousands of genes at a time – instead of one by one – microarray gene expression profiling can be used to determine which genes in a tissue are turned on and which are turned off – and how actively the genes are producing proteins.

Such gene "portraits" can identify patients with early stage diseases as compared to no disease or late stage disease, to distinguishing patients with different diseases, or patients with different stages of disease for disease prognosis, drug effect monitoring and other clinical applications. As microarray technology advances researchers will be able to ask increasingly probing and important biologic questions.

Mutation

A mutation is a change in the DNA sequence of a gene. If the mutation is significant, then the protein produced by the gene will be defective in some way and unable to function properly. Not all mutations are harmful; some may be beneficial and some may have no discernible effect.

There are different types of mutations: base substitution, in which a single base is replaced by another: deletion, in which base(s) are left out; or insertion in which base(s) are added.

Mutations can be caused by radiation, chemicals or may occur during the process of DNA replication. Some mutations can be passed on through generations.

Protein

A protein is a large molecular chain of amino acids. Proteins are the cell's main structural building blocks and proteins are involved in all cellular functions. Information in the gene encodes for the protein and most of the genes of living organisms produce proteins. Humans are calculated to have about 400,000 proteins, far more than our 20,000 or so genes.

Proteomics

An understanding of cellular biology depends fundamentally on understanding protein structure and behaviour. Proteomics is the large-scale comprehensive study of the proteome, the complement of all of the proteins expressed in a cell, a tissue or an organism. Proteomics uses technology similar to genomics technologies, such as protein microarrays, to explore the structure and function of proteins and protein behaviour in response to changing environmental signals.

RNA

The relationship between a gene and its protein is not straightforward. DNA does not construct proteins directly; rather, genes set in motion intermediate processes that result in amino acid chains. The main molecule involved in this process is called ribonucleic acid (RNA). RNA nucleotides contain bases: adenine (A) uracil (U) guanine (G) and cytosine (C). Thus, RNA is chemically very similar to DNA, except that RNA has a uracil base rather than thymine.

The process of producing a protein from DNA template begins in the cell nucleus via the intermediary messenger (m) RNA. mRNA copies the relevant piece of DNA in a process called transcription. The short, single-stranded mRNA transcript is then transported out of the cell nucleus by transfer RNA and into the cytoplasm where it is translated into a protein by the ribosome. (Ribosomal RNA (rRNA), is involved in constructing the ribosomes.) Since the 1990s many new noncoding RNA genes have been discovered, such as microRNA.

Single nucleotide polymorphism

A single nucleotide polymorphism (SNP) is a base alteration in a single nucleotide in the genome. Unlike mutations, which are rare, single nucleotide polymorphisms are common alterations in populations, occurring in at least 1% of the population. SNPs make up 90% of all human genetic variation and occur every 100–300 human genome bases. In time researchers hope to be able to develop SNP patterns that can be used to test individuals for disease susceptibility or drug response.

Further information

National Center for Biotechnology Information. A Science Primer. http://www.ncbi.nlm.nih.gov/About/primer/genetics_molecular.html

National Institutes of Health. NIGM. Genetics Basics http://publications.nigms.nih.gov/genetics/science.html

Welcome Trust. Gene Structure. http://genome.wellcome.ac.uk/doc_WTD020755.html

Human Genome Project. http://www.google.ca/search?q=%22%22+what+is+a+gene%22+&hl=en&lr=&c2coff=1&start=30&sa=N

Microarrays http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html

Introduction to proteomics: http://www.childrenshospital.org/cfapps/research/data_admin/Site602/mainpageS602P0.html

Bio-pro. Proteomics http://www.bio-pro.de/en/life/thema/01950/index.html

The human metabolome project http://www.metabolomics.ca/

References

1 Mendel G. Experiments in Plant Hybridization (1865) Read at the meetings of the Brünn Natural History Society, February 8 and March 8, 1865. (Available online at www.mendelweb.org)

2 Correns C. Mendel's law concerning the behavior of progeny of varietal hybrids. (Trans: Piernick LK) Electronic Scholarly Publications. 2000. http://www.esp.org/foundations/genetics/classical/holdings/c/cc-00.pdf (Originally: Mendel's Regel über das Verhalten der Nachkommenschaft der Rassenbastarde. *Ber Dtsch Botanisch Gesellschadt* 1900; 18: 158–168.)

3 De Vries H. Concerning the law of segregation of hybrids. (Trans: Hannah A.) Electronic Scholarly Publications. 2000 (http://www.esp.org/foundations/genetics/classical/holdings/v/hdv-00.pdf) (originally: Das Spaltungsgesetz der Bastarde. *Ber Dtsch Botanisch Gesellschaft* 1900; 18: 83–90.)

4 Tschermak E. Concerning artificial crossing in Pisum sativum. *Genetics* 1950: 35: 42–47. (Originally: Über Künstliche Kreuzung bei Pisum sativum. *Ber Dtsch Botanisch Gesellschaft* 1900; 18: 232–239.)

5 De Vries H. Intracellular pangenesis. Including a paper on Fertilization and Hybridization. (Trans: Gager CS.) Open Court Publishing, Chicago, 1910. http://www.esp.org/books/devries/pangenesis/facsimile/title3.html

6 Bateson W. The progress of genetic research. In: *Report of the Third International Conference on Genetics 1906*. Royal Horticultural Society. London. 1907; 90–97.

7 Johannsen W. *Elemente der Exakten Erblichkeitslehre*. Gustav Fischer, Jena, 1909.

8 Carlson EA. *Mendel's legacy: the origin of classical genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2004.

9 Sturtevant AH. *A History of Genetics*. Harper and Row, New York, 1965.

10 Sturtevant AH. The linear arrangement of six sex linked traits in drosophila, as shown by their mode of association. *J Exp Zool* 1913; 14: 43–59.

11 Bridges CE. Salivary chromosome maps: with a key to the banding of the chromosomes of *Drosophila melanogaster*. *J Hered* 1935; 26: 60–64.

12 Muller HJ. Artificial transmutation of the gene. *Science* 1927; 66: 84–87.

13 Muller HJ. The gene as the basis of life. *Proc Int Cong Plant Science* 1929; 1: 897–921.
14 Avery OT, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumoccocal types. *J Exp Med* 1944; 79: 137–158.
15 Hershey AD, Chase M. Independent functions of viral proteins and nucleic acid in growth of bacteriophage. *J Gen Physiol* 1952: 36: 39–56.
16 Watson JD, Crick FHC. Molecular structure of nucleic acids. *Nature* 1953; 171: 737–738.
17 Olby RC. *The Path to the Double Helix*. Macmillan, London, 1974.
18 Garrod AE. The incidence of alkaptonuria: A study in chemical individuality. *Lancet* 1902; ii: 1616–1620.
19 Garrod AE. *Inborn Errors of Metabolism*, 2nd edn. Henry Frowde and Hodder & Stoughton, London, 1923.
20 Beadle G, Tatum E. Genetic control of biochemical reactions in Neurospora. *Proc Natl Acad Sci USA* 1941; 27: 499–506.
21 Neel JV. The inheritance of sickle cell anemia. *Science* 1949; 110: 64–66.
22 Pauling L, Itano H, Singer SJ, Wells I. Sickle cell anemia, a molecular disease. *Science* 1949; 110: 543–548.
23 Ingram VM. Gene mutations in human haemoglobin: the chemical difference between normal and sickle-cell haemoglobin. *Nature* 1957; 180: 326–328.
24 Crick FHC. On protein synthesis. *Symp Soc Exp Biol* 1958; XII: 139–163.
25 Keller EF. *The Century of the Gene*. Harvard University Press, Cambridge, 2000.
26 Morange M. Century of the gene. Isuma. *Can J Policy Res* 2001; 2: 22–27.
27 Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 1961; 3: 318–356.
28 Morange M. What history tells us. The operon model and its legacy. *J Biosci* 2005; 30: 313–316.
29 Morange M. *The Misunderstood Gene*. Harvard University Press, Cambridge, 2001.
30 Portin P. The concept of the gene: Short history and present status. *Q Rev Biol* 1993; 68: 173–223.
31 Maas WK. *Gene Action*. Oxford University Press, Oxford, 2001.
32 Temin HM, Mizutani S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 1970: 226: 1211–1213.
33 Baltimore D. RNA dependent DNA polymerase in virions of RNA tumor viruses. *Nature* 1970; 226: 1209–1211.
34 Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 1977; 12: 1–8.
35 Berget SM, Moore C, Sharp PA. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* 1977; 74: 3171–3175.
36 Gilbert W. Why genes in pieces? *Nature* 1978; 271: 501.
37 Lopez AJ. Alternative splicing of pre-mRNA: Developmental consequences and mechanisms of regulation. *Annu Rev Genet* 1998; 32: 279–305.
38 McClintock B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA* 1950; 36: 344–355.
39 Britten RJ, Kohne DE. Repeated sequences in DNA. *Science* 1968; 161: 529–540.
40 International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431: 931–945.
41 Gibbs WW. The unseen genome: gems among the junk. *Sci Am* 2003; 289: 46–53.
42 Mattick JS. The hidden genetic program of complex organisms. *Sci Am* 2004; 4: 60–67.
43 Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2001; 2: 919–929.
44 Storz G, Altuvia S, Wassarman KM. An abundance of RNA regulators. *Annu Rev Biochem* 2005; 74: 199–217.
45 Eddy SR. Computational genomics of noncoding RNA genes. *Cell* 2002; 109: 137–40.
46 Mattick JS. The functional genomics of noncoding RNA. *Science* 2005; 309: 1527–1528.
47 Noller HF. RNA structure: reading the ribosome. *Science* 2005; 309: 1508–1514.
48 Zamore PD, Haley B. Ribo-gnome: the big world of small RNAs. *Science* 2005; 309: 1519–1524.
49 Ambros V. The functions of animal microRNAs. *Nature* 2004; 431: 350–355.
50 Bentwich I, Avniel A, Karov Y *et al.* Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 2005; 37: 766–770.
51 Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; 120: 15–20.
52 Chen CZ. MicroRNAs as oncogenes and tumor suppressors. *N Engl J Med* 2005; 353: 1768–1771.
53 Calin GA, Ferracin M, Cimmino A *et al.* A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* 2005; 353: 1793–1801.
54 Falk R. Long live the genome! So should the gene. *Hist Philos Life Sci* 2004; 26: 105–121.
55 Venter JC, Adams MD, Myers EW *et al.* The sequence of the human genome. *Science* 2001; 291: 1304–1351.
56 International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921.
57 Dempsey AA, Dzau VJ, Liew CC. Cardiovascular genomics: Estimating the total number of genes expressed in

the human cardiovascular system. *J Mol Cell Cardiol* 2001; 33: 1879–1886.

58 Hwang DM, Dempsey AA, Wang RX *et al.* A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. *Circulation* 1997; 96: 4146–4203.

59 Pennisi E. The human genome. *Science* 2001; 291: 1177–1180.

60 Adams MD, Kelley JM, Gocayne JD *et al.* Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991; 252: 1651–1656.

61 Roberts L. Gambling on a shortcut to genome sequencing. *Science* 1991; 252: 1618–1619.

62 Liew CC. A human heart cDNA library: the development of an efficient and simple method for automated DNA sequencing. *J Mol Cell Cardiol* 1993; 25: 891–894.

63 Liew CC, Hwang DM, Fung YW *et al.* A catalogue of genes in the cardiovascular system as identified by expressed sequence tags (ESTs). *Proc Natl Acad Sci USA* 1994: 91: 10645–10649.

64 Roberts L. Genome patent fight erupts. *Science* 1991: 254: 184–186.

65 Marshall E. The company that genome researchers love to hate. *Science* 1994; 266: 1800–1802.

66 Levy-Lahad E, Wasco W, Poorkaj P, *et al.* Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* 1995; 269: 973–977.

67 Schena M, Shalon D, Davis RW *et al.* Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270: 467–470.

68 Lockhart DJ, Dong H, Byrne MC *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996; 14: 1675–1680.

69 Bowtell DDL. Options available – from start to finish – for obtaining expression data by microarray. *Nat Genet* 1999; Supplement 21: 25–32.

70 Lipshutz RJ, Fodor SPA, Gingeras TR *et al.* High density synthetic oligonucleotide arrays. *Nature Genetics.* 1999; Supplement 21: 20–24.

71 Perou CM, Jeffrey SS, van de Rijn M *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci USA* 1999; 96: 9212–9217.

72 Rhodes DR, Chinnaiyan AM. Integrative analysis of the cancer transcriptome. *Nat Genet* 2005; 37 Supplement: S31–S37.

73 Segal E, Friedman N, Kaminski N *et al.* From signatures to models: understanding cancer using microarrays. *Nat Genet* 2005; 37 Supplement: S38–S45.

74 Mohr S, Leikauf GD, Keith G, Rihn BH. Microarrays as cancer keys: an array of possibilities. *J Clin Oncol* 2002; 20: 3165–3175.

75 Liew CC, Dzau VJ. Molecular genetics and genomics of heart failure. *Nat Rev Genet* 2004; 5: 811–825.

76 Friddle CL, Koga T, Rubin EM, Bristo J. Expression profiling reveals distinct sets of genes altered during induction and regression of cardiac hypertrophy *Proc Natl Acad Sci USA* 2000; 97: 6745–6750.

77 Yang J, Moravec CS, Sussman MA. Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. *Circulation* 2000; 102: 3046–3052.

78 Barrans JD, Stamatiou D, Liew CC. Construction of a human cardiovascular cDNA microarray: portrait of a failing heart. *Biochem Biophys Res Commun* 2001; 280: 964–969.

79 Barrans JD, Allen PD, Stamatiou D *et al.* Global gene expression profiling of end stage dilated cardiomyopathy using a human cardiovascular based cDNA microarray. *Am J Pathol* 2002; 160: 2035–2043.

80 Liew CC. Expressed genome molecular signatures of heart failure. *Clin Chem Lab Med* 2005; 43: 462–469.

81 Whitney AR, Diehn M, Popper SJ *et al.* Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci USA* 2003; 100, 1896–1901.

82 Radich JP, Mao M, Stepaniants S *et al.* Individual-specific variation of gene expression in peripheral blood leukocytes. *Genomics* 2004; 83: 980–988.

83 Barnes MG, Aronow BJ, Luyrink LK *et al.* Gene expression in juvenile arthritis and spondyloarthropathy: pro-angiogenic ELR+ chemokine genes relate to course of arthritis. *Rheumatol (Oxf)* 2004; 43: 973–979.

84 Okuda T, Sumiya T, Mizutani K *et al.* Analyses of differential gene expression in genetic hypertensive rats by microarray. *Hypertens Res* 2002; 25: 249–255.

85 Chon H, Gaillard CA, van der Meijden BB *et al.* Broadly altered gene expression in blood leukocytes in essential hypertension is absent during treatment. *Hypertension* 2004; 43: 947–951.

86 Bull TM, Coldren CD, Moore M *et al.* Gene microarray analysis of peripheral blood cells in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2004; 170: 827–828.

87 DePrimo SE, Wong LM, Khatry DB *et al.* Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. *BMC Cancer* 2003; 3: 3.

88 Whistler T, Unger ER, Nisenbaum R, Vernon SD. Integration of gene expression, clinical, and epidemiologic data to characterize chronic fatigue syndrome. *J Transplant Med* 2003; 1: 10.

89 Tang Y, Lu A, Aronow BJ, Sharp FR. Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: blood genomic fingerprints of disease. *Ann Neurol* 2001; 50: 699–707.

90 Tang Y, Nee AC, Lu A *et al.* Blood genomic expression profile for neuronal injury. *J Cereb Blood Flow Metab* 2003; 23: 310–319.

91 Connolly PH, Caiozzo VJ, Zaldivar F *et al.* Effects of exercise on gene expression in human peripheral blood mononuclear cells. *J Appl Physiol* 2004; 97: 1461–1469.

92 Ezendam J, Staedtler F, Pennings J *et al.* Toxicogenomics of subchronic hexachlorobenzene exposure in Brown Norway rats. *Environ Health Perspect* 2004; 112: 782–791.

93 Wu MM, Chiou HY, Ho IC *et al.* Gene expression of inflammatory molecules in circulating lymphocytes from arsenic-exposed human subjects. *Environ Health Perspect* 2003; 111: 1429–1438.

94 Ryder MI, Hyun W, Loomer P, Haqq C. Alteration of gene expression profiles of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. *Oral Microbiol Immunol* 2004; 19: 39–49.

95 Ma J, Liew CC. Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. *J Mol Cell Cardiol* 2003; 35: 993–998.

96 Tsuang MT, Nossova N, Yager T *et al.* Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet* 2005; 133: 1–5.

97 Glatt SJ, Everall IP, Kremen WS *et al.* Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 upregulation in schizophrenia. *Proc Natl Acad Sci USA* 2005; 102: 15533–15538.

98 Marshall KW, Zhang H, Yager T *et al.* Blood-based biomarkers for detecting mild osteoarthritis in the human knee. *Osteoarthritis Cartilage* 2005; 13: 861–871.

99 Osman I, Bajorin D, Sun TT *et al.* Novel blood biomarkers of human urinary bladder cancer. *Clin Cancer Res* 2006; 12: 3374–3380.

100 Gilbert, W. Origin of life: The RNA world. *Nature* 1986; 319: 618.

101 Brown PO. Website (http://biochemistry.stanford.edu/research/brown.html)

102 Pennisi E. Why do humans have so few genes? *Science* 2005; 309: 80.

103 Ruddle F. Hundred-year search for the human genome. *Annu Rev Genomics Hum Genet* 2001; 2: 1–8.

104 Hughes SE. The pathology of hypertrophic cardiomyopathy. *Histopathology* 2004; 44: 412–427.

105 Seidman JG, Seidman C. The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. *Cell* 2001; 104: 557–567.

106 Towbin JA, Bowles NE. The failing heart. *Nature* 2002; 415: 227–233.

107 Bonne G, Carrier L, Richard P. Familial hypertrophic cardiomyopathy: from mutations to functional defects. *Circ Res* 1998; 83: 580–593.

108 Weiss A, Leinwald LA. The mammalian myosin heavy chain gene family. *Annu Rev Cell Dev Biol* 1996; 12: 417–439.

109 Liew CC, Jandreski MA. Construction and characterization of the α form of a cardiac myosin heavy chain cDNA clone and its developmental expression in the Syrian hamster. *Proc Natl Acad Sci USA* 1986; 83: 3175–3179.

110 Jandreski MA, Liew CC. Construction of a human ventricular cDNA library and characterization of a β-myosin heavy chain cDNA clone. *Hum Genet* 1987; 76: 47–53.

111 Jandreski MA, Sole MJ, Liew CC. Two different forms of β-myosin heavy chain are expressed in human striated muscle. *Hum Genet* 1987; 77: 127–131.

112 Strehler EE, Strehler-Page MA, Perriard JC *et al.* Complete nucleotide and encoded amino acid sequence of a mammalian myosin heavy chain gene. Evidence against intron dependent evolution of the rod. *J Mol Biol* 1986; 190: 291–317.

113 Yamauchi-Takihara K, Sole MJ, Liew J *et al.* Characterization of human cardiac myosin heavy chain genes. *Proc Natl Acad Sci USA* 1989: 86: 3504–3508.

114 Morkin E. Control of cardiac myosin heavy chain gene expression. *Microsc Res Tech* 2000; 50: 522–531.

115 Lyons GE, Ontell M, Cox R *et al.* The expression of myosin genes in developing skeletal muscle in the mouse embryo. *J Cell Biol* 1990; 111: 1465–1476.

116 Lompre AM, Nadal-Ginard B, Mahdavi V. Expression of the cardiac ventricular α- and β-myosin heavy chain genes is developmentally and hormonally regulated. *J Biol Chem* 1984; 259: 6437–6446.

117 Geisterfer-Lowrence AA, Kass S, Tanigawa G *et al.* A molecular basis for familial hypertrophic cardiomyopathy a β cardiac myosin heavy chain gene mis-sense mutation. *Cell* 1990; 62: 999–1006.

118 Seidman C. For an updated list go to: Sarcomere Protein Gene Mutation Database. (http://genetics.med.harvard.edu/~seidman/cg3/)

119 Ahmad F, Seidman JG, Seidman C. The genetic basis for cardiac remodeling. *Annu Rev Genomics Hum Genet* 2005; 6: 185–216.