Published in final edited form as:
Nat Rev Genet. 2010 October ; 11(10): 723–733. doi:10.1038/nrg2878.

Ten years of genetics and genomics: what have we achieved and where are we heading?

Edith Heard,
Mammalian Developmental Epigenetics Group, Unité de Génétique et Biologie du Développement, INSERM U934/CNRS UMR3215, Institut Curie – Centre de Recherche, 26, rue d'Ulm, 75248 Paris Cedex 05, France. Edith.Heard@curie.fr

Sarah Tishkoff,
Departments of Genetics and Biology, University of Pennsylvania School of Medicine, 428 Clinical Research Building, 415 Curie Boulevard, Philadelphia, Pennsylvania 19104-6145, USA. tishkoff@mail.med.upenn.edu

John A. Todd,
Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0XY, UK. john.todd@cimr.cam.ac.uk

Marc Vidal,
Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute; and the Department of Genetics, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. marc_vidal@dfci.harvard.edu

Günter P. Wagner,
Yale Systems Biology Institute, Yale University, POB 208106, New Haven, Connecticut 06520-8106, USA. gunter.wagner@yale.edu

Jun Wang,
BGI-Shenzhen, Shenzhen 518083, China; and the Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark. wangj@genomics.org.cn

Detlef Weigel, and
Department of Molecular Biology, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany. weigel@weigelworld.org

Richard Young

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Edith Heard's homepage: http://www.curie.fr/recherche/themes/detail_equipe.cfm/lang/_gb/id_equipe/63.htm
Sarah Tishkoff's homepage: http://www.med.upenn.edu/tishkoff/Lab/Tishkoff/Tishkoff.html
John A. Todd's homepage: http://www.cimr.cam.ac.uk/investigators/todd/index.html
Marc Vidal's homepage: http://ccsb.dfci.harvard.edu
Günter P. Wagner's homepage: http://pantheon.yale.edu/~gpwagner/index.html
Detlef Weigel's homepage: http://www.weigelworld.org
Richard Young's homepage: http://web.wi.mit.edu/young
1000 Genomes Project: http://www.1000genomes.org
The BGI homepage: http://www.genomics.cn
All LINKS ARE ACTIVE IN THE ONLINE PDF

Whitehead Institute for Biomedical Research, MIT, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. young@wi.mit.edu

Abstract

To celebrate the first 10 years of *Nature Reviews Genetics*, we asked eight leading researchers for their views on the key developments in genetics and genomics in the past decade and the prospects for the future. Their responses highlight the incredible changes that the field has seen, from the explosion of genomic data and the many possibilities it has opened up to the ability to reprogramme adult cells to pluripotency. The way ahead looks similarly exciting as we address questions such as how cells function as systems and how complex interactions among genetics, epigenetics and the environment combine to shape phenotypes.

What advances in the past decade have particularly excited and surprised you?

Edith Heard

It is amazing to see how epigenetics has been propelled into the headlines over the past decade. On the one hand, it has been hailed as an explanation for inter- and intra-individual diversity, and on the other, as a purveyor of hidden information — beyond genes — that can be influenced by intrinsic and extrinsic fluctuations. The ‘hype’ surrounding epigenetics may partly be due to the fact that, since the human genome was sequenced 10 years ago, we have been confronted with the reality, and perhaps inevitability, of our genetic constitution. Epigenetics may provide hope that we are more than just the sequence of our genes — and that our destiny and that of our children can be shaped, to some extent, by our lifestyle and environment. The recent groundbreaking discoveries on induced pluripotency have also brought the reversible nature of epigenetic states to the forefront. Such reversibility brings much hope for treating diseases such as cancer, which have not just a genetic but also an epigenetic basis, for which ‘epidrugs’ can be used to reverse aberrant epigenetic changes (epimutations).

What has caused this excitement? Major advances in our understanding of epigenetic mechanisms have been made in the past decade. Our grasp of the molecular basis of epigenetics has expanded well beyond DNA methylation, which was previously the best known epigenetic mark. It now includes other chromatin components, such as histone variants and histone modifications, as well as associated protein complexes.

The nature and mechanism of action of complexes such as Polycomb and Trithorax — which can modify histones, alter chromatin structure and shape the epigenome, as well as ensuring heritability of gene expression states — are starting to be unravelled[1]. One surprising advance has been the finding that chromatin states can be incredibly labile. For example, histone methylation was initially believed to be a very stable modification, but the past decade has seen the discovery not only of many of the enzymes that lay down such histone modifications[1,2] but also of those that remove them[3]. Even DNA methylation, long thought to be stable and reversible only through passive loss during DNA replication, may be subject to active removal under some circumstances[4].

Advances over the past decade have also highlighted the fact that epigenetic memory may consist of more than one modification and has to be considered in the context of time — cell cycle, developmental and generational. Different molecular marks may act in a cascade of events to ensure heritability of a particular state. An elegant illustration of this has come from the exciting findings in *Schizosaccharomyces pombe*, which reveal that transcription

and RNAi work together with histone modifying and binding complexes to ensure the accurate propagation of inactivity during the cell cycle[5]. In a more developmental context, recent work on X-chromosome inactivation illustrates how a cascade of epigenetic changes can ensure reversible silencing of a chromosome during early mouse embryogenesis[6]. Developmental plasticity of epigenetic marks is a recurring theme.

These studies have started to provide answers to some fundamental questions in epigenetics: how is the memory of a particular state ensured through cell division or even across generations? And how is this memory forgotten or erased under particular circumstances (for example, in the germ line) or by accident (for example, in cancer)?

Sarah Tishkoff

One of the most exciting advances has been the development of low-cost, high-throughput methodologies for studying human genome-scale variation. These technologies have provided unprecedented knowledge about the structure of the human genome and human origins. They have also led to the identification of genetic variants with roles in human phenotypic variation, both in relation to disease susceptibility and evolutionary adaptation.

Advances in genome-wide association (GWA) studies — which have used microarray technology for SNP genotyping and have included meta-analyses of tens of thousands of individuals — have identified loci associated with common variable traits including skin, hair and eye colour and height (for example, REFS 7,8), which has shed light on human adaptation and evolution. Several approaches have also been developed for scanning the genome for targets of natural selection and have identified regions that have important roles in adaptation to diverse environments during human evolution. These include loci that have a role in lactose tolerance9,10, skin pigmentation[11], adaptation to high-altitude[12] and malaria resistance[13,14]. Some variants that are targets of natural selection also have a role in susceptibility to common disease; for example, apolipoprotein l, 1 (*APOL1*) variants associated with resistance to sleeping sickness are also associated with risk for kidney disease in individuals of African ancestry[15].

The development of statistical approaches for analysing genome-scale variation has also been crucially important. For example, methods based on principal components analysis (PcA)[16] and STRUCTURE17 analysis have facilitated studies of fine-scale population structure and individual ancestry. The results have demonstrated a remarkable correlation between genetic variation and the geographical origin of individuals18, even at the regional level in Europe19.

Sequencing technologies have also made a strong contribution to understanding human genetic variation. The first two draft human genome sequences were completed in 2001 (REFS 20,21), and the first human genome was resequenced using next-generation sequencing (NGS) technology, at a fraction of the cost, in 2008 (REF. 22). As of 2010 more than 15 complete human genomes — originating from Europe, Asia and Africa — have been resequenced, and that number is likely to rise exponentially in the next several years as costs fall. These studies have identified millions of SNPs and indicate that individual genomes may differ by megabases of sequence because of structural variation, including insertions, deletions and inversions.

Furthermore, low-coverage sequencing of the Neanderthal genome[23] has demonstrated that it has remarkable similarity to the modern human genome, and that the two species diverged less than 500,000 years ago. Those genomic regions that differ between the two appear to be involved in skin physiology and, possibly, social behaviour[23]. These studies have also raised the intriguing possibility that there may have been low levels of gene flow between

Neanderthals and modern humans in Eurasia, suggesting that our genomes may contain mosaic sequences.

The contributors*

Edith Heard is a Centre National de la Recherche Scientifique (CNRS) research director and director of the Unit of Genetics and Developmental Biology at the Institut Curie in Paris, France. She was trained in genetics at Cambridge University, UK, carried out her Ph.D. at the Imperial Cancer Research Fund, where she worked on gene amplification mechanisms, and then moved to Paris for her postdoctoral studies at the Pasteur Institute, Paris, France, where she worked on X-chromosome inactivation. Since 2001, her laboratory has been best known for its studies on epigenetic changes during development and the role of nuclear organization in gene regulation using X inactivation as a model system. She has been awarded several prizes was elected as a member of EMBO in 2005.

Sarah A. Tishkoff is the David and Lyn Silfen University Associate Professor in Genetics and Biology at the University of Pennsylvania, Philadelphia, USA, and holds appointments in the School of Medicine and the School of Arts and Sciences. She has won a Packard Career Award and a Burroughs/Wellcome Fund Career Award and was named one of *Popular Science* magazine's 'Brilliant Ten' American scientists in 2003. In 2009 she received a US National Institutes of Health (NIH) Pioneer Award. She is currently a member of the editorial board of *Genome Research* and an associate editor of *Molecular Biology and Evolution*. Her work focuses on genomic variation in Africa, human evolutionary history, the genetic basis of adaptation and phenotypic variation in Africa and the genetic basis of susceptibility to infectious disease in Africa.

John A. Todd is Professor of Medical Genetics at Cambridge University and Director of the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory in the University's Cambridge Institute for Medical Research. He researches type 1 diabetes genetics and disease mechanisms in collaboration with Linda Wicker and David Clayton and as part of the Wellcome Trust Case Control Consortium and the Type 1 Diabetes Genetics Consortium. Previously, he was Professor of Human Genetics at Oxford University, UK, and a Wellcome Trust Principal Research Fellow. He received his B.Sc. from Edinburgh University, UK, in biological sciences, and his Ph.D. from Cambridge University in biochemistry. He undertook his postdoctoral training at Cambridge University and Stanford University, California, USA. He has over 380 publications and more than 25,000 citations with an h factor of 80. He has received several awards and honours for his research, and his main goal is to provide knowledge based on genetic aetiological findings that can inform in the primary prevention of autoimmunity in type 1 diabetes and in other immune-mediated diseases.

Marc Vidal is Professor of Genetics at Harvard Medical School, Boston, Massachusetts, USA, and Founding Director of the Center for Cancer Systems Biology (CCSB) at the Dana-Farber Cancer Institute, Boston, Massachusetts, USA. His thesis work provided one of the crucial steps in the discovery that histone modifications are key to transcriptional regulation. Originally trained as a bioengineer and a geneticist, he pioneered the modelling of interactome networks, which are complex systems of interacting macromolecules operating inside cells. He is an associate member of the Royal Academy for Sciences, Letters and Arts of Belgium, his native country, and a Chercheur Qualifié Honoraire from the Fonds de la Recherche Scientifique (FRS-FNRS) of the French Community of Belgium.

Günter P. Wagner is the Alison Richard Professor of Ecology and Evolutionary Biology at Yale University, New Haven, Connecticut, USA, and the acting director of the Yale

Systems Biology Institute. He is an elected fellow of the American Association for the Advancement of Science and the American Academy of Arts and Sciences as well as a foreign member of the Austrian Academy of Sciences. He is the recipient of a MacArthur fellowship and is also the editor in chief of the *Journal of Experimental Zoology*. He studied biochemical engineering and later zoology and mathematical logics in Vienna and was appointed as Professor of Biology at Yale University in 1991. His work focuses on the biological basis of homology and character identity through studies of gene regulatory network evolution.

Jun Wang is the Executive Director of the BGI (previously known as the Beijing Genomics Institute). He was instrumental in the founding and growth of the BGI Bioinformatics Department, which is now widely recognized as one of world's premier research facilities committed to excellence in genome sciences. He also holds a position as an Ole Rømer professor at the University of Copenhagen, Denmark. He has been recognized with an award from The His Royal Highness Prince Foundation in Denmark, an Outstanding Science and Technology Achievement from the Chinese Academy of Sciences, from Top 10 Scientific Achievements in China, and the prize for Important Innovation and Contribution from the Chinese Academy of Sciences. His research focuses on genomics and related bioinformatics analysis of common diseases and agricultural crops, with the goal of developing applications using this genomic information.

Detlef Weigel is Director of the Department of Molecular Biology at the Max Planck Institute for Developmental Biology in Tübingen, Germany. He has been elected as a Member of the US National Academy of Sciences and the German National Academy of Sciences Leopoldina, and is a Foreign Member of the Royal Society of London. His work has been recognized with the Young Investigator Award of the US National Science Foundation, the Charles Albert Shull Award of the American Society of Plant Biologists, the Gottfried Wilhelm Leibniz Award of the Deutsche Forschungsgemeinschaft and the Otto Bayer Award of the Bayer Foundations. He uses both bottom-up, forward-genetic approaches and top-down, whole-genome approaches to study genetic variation in plants. His laboratory is also known for the generation of widely used community resources and for its contributions to understanding the role of microRNAs in plant development.

Richard Young is a Member of the Whitehead Institute and a Professor of Biology at the Massachusetts Institute of Technology (MIT), Cambridge, USA. He received his Ph.D. in molecular biophysics and biochemistry at Yale University and his postdoctoral training at Stanford University School of Medicine. His honours include a Burroughs Wellcome Scholarship, the Chiron Corporation Biotechnology Research Award and Yale's Wilbur Cross Medal, and *Scientific American* recognized him as one of the top 50 leaders in science, technology and business in 2006. He has served as an advisor to *Science* magazine, the NIH and the World Health Organization. His laboratory studies the genetic programmes that control cell state and differentiation in mice and humans.

*Listed in alphabetical order.

John A. Todd

It has been a decade of relief and enlightenment for the study of the genetics of common, multifactorial diseases. It was a long wait. We realized in the early 1990s that linkage studies, which work well in highly penetrant, rare, Mendelian diseases, weren't going to be very helpful. What large linkage studies did tell us about these diseases was that there are no, or extremely few, causal loci that have large individual effects on phenotype. This is true no matter what the sequence complexity of the causal loci, from genes with multiple causal

alleles[24] to structural variants, such as highly repetitive DNA sequences[25], that may be inaccessible to the current high-throughput genotyping platforms.

Even in the run-up to affordable GWA studies in sufficiently large numbers of samples, there were (mainly futile) debates on genetic models: were common variants responsible for the strong familial clustering of many common diseases? Or were low-frequency to rare variants involved? However, it was obvious that wide spectrums of allele effect sizes and frequencies would underlie the inheritance of common disease. It was also obvious that, when sequencing technology at last broke free from the small-scale capillary format, variants of increasingly low frequencies would be discovered that contribute to disease risk[24]. GWA studies confirmed (and extended) Fisher's conclusion of 1918 that many common diseases and other human characteristics are determined by inheriting a sufficient dose of susceptibility alleles from a large number of loci across the genome in a disease-permissive environment26. For example, the familial clustering of type 1 diabetes27, schizophrenia28 and height8 can be explained almost completely by common variation.

Nevertheless, largely specious debates rage on about 'missing heritability'. unfortunately, many authors are actually referring to familial clustering (based on similarity or co-occurrence of traits among first-degree family members), which is caused by intrafamilial sharing of numerous alleles and environmental factors. Familial clustering could be overestimated[27], and in the future, with larger sample sizes and better genetic coverage, we will begin to define how much of it is due to environmental exposures. One intriguing possibility is that a small portion of familial clustering is due to transgenerational effects — gene–environmental interaction events in parents, grandparents or beyond that alter the phenotypes and behaviours of the current generation29.

Regardless of genetic models or statistical methods, the most important outcome of reproducible GWA study results is what they have told us about biology and mechanisms of disease. GWA studies have been, as expected, particularly revealing in two main aspects. First, common alterations to gene expression and function do not map to obviously functional DNA changes, illustrating how little we know about genome function. Second, from the genes and pathways revealed by GWA studies, major mechanistic insights have been gained. To take a recent example, a GWA study of lung cancer shows that this disease involves genetic susceptibility to behaviour with regard to smoking, such as its initiation, dependency on smoking and even the ability to stop the habit[30].

For me, the surprise of the decade has been the discovery of an entire parallel universe of small (and large) non-protein-coding RNAs and their regulation of protein-coding gene expression[31]. A recent, beautiful example32 is the alteration of the critical tumour suppressor gene phosphatase and tensin homolog (*PTEN*) through competition for the regulatory microRNA (miRNA) between the classic *PTEN* gene and the mRNA of its pseudogene (yes, I said pseudogene), *PTENP1*. Now, proteins are being discovered — such as ROQUIN (also known as Rc3H1)33 and dead end homologue 1 (DND1)29 — that bind RNA and modulate miRNA function. What I find particularly fascinating about these two proteins is their potential effects on gene–environment interactions. They are both localized in stress granules, which are cytoplasmic sites involved in mRNA metabolism that are formed in response to stress, including in germ cells (oocytes and sperm). The functions of the non-coding RNAs, mRNAs and RNA-binding proteins in these granules could underlie transgenerational effects transmitted by germ cells29. Transgenerational phenotypes are now being documented convincingly in mouse models, including the *DND1* gene itself[29]. Transgenerational effects, including epigenetic variability[34], could help to explain the rapid increases in certain common diseases in countries with modern health care and lifestyles, including interactions between many environmental factors, not least common viral

infections that remain largely uncontrolled, our gut microbiomes and our highly polymorphic genomes[26].

Marc Vidal

At the end of a decade of unprecedented productivity in biological data accumulation, the excitement generated is somewhat tempered by a surprising level of confusion.

The most compelling and surprising excitement has emerged from the incredible amount of empirical information that has been gathered on genomes, transcriptomes, proteomes and interactomes. A recurrent theme emerging from such systematic, unbiased, high-throughput analyses is how much more complicated the molecular organization of the cell appears to be relative to what was originally revealed using focused, hypothesis-driven, reductionist molecular biology approaches. Transcriptomes, both protein-coding and non-coding, are more complex than initially thought, by at least an order of magnitude, with more than 80% of the genome seemingly transcribed in humans. The proteome is also more difficult to grasp than originally anticipated, considering the potential amount of alternative splicing, posttranslational modifications, protease-induced cleavages, disordered domains and allosteric changes that have been found to take place. And last, the landscape of macromolecular interactions is richer than imagined. For example, estimates of the numbers of protein–protein and DNA–protein interactions that can happen in a single cell are in the range of 10^5 to 10^6. Interactome maps, together with genome-wide gene knockout and knockdown analyses that are coupled to transcriptomic measurements, have revealed that molecular pathways, rather than being linear, independent from or parallel to each other, are instead incredibly intertwined and form very complex biological networks.

In an odd twist of events, these substantial advances, which were originally aimed at ‘finding and understanding all genes’, are somewhat undermining the concept of the term ‘gene’ itself. I find this of crucial importance, because the gene is often considered the fundamental unit of biology. The ‘gene number paradox’, according to which organisms of different complexity have similar numbers of protein-coding genes (for example, humans21 and worms35 both have approximately 20,000 genes), illustrates how a certain perception of biology revolves around the notion that to understand an organism, we need only ‘to understand all its genes’. consider the ambitious, exciting project framed in January 2000 and funded by the US National Science Foundation “… to understand a function for all genes of *Arabidopsis* by 2010.”[36] Ten years on, things appear more complicated.

Part of the confusion goes back to how we interchangeably use the concept of the gene as both a molecular encoding unit and a unit of heredity[37]. but this classical problem has been enormously amplified in the past decade. At the molecular level, it is increasingly harder to define the exact boundaries of both protein- and non-coding genes. At the level of heredity, GWA studies now identify genes whose relative contribution to a particular phenotype of interest can be no more than a few per cent.

The word ‘gene’ was coined 101 years ago by Wilhelm Johannsen[38]. This centennial anniversary was largely ignored last year compared to the publicity surrounding the 150th anniversary of Darwin’s *On the Origin of Species* or the 100th anniversary of Einstein’s paper on the theory of relativity. could powerful far-reaching notions that emerged at the dawn of molecular biology — such as beadle and Tatum’s ‘one gene, one enzyme, one function’ hypothesis[39] or crick’s ‘DNA makes RNA makes protein’ central dogma[40] — be running out of steam? What could replace, or at least complement, such powerful notions to improve our understanding of biology in the next decade and beyond?

Günter P. Wagner

Genomics, though a thoroughly descriptive discipline, is revolutionizing our understanding of the nature of evolutionary change in a way that was unexpected to me. Most amazing to me is the emerging role of transposable elements in evolutionary change. This became apparent through the comparative sequencing of many related genomes, both revealing the dynamical nature of genomes as well as the parts transposable elements play in that process. Certainly, the idea that transposable elements may have an important role in the evolution of gene regulatory networks is not new, but it is fair to say that this idea had only a limited influence on mainstream evolutionary genetics for most of its existence.

What is different now is that we can see exactly what transposable elements are contributing to evolutionary change. They are not just another source of random perturbation of the genetic material, they also have a constructive role, one that single-nucleotide substitutions do not have. The evidence is now pouring in, and it is greatly acknowledged in the field of genomics, but the message has not yet reached evolutionary genetics, not to speak of other areas of biology.

There is a long list of novel genetic elements that transposable elements can create[41]. To me, the most surprising is their role in the creation of novel transcription factor proteins. The poster child of a key regulatory gene is paired box gene 6 (*PAX6*), and it has been found to be derived from a transposase[41]! But the major constructive role of transposable elements probably is their ability to create new *cis*-regulatory elements. The reason is that transposable elements come prefabricated with a large number of transcription factor binding sites in their sequence, which predisposes them to become *cis*-regulatory elements. This opens the door to major and fast rewiring of gene regulatory networks, and transposable elements are thus important agents in the origin of major innovations[42].

Why are these findings leading to a revolution in evolutionary biology? One of the basic ideas at the root of the Darwinian theory of evolution is the idea of uniformitarianism. It says that past change is due to exactly the same kinds of processes that we can observe today. This makes a lot of sense and allows a mechanistic understanding of past events. But what the new research about transposable elements tells us is that this principle is true only to a certain degree. It turns out that the kinds of genetic changes possible at any particular time in history and in each lineage depend on what kinds of transposable elements are present and active[43]. Because of the biased nature of the changes transposable elements cause, certain changes are more likely at certain times in history, and in certain lineages, than at other times or in other lineages. That means that the potential for evolutionary innovations is heterogeneous in time, among lineages and among tissue types. Transposable elements are active in a lineage for only a limited amount of time, and evolutionary innovation seems to be particularly intense during these periods in history. Since each lineage has its own transposable elements, it is also clear that the likelihood of genetic innovations differs simply because one lineage has a certain kind of transposable element and others do not. But all this flies into the face of uniformitarianism and thus forces us to rethink how genetic change is happening in evolution.

Jun Wang

Within the 10 years following the announcement of the landmark completion of the Human Genome Project[21] there have been an amazing number of revolutionary innovations in the field of genomics that have aimed at answering the question, ‘is deciphering the genome useful?’ That is, will genomics help to improve our health and our everyday life? Although there is still no consensus, scientists have completed numerous milestone projects that indicate the answer is probably a qualified yes.

The first set of goals was aimed at ‘defining genomes’ — establishing data to serve as a new and fundamental resource for genetics. Scientists have sequenced an impressive list of genomes, including those of humans20،44, standard model animals (for example, mice45) and plants (for example, *Arabidopsis thaliana*46), economical animals (for example, the bovine genome47) and crop plants (for example, rice48) and many others. Large-scale efforts aiming to generate reference data sets, such as the HapMap project49, are also a part of this — that is, defining the genomes of multiple individuals.

The process of defining genomes has been accompanied by other crucial advances: the development, organization and management of consortia to carry out these large-scale scientific projects; the establishment of data access conventions; and the development of many novel approaches and tools for data analysis. These factors have jointly established a maturing model for defining genomes and for making that information broadly available. They have also had a surprising and profound impact on the entire biological community by promoting collaboration and establishing goals for quantifying, standardizing and sharing all types of biological and medical data[50].

The second set of goals has been in ‘understanding genomes’ — making sense of each nucleotide of these sequences. This has primarily involved large-scale projects doing association studies51 and quantitative trait loci mapping52 or direct functional experiments on specific genes and genetic variations[53]. The findings have demonstrated possibilities for linking the genetic code to a phenotype (but not vice versa). Another focus has been investigating the relationship between genotypes and gene expression[54] and between epigenetics and expression55 to understand the intermediate levels of the organization of life that lie between genome and phenotype. Altogether, this work has begun to characterize the digital information underlying the mechanisms of life, which sheds light on how to go forward after the era of defining genomes.

Still, after a decade of criticism, controversy and success, the interpretation of genomes (as expected) lags far behind the progress in genome sequencing and, with a few exceptions, surprisingly little has been done to develop applications using genomes. That the genome is useful has been a basic, but uncertain, tenet over the past 10 years. While it is now one that seems far clearer, it is time for the genome to fulfil its promise.

Detlef Weigel

Foremost is the phenomenal increase in DNA sequencing capacity, which very few, if any, of us anticipated 10 years ago, but I would like to emphasize a few areas in which plant biology has been in the vanguard of genetics or has made unique contributions.

An excellent example is the realization that small RNAs are ubiquitous, that they come in many different flavours, that they affect an enormously broad range of biological processes, that they do so through many different mechanisms and that they themselves are subject to multiple modes of regulation. As in animals, plant miRNAs can control target genes by preventing translation of their mRNAs. In addition, they often guide mRNA cleavage, either directly or indirectly through secondary small RNAs[56]. Furthermore, miRNAs themselves are not only regulated transcriptionally but also through decoys called target mimics — a mechanism discovered in plants[57].

The first whole-genome maps of small RNAs were produced in plants[58]; most recently, this has led to the remarkable finding that plants ‘allow’ transposons to become active in vegetative tissue so that they can make use of the resulting small RNAs in order to silence these harmful elements in generative cells[59]. Plant genomics has also led the way in linking small interfering RNAs and DNA methylation on a whole-genome scale. Ironically, the

breakthrough advances in producing whole-genome DNA methylation maps were widely noticed only once the methods developed for plants were applied to human cells[60]. In the general area of gene silencing, the flowering regulator *FLC* has provided a powerful platform for discovering several principles of chromatin modification, including ones involving non-coding RNAs, that also apply to animals[61].

There are several other areas in which plant geneticists have made dramatic progress during the past decade, not least in identifying the receptors for all of the major plant hormones, most of which operate through mechanisms that are rather different from those employed in animal signal transduction[62]. And in those cases in which similar types of molecules are recognized, as with steroids[63], plants have found unique solutions to the problem of relaying information between cells and organs.

Richard Young

The most exciting advances have been in the area of cellular reprogramming with defined transcription factors. In 2006 Yamanaka and colleagues showed that transient expression of four transcription factors could reprogramme somatic cells to a pluripotent embryonic stem (ES) cell-like state[64]. Before this discovery, it was evident that the nuclei of differentiated cells could be reprogrammed when transferred into an oocyte, but the nature of the reprogramming factors was a mystery. The discovery that a few key transcription factors are sufficient to reprogramme the vertebrate genome has profoundly affected many fields of biomedical research and the quest for regenerative medicine.

The ability to generate induced pluripotent stem (iPS) cells has allowed investigators to create disease- and patient-specific embryonic stem cells[65]. Disease-specific iPS cells have already been generated from patients with various genetic disorders, allowing researchers to study the effects of specific mutations on cellular function and providing cellular reagents for drug screening. Patient-specific iPS cells may provide the raw material for tissue-replacement therapies that should circumvent immune rejection. The promise of regenerative medicine seems more tangible with iPS cells, prompting discussion of the challenges that must be addressed before therapies based on these or other stem cells are used in the clinic[65–68].

Recent studies have shown that some differentiated cell types can be directly reprogrammed into other differentiated cell types. For example, forced expression of three transcription factors induces the direct conversion of pancreatic exocrine cells into insulin-producing endocrine cells[69]. Similarly, transient expression of three transcription factors is sufficient to reprogramme fibroblasts into neurons[70].

The ability to reprogramme cells has also improved our understanding of how cell state and differentiation are controlled. The reprogramming experience suggests that both embryonic and differentiated cells rely on a small collection of key transcription factors to establish and maintain cell state[71]. In the best-studied cases, the exogenous reprogramming transcription factors jump-start their endogenous counterparts, which form a positive feedback loop that maintains the cellular control circuitry after the exogenous factors are silenced or removed[72]. It is conceivable that many cell types rely on a small set of these ‘master’ transcription factors to maintain a stable and specific cell state, and that reprogramming into a broad spectrum of clinically useful cell types simply awaits the identification of these factors.

What are the key prospects and challenges in the next few years?

E.H.

One major challenge that cuts across many fields will be to integrate the massive data sets that are being generated by NGS technology. We now have access to epigenomes, transcriptomes, small RNA and replication timing profiles. We also know a lot more about nuclear organization and chromosome structure thanks to chromosome conformation capture technologies. Furthermore, such data sets can be analysed in the context of allelic differences, providing important insights into regulatory variation. These studies represent a gold mine of information and have already provided a glimpse of the sophistication with which the genome is organized and interpreted. For example, we can at last define the molecular signatures associated with different expression states and thus understand the nature of euchromatin and heterochromatin. Based on studies in lower eukaryotes, integrating these huge data sets, modelling the results and obtaining testable predictions should be achievable aims for systems biologists. The question now is, how easy will this be in higher eukaryotes? For example, repetitive sequences — long considered as junk DNA but now recognized as key structural and functional components of the genome — are an important feature of higher eukaryotes but represent a major bioinformatics challenge.

Another challenge is to go beyond the static view that NGS data provide, as well as the heterogeneity that they can hide (in populations of cells). NGS-based approaches are starting to be applied to specific developmental stages, or to cells obtained by sorting, grown under different conditions or taken from mutant backgrounds. Single-cell approaches are also coming of age for transcriptome analyses[103], and high-resolution microscopy and live cell-imaging techniques are providing unprecedented views of nuclear dynamics and gene regulation73. Obtaining an integrated four-dimensional view of nuclear structure and function should at last be possible.

In the field of epigenetics, many challenges remain. One issue concerns the extent to which gene epigenetics relies on negative and positive feedback loops. We also have yet to understand the detailed molecular mechanisms by which an expressed or silent state can be transmitted through cell division. How exactly are chromatin states replicated[74]? To what extent are histones themselves the carriers of epigenetic information? What is the exact role of Polycomb or Trithorax complexes in propagating epigenetic states? Is RNA involved and, if so, how? Is transcription itself an epigenetic propagator? How does nuclear organization influence chromatin heritability[1,74]?

Perhaps one of the most important and exciting challenges over the next few years will be to determine the extent to which epigenetic modifications can be transgenerational and the impact that such heritable epimutations or epivariants might have on quantitative traits, or on disease predisposition[75]. The potential implications of this ‘hidden’ heritability are enormous. understanding the molecular mechanisms by which transgenerational heritability of chromatin states occurs is clearly a tenable prospect in the next few years for plants[76], but is likely to be more of a challenge in mammals, in which germline erasure predominates and functional studies are more laborious[77].

S.T.

Despite the advances in genomics technologies and genome-wide scans of selection and disease association over the past decade, there have been few success stories identifying functional genetic variants that have a role in complex traits such as height, diabetes, cardiovascular disease or autism. Many of the loci identified by GWA studies account for just a fraction of the phenotypic variation and disease risk and, in the majority of cases, biologically functional variants have not yet been identified. Thus, a key challenge in the

future will be to detect the ‘missing heritability’ in GWA studies. Strategies for tackling this issue will include the determination of gene–environment and gene–gene interactions, exploration of how rare variants contribute to phenotypic variability, and determining the extent of structural variation in ethnically diverse populations and its role in producing phenotypic diversity.

Importantly, although many of the alleles underlying complex trait variation are likely to influence gene expression, the identification of regulatory variants is particularly challenging. This is partly because we are just beginning to understand the important effects of small RNAs and epigenetic factors on gene regulation. However, recent advances in the development of high-throughput genomics technologies now facilitate approaches that integrate genomics (for example, DNA sequencing and chromatin immunoprecipitation followed by sequencing (chIP–seq)), transcriptomics (for example, RNA sequencing), proteomics and epigenomic data from assorted tissues to identify both coding and non-coding variants that impact variable traits. A systems biology approach that explores how naturally occurring genetic variants perturb genetic and transcriptional networks may be particularly successful for dissecting the genetic architecture of complex variable traits. Equally important will be an understanding of how these networks are influenced by environmental factors, including diet, lifestyle and disease status.

As the cost of resequencing decreases, the sequencing of exomes, and eventually whole genomes, at the population level is becoming feasible. The 1000 Genomes Project has already initiated population-level analyses of whole-genome variation. In the future, it will be crucial to include ethnically diverse populations, which may have population-specific variants owing to their demographic history or to local adaptation. Africa, in particular, has been underrepresented in genomics and GWA studies. It will be important to tackle this imbalance, given that African populations contain the greatest levels of human genetic variation and high levels of genetic substructure, and because this region is the source of the world-wide range expansion of all modern humans over the past 100,000 years. Africa also has high levels of both communicable and non-communicable disease. In order to facilitate GWA studies in Africans, SNP arrays will need to be developed based on SNPs identified through resequencing studies in Africans. In general, population-level resequencing studies will require the development of computational methods for storing and analysing large amounts of data and for determining which of the millions of SNPs likely to be identified are functionally relevant.

Novel approaches will also need to be developed to identify genetic variants that are targets of natural selection and have a role in adaptation. Current approaches are most effective for identifying recent strong selection of novel adaptive variation. One of the greatest challenges in the future will be to identify more complex signatures of natural selection, as would be expected for genetic variants that have a role in complex adaptive traits that are influenced by multiple loci and the environment, by selection acting on standing variation and by fluctuating selection that changes with the environment. Additionally, methods are needed for detecting natural selection acting on pathways rather than single loci. Thus, the integration of systems biology approaches and studies of natural selection may be particularly informative. Finally, but crucially, the identification of causal variants will require *in vitro* and *in vivo* functional assays in order to elucidate the biological roles of adaptive variants.

J.A.T.

A major task will be to correlate gene expression and function with genome variation. Small differences in the expression and function of certain genes can result in specific cell phenotypes that can have highly significant, pleiotropic effects on development as well as on

susceptibility and resistance to disease32,78. Fortunately, we are not without tools to explore these links, many of them recently developed or enhanced significantly by NGS[78]. For example, we can survey, on a genome-wide basis, which sequences bind which transcription factors (chIP–seq); which sequences are conserved or not between species; which enhancer sequences loop round to contact promoter sequences (chromosome conformational capture); which regions are differentially methylated; and which genome variants correlate with gene splicing, transcription and translation, and with susceptibility to disease.

We need to further improve these methods by making them as quantitative as possible. We also need to invent and improve similarly enabling tools for protein and cell analyses in order to correlate the presence of certain alleles with protein and cellular phenotypes. High-affinity monoclonal antibodies for each gene product and their splice isoforms would be incredibly helpful. We need to apply these methods to cell populations derived from healthy individuals and those with disease[79] or disease risk factors. A particular focus should be on patients undergoing treatment, and on the evaluation of new therapies and risk-lowering prevention strategies. Once we have identified the inherited phenotypes that precede disease diagnosis, we can begin to unravel disease mechanisms that could be targeted specifically by future therapeutics.

In general, however, we cannot expect health benefits from genetic- and genomics-derived knowledge of mechanisms of disease to happen any faster than the 20 years it can take from molecule to approved drug in the traditional pharmaceutical pipeline. For example, 15 years since they were first developed, microarrays for gene expression analysis are only just yielding their first clinical applications in cancer and in heart transplant rejection[80]. Modification of environmental factors and behaviour may provide substantial health benefits in the future, with GWA study results providing key guidelines about which ones to modify[24,26,30].

M.V.

The complexity of transcriptomes, proteomes and interactomes revealed in the past decade relates back to classical problems. In addition to coining the word ‘gene’, Johannsen defined the words ‘genotype’ and ‘phenotype’[38] from his observation that self-fertilized inbred ‘pure lines’ of bean plants (read ‘isogenic’) followed a normal distribution of pod sizes. The underappreciated take-home message was that identical genotypes do not always give rise to identical phenotypes.

Of course, the environment comes to mind as a mediator between genotype and phenotype, but there is clearly more to it. A century after Johannsen, genotype–phenotype relationships are far from being fully understood. Identical twins have different fingerprints. Incomplete pene-trance and variable expressivity are observed more often than not, even in genetic experiments that analyse the effects of mutations under well-controlled conditions. Only relatively small proportions of genes appear to be essential, at least in simple unicellular organisms such as yeast[81,82]. On the other hand, powerful GWA studies aimed at understanding complex traits in humans are revealing more contributing loci than were originally anticipated[83].

The next decade will bring new inroads to the challenge of relating genotype to phenotype. Paradoxically, from an arguably confusing situation concerning what genes ‘are’ and ‘do’ (if they do anything) at the end of the last decade, the 2010s are poised to deliver fundamentally new ways of comprehending biological problems, further clarifying the understanding of genetics, and perhaps treating and curing disease.

Significant answers will probably come from the increasingly clear perception that virtually no gene product functions in isolation. Proteins, miRNAs, regulatory DNA sites — that is, all macromolecular entities — need to physically, biochemically or functionally interact with others to perform their cellular functions. Intimately connected biological components form complex, dynamic and logical networks or systems, which exhibit emergent and non-intuitive properties[84]. A systems approach to biological questions is likely to illuminate genotype– phenotype relationships. Observable phenotypes should be considered as the manifestation of precisely those systems properties, rather than simply the result of genomic variations. Phenotypes are probably more directly related to systems properties than they are to DNA and 'genes'. Mutations affect encoded macromolecular components of the cell, which in turn affect the properties of systems.

The idea that multi-scale, complex systems formed by interacting macromolecules and metabolites, cells, organs and organisms underlie life is not new. Visionaries such as Delbrück[85], Waddington[86], Monod and Jacob[87], among others, proposed such thinking in the mid-twentieth century. There was a powerful resurgence of these concepts over the past decade, together with the development of sophisticated informatic and imaging tools, combined with the engineering and physics concepts of control and graph theory. The resulting systems-level understanding of what life is may materialize as one of the major ideas of biology[88].

Provided that enough funding can be dedicated to mapping, modelling, perturbing and synthetically reconstituting biological networks and systems, the next decade should give fresh insights into the fundamental question of genotype–phenotype relationships. Perturbations of systems properties will probably be found to underlie many complex traits.

G.P.W.

One important outstanding issue to understand is the origin of the networks that determine cell type identity. In part due to the efforts in stem cell research, we now have a fairly good understanding of the mechanistic basis of cell type identity[71]. It turns out that a cell's identity is mediated by a core gene regulatory network that cooperatively regulates target genes that, in turn, cause the cell-type-specific phenotype. This means that the hierarchy is very shallow, with one level comprising the core network and a second level made up of the target genes. There is preliminary evidence that this organization might also be applicable to the identity of multicellular characters[89]. Hence, character identity, and thus homology — one of the most difficult concepts in evolutionary biology — might be understandable at the molecular level. Now the challenge is to understand the genetic events that lead to the origin of novel cell type identity networks[90]. This is a largely untouched area of research.

In my laboratory we think that the origin of these networks involves molecular mechanisms that are different from nucleotide substitutions in existing *cis*-regulatory elements[42]. Among the molecular changes that are likely to be involved in the evolution of novel gene regulatory networks are new *cis*-regulatory elements derived from transposable elements, as discussed above, and the evolution of novel transcription factor complexes[91]. The latter seems to be important since cell-type-specific gene regulation, at least in vertebrates, seems to depend on the formation of transcription factor complexes that mediate the cooperative action of the participating transcription factors[71]. How the origin of novel *cis*-regulatory elements derived from transposable elements and the evolution of novel transcription factor complexes ultimately lead to innovations, such as novel cell types, is currently unclear.

The challenge is to find examples in which the biological significance of the genetic changes is certain; that is, what was the ancestral situation, what is the function of the derived cell type and when did it happen in evolution? It is hard to make sense of the vast amount of

differences one can identify by comparing the genomes and transcriptomes of two or three randomly chosen organisms, such as humans, mice and zebrafish. Research has to focus on specific evolutionary events in which novel characters, such as novel cell types, hair or pregnancy, evolved. The experimental work has to focus on organisms that are informative about these transitions; that is, they need to be well placed on the phylogeny to reveal the relevant changes. The threshold of doing that, that is, experiments with minimal reliance on traditional model organisms, is becoming lower with new technology. Now one needs only a genome sequence to do transcriptomics (instead of investing in a microarray), and methods for manipulating gene expression, such as RNAi, are more and more universal. For this research, what an appropriate model organism is has to be redefined in terms of how informative a species is in a phylogenetic context.

J.W.

The past decade has witnessed the impact of genomes on scientific research. We also have more confidence than we had 10 years ago in the premise that genomes will bring new applications and be useful to the general public. Only a few applications are currently being developed by virtue of genomics, but I feel we are now entering the stage at which scientists will be asking, ‘How can we use the genome?’ In the coming 2–3 years, I believe genomics work will still primarily concentrate on deciphering and understanding, but using genomes in a broader context than basic research will become especially important. To be able to do this, it is essential that work focuses on innovation in experimental technologies and computational methodologies.

Collecting biological information (deciphering) using high-throughput, precise profiling of data on multiple biological levels will be crucial. Although second-generation technology has solved many throughput issues for nucleotide sequencing, several problems remain. The refinement of computational algorithms for nucleotide sequencing and profiling is required to ease downstream analyses. Issues related to wet-lab protocols, such as obtaining adequate starting material and the presence of experimental bias, still create serious limitations. Additionally, the throughput levels and bioinformatics tools for non-nucleotide biological information — such as proteins, metabolites, phenotypes and even environmental factors — are currently no match for those for nucleotide sequencing. It is exciting to note that efforts to address this issue have already begun (for example, studies relating to the impact of environment on complex disease92), but more needs to be done to tackle this challenge.

Interpreting these disparate types of biological information (‘understanding’), given the complexity of biological interactions, will require truly integrative analyses from a more mature form of systems biology that deals with combinations of genomic, epigenetic, expression, metabolism and phenotypic data. Reaching this goal will take far more than 2–3 years; thus, the challenge over these coming years will be to develop proper algorithms and analytical tools to handle such unprecedented large-scale multidimensional networked data. This is a highly achievable goal.

In the field of medicine, with the technology we have today we should be able to decipher nearly all confirmed monogenic disorders and, with a focus on developing better integrative analyses and tools, we will begin to gain a more complete understanding of complex disorders, including cancer. New and improved medications will come from the identification of a wide range of novel druggable targets. Of greatest interest will be progress in medical genomics and personal genomics, which may shift some of our focus from drug development to a less expensive and more productive predictive and preventive health care system. In agriculture and ecology, we will see the identification and analysis of key genes and their activities in animals, plants and microbes that contribute to important economical and ecological phenotypes. We will then be able to establish methods for

applying this information to guide the breeding and cultivation of new varieties or strains in an environmentally safe way.

As I see it, the coming 2–3 years will be the incubation period for large-scale applied genomics and bio-industry. It will certainly be a historical time point when scientists and the public alike are convinced that, after years of effort, genomics is useful.

D.W.

By studying genetic variation, we can begin to interpret the results from mechanistic approaches in an evolutionary and ecological context. In humans, the excitement about the use of GWA studies for the discovery of alleles that affect disease susceptibility or other important traits has recently become a bit muted because the effect sizes and the fraction of explained variation tend to be rather small.

By contrast, proof-of-principle studies in plants have already shown that GWA studies, even with relatively small sample sizes, can pick up loci that account for a large fraction of phenotypic variation, and these approaches are further enhanced by combining them with association analyses in populations derived from crosses[93]. Moreover, the inbred nature of many species, particularly of crops, greatly facilitates the comprehensive discovery of sequence variants and thus should speed up the identification of polymorphisms that are causal for phenotypic differences.

Perhaps the most promising short-term prospect is in exploiting new sequencing technologies for high-precision linkage mapping in experimental populations that segregate for a trait of interest. In model organisms, one can use pools of mutant individuals to both map and identify in a single sequencing reaction sequence polymorphisms that are causal for mutant phenotypes94. This approach can be readily extended to polygenic traits, which typically underlie phenotypic differences between wild strains[95]. In principle, this makes any sexually reproducing organism amenable to genetic analysis. We will see a new era in which the proverbial ‘awesome power of genetics’ is brought to bear on any species that has at least one special trait of interest, such as production of a valuable chemical or resistance to a dangerous crop pathogen. All that one needs is individuals that differ in this trait, crosses between these individuals and, ideally, the ability to phenotype large segregating populations with thousands of individuals, so that high-resolution mapping of recombinants allows very accurate triangulation of the causal gene (or genes). Performing genetic studies in much broader groups of species will allow us to harness the evolutionary and ecological history embedded in the genomes of all sorts of organisms that survive under an incredibly broad range of environmental conditions.

Sequencing will also play an important part in understanding the extended phenotype, through the analysis of microbial communities. Most of us have suffered from the consequences of having the wrong type of bacteria thrive in our intestines, which dramatically underscores how important microbes are for our wellbeing. Indeed, we have already learned how diverse the bacteria are that live in the human gut[96]. Similarly, there are many reasons to believe that bacteria, fungi, oomycetes and other microbes play an essential, yet poorly understood, part in helping plants to extract resources from their environment. Moreover, as with animals, most microbes do not cause disease on most plants because they express conserved pathogen-associated molecular patterns (PAMPs). To overcome PAMP-triggered basal defences that limit microbial growth, successful pathogens need to deploy virulence proteins[97]. Yet, despite PAMP recognition, huge numbers of microbes live relatively peacefully on and in multicellular organisms. I am certain that we will soon find out how the host distinguishes beneficial or innocuous microbes from harmful ones. An exciting opportunity is offered by the prospect of exploiting this knowledge to

engineer new, advantageous plant–microbe associations to enhance the performance of crops and trees.

Finally, there are two related areas in which I believe plants will lead the way. One is in revealing cell-type-specific gene expression patterns[98], a key step for deciphering the transcriptional code embedded in *cis*-regulatory elements. (I admit to having believed 10 years ago that we would have mastered this challenge by now.) In addition, physical models of growing plant organs have recently become impressively sophisticated[99], and integrating them with cell-type-specific models of gene regulatory networks holds exceptional promise for being able to predict the development of an entire organism. Combining these, in turn, with knowledge of genetic variation will usher in a new era of true systems biology.

R.Y.

I believe we will see some very exciting and important advances in our understanding of development and disease that will challenge us in new ways.

I anticipate substantial advances in developmental biology, in which studies of the regulatory mechanisms that control cell state will continue to produce new and surprising insights. For example, recent insights into gene silencing, which is crucial for establishing and maintaining cell states, suggest that non-coding RNA (ncRNA) molecules are frequently involved[100]. Much remains to be learned about ncRNA genes in humans and the functions of various classes of ncRNAs in different cell types. Because some regulatory ncRNAs are short-lived, it will be a challenge to identify all of these and ascertain their functions.

The study of disease-specific iPS cells will lead to new insights into disease mechanisms, and these should ultimately facilitate the development of novel therapeutic interventions. Novel therapies generally take at least a decade to develop, so it will continue to be a major challenge to accelerate delivery of useful therapies to the clinic. After extensive preclinical tests, a substantial fraction of drugs fail to show efficacy in the clinic. Clinical trial design is now beginning to take advantage of the genetics of individuals, which may increase the success rate of novel therapeutics.

The cost of genome sequencing will become so low in the next 2–3 years that thousands of individuals will be sequenced. Consequently, we will begin to reclassify disease based on genome sequence. The combination of information from genome sequencing and iPS cell studies will significantly accelerate our understanding of disease mechanisms. Screening of chemical compound libraries for molecules that affect iPS cell phenotypes will become common, providing hits that may accelerate the development of lead compounds for new therapeutics. Collaborations between biologists, chemists, computer scientists and clinicians will become ever more valuable and essential.

The concept that transcription factors control gene expression programmes was established half a century ago[101], at the same time that Gurdon showed by nuclear transfer that differentiated intestinal cells in amphibians retain all of the genetic information needed to produce an entire frog[102]. Transcription factors recognize specific DNA sequences and instruct the transcription apparatus to produce RNA, thus controlling gene expression programmes, development and cell states. The ability of specific transcription factors to reprogramme cell states is the ultimate demonstration of these concepts. In this important area, at least two major challenges face the biomedical community. The first is to begin the process of discovering how the ~1,500 human transcription factors control the hundreds of cell states that exist in humans. The second is to learn how to modify the activities of transcription factors directly using small-molecule chemistry. We will begin to address these challenges in the next 2 years.

Acknowledgments

S.T. thanks members of her laboratory for helpful discussions. J.A.T. thanks the Wellcome Trust, the Juvenile Diabetes Research Foundation International and the National Institute for Health Research Cambridge Biomedical Research Centre for funding, and J. Nadeau for sharing unpublished information. The Cambridge Institute for Medical Research is a recipient of a Wellcome Trust Strategic Award (079895). M.V. would like to thank M. Walhout, J. Dekker and J. Vandenhaute for helpful conversations on the subject discussed here.

References

1. Margueron R, Reinberg D. Chromatin structure and the inheritance of epigenetic information. Nature Rev. Genet 2010;11:285–296. [PubMed: 20300089]
2. Jenuwein T, Allis CD. Translating the histone code. Science 2001;293:1074–1080. [PubMed: 11498575]
3. Cloos PA, Christensen J, Agger K, Helin K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev 2008;22:1115–1140. [PubMed: 18451103]
4. Baumann K. Epigenetics: Unravelling demethylation. Nature Rev. Mol. Cell Biol 2010;11:87.
5. Kloc A, Martienssen R. RNAi, heterochromatin and the cell cycle. Trends Genet 2008;24:511–517. [PubMed: 18778867]
6. Okamoto I, Heard E. Lessons from comparative analysis of X-chromosome inactivation in mammals. Chromosome Res 2009;17:659–669. [PubMed: 19802706]
7. Sulem P, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nature Genet 2007;39:1443–1452. [PubMed: 17952075]
8. Yang J, et al. Common SNPs explain a large proportion of the heritability for human height. Nature Genet 2010;42:565–569. [PubMed: 20562875]
9. Bersaglieri T, et al. Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet 2004;74:1111–1120. [PubMed: 15114531]
10. Tishkoff SA, et al. Convergent adaptation of human lactase persistence in Africa and Europe. Nature Genet 2007;39:31–40. [PubMed: 17159977]
11. Pickrell JK, et al. Signals of recent positive selection in a worldwide sample of human populations. Genome Res 2009;19:826–837. [PubMed: 19307593]
12. Simonson TS, et al. Genetic evidence for high-altitude adaptation in Tibet. Science 2010;329:72–75. [PubMed: 20466884]
13. Sabeti PC, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature 2002;419:832–837. [PubMed: 12397357]
14. Tishkoff SA, et al. Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. Science 2001;293:455–462. [PubMed: 11423617]
15. Genovese G, et al. Association of trypanolytic *APOL1* variants with kidney disease in African-Americans. Science 2010;329:841–845. [PubMed: 20647424]
16. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2:e190. [PubMed: 17194218]
17. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945–959. [PubMed: 10835412]
18. Rosenberg NA, et al. Genetic structure of human populations. Science 2002;298:2381–2385. [PubMed: 12493913]
19. Novembre J, et al. Genes mirror geography within Europe. Nature 2008;456:98–101. [PubMed: 18758442]
20. Venter JC, et al. The sequence of the human genome. Science 2001;291:1304–1351. [PubMed: 11181995]
21. Lander ES, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921. [PubMed: 11237011]
22. Wheeler D, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008;452:872–876. [PubMed: 18421352]

23. Green RE, et al. A draft sequence of the Neandertal genome. Science 2010;328:710–722. [PubMed: 20448178]
24. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 2009;324:387–389. [PubMed: 19264985]
25. Lupski JR. Retrotransposition and structural variation in the human genome. Cell 2010;141:1110–1112. [PubMed: 20602993]
26. Todd JA. D'oh! genes and environment cause Crohn's disease. Cell 2010;141:1114–1116. [PubMed: 20602995]
27. Clayton DG. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. PLoS Genet 2009;5:e1000540. [PubMed: 19584936]
28. Purcell SM, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748–752. [PubMed: 19571811]
29. Nadeau JH. Transgenerational genetic effects on phenotypic variation and disease risk. Hum. Mol. Genet 2009;18:R202–R210. [PubMed: 19808797]
30. Amos CI, Spitz MR, Cinciripini P. Chipping away at the genetics of smoking behavior. Nature Genet 2010;42:366–368. [PubMed: 20428092]
31. Brown MS, Ye J, Goldstein JL. Medicine. HDL miR-ed down by SREBP introns. Science 2010;328:1495–1496. [PubMed: 20558698]
32. Poliseno L, et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010;465:1033–1038. [PubMed: 20577206]
33. Athanasopoulos V, et al. The ROQUIN family of proteins localizes to stress granules via the ROQ domain and binds target mRNAs. FEBS J 2010;277:2109–2127. [PubMed: 20412057]
34. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature 2010;465:721–727. [PubMed: 20535201]
35. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 1998;282:2012–2018. [PubMed: 9851916]
36. Somerville C, Dangl J. Plant biology in 2010. Science 2000;290:2077–2078. [PubMed: 11187833]
37. Falk R. The gene in search of an identity. Hum. Genet 1984;68:195–204. [PubMed: 6389318]
38. Johannsen, W. Elemente der exakten Erblichkeitslehre. Jena: Gustav Fischer; 1909.
39. Beadle GW, Tatum EL. Genetic control of biochemical reactions in *Neurospora*. Proc. Natl Acad. Sci. USA 1941;27:499–506. [PubMed: 16588492]
40. Crick F. Central dogma of molecular biology. Nature 1970;227:561–563. [PubMed: 4913914]
41. Feschotte C. Transposable elements and the evolution of regulatory networks. Nature Rev. Genet 2008;9:397–405. [PubMed: 18368054]
42. Wagner GP, Lynch VJ. Evolutionary novelties. Current Biol 2010;20:R48–R52.
43. Oliver KR, Greene WK. Transposable elements: powerful facilitators of evolution. BioEssays 2009;31:703–714. [PubMed: 19415638]
44. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931–945. [PubMed: 15496913]
45. Waterston RH, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520–562. [PubMed: 12466850]
46. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000;408:796–815. [PubMed: 11130711]
47. Elsik CG, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 2009;324:522–528. [PubMed: 19390049]
48. Yu J, et al. A draft sequence of the rice genome (*Oryza sativa*, L. ssp. *indica*). Science 2002;296:79–92. [PubMed: 11935017]
49. International HapMap Consortium. The International HapMap Project. Nature 2003;426:789–796. [PubMed: 14685227]

50. Gerdsen F, Mueller S, Jablonski S, Prokosch HU. Standardized exchange of medical data between a research database, an electronic patient record and an electronic health record using CDA/SCIPHOX. AMIA Annu. Symp. Proc 2005;2005:963. [PubMed: 16779250]
51. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nature Rev. Genet 2005;6:109–118. [PubMed: 15716907]
52. Andersson L, et al. Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science 1994;263:1771–1774. [PubMed: 8134840]
53. Li C, Zhou A, Sang T. Rice domestication by reducing shattering. Science 2006;311:1936–1939. [PubMed: 16527928]
54. West MA, et al. Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. Genetics 2007;175:1441–1450. [PubMed: 17179097]
55. Xiang H, et al. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. Nature Biotech 2010;28:516–520.
56. Allen E, Xie Z, Gustafson AM, Carrington JC. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. Cell 2005;121:207–221. [PubMed: 15851028]
57. Franco-Zorrilla JM, et al. Target mimicry provides a new mechanism for regulation of microRNA activity. Nature Genet 2007;39:1033–1037. [PubMed: 17643101]
58. Lu C, et al. Elucidation of the small RNA component of the transcriptome. Science 2005;309:1567–1569. [PubMed: 16141074]
59. Slotkin RK, et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 2009;136:461–472. [PubMed: 19203581]
60. Lister R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009;462:315–322. [PubMed: 19829295]
61. Liu F, Marquardt S, Lister C, Swiezewski S, Dean C. Targeted 3′ processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. Science 2010;327:94–97. [PubMed: 19965720]
62. Tan X, et al. Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 2007;446:640–645. [PubMed: 17410169]
63. Wang ZY, Seto H, Fujioka S, Yoshida S, Chory J. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. Nature 2001;410:380–383. [PubMed: 11268216]
64. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006;126:663–676. [PubMed: 16904174]
65. Yamanaka S. A fresh look at iPS cells. Cell 2009;137:13–17. [PubMed: 19345179]
66. Daley GQ, Scadden DT. Prospects for stem cell-based therapy. Cell 2008;132:544–548. [PubMed: 18295571]
67. Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell 2008;132:661–680. [PubMed: 18295582]
68. Saha K, Jaenisch R. Technical challenges in using human induced pluripotent stem cells to model disease. Cell Stem Cell 2009;5:584–595. [PubMed: 19951687]
69. Zhou Q, et al. *In vivo* reprogramming of adult pancreatic exocrine cells to β-cells. Nature 2008;455:627–632. [PubMed: 18754011]
70. Vierbuchen T, et al. Direct conversion of fibroblasts to functional neurons by defined factors. Nature 2010;463:1035–1041. [PubMed: 20107439]
71. Graf T, Enver T. Forcing cells to change lineages. Nature 2009;462:587–594. [PubMed: 19956253]
72. Jaenisch R, Young RA. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 2008;132:567–582. [PubMed: 18295576]
73. Janicki SM, Spector DL. Nuclear choreography: interpretations from living cells. Curr. Opin. Cell Biol 2003;15:149–157. [PubMed: 12648670]
74. Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. Nature Rev. Mol. Cell Biol 2009;10:192–206. [PubMed: 19234478]
75. Johannes F, Colot V, Jansen RC. Epigenome dynamics: a quantitative genetics perspective. Nature Rev. Genet 2008;9:883–890. [PubMed: 18927581]

76. Teixeira FK, Colot V. Repeat elements and the *Arabidopsis* DNA methylation landscape. Heredity 2010;105:14–23. [PubMed: 20461104]
77. Eichler EE, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Rev. Genet 2010;11:446–450. [PubMed: 20479774]
78. Heinig M, et al. A *trans*-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature. (in the press).
79. McKinney EF, et al. A $CD8^+$ T cell transcription signature predicts prognosis in autoimmune disease. Nature Med 2010;16:586–591. 1p following 591. [PubMed: 20400961]
80. Pham MX, et al. Gene-expression profiling for rejection surveillance after cardiac transplantation. N. Engl. J. Med 2010;362:1890–1900. [PubMed: 20413602]
81. Winzeler EA, et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 1999;285:901–906. [PubMed: 10436161]
82. Giaever G, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 2002;418:387–391. [PubMed: 12140549]
83. Manolio TA. Genomewide association studies and assessment of the risk of disease. N. Engl. J. Med 2010;363:166–176. [PubMed: 20647212]
84. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature Rev. Genet 2004;5:101–113. [PubMed: 14735121]
85. Delbrück, M. Unités Biologiques Douées De Continuité Génétique. Editions du Centre National de la Recherche Scientifique Paris; 1949. p. 33-35.
86. Waddington, CH. The Strategy of the Genes. London: Geo. Allen & Unwin; 1957.
87. Monod J, Jacob F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harb. Symp. Quant. Biol 1961;26:389–401. [PubMed: 14475415]
88. Nurse P. The great ideas of biology. Clin. Med 2003;3:560–568. [PubMed: 14994716]
89. Wagner GP. The developmental genetics of homology. Nature Rev. Genet 2007;8:473–479. [PubMed: 17486120]
90. Arendt D. The evolution of cell types in animals: emerging principles from molecular studies. Nature Rev. Genet 2008;9:868–882. [PubMed: 18927580]
91. Lynch VJ, et al. Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. Proc. Natl Acad. Sci. USA 2008;105:14928–14933. [PubMed: 18809929]
92. Patel CJ, Bhattacharya J, Butte AJ. An environment-wide association study (EWAS) on type 2 diabetes mellitus. PLoS ONE 2010;5:e10746. [PubMed: 20505766]
93. Buckler ES, et al. The genetic architecture of maize flowering time. Science 2009;325:714–718. [PubMed: 19661422]
94. Schneeberger K, et al. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nature Methods 2009;6:550–551. [PubMed: 19644454]
95. Ehrenreich IM, et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 2010;464:1039–1042. [PubMed: 20393561]
96. Qin J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010;464:59–65. [PubMed: 20203603]
97. Dangl JL, Jones JD. Plant pathogens and integrated defence responses to infection. Nature 2001;411:826–833. [PubMed: 11459065]
98. Brady SM, et al. A high-resolution root spatiotemporal map reveals dominant expression patterns. Science 2007;318:801–806. [PubMed: 17975066]
99. Chickarmane V, et al. Computational morphodynamics: a modeling framework to understand plant growth. Annu. Rev. Plant Biol 2010;61:65–87. [PubMed: 20192756]
100. Guenther M, Young RA. Repressive transcription. Science 2010;329:150–151. [PubMed: 20616255]
101. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol 1961;3:318–356. [PubMed: 13718526]
102. Gurdon JB. Adult frogs derived from the nuclei of single somatic cells. Dev. Biol 1962;4:256–273. [PubMed: 13903027]

103. Tang F, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis. Cell Stem Cell 2010;6:468–478. [PubMed: 20452321]