nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

# THE FUTURE IS BRIGHT

**Reflections on the first ten years of the human genomics age**

GENOMICS

## THE END OF THE BEGINNING

*Eric Lander on the impact of the human genome sequence*

PAGE 187

METHODS

## MORE BASES PER DOLLAR

*Elaine Mardis on the march of sequencing technology*

PAGE 198

HEALTH

## FROM LAB TO CLINIC

*A road map to genomic medicine*

PAGE 204

NATURE.COM/NATURE

10 February 2011

Vol. 470, No. 7333

PERSPECTIVE

doi:10.1038/nature09764

# Charting a course for genomic medicine from base pairs to bedside

Eric D. Green[1], Mark S. Guyer[1] & National Human Genome Research Institute*

**There has been much progress in genomics in the ten years since a draft sequence of the human genome was published. Opportunities for understanding health and disease are now unprecedented, as advances in genomics are harnessed to obtain robust foundational knowledge about the structure and function of the human genome and about the genetic contributions to human health and disease. Here we articulate a 2011 vision for the future of genomics research and describe the path towards an era of genomic medicine.**

Since the end of the Human Genome Project (HGP) in 2003 and the publication of a reference human genome sequence[1,2], genomics has become a mainstay of biomedical research. The scientific community's foresight in launching this ambitious project[3] is evident in the broad range of scientific advances that the HGP has enabled, as shown in Fig. 1 (see rollfold). Optimism about the potential contributions of genomics for improving human health has been fuelled by new insights about cancer[4–7], the molecular basis of inherited diseases (http://www.ncbi.nlm.nih.gov/omim and http://www.genome.gov/GWAStudies) and the role of structural variation in disease[8], some of which have already led to new therapies[9–13]. Other advances have already changed medical practice (for example, microarrays are now used for clinical detection of genomic imbalances[14] and pharmacogenomic testing is routinely performed before administration of certain medications[15]). Together, these achievements (see accompanying paper[16]) document that genomics is contributing to a better understanding of human biology and to improving human health.

As it did eight years ago[17], the National Human Genome Research Institute (NHGRI) has engaged the scientific community (http://www.genome.gov/Planning) to reflect on the key attributes of genomics (Box 1) and explore future directions and challenges for the field. These discussions have led to an updated vision that focuses on understanding human biology and the diagnosis, prevention and treatment of human disease, including consideration of the implications of those advances for society (but these discussions, intentionally, did not address the role of genomics in agriculture, energy and other areas). Like the HGP, achieving this vision is broader than what any single organization or country can achieve—realizing the full benefits of genomics will be a global effort.

This 2011 vision for genomics is organized around five domains extending from basic research to health applications (Fig. 2). It reflects the view that, over time, the most effective way to improve human health is to understand normal biology (in this case, genome biology) as a basis for understanding disease biology, which then becomes the basis for improving health. At the same time, there are other connections among these domains. Genomics offers opportunities for improving health without a thorough understanding of disease (for example, cancer therapies can be selected based on genomic profiles that identify tumour subtypes[18,19]), and clinical discoveries can lead back to understanding disease or even basic biology.

The past decade has seen genomics contribute fundamental knowledge about biology and its perturbation in disease. Further deepening this understanding will accelerate the transition to genomic medicine (clinical care based on genomic information). But significant change rarely comes quickly. Although genomics has already begun to improve diagnostics and treatments in a few circumstances, profound improvements in the effectiveness of healthcare cannot realistically be expected for many years (Fig. 2). Achieving such progress will depend not only on research, but also on new policies, practices and other developments. We have illustrated the kinds of achievements that can be anticipated with a few examples (Box 2) where a confluence of need and opportunities should lead to major accomplishments in genomic medicine in the coming decade. Similarly, we note three cross-cutting areas that are broadly relevant and fundamental across the entire spectrum of genomics and genomic medicine: bioinformatics and computational biology (Box 3), education and training (Box 4), and genomics and society (Box 5).

## Understanding the biology of genomes

Substantial progress in understanding the structure of genomes has revealed much about the complexity of genome biology. Continued acquisition of basic knowledge about genome structure and function will be needed to illuminate further those complexities (Fig. 2). The contribution of genomics will include more comprehensive sets (catalogues) of data and new research tools, which will enhance the capabilities of all researchers to reveal fundamental principles of biology.

### Comprehensive catalogues of genomic data

Comprehensive genomic catalogues have been uniquely valuable and widely used. There is a compelling need to improve existing catalogues and to generate new ones, such as complete collections of genetic variation, functional genomic elements, RNAs, proteins, and other biological molecules, for both human and model organisms.

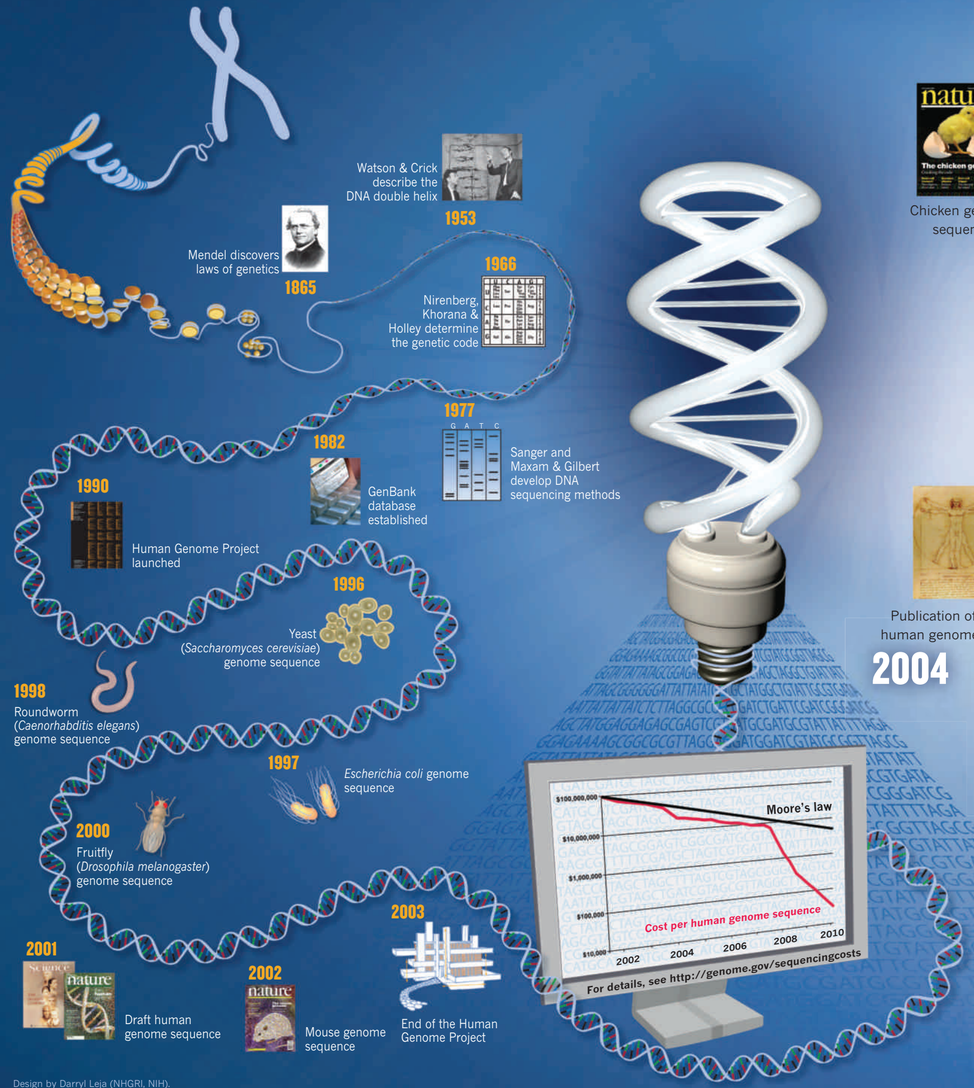
Genomic studies of the genes and pathways associated with disease-related traits require comprehensive catalogues of genetic variation, which provide both genetic markers for association studies and variants for identifying candidate genes. Developing a detailed catalogue of variation in the human genome has been an international effort that began with The SNP Consortium[20] and the International HapMap Project[21] (http://hapmap.ncbi.nlm.nih.gov), and is ongoing with the 1000 Genomes Project[22] (http://www.1000genomes.org).

Over the past decade, these catalogues have been critical in the discovery of the specific genes for roughly 3,000 Mendelian (monogenic) diseases

**Figure 1 | Genomic achievements since the Human Genome Project (see accompanying rollfold).** ▶

[1]National Human Genome Research Institute, National Institutes of Health, 31 Center Dr., Bethesda, Maryland 20892-2152, USA.
*Lists of participants and their affiliations appear at the end of the paper.

# Genomic achievements since the Human Genome Project





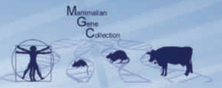
Chicken genome sequence


Phase I HapMap

NCBI's Database of Genotypes and Phenotypes (dbGaP) launched


Rhesus macaque genome sequence


First personal genome sequenced using new technologies



Completion of the Mammalian Gene Collection (MGC)

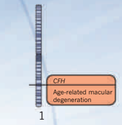

Southern African genome sequences


500th genome-wide association study published


Publication of finished human genome sequence

2004


Rat genome sequence



First genome-wide association study published


Honeybee genome sequence


Human genetic variation is breakthrough of the year


Yoruba genome sequence


International data release workshop

UK Biobank reaches 500,000 participants


Wellcome Trust Case Control Consortium publication


Platypus genome sequence

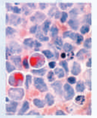

>1,000 mouse knockout mutations




Chimpanzee genome sequence


Sea urchin genome sequence


ENCODE pilot project complete


First cancer genome sequence (AML)


Bovine genome sequence

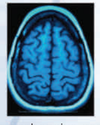

modENCODE publications

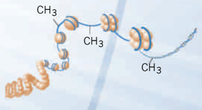

Dog genome sequence

2005


First direct-to-consumer whole-genome test

First personal genome sequenced


Comprehensive genomic analysis of glioblastoma



First human methylome map


Neanderthal genome sequence

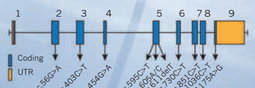
2006


Han Chinese genome sequence


1000 Genomes pilot project complete

2007


Genetic Information Nondiscrimination Act (GINA) passed in US



Miller syndrome gene discovered by exome sequencing

2008


Korean genome sequence


Nuffield Council on Bioethics publication on personalized healthcare

2009

2010

Design by Darryl Leja (NHGRI, NIH).
Watson and Crick photograph: A. Barrington Brown/Photo Researchers; images of *Science* covers courtesy of AAAS.

**BOX 1**

## The essence of genomics

Genomics grew primarily out of human genetics and molecular biology. Although the fields have much in common, genomics has several distinguishing characteristics.

**Comprehensiveness.** Genomics aims to generate complete data sets. Although relatively easy to define and measure for a genome sequence, attaining comprehensiveness can be more challenging for other targets (for example, functional genomic elements or the 'proteome').

**Scale.** Generation of comprehensive data sets requires large-scale efforts, demanding attention to: (1) organization, often involving large interdisciplinary consortia; (2) robust data standards, to ensure high-quality data and broad utility; and (3) computational intensity (see Box 3).

**Technology development.** Genomics demands high-throughput, low-cost data production, and requires that resources be devoted to technology development.

**Rapid data release.** Large data catalogues and analytical tools are community resources. This calls for policies that maximize rapid data release (harmonized internationally), while respecting the interests of the researchers generating the data and the human participants involved in that research[99,100].

**Social and ethical implications.** Genomics research and the many ways in which genomic data are used have numerous societal implications that demand careful attention (Box 5).

(http://www.ncbi.nlm.nih.gov/omim) and in establishing genetic associations between more than 900 genomic loci and complex (multigenic) traits, many of them diseases (http://www.genome.gov/GWAStudies). New genes and pathways have been implicated in disease, unexpected genetic connections among diseases have been identified, and the importance of non-coding variants in human disease has been highlighted. Together, these findings have accounted for a portion, but not all, of the heritability for many complex diseases[23]. Complete characterization of the genetics of complex diseases will require the identification of the full spectrum of human genomic variation in large, diverse sample sets.

Comprehensive catalogues of genetic variation in non-human species are similarly valuable. For example, understanding genetic variation in insect disease vectors may help inform the development of new strategies to prevent disease transmission, whereas knowledge about variation among microbial pathogens may lead to more robust vaccine-design strategies and novel therapeutics.

Catalogues of functional elements in the human genome, and the genomes of other species, are also being developed ('functional elements' include genes that encode proteins and non-coding RNAs; transcripts, including alternative versions; protein–nucleic-acid interaction sites; and epigenomic modifications). The ENCylopedia Of DNA Elements (ENCODE)[24] (http://genome.gov/encode) and modENCODE (http://genome.gov/modencode) projects are developing catalogues of functional elements in the human genome and in the genomes of *Caenorhabditis elegans*[25] and *Drosophila melanogaster*[26], respectively. But building a truly comprehensive catalogue of functional elements for any multicellular organism will require analysis of a large number of biological samples using many assays. Novel high-throughput, cost-effective technologies, and new reagents (see below), are needed to complete the human, fly and worm catalogues and compile catalogues of other genomes (for example, mouse and rat).

Biomedical research would benefit immensely from the availability of additional catalogues, for example of DNA modifications (epigenomics), gene products such as RNAs (transcriptomics) and proteins (proteomics), and indirect products of the genome such as metabolites (metabolomics) and carbohydrates (glycomics). Undertaking such large efforts will depend on both demand and the opportunity to cost-effectively assemble data sets of higher quality and greater comprehensiveness than would otherwise emerge from the combined output of individual research projects. Although the generation of some of these catalogues has already begun, major advances in technologies and data analysis methods are needed to generate, for example, truly comprehensive proteomic data sets and resources.

Additional insights will come from combining the information from different catalogues. For example, analysing genetic variation within functional elements will be particularly important for identifying such elements in non-coding regions of the genome. To this end, the GTEx (Genotype-Tissue Expression) project (http://www.commonfund.nih.gov/GTEx) has been established to map all sites in the human genome where sequence variation quantitatively affects gene expression.

### New tools for genomics research

Technology development has driven genomics. Both revolutionary (new methods, reagents and instruments) and evolutionary (incremental improvement in efficiency and output) technology development have been critical for achieving the remarkable increases in throughput and reductions in costs of DNA sequencing and other genomic methods. However, the inherent complexity of biology means that current technology is still not adequate for obtaining and interpreting the next generation of genomic data. Technological challenges include the design, synthesis and use of synthetic DNAs, and the measurement of cell- and organism-level phenotypes. Orders-of-magnitude improvements in throughput, cost-effectiveness, accuracy, sensitivity and selectivity of genomic technologies will require novel approaches[27,28].

Massively parallel DNA sequencing[29] has enabled a three-to-four orders-of-magnitude fall in the cost of genome sequencing (Fig. 1; see accompanying paper[30] and http://genome.gov/sequencingcosts). Nevertheless, sequencing a whole human genome remains much too expensive for most human disease studies, each of which can involve thousands or tens of thousands of individuals. Even in the case of well-understood coding regions (exons), sequencing errors complicate downstream analyses, and current sequencing error rates hinder reliable analysis of the remaining, poorly understood 98% of the genome. Perhaps most importantly, very low cost and extremely high accuracy will be critical for the routine clinical use of genome sequencing (for example, genetic screening of newborn babies[31,32]).

Structurally complex genomic regions, which are known to have a role in human disease[8], remain inherently difficult to sequence, even with the new DNA sequencing technologies. Additional technological improvements (for example, much longer read lengths) are needed to sequence such complex regions and to finish any specific region efficiently. Only with the ability to sequence entire genomes at very high accuracy, completeness and throughput will genome sequencing reach its full potential.

Some clinical applications (for example, rapid genomic analysis of tumours or microbiomes) may benefit from complete genomic sequencing in hours rather than weeks ('Making genomics-based diagnostics routine', Box 2). Although speed may be less important for research applications, it could have profound benefits in certain situations in the clinic. As genomics permeates clinical practice, point-of-care implementations will be needed, including in locations with minimal infrastructure. Separate technologies are likely to emerge for the research and clinical settings.

Analysis of functional genomic elements will require high-specificity affinity reagents (for example, antibodies or other tagging molecules) for all transcription factors, nucleic-acid-binding proteins, histone forms and chromatin modifications. These reagents must function well in a number of assays to be maximally useful. Several large-scale efforts to generate such reagents are under way[33,34] (see also http://www.commonfund.

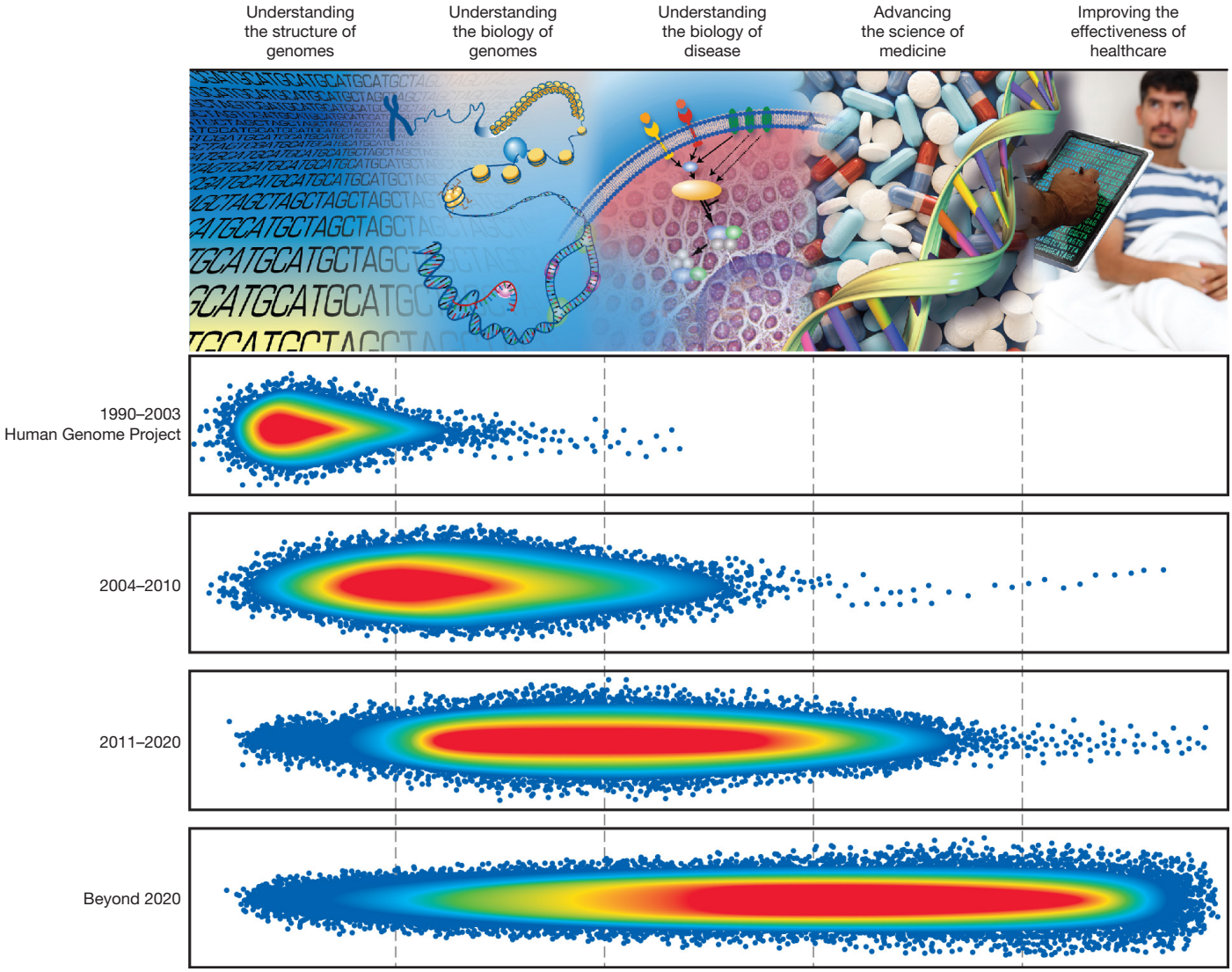



**Figure 2 | Schematic representation of accomplishments across five domains of genomics research.** The progression from base pairs to bedside is depicted in five sequential, overlapping domains (indicated along the top). Genomic accomplishments across the domains are portrayed by hypothetical, highly schematized density plots (each blue dot reflecting a single research accomplishment, with green, yellow and red areas reflecting sequentially higher densities of accomplishments). Separate plots are shown for four time intervals: the HGP; the period covered by the 2003 NHGRI vision for the future of genomics research[17]; the period described here (2011–2020); and the open-ended future beyond 2020.

nih.gov/proteincapture and http://antibodies.cancer.gov), but current approaches will probably not produce the full spectrum of reagents of the required specificity and utility. Suitable affinity reagents for larger-scale proteomic analyses pose an even greater challenge.

Most assays of functional genomic elements are currently limited by the need for a large number of cells, so many experiments are now performed with either tissue culture cells (which may not accurately reflect *in vivo* states) or heterogeneous tissue samples (in which sub-tissue-specific patterns may go undetected). Developing methods for producing accurate cell-specific profiles of single cells is a challenge. Analysing genomic data requires integration of multiple data types (Box 3). Robust analysis of promoters, for example, typically involves integration of data on transcription factor binding, protein complex formation, transcription start sites and DNase hypersensitive sites. Improved data integration approaches will require new algorithms and robust computational tools (Box 3).

The spatially and temporally dynamic nature of genomic regulation (see below) presents another formidable challenge to the comprehensive identification of all functional elements, as some critical regulatory processes only occur during brief developmental periods or in difficult-to-access tissues. New methods for *in situ* and real-time analysis will be necessary to understand fully the choreography of gene regulation.

Phenotypes arise from complex interactions among genes, cells, tissues, organs and the environment. Ultimately, the ability to co-analyse variation and phenotypic data will be critical for generating reliable inferences about disease-causing loci, genotype–phenotype correlations, and both gene–gene and gene–environment interactions. Therefore, better technologies for measuring phenotypes, behaviours, exposures and other environmental variables will be required.

## Understanding fundamental principles of biology

Comprehensive genomic catalogues are the 'parts lists', and just as such a list is not sufficient to understand how a machine functions, genomic catalogues are not sufficient to understand biological processes.

Recent advances have greatly improved our understanding of the importance of non-coding regions in the human genome for gene regulation, chromosome function and the generation of untranslated RNAs[35]. This is further supported by the finding that the great majority of trait-associated regions identified by genome-wide association studies fall in non-coding sequences[36]. New technologies, experimental strategies and

**BOX 2**

## Imperatives for genomic medicine

Opportunities for genomic medicine will come from simultaneously acquiring foundational knowledge of genome function, insights into disease biology and powerful genomic tools. The following imperatives will capitalize on these opportunities in the coming decade.

**Making genomics-based diagnostics routine.** Genomic technology development so far has been driven by the research market. In the next decade, technology advances could enable a clinician to acquire a complete genomic diagnostic panel (including genomic, epigenomic, transcriptomic and microbiomic analyses) as routinely as a blood chemistry panel.

**Defining the genetic components of disease.** All diseases involve a genetic component. Genome sequencing could be used to determine the genetic variation underlying the full spectrum of diseases, from rare Mendelian to common complex disorders, through the study of upwards of a million patients; efforts should begin now to organize the necessary sample collections.

**Comprehensive characterization of cancer genomes.** A comprehensive genomic view of all cancers[4–7] will reveal molecular taxonomies and altered pathways for each cancer subtype. Such information should lead to more robust diagnostic and therapeutic strategies and a roadmap for developing new treatments[74,75].

**Practical systems for clinical genomic informatics.** Thousands of genomic variants associated with disease risk and treatment response are known, and many more will be discovered. New models for capturing and displaying these variants and their phenotypic consequences should be developed and incorporated into practical systems that make information available to patients and their healthcare providers, so that they can interpret and reinterpret the data as knowledge evolves.

**The role of the human microbiome in health and disease.** Many diseases are influenced by the microbial communities that inhabit our bodies (the microbiome)[101]. Recent initiatives[102,103] (http://www.human-microbiome.org) are using new sequencing technologies to catalogue the resident microflora at distinct body sites, and studying correlations between specific diseases and the composition of the microbiome[104]. More extensive studies are needed to build on these first revelations and to investigate approaches for manipulating the microbiome as a new therapeutic approach.

computational approaches are needed to understand the functional role of non-coding sequences in health and disease.

Genomics will also contribute new technologies and resources to the analysis of gene-interaction networks. Network analysis will benefit from understanding the dynamics of gene expression, protein localization and modification, as well as protein–protein and protein–nucleic-acid associations. The ultimate challenge will be to decipher the ways that networked genes produce phenotypes. Genomics can contribute to solving this problem by providing data from systematic large-scale studies of gene expression that include determination of cellular responses to genetic changes, external perturbations and disease (see http://www.commonfund.nih.gov/LINCS). Here too, robust new computational tools are needed specifically to access, analyse and integrate large, complex data sets, and to develop predictive models, new visualization technologies and a 'knowledge base' of networks (Box 3).

Ultimately, human biology must be understood in the context of evolution. Comparative genomic studies have revealed the most highly conserved (and probably functional) portions of human, mammalian and vertebrate genomes[37]. Evolutionary relationships also underlie the use of model organisms in functional studies, and diverse data sets from unicellular organisms to mammals[38] will lead to key insights about genome function and biological pathways.

Despite their necessity, however, large-scale genomic studies alone will not be sufficient for gaining a fundamental understanding of biology. Most of the data analysis and interpretation will actually come from individual research efforts. Indeed, a primary motivation for the development of genomics (and other '-omics' disciplines) has been to generate data catalogues and technological tools that empower individual investigators to pursue more effective hypothesis-driven research.

## Understanding the biology of disease

All diseases are influenced by genetic variation (inherited and/or somatic), environmental agents and/or health behaviours, and there is increasing evidence for epigenomic contributions[39,40]. Using genomics is essential to understand both the normal and disease-related functions of the genetic and epigenetic[41] contributors to disease, and the cellular pathways and biological processes in which they are involved, an understanding that is critical to the development of improved strategies for diagnosis, prevention and therapeutic intervention.

The power of genomic approaches to elucidate the biology of disease is illustrated by the study of Crohn's disease. A decade ago, the mechanisms underlying this debilitating gastrointestinal disorder were opaque. Since then, genome-wide association studies have identified dozens of genomic regions harbouring genetic variants conferring risk for Crohn's disease[42]. Analyses of genes in these regions have revealed key, previously unappreciated roles in the disease for several physiological processes, including innate immunity, autophagy and interleukin (IL)-23R signalling[43,44]. Cellular models have been developed and used both to document the pathogenicity of specific mutations and to extend knowledge of the relevant biological pathways[42]. Chemical screens have been designed[42,45] to identify new candidate therapeutic agents. Furthermore, animal models have been developed that accurately model the effects of causal variants found in patients. In sum, the use of genomic approaches to identify risk-conferring variants has catalysed molecular, cell biological and animal model studies that have led to a better understanding of Crohn's disease and the development of novel therapies. This and other examples[11,46,47] justify the optimism about genomics' potential to accelerate the understanding of disease.

### Genetic and non-genetic bases of disease

Genomics will allow the compilation of rich catalogues spanning the full spectrum of germline variants (both common and rare) conferring risk for inherited disease ('Defining the genetic components of disease', Box 2). Catalogues of somatic mutations that contribute to all aspects of tumour biology for each major cancer type are under development[4,48] ('Comprehensive characterization of cancer genomes', Box 2; see also http://www.icgc.org and http://www.sanger.ac.uk/perl/genetics/CGP/cosmic). Effective partnerships among investigators with genomics expertise, those with in-depth knowledge of specific diseases, and patients will lead to the definition of the pathways from genetic variant to disease. Success will require improved definitions and measurements of phenotypes, new databases ('Practical systems for clinical genomic informatics', Box 2) and novel experimental strategies, such as studies of individuals associated with extremes of risk and phenotype and studies of risk-reducing variants (which may provide guides to new therapeutics and approaches to disease management).

Genome-wide association studies have implicated hundreds of non-coding genomic regions in the pathogenesis of complex diseases[49], creating a major challenge. Establishing disease causality of non-coding variants will be considerably more difficult than identifying causal variants in protein-coding sequences. In developing methods to characterize the functional landscape of non-coding DNA (discussed above), particular attention must be paid to establishing novel strategies for identifying non-coding variants that influence disease. Actually, disease research

**BOX 3**

## Bioinformatics and computational biology


The major bottleneck in genome sequencing is no longer data generation—the computational challenges around data analysis, display and integration are now rate limiting. New approaches and methods are required to meet these challenges.

**Data analysis.** Computational tools are quickly becoming inadequate for analysing the amount of genomic data that can now be generated, and this mismatch will worsen. Innovative approaches to analysis, involving close coupling with data production, are essential.

**Data integration.** Genomics projects increasingly produce disparate data types (for example, molecular, phenotypic, environmental and clinical), so computational approaches must not only keep pace with the volume of genomic data, but also their complexity. New integrative methods for analysis and for building predictive models are needed.

**Visualization.** In the past, visualizing genomic data involved indexing to the one-dimensional representation of a genome. New visualization tools will need to accommodate the multidimensional data from studies of molecular phenotypes in different cells and tissues, physiological states and developmental time. Such tools must also incorporate non-molecular data, such as phenotypes and environmental exposures. The new tools will need to accommodate the scale of the data to deliver information rapidly and efficiently.

**Computational tools and infrastructure.** Generally applicable tools are needed in the form of robust, well-engineered software that meets the distinct needs of genomic and non-genomic scientists. Adequate computational infrastructure is also needed, including sufficient storage and processing capacity to accommodate and analyse large, complex data sets (including metadata) deposited in stable and accessible repositories, and to provide consolidated views of many data types, all within a framework that addresses privacy concerns. Ideally, multiple solutions should be developed[105].

**Training.** Meeting the computational challenges for genomics requires scientists with expertise in biology as well as in informatics, computer science, mathematics, statistics and/or engineering. A new generation of investigators who are proficient in two or more of these fields must be trained and supported.

may have a leading role in illuminating the fundamental biology of non-coding sequence variation and its phenotypic implications.

A full understanding of disease will require capturing much of the genetic variation across the human population[50]. Accomplishing this will involve collaborations with relevant communities, taking into account how genomics is understood and perceived by different racial, ethnic and cultural groups, to form effective partnerships that will ensure that such research is sound and ethically conducted. Given the history of incidents leading to misunderstanding and mistrust[51], this is an area ripe for innovative approaches.

A complete understanding of disease also requires the annotation and correlation of genomic information with high-quality phenotypic data. Obtaining phenotypic data that are both thorough and accurate enough to be analysed in conjunction with high-quality genomic and environmental data requires meticulous application of phenotyping methods, improved definitions of phenotypes, new technologies, and the consistent use of data standards[52] (http://www.phenx.org). To interrogate this information effectively, widely accessible databases containing extensive phenotypic information linked to genome sequence data (genotype) are needed[53]. Such efforts will benefit greatly from the linkage of genomic information to data gathered in the course of actual clinical care, such as in electronic medical/health records. Research is needed to help formulate evidence-based solutions for the complex ethical, legal and regulatory challenges associated with generating and using such linkages.

The integration of genomic information and environmental exposure data can help to understand the links between biological factors and extrinsic triggers, providing a much fuller understanding of disease aetiology. Obtaining such integrated data sets can be immeasurably aided by large-scale prospective cohort studies, which allow robust analyses of genetic and environmental risks across the human lifespan, but present unique challenges in scale-up and implementation[54]. Several such cohort studies have been initiated (http://www.p3g.org/secretariat/memb.shtml and http://www.nationalchildrensstudy.org) or proposed[55].

Studies of non-human organisms can help to characterize disease-implicated variants and understand their biology, providing valuable insights about health and disease. Genomics has enhanced the utility of both widely used models (for example, yeast, fruitflies, worms, zebrafish, mice and rats) and less commonly used organisms that provide good models for human disease (for example, the ferret for studying influenza, the armadillo for leprosy, and the prairie vole for social behaviour, including autism). New animal models developed on the basis of genomic insights are enormously valuable and should be made broadly available. A particularly interesting application of genomics involves microbes. The biological relevance of human–microbe interactions is both obvious (in infectious diseases) and relatively unexplored (in the maintenance of human health). Advances in DNA sequencing technologies and new approaches for data analysis have contributed to the emergence of metagenomics, which offers unprecedented opportunities for understanding the role of endogenous microbes and microbial communities in human health and disease ('The role of the human microbiome in health and disease', Box 2).

### Human participants in genomics research

Effective genomics research needs continual, broad and representative public participation, and depends on developing trust and informed partnerships between researchers and different segments of society. In both genomics research and medicine, it is particularly important to recognize the need for balance among a range of competing considerations (Box 5). And as in all biomedical research, it is imperative to recognize and respect the distinctions between research and clinical care.

The oversight system for human subjects' protection is based on principles related to identifiability, risk-benefit assessment, equitable selection of participants and considerations of informed consent. However, genomics research can sometimes challenge our ability to apply these principles. For example, existing definitions of identifiability are problematic because even modest amounts of genomic sequence are potentially identifying and refractory to anonymization. Other types of genomic information (for example, transcript and microbiome profiles) may also be identifying. In addition, concepts of genomic privacy vary among individuals and cultures.

Genomics research challenges standard approaches to informed consent because it is necessary to design consent language that fully accounts for the broad utility that genomic data can offer beyond the immediate study. Such challenges are magnified in large studies that involve many thousands of participants. Studies that use archived samples pose distinct problems because such samples were often collected using consent processes that did not anticipate the potential identifiability of genomic data or the value of broad, long-term data sharing.

In consideration of the unique and potentially sensitive nature of genomic information, the framework for oversight of genomics research involving human subjects should be re-examined to ensure appropriate protections of all participants. Although legal protections to prevent inappropriate use of genetic information have been developed in some countries[56,57], best practices for informed consent processes and improved policies on the use of existing samples and data are needed[58].

**BOX 4**

## Education and training

Realizing the benefits of genomics will require an educated public who can understand the implications of genomics for their healthcare and evaluate the relevant public policy issues. Clinical professionals will need to be trained to work within interdisciplinary teams. The development of effective education and training efforts will require that diverse communities be engaged, so that all can appropriately benefit.

**Strengthening primary and secondary education.** If general science literacy is to improve, including an understanding of probability and risk that is relevant to genomic medicine, biological sciences curricula during primary and secondary education need to change. This, in turn, requires improvements in the training of science educators.

**Conducting public outreach.** Education programmes are needed to promote lifelong public understanding and awareness of the role of genomics in human health and other areas.

**Building healthcare providers' genomic competencies.** All healthcare providers must acquire competency in genomics to provide services appropriate for their scope of practice. Genomics needs to be better integrated into the curricula of healthcare professional education programmes, as well as their licensing and accrediting processes.

**Preparing the next generation of genomics researchers.** Many disciplines beyond bioinformatics/computational biology and medicine, including mathematics, public health, engineering and the humanities, have relevance to genomics and its uptake. The number of trainees acquiring expertise in both genomics and one or more related fields must increase. The diversity of the genomics workforce must also expand.

Another acute challenge arises from the fact that genomics research inevitably reveals information about participants' risk factors or disease status for disorders and traits not being directly studied (so-called incidental findings). Additional research and policies are needed to guide decisions about whether, when, and how to return individual research findings (especially incidental findings) to research participants[59–61]. Guidance is also needed to account for the likelihood that the interpretation of genomic information will evolve over time.

Identifiability, privacy, informed consent and return of results are not the only issues pertaining to research participants that are raised by genomics. Research is also needed to understand issues related to ownership of samples and data, data access and use, intellectual property, and benefit sharing, among others.

## Advancing the science of medicine

The science of medicine and the practice of medicine (that is, the provision of healthcare) are distinct domains. Our burgeoning knowledge of the human genome is beginning to transform the former, and there are already examples where genomic information is now part of the standard of care[62–64]. Genomic discoveries will increasingly advance the science of medicine in the coming decades (Fig. 2), as important advances are made in developing improved diagnostics, more effective therapeutic strategies, an evidence-based approach for demonstrating clinical efficacy, and better decision-making tools for patients and providers. Realistically, however, a substantial amount of research is usually needed to bring a genomic discovery to the bedside, as initial findings indicating potential benefits must be followed by clinical studies to demonstrate efficacy and effectiveness[65].

### Diagnostics

Over the next decade, the variant genes responsible for most Mendelian disorders will be identified and, for some number, such knowledge will lead to the development of practical treatments. A more immediate benefit will be an accurate diagnosis that, even in the absence of a treatment, can be clinically valuable. A rapid, accurate diagnosis cuts short the 'diagnostic odyssey' that often involves many false leads and ineffective treatments, can reduce healthcare costs, and provide psychological benefit to patients and families.

Beyond Mendelian disorders, a major benefit of genomic (and other '-omic') information will come from accurate subclassification of diseases. As shown for breast cancer[19], understanding the 'molecular taxonomy' of a disease can help distinguish different conditions that have common pathophysiological or morphological features, yet respond to different treatments.

### Therapeutics

Genomic information can be used in many ways for developing improved therapeutics. The following discussion focuses on pharmaceuticals, where genomic information can inform target identification, rational drug design, genomics-based stratification in clinical trials, higher efficacy and fewer adverse events from genotype-guided drug prescription (pharmacogenomics), as well as guide the development of gene therapy strategies. Genomic information will also inform therapeutic approaches based on dietary, behavioural and lifestyle interventions, modification of environmental exposures, and other population-based or societal interventions that have genotype-specific effects[66–68].

The systematic development of a pharmaceutical requires the discovery and validation of a disease-relevant target in the relevant cells. Traditionally, targets have been identified biochemically, one at a time. The more thorough understanding of disease potentiated by genomics will bring extraordinary opportunities for identifying new targets for drug development[69]. Using detailed information about a disease, candidate therapeutic agents (for example, small molecules, antibodies and other proteins, and small interfering RNAs) can be identified by high-throughput screening methodologies or developed by molecular design technologies. It must be noted, however, that many of the subsequent steps in drug development (for example, medicinal chemistry, pharmacokinetics and formulation) do not involve genomics, and cannot be expected to be improved by it. The development of new pharmaceuticals based on genomic knowledge of specific targets and their role in disease has already been markedly successful[70–73], and is becoming increasingly commonplace, particularly for cancer drug development[74,75].

At the same time, understanding the underlying disease biology based on genomic information does not guarantee new therapeutics. For example, although some human disease genes (such as those for sickle cell anaemia, Huntington's disease, and cystic fibrosis) were identified more than two decades ago, the development of suitable therapies for these disorders has been much slower than anticipated. Although there have been recent promising developments[13,76], success is by no means certain in all cases.

Another significant opportunity offered by genomics is improved design of clinical trials[77]. Currently, many clinical trials treat the tested population as genetically homogeneous. But stratification of trial participants using genomic information can allow the use of smaller numbers of participants and increase statistical power for establishing effectiveness and reducing morbidity. An example is gefitinib, for which survival benefit was only documented by analysis in a genomically selected population[78]. Genomics should also allow the identification of individuals genetically susceptible to adverse reactions[79]. Correlation of genomic signatures with therapeutic response will enable the targeting of appropriate patients at appropriate stages of their illness in clinical trials, resulting in more effective drugs as well as better dosing and monitoring. It will also significantly affect the information provided to prospective research participants regarding the potential for medical benefit directly related to trial participation, a topic of intense controversy for early-phase clinical trials[80].

**BOX 5**

## Genomics and society

Effectively examining the societal implications of genomic advances requires collaborations involving individuals with expertise in genomics and clinical medicine and experts in bioethics, psychology, sociology, anthropology, history, philosophy, law, economics, health services research and related disciplines.

**Psychosocial and ethical issues in genomics research.** These include ensuring appropriate protection of human research participants and addressing the perceptions of risks and benefits of participating in genomic studies; expanding the diversity of research cohorts; incorporating biological ancestry markers and self-identified race and ethnicity as variables in genomic studies; accomplishing effective community engagement; and including vulnerable populations (for example, children and the disabled) and deceased individuals in genomics research.

**Psychosocial and ethical issues in genomic medicine.** These include communicating with patients about the uncertainty and evolving nature of predictions based on genomic information; interpreting information from direct-to-consumer genetic tests; ensuring fair access to genomic medicine; assessing the effectiveness of genomically informed diagnostics and therapeutics; using genomic information to improve behaviour change interventions; addressing issues associated with pre-implantation, prenatal and postnatal genetic diagnoses; and determining how constructs of race and ethnicity relate to the biology of disease and the potential to advance genomic medicine.

**Legal and public policy issues.** These include intellectual property in genomics; insurance reimbursement for genomic services; regulation of genetic testing; regulatory and non-regulatory approaches for dealing with direct-to-consumer genetic testing; the regulation of pharmacogenomics and genomics-based therapeutics; protection against genetic discrimination and stigmatization; and uses of genomics in non-medical settings.

**Broader societal issues.** These include the implications of increasing genomic knowledge for conceptualizing health and disease; for understanding identity at the individual and group levels, including race and ethnicity; for gaining insights about human origins; and for considering genetic determinism, free will and individual responsibility.

Pharmacogenomics is another direct clinical application of genomic medicine. Genetically guided prescription of the antiretroviral drug abacavir is now the standard of care for HIV-infected patients[81], and it is likely that the use of tamoxifen[82], clopidogrel[83] and possibly warfarin[84] will soon benefit from genetic considerations. Realistically, however, pharmacogenomics will not be useful for all drugs, such as those for which metabolism is not affected by genetic variation or for which there are redundant metabolic pathways. As in any other area of medicine, actual patient benefit must be demonstrated before routine clinical use of a pharmacogenomic test[65].

### An evidence base for genomic medicine

The effectiveness of genomic information in tailoring interventions and, ultimately, improving health outcomes must be demonstrated. Genomically informed interventions (for example, pharmacogenomic tests or the use of genomics-based information to change risk behaviour) must be evaluated with a portfolio of research approaches, including retrospective analyses, prospective studies, clinical trials and comparative effectiveness studies, to evaluate their impact on decision making, health outcomes and cost. This will also help to avoid harm to patients or the wasting of time and resources[68]. However, although a substantial evidence base before clinical introduction is ideal, there can be costs in delaying the implementation of useful genomics-based strategies. In some situations, genomic information may provide opportunities to develop and use innovative clinical trial designs that lead to provisional approval with continued study. Informed and nuanced policies for healthcare payer coverage could also facilitate provisional implementation while definitive data are accrued.

### Genomic information and the reduction of health disparities

Most documented causes of health disparities are not genetic, but are due to poor living conditions and limited access to healthcare. The field of genomics has been appropriately cautioned not to overemphasize genetics as a major explanatory factor in health disparities[85]. However, genomics research may still have a role in informing the understanding of population differences in disease distribution, treatment response and the influence of gene–environment interaction and epigenomics on disease and health[86,87]. For example, a few genetic variants can be correlated with population differences associated with an increased risk for several diseases with documented prevalence disparities, such as prostate cancer[88] and kidney disease[89]. Although the results of most genomic studies will apply broadly, it is important to identify any specific genetic factors that may be associated with disparate disease risk, incidence, or severity among population groups.

Barriers to obtaining the benefits of genomics need to be identified and addressed. It will be important to recognize and understand how genomics researchers and research participants conceptualize and characterize human groups and whether or how such categorizations shape research outcomes. Many group-based social identities, most notably those reflecting race, ethnicity and nationality, include ancestry and morphology as bases of categorization[90]. When analysing phenotypic data, innovative approaches will be needed to tease apart the many confounders that co-vary with social identity. Progress in parsing the interactions among multiple genetic, environmental and social factors promises to provide more accurate predictions of disease risk and treatment response. Most importantly, as genomics continues to be applied in global healthcare settings, it must not be mistakenly used to divert attention and resources from the many non-genetic factors that contribute to health disparities, which would paradoxically exacerbate the problem.

### Delivering genomic information to patients

The routine use of genomics for disease prevention, diagnosis and treatment will require a better understanding of how individuals and their healthcare providers assimilate and use such information. The amount and heterogeneous nature of the data, which will include both expected and unexpected results, will antiquate current mechanisms for delivering medical information to patients.

Healthcare professionals will need to be able to interpret genomic data, including those from direct-to-consumer services, that are relevant to their scope of practice and to convey genetic risk to their patients. Patients will need to be able to understand the information being provided to them and to use that information to make decisions. Implementation research will help define the best ways to convey the uncertainties and complexities of genomics-based risk information to individuals and their families, how such information is understood, and how it influences health-related behaviour. Principles should be developed for guiding decisions about acquiring genomic information. These principles will have to balance the potential benefits of new preventive measures and therapeutics with economic impact and the potential for harm.

Achieving effective information flow will require an understanding of the issues related to achieving genomic medicine literacy by healthcare providers and consumers (Box 4) and the influence of genomic information on an array of health behaviours[68]. Additional research should

investigate the impact of various factors (for example, family history and underlying motivations) on patients' ability to reduce their risk. Here too, evidence-based best practices are needed to ensure that patients have adequate information, access to appropriate healthcare services, and suitable follow-up to help them use their genomic information. These best practices should also inform the development and implementation of evidence-driven regulatory policies that enhance the public benefit of genomics, but at the same time protect the public from inaccurate claims and the dissemination of unreliable information.

Additional challenges will arise as genomics becomes part of global medicine. Strategies that take into account differences in healthcare practices and systems will be required to realize the potential of genomics to prevent and treat disease around the world.

## Improving the effectiveness of healthcare

Clinical deployment of genomics has already begun in a small number of cases; widespread implementation, however, will take many years (Fig. 2) and must be an iterative process that continually incorporates new findings. To obtain the healthcare benefits of genomics, various important issues need to be considered.

### Electronic medical/health records

Viable electronic medical/health records systems capable of handling family history and genomic data are required to fully utilize genomic information for patient care. Existing clinical informatics architectures are largely incapable of storing genome sequence data in a way that allows the information to be searched, annotated and shared across healthcare systems over an individual's lifespan. Innovative approaches are needed to assimilate a patient's genomic information[91,92], as are user-friendly systems that permit retrieval and queries by healthcare providers[93]. There are intensive efforts to create new technologies and systems that bring the electronic medical/health record into routine use[94]. The value of such records for genomics research has been demonstrated[95,96]. In developing these systems, close attention to the ethical, legal and regulatory complexities is essential. Public concern about health information privacy is already widespread. Although the concern may be greater for genomic information, it is inherent to medical information and can be addressed[97] through the interaction of genomics experts with the medical informatics and policy communities.

### Demonstrating effectiveness

Demonstrating utility will be critical for the widespread adoption of genomic medicine, including reimbursement for services. The thresholds for evidence of benefit and harm vary across stakeholders, and defining robust metrics for measuring utility is an important research objective. Such studies will need to assess patient outcomes (including morbidity and mortality or, minimally, widely accepted surrogate health markers).

The effective uptake of genomic medicine will require productive interactions with the regulatory systems in each country. Addressing these and other rapidly emerging issues will require sustained, yet agile, collaborative efforts by the research, regulatory and healthcare communities, as well as new research models that involve rapid iterative cycles. Rather than using traditional clinical trials, such an approach could involve practice-based interventions spanning the range of clinical, patient-reported and economic outcomes measured at the level of individuals, practices and systems.

### Educating healthcare professionals, patients and the public

Education at many levels will be critical for the successful introduction of genomics into healthcare (Box 4). Genomics-based healthcare is no different from standard healthcare in being a combined responsibility of the patient and medical professionals, and all must be well informed. As genomics moves into routine clinical practice, innovative methods will be needed to provide healthcare practitioners with the ability to interpret genomic data and make evidence-based recommendations. Research is needed to establish appropriate competencies and on making the necessary educational opportunities available to all healthcare providers effectively, appropriately, and in culturally and linguistically relevant ways across diverse patient populations. Point-of-care clinical decision-support processes are also required. The challenge will be to develop models that can be implemented at the time, place and knowledge level needed to provide effective care.

Equally important is a well-informed public that is supportive of genomics research and appreciates the value of research participation. Consumers will need tools to assess the promises and claims of genomic testing services. Development and implementation of appropriate healthcare policies will depend on educated policy makers. Research is needed to determine the knowledge necessary for making genomically informed clinical decisions at both the individual and societal levels. A variety of pilot efforts should be developed, tested and assessed for their effectiveness in engendering genomically (and more broadly, scientifically and statistically) literate healthcare providers, patients and the general public.

### Increasing access to genomic medicine

Genomics will only achieve its full potential to improve health when the advances it engenders become accessible to all. The development of novel and effective mechanisms for involving diverse stakeholder groups is needed to maximize the relevance of genomics to different healthcare systems.

Many existing healthcare infrastructures are poorly suited for the delivery of genomic medicine to all segments of the population. Optimal models for ensuring that the best practices in genomic medicine become available to all at-risk patient populations have yet to be defined. Some possibilities for new approaches include reliance on non-geneticist healthcare providers guided by informatics support, increased use of telemedicine and enhanced genomics education for future generations of healthcare providers. All of these must be pursued.

## Concluding comments

The discussions in the 1980s that led to the HGP were motivated by a vision that knowing the human genome sequence would be extraordinarily useful for understanding human biology and disease. For example, Dulbecco wrote[98] in 1986 that "If we wish to learn more about cancer, we must now concentrate on the cellular genome," and he advocated sequencing "the whole genome of a selected animal species," specifically, the human genome. In 1988, a US National Research Council (NRC) report[3] articulated a bold plan for an effort that would culminate in sequencing the human genome; the report stated that such a "project would greatly increase our understanding of human biology and allow rapid progress to occur in the diagnosis and ultimate control of many human diseases." In the past quarter-century, the prescience of this audacious vision has been confirmed. Progress in genomics has been monumental. Although staggering challenges remain, the fundamental goals have not changed—genomics and related large-scale biological studies will, in ways not previously available, lead to a profound understanding about the biology of genomes and disease, to unimagined advances in medical science, and to powerful new ways for improving human health.

Achieving these goals will continue to rely on new technologies, large-scale collaborative efforts, multidisciplinary and international teams, comprehensiveness, high-throughput data production and analysis, computational intensity, high standards for data quality, rapid data release, and attention to societal implications. The perfusion of genomics into other areas of biomedical research will enable these disciplines to make advances far beyond what is possible today. Achieving such a pervasive positive influence on biomedicine is one of the most gratifying aspects of genomics, as anticipated by the NRC report's detailed 'call to action' blueprint for the HGP[3]. It is thus with a continuing sense of wonder, a continuing need for urgency, a continuing desire to balance ambition with reality, and a continuing responsibility to protect individuals while maximizing the societal benefits of genomics that we have discussed here some of the many compelling opportunities and significant challenges for the next decade of genomics research. This new vision is ambitious and far-reaching, both in scope and timing. It goes

well beyond what any one organization can realistically support, and will (once again) require the creative energies and expertise of genome scientists around the world and from all sectors, including academic, government and commercial.

Successfully navigating a course from the base pairs of the human genome sequence to the bedside of patients seems within reach, would usher in an era of genomic medicine, would fulfil the promise originally envisioned for the HGP and, most importantly, would benefit all humankind.


1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
3. National Research Council. *Mapping and Sequencing the Human Genome* (National Academy Press, 1988).
4. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
5. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456,** 66–72 (2008).
6. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463,** 893–898 (2010).
7. Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nature Rev. Genet.* **11,** 685–696 (2010).
8. Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61,** 437–455 (2010).
9. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from ''never smokers'' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl Acad. Sci. USA* **101,** 13306–13311 (2004).
10. Yan, H. *et al. IDH1* and *IDH2* mutations in gliomas. *N. Engl. J. Med.* **360,** 765–773 (2009).
11. Brooke, B. S. *et al.* Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. *N. Engl. J. Med.* **358,** 2787–2795 (2008).
12. D'Hulst, C. & Kooy, R. F. Fragile X syndrome: from molecular genetics to therapy. *J. Med. Genet.* **46,** 577–584 (2009).
13. Dietz, H. C. New therapeutic approaches to mendelian disorders. *N. Engl. J. Med.* **363,** 852–863 (2010).
14. Miller, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86,** 749–764 (2010).
15. Frueh, F. W. *et al.* Pharmacogenomic biomarker information in drug labels approved by the United States Food and Drug Administration: prevalence of related drug use. *Pharmacotherapy* **28,** 992–998 (2008).
16. Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* doi:10.1038/nature09792 (this issue).
17. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422,** 835–847 (2003).
18. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009 (2002).
19. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351,** 2817–2826 (2004).
20. Thorisson, G. A. & Stein, L. D. The SNP Consortium website: past, present and future. *Nucleic Acids Res.* **31,** 124–127 (2003).
21. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467,** 52–58 (2010).
22. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
23. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461,** 747–753 (2009).
24. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
25. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330,** 1775–1787 (2010).
26. The modENCODE Consortium. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330,** 1787–1797 (2010).
27. Jain, K. K. Applications of nanobiotechnology in clinical diagnostics. *Clin. Chem.* **53,** 2002–2009 (2007).
28. Oita, I. *et al.* Microfluidics in macro-biomolecules analysis: macro inside in a nano world. *Anal. Bioanal. Chem.* **398,** 239–264 (2010).
29. Metzker, M. L. Sequencing technologies—the next generation. *Nature Rev. Genet.* **11,** 31–46 (2010).
30. Mardis, E. R. A decade's perspecitve on DNA sequencing technology. *Nature* doi:10.1038/nature09796 (this issue).
31. Hiraki, S. & Green, N. S. Newborn screening for treatable genetic conditions: past, present and future. *Obstet. Gynecol. Clin. North Am.* **37,** 11–21 (2010).
32. Levy, P. A. An overview of newborn screening. *J. Dev. Behav. Pediatr.* **31,** 622–631 (2010).
33. Berglund, L. *et al.* A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics* **7,** 2019–2027 (2008).
34. Bjorling, E. & Uhlen, M. Antibodypedia, a portal for sharing antibody and antigen validation data. *Mol. Cell. Proteomics* **7,** 2028–2037 (2008).
35. Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M. B. Annotating non-coding regions of the genome. *Nature Rev. Genet.* **11,** 559–571 (2010).
36. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
37. Margulies, E. H. *et al.* Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res.* **17,** 760–774 (2007).
38. Austin, C. P. *et al.* The Knockout Mouse Project. *Nature Genet.* **36,** 921–924 (2004).
39. Feinberg, A. P. Epigenomics reveals a functional genome anatomy and a new approach to common disease. *Nature Biotechnol.* **28,** 1049–1052 (2010).
40. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nature Biotechnol.* **28,** 1057–1068 (2010).
41. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnol.* **28,** 1045–1048 (2010).
42. Rioux, J. D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genet.* **39,** 596–604 (2007).
43. Van Limbergen, J., Wilson, D. C. & Satsangi, J. The genetics of Crohn's disease. *Annu. Rev. Genomics Hum. Genet.* **10,** 89–116 (2009).
44. Brest, P. *et al.* Autophagy and Crohn's disease: at the crossroads of infection, inflammation, immunity, and cancer. *Curr. Mol. Med.* **10,** 486–502 (2010).
45. Cadwell, K. *et al.* A key role for autophagy and the autophagy gene *Atg16l1* in mouse and human intestinal Paneth cells. *Nature* **456,** 259–263 (2008).
46. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308,** 385–389 (2005).
47. Smushkin, G. & Vella, A. Genetics of type 2 diabetes. *Curr. Opin. Clin. Nutr. Metab. Care* **13,** 471–477 (2010).
48. The International Cancer Genome Consortium. International network of cancer genome projects. *Nature* **464,** 993–998 (2010).
49. Manolio, T. A. Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363,** 166–176 (2010).
50. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* **25,** 489–494 (2009).
51. Couzin-Frankel, J. Ethics. DNA returned to tribe, raising questions about consent. *Science* **328,** 558 (2010).
52. Stover, P. J., Harlan, W. R., Hammond, J. A., Hendershot, T. & Hamilton, C. M. PhenX: a toolkit for interdisciplinary genetics research. *Curr. Opin. Lipidol.* **21,** 136–140 (2010).
53. Mailman, M. D. *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nature Genet.* **39,** 1181–1186 (2007).
54. Manolio, T. A. & Collins, R. Enhancing the feasibility of large cohort studies. *J. Am. Med. Assoc.* **304,** 2290–2291 (2010).
55. Collins, F. S. The case for a US prospective cohort study of genes and environment. *Nature* **429,** 475–477 (2004).
56. Gulcher, J. R. & Stefansson, K. The Icelandic Healthcare Database and informed consent. *N. Engl. J. Med.* **342,** 1827–1830 (2000).
57. Hudson, K. L., Holohan, M. K. & Collins, F. S. Keeping pace with the times—the Genetic Information Nondiscrimination Act of 2008. *N. Engl. J. Med.* **358,** 2661–2663 (2008).
58. Ludman, E. J. *et al.* Glad you asked: participants' opinions of re-consent for dbGap data submission. *J. Empir. Res. Hum. Res. Ethics* **5,** 9–16 (2010).
59. McGuire, A. L., Caulfield, T. & Cho, M. K. Research ethics and the challenge of whole-genome sequencing. *Nature Rev. Genet.* **9,** 152–156 (2008).
60. Wolf, S. M. *et al.* Managing incidental findings in human subjects research: analysis and recommendations. *J. Law Med. Ethics* **36,** 219–248 (2008).
61. Kohane, I. S. & Taylor, P. L. Multidimensional results reporting to participants in genomic studies: getting it right. *Sci. Transl. Med.* **2,** 37cm19 (2010).
62. Harris, L. *et al.* American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J. Clin. Oncol.* **25,** 5287–5312 (2007).
63. Allegra, C. J. *et al.* American Society of Clinical Oncology provisional clinical opinion: testing for *KRAS* gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J. Clin. Oncol.* **27,** 2091–2096 (2009).
64. Panel on Antiretroviral Guidelines for Adults and Adolescents. *Guidelines for the Use of Antiretroviral Agents in HIV-1 Infected Adults and Adolescents* 1–161 ⟨http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf⟩ (Department of Health and Human Services, 2009).
65. Burke, W., Laberge, A. M. & Press, N. Debating clinical utility. *Public Health Genomics* **13,** 215–223 (2010).
66. Shiwaku, K. *et al.* Difficulty in losing weight by behavioral intervention for women with Trp64Arg polymorphism of the $\beta_3$-adrenergic receptor gene. *Int. J. Obes. Relat. Metab. Disord.* **27,** 1028–1036 (2003).
67. Feldstein Ewing, S. W., LaChance, H. A., Bryan, A. & Hutchison, K. E. Do genetic and individual risk factors moderate the efficacy of motivational enhancement therapy? Drinking outcomes with an emerging adult sample. *Addict. Biol.* **14,** 356–365 (2009).
68. McBride, C. M. *et al.* Future health applications of genomics: priorities for communication, behavioral, and social sciences research. *Am. J. Prev. Med.* **38,** 556–565 (2010).
69. Chan, J. N., Nislow, C. & Emili, A. Recent advances and method development for drug target identification. *Trends Pharmacol. Sci.* **31,** 82–88 (2010).
70. Druker, B. J. *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **344,** 1031–1037 (2001).

71. Druker, B. J. *et al.* Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N. Engl. J. Med.* **344,** 1038–1042 (2001).
72. Brufsky, A. Trastuzumab-based therapy for patients with HER2-positive breast cancer: from early scientific development to foundation of care. *Am. J. Clin. Oncol.* **33,** 186–195 (2010).
73. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363,** 809–819 (2010).
74. Sawyers, C. Targeted cancer therapy. *Nature* **432,** 294–297 (2004).
75. Stuart, D. & Sellers, W. R. Linking somatic genetic alterations in cancer to therapeutics. *Curr. Opin. Cell Biol.* **21,** 304–310 (2009).
76. Grasemann, H. & Ratjen, F. Emerging therapies for cystic fibrosis lung disease. *Expert Opin. Emerg. Drugs* **15,** 653–659 (2010).
77. Simon, R. The use of genomics in clinical trial design. *Clin. Cancer Res.* **14,** 5984–5993 (2008).
78. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N. Engl. J. Med.* **361,** 947–957 (2009).
79. Caskey, C. T. Using genetic diagnosis to determine individual therapeutic utility. *Annu. Rev. Med.* **61,** 1–15 (2010).
80. Henderson, G. E. *et al.* Clinical trials and medical care: defining the therapeutic misconception. *PLoS Med.* **4,** e324 (2007).
81. Tozzi, V. Pharmacogenetics of antiretrovirals. *Antiviral Res.* **85,** 190–200 (2010).
82. Goetz, M. P. *et al.* The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res. Treat.* **101,** 113–121 (2007).
83. Angiolillo, D. J. *et al.* Contribution of gene sequence variations of the hepatic cytochrome P450 3A4 enzyme to variability in individual responsiveness to clopidogrel. *Arterioscler. Thromb. Vasc. Biol.* **26,** 1895–1900 (2006).
84. Rieder, M. J. *et al.* Effect of *VKORC1* haplotypes on transcriptional regulation and warfarin dose. *N. Engl. J. Med.* **352,** 2285–2293 (2005).
85. Sankar, P. *et al.* Genetic research and health disparities. *J. Am. Med. Assoc.* **291,** 2985–2989 (2004).
86. Ramos, E. & Rotimi, C. The A's, G's, C's, and T's of health disparities. *BMC Med. Genomics* **2,** 29 (2009).
87. Rotimi, C. N. & Jorde, L. B. Ancestry and disease in the age of genomic medicine. *N. Engl. J. Med.* **363,** 1551–1558 (2010).
88. Freedman, M. L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl Acad. Sci. USA* **103,** 14068–14073 (2006).
89. Kao, W. H. *et al. MYH9* is associated with nondiabetic end-stage renal disease in African Americans. *Nature Genet.* **40,** 1185–1192 (2008).
90. Race, Ethnicity, and Genetics Working Group. The use of racial, ethnic, and ancestral categories in human genetics research. *Am. J. Hum. Genet.* **77,** 519–532 (2005).
91. Murphy, S. *et al.* Instrumenting the health care enterprise for discovery research in the genomic era. *Genome Res.* **19,** 1675–1681 (2009).
92. Murphy, S. N. *et al.* Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J. Am. Med. Inform. Assoc.* **17,** 124–130 (2010).
93. Del Fiol, G. *et al.* Integrating genetic information resources with an EHR. *AMIA Annu. Symp. Proc.* **2006,** 904 (2006).
94. Buntin, M. B., Jain, S. H. & Blumenthal, D. Health information technology: laying the infrastructure for national health reform. *Health Aff. (Millwood)* **29,** 1214–1219 (2010).
95. Denny, J. C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26,** 1205–1210 (2010).
96. Ritchie, M. D. *et al.* Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am. J. Hum. Genet.* **86,** 560–572 (2010).
97. Loukides, G., Gkoulalas-Divanis, A. & Malin, B. Anonymization of electronic medical records for validating genome-wide association studies. *Proc. Natl Acad. Sci. USA* **107,** 7898–7903 (2010).
98. Dulbecco, R. A turning point in cancer research: sequencing the human genome. *Science* **231,** 1055–1056 (1986).
99. Toronto International Data Release Workshop Authors. Prepublication data sharing. *Nature* **461,** 168–170 (2009).
100. Contreras, J. L. Information access. Prepublication data release, latency, and genome commons. *Science* **329,** 393–394 (2010).
101. Robinson, C. J., Bohannan, B. J. & Young, V. B. From structure to function: the ecology of host-associated microbial communities. *Microbiol. Mol. Biol. Rev.* **74,** 453–476 (2010).
102. The NIH HMP Working Group. The NIH Human Microbiome Project. *Genome Res.* **19,** 2317–2323 (2009).
103. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
104. Kuczynski, J. *et al.* Direct sequencing of the human microbiome readily reveals community differences. *Genome Biol.* **11,** 210 (2010).
105. Stein, L. D. The case for cloud computing in genome informatics. *Genome Biol.* **11,** 207 (2010).


**Acknowledgements** This paper is endorsed by the National Advisory Council for Human Genome Research, whose members (including ad hoc members (indicated by an asterisk) at the September 2010 meeting) are M. Boehnke, M. Chee, R. Chisholm, J. Contreras, R. Cooper, C. Fraser-Liggett, R. Gibbs, G. Ginsberg, R. Hardison*, H. McLeod*, J. Mesirov*, R. Myers, P. O'Rourke, P. Ossorio, P. Sankar*, D. Valle, R. Weinshilboum, D. Williams* and R. Wilson*. The dedicated effort of NHGRI staff was instrumental for the strategic planning process that culminated in this paper. The final stages of this process were led by E. Green and M. Guyer, and involved extensive contributions from K. Wetterstrand, A. Bailey and S. Vasquez. We also thank the additional NHGRI staff who made important contributions to the writing of this paper: C. Austin, L. Biesecker, V. Bonazzi, V. Bonham, J. Boyer, L. Brody, L. Brooks, G. Feero, E. Feingold, A. Felsenfeld, P. Frosst, P. Good, B. Graham, D. Hadley, S. Hart, L. Hindorff, C. McBride, J. McEwen, M. Muenke, J. Mullikin, E. Ostrander, A. Pillai, R. Pozzatti, L. Proctor, E. Ramos, L. Rodriguez, J. Schloss, J. Segre, J. Struewing, L. Thompson and A. Wilson. We thank the following for editorial input: D. Altshuler, M. Boehnke, D. Botstein, J. Evans, C. Fraser-Liggett, E. Lander, R. Lifton, R. Myers, D. Valle, R. Waterston and H. Varmus. We acknowledge the graphical contributions of D. Leja and J. Latman, as well as the logistical support of P. Klein, J. Mullaney and E. Rolfes. We also thank H. Dietz and S. Terry who, along with some of those previously listed, crafted topical white papers for the strategic planning process (see http://genome.gov/Planning). Finally, we thank the many participants at various NHGRI strategy meetings, as well as F. Collins and A. Guttmacher for their early leadership of this strategic planning process.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to E.D.G. (egreen@nhgri.nih.gov).

**National Human Genome Research Institute**

**Overall leadership** Eric D. Green[1] & Mark S. Guyer[1]

**Coordination of writing contributions (see Acknowledgements for list of other contributors)** Teri A. Manolio[1] & Jane L. Peterson[1]

[1]National Human Genome Research Institute, National Institutes of Health, 31 Center Dr., Bethesda, Maryland 20892-2152, USA.