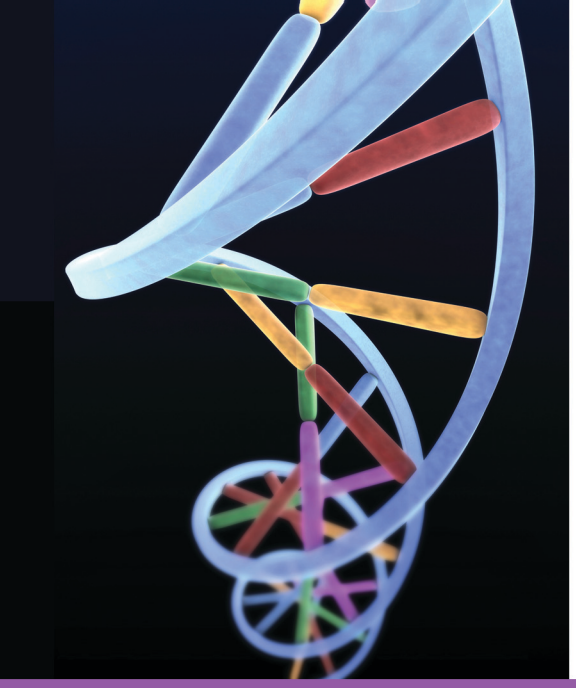


Economic Impact of the Human Genome Project

How a $3.8 billion investment drove $796 billion in economic impact, created 310,000 jobs and launched the genomic revolution

Prepared by Battelle Technology Partnership Practice

May 2011

Battelle
The Business of Innovation

Contents

Certain technologies discussed in this report have not received regulatory approval or clearance for clinical uses. In this regard, sequencers are currently intended for research or investigational uses.

Executive Summary

Introduction

The sequencing of the human genome represented the largest single undertaking in the history of biological science and stands as a signature scientific achievement.[1] All of history in the making, human DNA took just 13 years to sequence under the Human Genome Project (HGP), an international public project led by the United States, and a complementary private program. Sequencing the human genome—determining the complete sequence of the 3 billion DNA base pairs and identifying each human gene—required advanced technology development and the assembly of an interdisciplinary team of biologists, physicists, chemists, computer scientists, mathematicians and engineers.

Scientists are using the reference genome, the knowledge of genome structure, and the data resulting from the HGP as the foundation for fundamental advancements in medicine and science with the goals of preventing, diagnosing, and treating human disease. Also, while foundational to the understanding of human biological systems, the knowledge and advancements embodied in the human genome sequencing, and the sequencing of model organisms, are useful beyond human biomedical sciences.

The resulting "genomic revolution" is influencing renewable energy development, industrial biotechnology, agricultural biosciences, veterinary sciences, environmental science, forensic science and homeland security, and advanced studies in zoology, ecology, anthropology and other disciplines.

In the ten years since the first sequences were published, much has been written about the scientific consequences of mapping the genome but little analysis has been done of the economic significance of the achievement. For this reason, Battelle has produced, with support from Life Technologies Foundation, this first comprehensive analysis of the economic impact of the HGP.

The full report provides multiple analyses and information resources to spur consideration and discussion. It includes

- A quantitative measurement of the direct and indirect economic impacts in the United States derived from actual expenditures of the HGP project and follow-on federal expenditures in major genomic science programs. This is quantified using the regional economic analysis technique of input/output analysis.
- A quantitative estimate of the economic impact of the U.S. "genomics and genomics-enabled industry", with acknowledgement that those within the industry credit the HGP and related programs as being integral components in the development of the industry.
- Specific examples and case studies of "genomics in action"—highlighting the functional application of genomics tools, technologies and knowledge in a range of economically relevant fields such as healthcare, veterinary medicine, agriculture, the environment, industrial biotechnology, and security and justice.
- A look into the potential for future impacts along various genomic pathways. The first human genome sequences were an important step in the development of these pathways, paving the way for much of what is expected in the "post-genomic" future.

The general structure of impacts is shown in Figure ES-1.

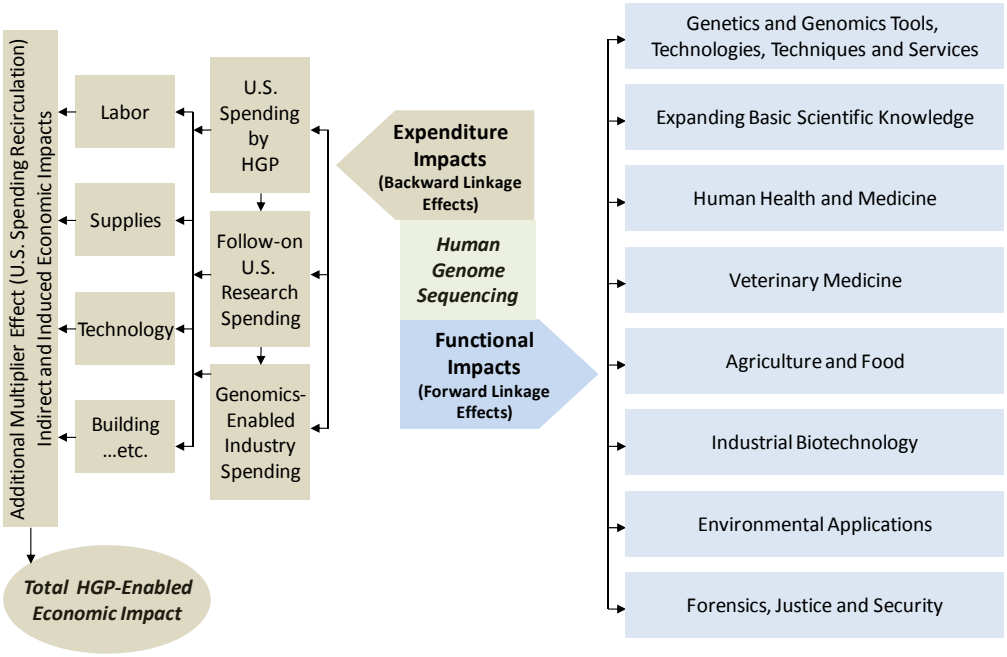
This report aims to fill a gap in the literature regarding the Human Genome Project by assessing its economic and functional impacts.

The HGP required the development of advanced equipment, technologies, data analysis tools, and specialized analysis techniques that has facilitated the growth of an expanding "genomics industry."

Today this industry is empowering further scientific discovery, progress and commercial innovation on a broad range of fronts. From human healthcare to veterinary medicine, from industrial biotechnology to high productivity agriculture, the knowledge, tools and technologies supported through the sequencing of the human genome form a foundation of advanced economic and social progress for the United States and humankind.

[1] The Human Genome Project (HGP) typically refers to the federally-funded program that ran from 1990 to 2003 (though it also included initial funding in 1988 and 1989). Part of our economic analysis focuses just on this essential and galvanizing component of the total effort to unravel the human genome. But the total effort to decode the human genome involved many public and private players over many more years, and the work of entities such as Celera Genomics played an important part. Our analysis of the functional impacts of the HGP necessarily includes all these contributions, and in general our reference to the effort to map the genome should be understood to encompass numerous contributions made by multiple parties to our current understanding.

Figure ES-1: The Structure of Forward and Backward Linkage Impacts Associated with the Human Genome Sequencing



Battelle's detailed analysis produced five overarching conclusions:

1. **The economic and functional impacts generated by the sequencing of the human genome are already large and widespread. Between 1988 and 2010 the human genome sequencing projects, associated research and industry activity—directly and indirectly—generated an economic (output) impact of $796 billion, personal income exceeding $244 billion, and 3.8 million job-years of employment.[2]**
2. **The federal government invested $3.8 billion in the HGP through its completion in 2003 ($5.6 billion in 2010 $). This investment was foundational in generating the economic output of $796 billion above, and thus shows a return on investment (ROI) to the U.S. economy of 141 to 1—every $1 of federal HGP investment has contributed to the generation of $141 in the economy.**
3. **In 2010 alone, the genomics-enabled industry generated over $3.7 billion in federal taxes and $2.3 billion in U.S. state and local taxes. Thus in one year, revenues returned to government nearly equaled the entire 13-year investment in the HGP.**
4. **Overall, however, the impacts of the human genome sequencing are just beginning—large scale benefits in human medicine, agriculture, energy, and environment are still in their early stages. The best is truly yet to come.**
5. **The HGP is arguably the single most influential investment to have been made in modern science and a foundation for progress in the biological sciences moving forward.**

Between 1988 and 2010 the human genome sequencing projects and associated research and industry activity directly and indirectly generated:

- *$796 billion in U.S. economic output*
- *$244 billion in personal income for Americans*
- *3.8 million job-years of employment.*

[2] A "job-year" is equivalent to one person employed full time for one year. Personal income consists of both employee and proprietor income and includes wages and other financial benefits (e.g., retirement funds).

The Economic Impacts of Human Genome Project

The human genome sequencing programs had direct expenditures totaling almost $3.8 billion ($5.6 billion in constant 2010 dollars). In addition, beyond the original HGP investments, both the National Institutes of Health (NIH) and the Department of Energy (DOE) made and continue to make significant investments in the science, instrumentation, and related applications stemming from the human genome work. In fact, in both current and constant 2010 dollars the federal investment in human genomics-related or enhanced research over the last seven years (2004–2010) has actually exceeded the federal investment in the HGP by 28 percent. These additional expenditures also generated direct and indirect (multiplier effect) economic impacts in the U.S economy.

The decoding of the human genome was both a technological as well as scientific achievement. The technologies that have empowered genome sequencing range from the gene sequencers themselves, to sample preparation technologies, sample amplification technologies, and a range of analytical tools and technologies. An industry has grown up to supply the scientific research community in the private sector, government and academia with the equipment, supplies and services required to conduct genomics research and development (R&D) and associated product development. This industry, of course, generates additional economic impacts.

To evaluate these genomics-enabled industry impacts in the U.S., Battelle constructed a "from the ground-up" database of individual companies engaged within the sector. The employment of this industry base was used as the foundation for an input/output analysis to quantify the total impacts of these firms (in terms of direct and indirect output and employment and their multiplier effect). The results of the analysis show that—spurred by the original investment and technological development impetus of the human genome sequencing projects—a substantial economic sector has developed thereby benefiting the U.S. economy in terms of business volume, jobs, and personal income supporting American families.

Battelle's analysis of the impact of human genome sequencing includes separate assessments of the HGP impacts, post-HGP expenditure impacts, and the impacts of the genomics-enabled industry sector. These separate results are contained in the full report. Table ES-1 provides a summation of the combined impacts from these analyses for the 1988–2010 period.[3]

In 2010 alone, the human genome sequencing projects and associated research and industry activity directly and indirectly generated:

- ***$67 billion in U.S. economic output***
- ***$20 billion in personal income for Americans***
- ***310 thousand jobs.***

Table ES-1: Cumulative Economic Impact of Human Genome Sequencing, 1988–2010 (in Billions, 2010 $)

Impact	Employment (Job-Years)	Personal Income	Output	State/Local Tax Revenue	Federal Tax Revenue
Direct Effect	710,819	71.4	264.8	3.5	13.0
Indirect Impacts	1,298,216	89.2	265.8	10.8	18.0
Induced Impacts	1,818,459	83.3	265.7	15.2	17.9
Total Impact	**3,827,495**	**243.9**	**796.3**	**29.5**	**48.9**
Impact Multiplier	5.38	3.42	3.01	8.37	3.75

The direct impacts include nearly 711,000 direct job-years, a combined personal income direct impact of more than $71 billion, and direct genomics-driven output of nearly $265 billion. In combined direct tax revenue more than $3.5 billion has been generated in state/local taxes and more than $13 billion in federal taxes. From just the perspective of federal revenue and expenses,

[3] While 1988 and 1989 are outside of the official time period of the HGP, federal funding statistics record related expenditures for these years and thus are included in this analysis. Additionally, federal HGP funded continued until 2003.

The federal investment of $3.8 billion in the Human Genome Project ($5.6 billion in 2010 $) enabled the generation of more than $796 billion in economic output for a return on investment to the U.S. economy of 141 to 1—every $1 of investment has helped to generate $141 in the economy.

the direct federal taxes generated to date (in 2010 $) have exceeded the HGP and post-HGP federal investments ($12.8 billion) by $166 million.

The total employment impact exceeds 3.8 million job-years over the 23-year period examined. With an employment multiplier of 5.38, genomics research and associated industries generated an additional 4.38 jobs in the U.S. economy for every "genomics" job. This total level of impacted employment generated nearly $244 billion in personal income over the period—in 2010 dollars amounting to an average of $63,700 in personal income per job-year. The genome sequencing projects, associated research and industry activity generated a total economic (output) impact of more than $796 billion over the 1988–2010 period. Considering the federal investment of $3.8 billion in the Human Genome Project from 1988–2003 ($5.6 billion in 2010 $) the HGP generated a return on investment (ROI) to the U.S. economy of 141 to 1—every $1 of HGP investment has helped to generate $141 in the economy.

The Functional Impacts of the Human Genome Project

In medicine, genomics has provided the first systematic approaches to discover the genes and cellular pathways underlying disease. Whereas candidate gene studies yielded slow progress, comprehensive approaches have resulted in identification of ~2,850 genes underlying rare Mendelian diseases, ~1,100 loci affecting common polygenic disorders and ~150 new recurrent targets of somatic mutation in cancer. These discoveries are propelling research throughout academia and industry

Eric Lander
Nature. Vol. 470. 2011

While the ***economic*** impacts of the sequencing of the human genome are impressive, the ultimate goal of the HGP was to improve our biological understanding, and our health and wellbeing. The economic impact may be viewed as a "bonus" that has occurred in addition to the primary ***functional*** impacts. The programs sought to create benefits for humankind by elucidating basic molecular processes governing life and via the application of knowledge gained to human healthcare and multiple other fields that would benefit from advancements in genomics knowledge and technology.

Figure ES-1 also illustrates Battelle's representation of the structure of functional impacts generated by the sequencing of the human genome. These functional impacts include the development of genomics tools, technologies and techniques—an area of fundamental importance to making the sequencing feasible. Significant advancement was made over the course of human genome sequencing across a broad range of technology fronts and these have, in many cases, been commercialized and form the foundation for a highly active and growing commercial genomics-based industry.

The application of these genomic tools, technologies and techniques resulted in a truly dramatic expansion of basic biological knowledge. The full human genome sequence unveiled a complex biological system unanticipated by most in science, and has been a paradigm shifting event for biology.

As the sequencing information from the HGP was posted publically on a daily basis to the World Wide Web, the data contained could be immediately put to use advancing application of knowledge across a broad range of scientific disciplines and applied fields. Revealing the human genome structure and sequence had a direct impact on biomedical science. It also had a transformational effect on the identification of mechanisms of disease, diagnosis of genetic diseases and disorders, elucidation of pathways and potentially involved biological elements in complex diseases and disorders, and detection of specific targets for drugs and biologics. Furthermore, having the full human sequence has served to advance additional human biomedical applications in R&D in areas such as gene therapy, vaccine development, regenerative medicine and stem cell therapy, and the refined matching of organs and tissue from donors to patients.

From an impact standpoint, developments in pharmacogenomics and personalized medicine hold significant promise for more accurate diagnosis of disease, elucidation of preventive measures to avoid disease, more refined disease treatment and avoidance of costs associated with adverse drug reactions.

With the first full draft of the human sequence in hand, new areas of biomedical science have opened—with pharmacogenomics and companion diagnostics, for example, being used to identify appropriate drugs based on patient genomic profiling and to refine the dosing of therapies, thereby maximizing effectiveness and reducing negative side-effects. Already being applied in the treatment of some cancers, pharmacogenomics, empowered by the sequencing of the human genome, holds promise for revolutionizing the practice of medicine in the 21st Century.

Modern genomics, the reference human genome sequences, and comparative organism sequences, have empowered a new view of biological processes. In turn, these new views are being translated into biomedical applications, such as those shown in Table ES-2.

Table ES-2: Biomedical Applications of Genomics and Human Genome Sequencing

Potential Application	Genomics Advances Today	Hope for the Future
Diagnosis of single gene, Mendelian diseases and disorders	Specific genes for over 3,000 Mendelian monogenic diseases discovered. Genomic tests are being used to accurately diagnose rare diseases and disorders, many of which were previously misdiagnosed with inappropriate courses of treatment prescribed. Prenatal genetic screening is being performed to inform potential parents of risks for catastrophic inheritable disorders.	Gene therapies will achieve success in repairing genetic abnormalities leading to diseases and disorders. Custom therapeutic products will block or change expressed activity of defective genes.
Knowledge of predisposition towards specific diseases	Multiple genes and biomarkers have been identified for predisposition to multiple diseases such as cancers, neurological diseases, psychiatric disease and cardiovascular disease.	Understanding of risk for disease based upon multi-gene tests will likely lead to appropriate therapeutic interventions and personal behavior/lifestyle modification. Environmental components of disease emergence and progression will be teased-out from genomic factors and addressed appropriately.
Genomics driven drug discovery, known as rational drug development	New drug targets have been identified. Cancer drugs based on the genomics of tumors are on the market, including Gleevec (for chronic myelogenous leukemia), Herceptin (breast cancer), Tarceva (lung cancer) and Avastin (colon, lung and other cancers).	Many new drugs and biologics will be developed to successfully exploit elucidated drug targets.
Therapeutic products custom prescribed based on patient genomics, to maximize effect and reduce or eliminate side effects	Already being applied in the treatment of some forms of cancer and cardiovascular disease. Genetic tests are used for dosage levels in prescription of some drugs such as Coumadin (warfarin).	Routine sequencing of a patient's entire genome wil guide treatment selection and dose for the optimum response. Potential adverse reactions to drugs and treatment regimens, identified via genomic markers, will result in the avoidance of adverse events.
Repurposing or revitalization of some drugs shelved in development because of impact on a "genomic few"	Successful discovery of subpopulations for which previously unapproved drugs are efficacious. Iressa, for example, has been approved with patents testing positive for the EGFR mutation.	There will be a substantial volume of existing drugs found to be efficacious in selected sub-populations, and drug companies will have mined their previously "failed" R&D pipelines to bring forward previously non-marketable drugs to work in selected sub-populations.
Identification of means to combat infectious organisms	Multiple infectious organisms have had their whole genomes sequenced. Public health professionals sequence emerging infectious disease organisms to monitor migrations and mutations. Genetic testing already being applied to direct therapy for HIV/AIDS patients.	Rapid, real-time sequencing of pathogens twill direct public health efforts to combat disease outbreaks in humans and zoonotic disease outbreaks in animals. DNA vaccines will come to the market to impart immunity to infectious diseases. Genetic profiling of patient viruses will assist in customizing treatment regimens.
Gene therapies for inherited genetic diseases and disorders	After publicized setbacks, gene therapies are now achieving success. For example, the fatal brain disorder adrenoleukodystrophy has been treated, with progression stopped, in a sample of children.	Gene therapy may be routinely provided to newborns with identified genomic profiles to correct defective genes, particularly in conditions associated with devastating monogenic disorders.

*"Modern medicine arose when scientists learned to fight some of the worst infectious disease with vaccines and drugs. This strategy has not worked with AIDS, malaria, and a range of other diseases because of their complexity and the way they infiltrate processes in cells. Curing such infectious diseases, cancer, and the health problems that arise from defective genes will require a new type of medicine based on a thorough understanding of how cells work and the development of new methods to manipulate what happens inside them."**

The HGP was and is one of the central projects leading to this "understanding of how cells work" and opening the way for new applications of molecular medicine.

* *Russ Hodge. 2010. "The Future of Genetics: Beyond the Human Genome."*

We are on the leading edge of a true revolution in medicine, one that promises to transform the traditional "one size fits all" approach into a much more powerful strategy that considers each individual as unique and as having special characteristics that should guide an approach to staying healthy.

Francis Collins,
"The Language of Life,"
New York: Harper Collins, 2010

Beyond human healthcare, the technologies, tools and basic biological knowledge embodied in the human genome sequencing programs have found applications across a wide range of life-science and related disciplines. Agriculture and global food security are being significantly enhanced through the

application of genomics to plant and livestock improvement—positively altering both organismal input traits (such as nutrient uptake, pest resistance, drought tolerance, etc.) and output traits (such as nutrient content, chemical composition, quality and quantity of food produced). Genomics is also being applied in the tracing of food contamination and associated pathogenic events.

The benefits being felt in human healthcare also translate into veterinary medicine. The concurrent sequencing of various animal genomes during the HGP, and subsequent whole animal genome sequencing undertaken after completion of the human genome, has been as revolutionary for veterinary medicine as it has been for human biomedical sciences.

Early in the HGP, the U.S. Department of Energy began applying genomics tools and technologies resulting from the human genome sequencing program to the sequencing of microbes for functionality in environmental remediation and for investigation of their use in other applications such as bioprocessing and biofuels production. The Microbial Genome Project and the recent Genomes to Life Project have specifically applied genomics to a range of environmental and industrial applications. Likewise commercial enterprises across a series of product categories in biotechnology, biofuels, food processing, drug and vitamin production, and biobased materials are applying advanced genomic knowledge and technologies to bring to market new and more efficient industrial processes to power the U.S. and global economies.

Genomics has also become a tool for applications in the field of justice and security. For homeland security, the ability to genotype suspicious infectious pathogens and trace their origin is a national security priority. Law enforcement is also using genomics in identification of human remains in instances of natural disasters (such as the Haitian and Japanese earthquakes) or terrorist events (such as 9/11's ground zero), and for tracing illegal trade in endangered animals.

It would be impossible to quantify each and every impact being generated across the field of genomics and its application to such a broad range of scientific fields and technological disciplines. Battelle has not attempted to do that. Rather Battelle examined each of the key fields of application shown in Figure ES-1 and discusses within the full report the ways in which genomics is being applied within these fields, and provides case studies of genomics in action within these fields.

Into the Future

The reference human genome is akin to chemistry's periodic table, a perpetually useful fundamental platform for understanding and advancing science.

The advancement of knowledge and the technologies resulting from human genome sequencing have formed the platform for nothing less than a medical revolution. The primary impacts of this revolution in quantitative and personalized medicine may not yet be felt in daily clinical practice (although in some areas of cancer it is), but that day is accelerating towards us.

One of the key realizations that must be understood regarding the human genome sequencing is that its usefulness is perpetual. While other major big science projects have a life attached to them the human genome sequence will not wear out or become obsolete.[4] Rather, the reference human genome is akin to chemistry's periodic table, a perpetually useful fundamental platform for understanding and advancing science.

The impact on human medicine and health is profound and important but, as shown, the benefits of performing the HGP and related projects extend into areas of far beyond human health. Our fundamental understanding of genomics, the sequencing and genomic technologies whose development was spurred by the HGP, and the advanced "omics" disciplines created through these efforts, will continue to make contributions on a broad range of fronts. We can expect a variety of genomics-enabled enhancements, for example:

[4] For example, the $11 billion Superconducting Super Collider has an estimated life span of 30 years and the $1.5 billion Hubble Space Telescope has an estimated 15–20 year life.

- Agricultural productivity to increase considerably, working towards the challenge of feeding the world's rapidly expanding population in a sustainable manner.
- Not only will food availability increase, but the impact of its production on the global environment will reduce as crops and livestock are developed with traits suited to nitrogen use efficiency, no-till agriculture, water use efficiency and reduced waste production.
- Currently low-value biomass, especially low-value cellulosic biomass, will be converted into higher-value liquid fuels, energy sources, biobased chemicals, plastics and materials. These products will increasingly displace petroleum and other fossil-based inputs, contributing to reduced carbon emissions and associated climate and environmental benefits.
- An increasingly two-way flow of diagnostics, therapeutics and prevention tools will move between human medicine, veterinary medicine and agriculture as the cost of genomic technologies reduces and the applications of discoveries in one area can be applied to another because of comparative genomics and other genomic advancements.
- The legacy of pollution on the planet caused by human activity will be addressed increasingly through the application of genetically engineered, modified or synthetic organisms designed to perform remediation and mitigation functions.

From an economic standpoint, the impact of applied genomics will be equally profound. Modern developed societies are driven forward by innovation—typically technological innovation. The HGP and associated sequencing projects have advanced the state of innovation and technology and resulted in broad economic impacts as highlighted in the report. These positive economic impacts will continue to ripple outwards and expand. Personalized medicine approaches for the prevention and optimized treatment of diseases and disorders holds promise for reducing the future burden of disease and associated healthcare costs.

In the developing world the promise contained within applied genomics is fundamental to basic economic progress. Without affordable and sustainable food supplies, development will be highly limited as malnutrition reduces developing societies' productivity and incomes. Without higher productivity crops, we will see more and more marginal lands pressed into production and environmental degradation will become increasingly more widespread. Without enhanced abilities to combat current and emerging infectious and endemic diseases, developing societies will again be held back and progress limited. Applied genomics, built upon the foundation of the HGP and associated projects, holds promise for the provision of solutions to each and every one of these challenges and more.

In Battelle's interview with Lee Hood he called the HGP "the single most transformative event in the history of biological science." It would be difficult not to concur with that statement. In advancing basic scientific knowledge; in opening new pathways to enhanced human health; in spurring technologies for economic development; in providing solutions to widespread challenges in global food production and environmental sustainability, the $3.8 billion spent on the HGP may well represent the best single investment ever made in science.

Chapter I: Introduction

A. Background

The sequencing of the human genome and the mapping of human genes represented the largest single undertaking in the history of biological science and stands as a preeminent scientific achievement.

> *Along with Bach's music, Shakespeare's sonnets, and the Apollo Space Programme, the Human Genome Project is one of those achievements of the human spirit that makes me proud to be human.*
> **Richard Dawkins**

> *It is humbling for me and awe inspiring to realize that we have caught the first glimpse of our own instruction book, previously known only to God.*
> **Francis Collins, at the White House launch of the human genome draft sequences.**

The Human Genome Project (HGP) was an international public project led by the United States, coordinated by the National Institutes of Health (NIH) and U.S. Department of Energy (DOE), and complemented by the private efforts of Celera Genomics. The goal of the project was to discover the thousands of human genes, complete the sequence of the billions of DNA subunits, and make these results accessible to all researchers for further scientific study.

DNA's chemical language of four repeated letters A, T, G and C elegantly produces the complete compendium of life, staggering in its scale, diversity and complexity. From single cell organisms drifting in our oceans to sentient self-aware human beings, DNA codes all life on Earth. Mapping the complete genome of a human would provide a new view into the inner workings of the most complex and advanced organism on the planet. Some thought it simply could not be done.

The task of decoding the genome dictated a new approach to biological science—a large-scale, "big science" team approach. Requiring $3.8 billion in total public funding, the HGP sought to:

- Identify all the approximately 20,000–25,000 genes in human DNA,
- Determine the sequences of the 3 billion chemical base pairs that make up human DNA,
- Store this information in databases,
- Improve tools for data analysis,
- Transfer related technologies to the private sector, and
- Address the ethical, legal, and social issues that may arise from the project.[5]

It achieved all of these goals, with only the last one being still in progress. It also went further, generating sequences of other scientifically-valuable organisms (the *E. coli* bacterium, the fruit fly and the laboratory mouse) to enable comparative genomics studies to be performed which would facilitate the identification of gene function. Likewise, Celera met its goals to validate a new approach to sequencing (shotgun sequencing), prove the technique via the sequencing of model organism genomes, and concurrently produce a draft human sequence of its own alongside the public sequencing effort.

"The Human Genome Project (HGP) refers to the international 13-year effort, formally begun in October 1990 and completed in 2003, to discover all the estimated 20,000–25,000 human genes and make them accessible for further biological study. Another project goal was to determine the complete sequence of the 3 billion DNA subunits (bases in the human genome). As part of the HGP, parallel studies were carried out on selected model organisms such as the bacterium E. coli and the mouse to help develop the technology and interpret human gene function. The DOE Human Genome Program and the NIH National Human Genome Research Institute (NHGRI) together sponsored the U.S. Human Genome Project."

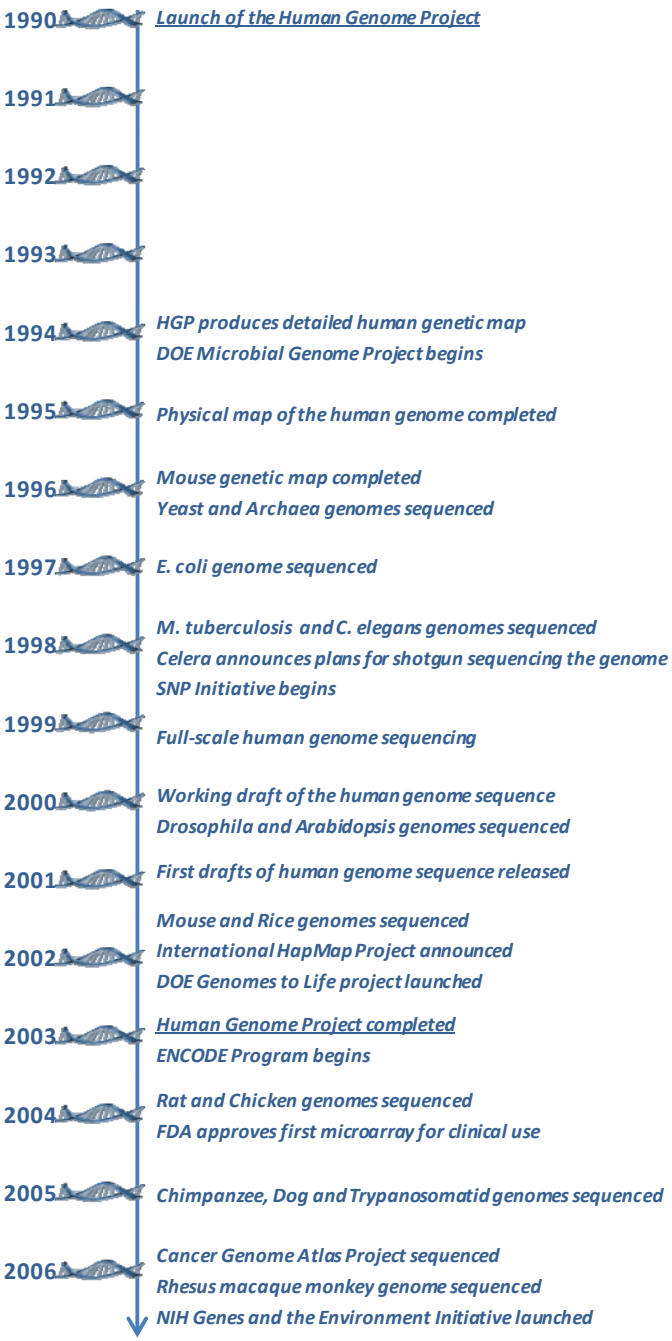
Human Genome Project Information at www.genomics.energy.gov. Oak Ridge National Laboratory

"It's one small piece of man... one giant leap for mankind."

The Mirror
UK newspaper headline upon announcement of the draft sequence.

[5] U.S. Department of Energy – Human Genome Project Information. *"About the Human Genome Project."* Accessed online at http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml

Figure 1: Timeline of Genome Sequencing



Source: National Human Genome Research Institute

Not only did the HGP and Celera programs achieve their goals, they generated and heralded unforeseen breakthroughs in scientific knowledge and technological development. Consider "just" the following, for example:

- While finding that the human genome comprises fewer genes than had previously been estimated (circa 21,000 distinct protein coding genes[6] versus some previous estimates of more than 100,000), the project enabled scientists to find unforeseen complexity in the genome—with individual genes interacting with one another and reading into multiple RNAs to produce more protein products than previously imagined.[7] Rather than the expected "one-gene one-protein" relationship, it was found that individual genes are able to encode multiple proteins.

- Non-coding DNA, previously termed "junk DNA" because it was thought to be a relic of evolution with little biological function, was instead confirmed to have specific functionality in transcription and translational regulation of protein-coding—i.e., most of it is not junk at all, it is central to life functions. This finding alone supports the vision of undertaking whole genome sequencing, since prior to the HGP some detractors argued that the budget would be better spent simply studying known protein coding genes and ignoring the rest. Eric Lander points out that "*it has reshaped our view of genome physiology, including the role of proteins-coding genes, non-coding RNAs and regulatory sequences*."[8] Leroy Hood, a pioneer in sequencing technology, notes that human genome sequencing brought to biology a new paradigm of discovery-driven science, stepping outside the bounds of traditional hypothesis-driven science and enabling a new view.[9]

- The HGP's technology development program, and the work of Celera, planted the seeds for, and accomplished, advancements in genome and gene sequencing technology, to the extent that acceleration of sequencing speeds in successive generations of equipment has exceeded even computer processing's Moore's Law.

Scientists in government laboratories, academe and industry will be using the reference genome, the knowledge of genome structure, and the publically published genome data resulting from the HGP (as well as Celera's data) as the foundation for fundamental advancements in science and the development of applied genomics and genetics tools, techniques and technologies.

[6] Clamp, M. et al. 2007. "Distinguishing protein-coding and noncoding genes in the human genome." Proceedings of the National Academy of Sciences USA. 104. 2007.

[7] Stewart Scherer. 2008. "A Short Guide to the Human Genome." Section on how many protein-coding genes are present in the genome. Cold Spring Harbor Laboratory Press, New York.

[8] Eric S. Lander. 2011. "Initial impact of the sequencing of the human genome." *Nature*. Volume 470, February 10, 2011.

[9] Leroy Hood. 2002. "After the Genome: Where Should We Go?" Chapter in Michael Yudell and Rob DeSalle (editors) "*The Genomic Revolution: Unveiling the Unity of Life.*" 2002. Joseph Henry Press, Washington DC with the American Museum of Natural History.

It is also very important to note that, while foundational to the understanding of human biological systems, the knowledge and advancements embodied in the human genome sequencing and the sequencing of model organisms are not only applicable to human biomedical sciences. The scientific and technological impact extends far beyond this, and the resulting "genomic revolution" is influencing agricultural biosciences, veterinary sciences, environmental science, forensic science, renewable energy development, industrial biotechnology and advanced studies in evolution, zoology, anthropology and other academic disciplines.

As a foundation for knowledge expansion, scientific advancement, and technological development few would argue that the sequencing of the human genome represents a pinnacle of science and a milestone in human history. The statements of early proponents of whole human genome sequencing about the likely growth of a genomics industry are now holding-up to scrutiny. Noted venture capitalist G. Steven Burrill states that "the genomics revolution is here now."[10] There should be little doubt that the "post-genomic" era is here, that significant impacts are being realized in the present, and tremendous impacts will be realized in the future.

Using the latest machines from genomics companies, an entire human genome can now be sequenced in under a day for less than $3,000.
The efficiency gains in gene mapping are leading to new uses for the technology. Agricultural companies use genomics to help design genetically modified plants that are sturdier or more resistant to parasites. Cleantech entrepreneurs are splicing genes in search of greener fuels. And the U.S. Food and Drug Administration is ordering a $695,000 sequence… to compare E.coli and other bacteria from across the country. The machine will help agency scientists determine the source of outbreaks

Rob Waters

B. Purpose of This Report

This report aims to fill a gap in the literature regarding human genome sequencing and its outcomes– an assessment of the economic impact of sequencing. Beyond the achievement of knowledge embodied in the HGP and Celera projects, these programs required the development of advanced equipment, technologies, data analysis tools, specialized analysis techniques and know-how that has facilitated the growth of an expanding genomics industry. Today this industry is empowering further scientific discovery, progress and commercial innovation on a broad-range of fronts. From human healthcare to veterinary medicine, from industrial biotechnology to high productivity agriculture, the knowledge, tools and technologies supported through the sequencing of the human genome form a foundation of advanced economic and social progress for the United States and humankind.

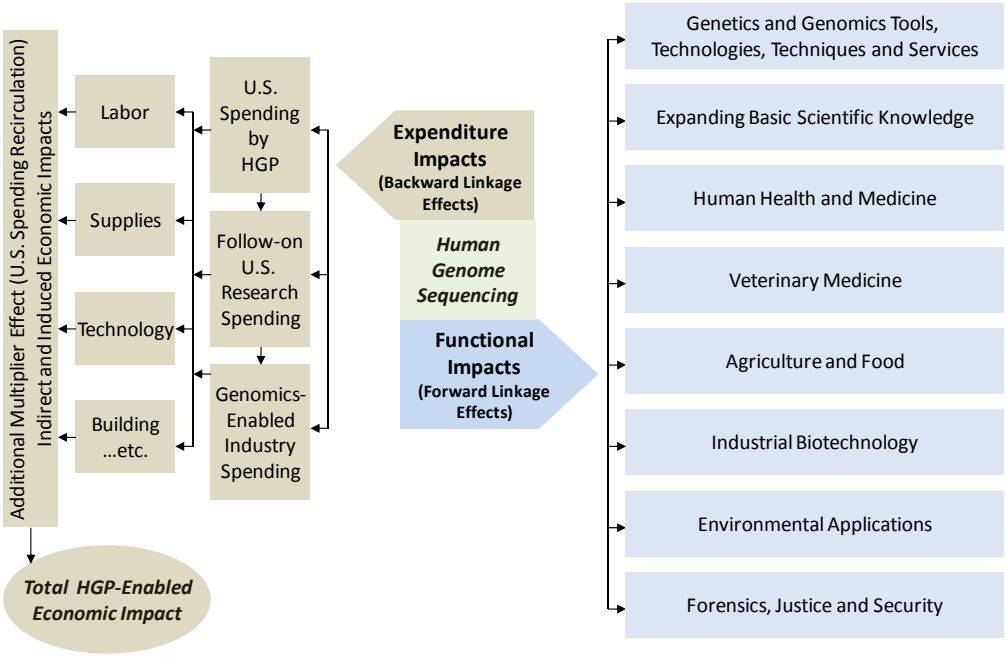
It should be noted that economic impact is just one measure of the "success" of a program. It may not necessarily be the best measure, and rarely should it be the sole measure of a publically supported program aiming at meeting public interests. The HGP in particular was not undertaken with the primary goal of generating economic impacts, although one of the program's stated goals, namely "*transfer related technologies to the private sector*" certainly does imply a commercial benefit in terms of business activity. There is always a risk in undertaking an economic impact study that, in doing so, a project may **only** be judged on its economic merits.

As such, Battelle has chosen herein not solely to address economic impacts in this report.[11] Rather, this report takes a more holistic view of human genome sequencing impacts—addressing the HGP's ***expenditure impacts****(or backward linkage effects)* certainly, but also examining in detail the HGP's ***functional impacts*** *(or forward linkage effects)* and implications of these impacts across a broad range of fields as illustrated in Figure 2.

[10] Rob Waters. 2011. "Boom Times for Genomic Startups: Investment is picking up as sales of gene sequencing machines surge." *BusinessWeek*. March 17, 2011. Accessed online at http://www.businessweek.com/magazine/content/11_13/b4221058333104.htm

[11] Typically, when institutions or regions release an economic impact report, backward linkage or "expenditure" impacts are the most frequently quantified economic impacts.

Figure 2: The Structure of Forward and Backward Linkage Impacts Associated with the Human Genome Sequencing



Thus, this report provides multiple analyses and information resources to spur consideration and discussion. It provides:

- A quantitative measurement of the direct and indirect economic impacts in the United States derived from a) actual expenditures of the specific HGP project and its consortium institution, b) additional follow-on federal investments in genomics-related and enabled research, s and
- A quantitative estimate of the economic impact of the U.S. "genomics industry", with acknowledgement that those within the industry credit the HGP and Celera programs as being integral components in the development of the industry.
- Specific examples and case studies of "genomics in action"—the functional application of genomics tools, technologies and knowledge in a range of economically relevant fields such as healthcare, veterinary medicine, agriculture, the environment, industrial biotechnology, and security and justice.
- A look into the potential for future impacts along the various genomic pathways. The first human genome sequences were an important step in the development of these pathways, paving the way for much of what is expected in the "post-genomic" future.

Acknowledgements

Battelle thanks Life Technologies Foundation for the vision shown in initiating this project and for funding the performance of the impact study. Janet Lambert and Heather Virdo of the Life Technologies Foundation, in particular, contributed to the project's scope and direction. Battelle thanks the many individuals who took time to participate in interviews and who responded to information requests during performance of this project. Battelle also thanks all of those individuals who made up the Human Genome Project team and the project team at Celera Genomics—without their efforts on behalf of science and mankind, the impacts described in this report would be greatly lessened.

Chapter II: Economic Impacts of the Human Genome Project

Determing the full "current" economic impact of a program as complex and game-changing as the sequencing of the human genome requires analysis of a time series of investments and organizational information extending back to the formal initiation of the Human Genome Project in 1990 and even slightly before.

Table 1 shows the annual federal funding for the HGP.[12] The combined annual federal investment in the HGP reached $3.8 billion in current dollars, and when adjusted for inflation totals more than $5.6 billion in constant 2010 dollars.

Table 1: U.S. Human Genome Project Federal Funding (in Millions $)

Fiscal Year	DOE (Current $)	NIH (Current $)	U.S. Federal Total (Current $)	U.S. Federal Total (Constant 2010 $)
1988	10.7	17.2	27.9	54.0
1989	18.5	28.2	46.7	88.2
1990	27.2	59.5	86.7	160.0
1991	47.4	87.4	134.8	243.1
1992	59.4	104.8	164.2	289.5
1993	63.0	106.1	169.1	291.6
1994	63.3	127.0	190.3	321.2
1995	68.7	153.8	222.5	367.6
1996	73.9	169.3	243.2	390.5
1997	77.9	188.9	266.8	425.5
1998	85.5	218.3	303.8	448.5
1999	89.9	225.7	315.6	453.6
2000	88.9	271.7	360.6	490.1
2001	86.4	308.4	394.8	523.3
2002	90.1	346.7	434.3	548.4
2003	64.2	372.8	437.0	552.9
Totals	1,015.0	2,785.8	3,798.3	5,647.9

[12] See http://www.ornl.gov/sci/techresources/Human_Genome/project/whydoe.shtml#invest

A. Methodology and Assumptions

The expenditure impacts (or backward linkage effects) stemming from the HGP are those impacts generated in the economy from the direct investments in the HGP, the investments in follow-on HGP-enabled research, and through the genomics-enabled industry that has been developed and fostered through the science and technological requirements of the HGP.

These economic impacts are estimated in this chapter.

As illustrated in Figure 2, the expenditure impacts (or backward linkage effects) stemming from the HGP are those impacts generated in the economy from the direct investments in the HGP, the investments in follow-on HGP-enabled research, and through the genomics-enabled industry that has been developed and fostered through the science and technologicalrequirements of the HGP.

The economic impact analysis reported herein uses an I/O model to represent the interrelationships among economic sectors. I/O data show the flow of commodities to industries from producers and institutional consumers for any given region (in this case the United States as a whole). The data also model consumption activities by workers, owners of capital, and imports from outside the region. These trade flows built into the model permit estimating the impacts of one sector on other sectors. The impacts consist of three impact types: ***direct*** (the specific expenditures impact of the program and/or sector(s) in question, ***indirect*** (the impact on suppliers to the focus industry or program), and ***induced*** (the additional economic impact of the spending of these suppliers and employees in the overall economy).

The estimated impacts of sequencing the human genome were calculated using a U.S.-specific I/O model, the IMPLAN system, from MIG, Inc.[13] IMPLAN provides a software system and geographic specific data regarding economic sector interactions for calculating economic impacts. The model incorporates detail of more than 420 individual industry sectors that cover the entire national economy. With this coverage of 420 sectors, Battelle is able to model the cross-sector economic activity that occurred throughout the economy as a result of the human genome sequencing work. Additionally, the IMPLAN model has built in economic "inflators" and "deflators" to allow for the development of a cumulative multi-year impact estimation for the 23 years included in the analysis.

Six key economic sectors were used in the development of the overall impact analysis as shown in Table 2. The unique nature of federal R&D investment, and the associated economic transactions and impacts flowing from these investments, caused Battelle to select a single industry component "Scientific R&D services" as the most representative of these economic impacts.[14]

Table 2: Sectors Used in Overall Genomics Impact Assessment

"Genomics" Sectors	IMPLAN Modeling Sector (Sector #s)
Genomics-Related Bioinformatics	Custom computer programming services (371)
Genomic & Related Testing	Medical and diagnostic labs (396)
Genomic-Related Biologics & Diagnostic Substances	Biologics & diagnostics (133 & 134)
Genomic Instruments & Equipment	Analytical laboratory instrument mfg (254)
Genomics R&D/Genomics Biotech	Scientific R&D services (376)
Drugs & Pharmaceuticals	Drugs & pharmaceuticals (131 & 132)

Two metrics are used to drive the economic interactions within the models: R&D funding and investment (treated as "***output***" from the performing sectors) and ***employment***. The presentation of output values as part of an impact assessment are commonly referred to as the "economic impact" of an industry or activity.

[13] IMPLAN is one of the most widely used models in regional economics and can be used to analyze the economic impacts of companies, projects, or entire industries.

[14] The use of this sector was determined in consultation with Doug Olson, Founder and Principal of MIG, Inc. and is typically used as most representative of the expenditure of federal research funds by research organizations (whether private sector R&D firms, university research labs or federal intramural research funds).

Historical R&D funding data includes federal HGP investments (Table 1) and additional external and intramural federal investments in genomics areas after the 2003 end of the HGP by the NIH and DOE.[15] To conservatively estimate additional genomics research investments by the major pharmaceutical manufacturers, an increasing share of the sector's R&D funds was used over the 2004–2010 period.[16] This was determined to be the most appropriate approximation of the pharmaceutical industry's genomics involvement. While sales of genomics-based drugs (or their related diagnostics) were considered, this would leave out significant past efforts and pipeline investments that have yet to yield sales.

Historical employment data were developed using the unique properties of the National Establishment Time-Series (NETS) database developed by Walls & Associates (built upon annual records from Dun & Bradstreet). Using a variety of current and historical sources and listings of genomics firms, a database of firms (where genomics-specific activities, products, and services are the principal activity of the firm) was compiled. Using the NETS database a longitudinal set of employment metrics (1993–2010) were developed for each of the five"genomics" sectors listed in Table 2. The impact of genomics on the drugs and pharmaceuticals sector is estimated using only the R&D performance as described in the previous paragraph and not using employment. The NETS database allowed Battelle to capture the recent and historic employment of genomics firms, including those firms that were acquired or that have gone out of business over the last two decades. Another key component of the NETS database for this purpose is the "establishment" level information allowing for the inclusion of "genomic-specific" operations of larger multiestablishment firms to be included without having to include the firm's total employment. This feature was particularly important in the Genomic & Related Testing sector.

In the analysis that follows Battelle presents the direct effect values driving the model; additional estimated indirect, and induced impacts; and a summation of the total impacts (direct, indirect, and induced). The following data are provided for each impact estimation: employment, personal income (including both wages and benefits), economic output, state and local tax revenue (including income and property taxes), and federal tax revenue (including contributions to Social Security). An impact ***multiplier*** is also provided for each type of data—for every one (job or dollar) of direct effect, the multiplier number will equal the total (including the direct effect) number of jobs or dollars created in the U.S. economy.

Note: Unless specified, all dollar values in the following analyses and tables are in millions of 2010 dollars for comparability over time.

B. Impacts of the Direct NIH/DOE Funding of the HGP

This section examines the impacts generated by the entire federal HGP funding period, including 1988 and 1989 which are outside of the formal definition of the HGP, but nevertheless included in federal funding figures.

[15] For NIH these data were estimated using key word searches of extramural and intramural awards in the NIH RePORT database. For DOE these data were estimated using annual DOE budget documents for select genomics-related operations.

[16] Overall pharmaceutical industry R&D was estimated by Battelle for the post HGP years (2004–2010) using the National Science Foundation's industry R&D data as available supplemented by other sources. To establish a "genomics" share we set a conservative 3 percent share in 2004 based upon anecdotal research described in the *World Survey of Genomics* (2000) report from the Stanford-in-Washington program. The actual share in 2004 was likely higher. We set the current share (2010) at a conservative 20 percent based upon recent work by the Tufts Center for the Study of Drug Development that estimates that personalized medicine accounts for between 12 percent and 50 percent of pharmaceutical companies' R&D pipelines. We used a linear increase in share from 3 percent to 20 percent over the seven year period to estimate the growth in importance and impact of genomics on pharmaceutical R&D.

To provide context to the overall assessment of the economic impact of the HGP federal investment, we first examine the final year of the combined NIH and DOE investment (2003) which reached $437.0 million (or $552.9 million in 2010 $). Table 3 provides the impact analysis for this 2003 federal funding.[17]

Table 3: Economic Impact of Human Genome Project Final Year Federal Funding, 2003 (in Millions, 2010 $)

Impact	Employment (Jobs)	Personal Income	Output	State/Local Tax Revenue	Federal Tax Revenue
Direct Effect	5,025	310.3	552.9	8.9	51.0
Indirect Impacts	2,432	128.0	370.5	15.1	25.6
Induced Impacts	4,965	227.4	724.6	41.4	48.8
Total Impact	**12,422**	**665.7**	**1,648.1**	**65.4**	**125.5**
Impact Multiplier	2.47	2.15	2.98	7.36	2.11

As quantified by the IMPLAN model, this single year of investment (2003) was directly responsible for 5,025 jobs, $310.3 million in personal income, and $59.9 million of state/local and federal tax revenue. Through the project spending of these primarily academic and federal researchers, additional indirect impacts were generated in the U.S. economy. These impacts include additional employment generated of more than 2,400, personal income to employed persons of $128 million, and an additional $370.5 million in economic output. Finally, the personal spending of the researchers and suppliers' employees generated induced impacts of 4,965 jobs, $227 million in personal income, and output of nearly $725 million. Combined the direct, indirect, and induced impacts provide a total U.S. impact of more than 12,400 jobs, nearly $666 million in personal income, and more than $1.6 billion in economic output. In turn this combined economic activity generated more than $65 million in state/local taxes and more than $125 million in federal tax revenue.

This single year analysis is useful for describing one aspect of the longitudinal analysis presented below. Within the context of the IMPLAN model, the employment information generated consists of the number of jobs generated and existing in the specific year. When developing a cumulative analysis, the summation of a number of years of financial measures leads to a larger value that has meaning in its magnitude since each dollar is unique. When similarly adding together a number of years of employment, the value must be interpreted as ***job-years***. For example, an individual may be employed in Year 1 of the analysis and have the same job in Year 2. While everyone would consider this one job, from an impact analysis perspective it constitutes two job-years.

The federal HGP funding yields significant U.S. direct economic impacts beyond the direct HGP investment of $5.6 billion (Table 4). The federal HGP investment directly generated a total of 43,536 job-years—which averages out to more than 2,700 direct jobs per year of HGP federal funding. The direct impacts also reached nearly $3.2 billion in personal income. Finally, the HGP funding generated more than $610 million in direct effect state/local and federal tax revenues.

[17] Note: columns within the economic impact results tables in this chapter may not sum due to rounding.

Table 4: Cumulative Economic Impact of HGP Federal Funding, 1988–2003 (in Millions, 2010 $)

Impact	Employment (Job-Years)	Personal Income	Output	State/Local Tax Revenue	Federal Tax Revenue
Direct Effect	43,536	3,164.2	5,647.9	90.6	520.3
Indirect Impacts	24,842	1,307.7	3,785.1	154.3	261.4
Induced Impacts	50,660	2,319.7	7,393.1	422.3	498.3
Total Impact	**119,037**	**6,791.6**	**16,826.1**	**667.2**	**1,280.1**
Impact Multiplier	2.73	2.15	2.98	7.36	2.46

These impacts move throughout the economy—the multiplier effect (or ripple effect)—and generate additional indirect and induced economic impacts. **In total, more than 119,000 job-years were created by the HGP investment, leading to nearly $6.8 billion in personal income to these workers. Overall output exceeded $16.8 billion leading to an impact multiplier of 2.98—every $1 dollar of federal HGP investment led to an additional $1.98 of U.S. economic output.**

C. Impacts of follow-on NIH and DOE investments

Beyond the original HGP investments, both NIH and DOE made and continue to make significant investments in the science, instrumentation, and related applications stemming from the human genome work. These investments take the form of both intramural funding and extramural grant awards. In fact, in both current and constant 2010 dollars the federal investment in human genomics-related or enhanced research over the last seven years (2004–2010) has exceeded the 16-year federal investment in the HGP by 28 percent (Table 5).

Table 5: Estimates of U.S. Post Human Genome Project Federal Funding (in Millions $)

Fiscal Year	DOE (Current $)	NIH (Current $)	U.S. Federal Total (Current $)	U.S. Federal Total (Constant 2010 $)
2004	137.9	610.4	748.3	912.9
2005	145.8	628.0	773.8	899.9
2006	157.5	629.5	787.1	897.1
2007	217.2	681.7	898.9	984.1
2008	228.7	685.7	914.4	977.3
2009	234.9	1,038.6	1,273.5	1,316.6
2010	234.6	991.6	1,226.2	1,226.2
Totals	1,356.5	5,265.6	6,622.1	7,214.0

The cumulative total federal investment in human genome and related science reached more than $7.2 billion from 2004–2010. This level of investment led to the direct creation of more than 76,000 job-years and a resulting total personal income of more than $4.0 billion (Table 6).

Table 6: Cumulative Economic Impact of Federal (Post-HGP) Genomic Funding, 2004–2010 (in Millions, 2010 $)

Impact	Employment (Job-Years)	Personal Income	Output	State/Local Tax Revenue	Federal Tax Revenue
Direct Effect	76,146	4,048.6	7,214.0	115.9	665.8
Indirect Impacts	31,730	1,670.3	4,834.6	197.1	333.9
Induced Impacts	64,787	2,966.5	9,454.7	540.0	637.3
Total Impact	**172,663**	**8,685.5**	**21,503.3**	**853.1**	**1,637.0**
Impact Multiplier	2.27	2.15	2.98	7.36	2.46

Overall this post-HGP federal investment is responsible for the creation of nearly 173,000 job-years, and total personal income totaling nearly $8.7 billion and total output exceeding $21.5 billion.

D. Structure of the Enabled Industry

The decoding of the human genome was as much a triumph of technology as it was a triumph of scientific will. The technologies empowering genome sequencing are broad, ranging from the gene sequencers themselves, to sample preparation technologies, sample amplification technologies, and a host of analytical tools and technologies. An industry has grown up to supply the scientific research community in industry, government and academia with the equipment, supplies and services required to conduct genomics R&D and associated product development. This industry, of course, generates economic impact.

The Biotechnology Industry Organization (BIO) notes that:

> *"Research, innovation, creativity and the high technologies that result from these inventive activities are the most valuable assets in the U.S. economy. These high technologies…have a long history of delivering for the U.S. economy: between 1990 and 2003 U.S. high-technology generated a $243 billion trade surplus in high-tech goods—compared to a $3.4 trillion deficit in all other U.S. goods combined. The development, commercialization and deployment of advanced technology drive economic growth and opportunity."*[18]

Within the genomics field, there is little doubt among industry and scientific leaders that the Human Genome Project, and the Celera sequencing project, represented critically important stimuli for the genomics industry's development. In interviewing key leaders in the field for this project, the Battelle research team was told over and over again that the HGP, in particular, empowered the industry's development. Battelle asked the question "Without the HGP where do you think we might be today in terms of sequencing technologies and associated genomics technologies?"—some typical responses included:

> *"We would not even be close to where we are today. Would just be doing bits and pieces."*
>
> *"Would not have advanced practically at all. There has really been a dramatic influence here"*
>
> *"The industry would not exist. The HGP set up the competitive space that enabled sequencing and genomics companies to grow."*

Leah Eisenstadt writes that:

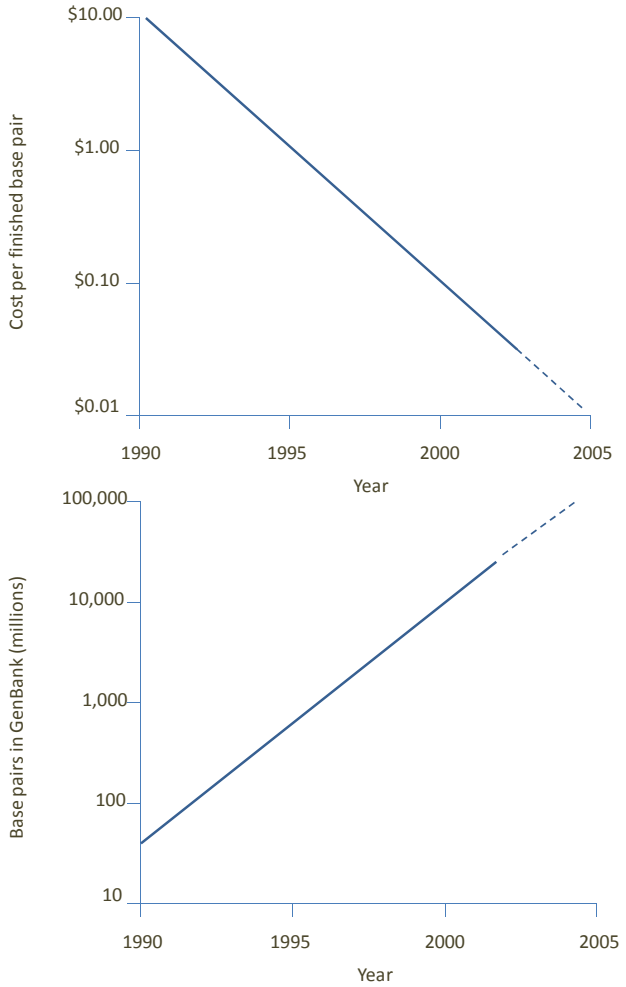
[18] BIO. 2010. *"Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life."* 2010. The Biotechnology Industry Organization. Accessed online at http://www.valueofbiotech.com/

"Once upon a time, sequencing the human genome took tens of millions of dollars and a warehouse full of DNA sequencing machines that analyzed samples throughout the day, and year after year. Now, less than a decade later, the same human genome sequence—the order of nucleotides or "letter"—can be generated using a single machine that analyzes samples for a few days, and for about 100-fold lower cost. The ability to sequence DNA faster and more cheaply comes from recent technological advancements representing the most significant technological metamorphosis in the history of modern genetics". [19]

The progress made in sequencing speed advancement has even outpaced that experienced in computer processors under Moore's Law (see Figure 3). Writing in *Nature*, Eric Lander notes that:

"The per-base cost of DNA sequencing has plummeted by ~100,000-fold over the past decade, far outpacing Moore's law of technological advance in the semiconductor industry. The current generation of machines can read ~250 billion bases in a week, compared to ~25,000 in 1990 and ~5 million in 2000.[20]

Figure 3: Dramatic Cost Reduction and Speed Increases in Sequencing Technology



Graphics adapted from Francis S. Collins, Michael Morgan and Aristides Patrinos. 2003. "The Human Genome Project: Lessons from large-Scale Biology." *Science.* Volume 300, 11 April 2003.

[19] Leah Eisenstadt. 2010. "Beyond the Genome: New Uses for DNA Sequencers."

[20] Eric S. Lander. 2011. "Initial impact of the sequencing of the human genome." *Nature*. Volume 470, February 10, 2011.

It is particularly noteworthy that the reference genome has become extremely important to the analysis of results from the latest generation of sequencing machines. While extremely fast, modern gene sequencing technologies produce reads that are considerably shorter than the 700 bases or so that the electrophoretic methods used in the HGP. It has been noted that "because it is challenging to assemble a genome sequence *de novo* from such short reads, most applications have focused on placing reads into the scaffold of the existing genome sequence to count their density or look for differences from a reference sequence."[21]

In addition to the production of equipment, tools, supplies and technologies for the practice of genomic analysis, there is also significant economic activity in the application of these tools to a range of functional needs in human and veterinary medicine, agriculture and food production, industrial biotechnology and biofuels development, environmental and justice applications. The programs to sequence the human genome thus represent a foundation upon which a diverse range of economic activity has been built.

In consideration of this burgeoning genomics-enabled industry, an assessment of the full impacts of the HGP would be remiss, and in fact severely limited, without including an estimation of this industry and its resulting economic impacts.

Using the methodology described in section A, Battelle calculated the annual employment of the five genomics-specific industry sectors for use as inputs to both an industry-specific impact assessment (Section D) as well as for inclusion in the overall analysis. The inclusion of this employment comes, however, with several assumptions. First, industry employment and related impacts are calculated starting in 1993. This year was chosen for its significance as the beginning of the industry's recognition of human genomics as a distinct opportunity as represented by the private sector engagement in sequencing activities (e.g., Celera Genomics, Human Genome Sciences, Incyte, Millennium, and other private sector investments). Second, the case could be made to begin including industry impacts only after the "completion" was announced, but that would disrespect the private sector technological advancements that occurred during the HGP project that have led to continued corporate successes and impact. Third, while a number of these firms existed prior to the completion and in some cases prior to the start of the HGP efforts, Battelle impact analysts worked to include only firms whose continued existence, products, and innovations were and are currently influenced by the scientific path the complete sequencing of the human genome has enabled. Finally, as discussed earlier, the impact of human genomics is easily recognized but more difficult to quantify within the pharmaceutical industry. For this reason, instead of working to locate specific genomics-related locations and/or to determine "genomics" employment figures for the drug and pharmaceutical manufacturers, we developed an estimation of R&D activities to drive the impacts from this sector. Table 7 provides the employment estimates for the five genomics sectors used in the impact analysis model for four key years.

Table 7: Employment Estimates of Genomics Sectors (Key Years)

"Genomics" Sectors	2010	2003	2000	1993
Genomics R&D/Genomics Biotech	13,323	13,140	8,275	2,378
Genomic Instruments & Equipment	11,704	15,724	10,957	9,917
Genomic-Related Biologics & Diagnostic Substances	7,234	9,427	7,145	2,243
Genomic & Related Testing	5,142	1,644	1,301	542
Genomics-Related Bioinformatics	797	1,430	667	174
Total Employment	37,200	41,365	28,345	15,254

[21] Ibid

The NETS data records illuminate unique trends and changes apparent in the genomics sectors. A decline in bioinformatics employment between 2003 and 2010 occurs due to both declines in the overall information technology industry, but also through the incorporation of early bioinformatics companies into larger entities. Similarly, a decline in genomic instruments and equipment appears to be due to the continued merger and acquisition activity in this sector. Also apparent is the growth in genomics and related testing that has occurred since the completion of the HGP. Finally, it is likely that the recent recession has also continued to dampen employment levels across the board in 2010. Sector employment peaked in 2004 at 42,288.

For the pharmaceutical industry, Table 8 projects the estimated genomics-related R&D as well as total pharmaceutical R&D for three key years. Given the magnitude of pharmaceutical R&D the Battelle team approached the use of these data with conservatism both in terms of the share of genomics R&D and the year in which these shares were first implemented in the model (2000).[22]

Table 8: Pharmaceutical R&D Estimates (Key Years, in Millions Current $)

Pharmaceutical Sector R&D	2010	2003	2000
Genomics-Related Pharmaceutical R&D	9,109	2,337	743
Total Pharmaceutical R&D	45,547	28,853	24,754

Together, these genomics sectors combined with the genomics-related share of the pharmaceutical industry are referred to herein as the ***genomics-enabled industry***.

E. Results of the Input/Output Analysis of the Combined Impact of the Genomics-enabled Industry Base

Based upon the genomic sector employment estimates and the genomics-related pharmaceutical R&D, an industry-specific impact model was developed separate from the federal investments. Table 9 details the results of the most recent year (2010). The impact of human genome sequencing has led to a current industry employing more than 44,000, generating nearly $4.9 billion in annual personal income, and with a genomics-related industry output exceeding $21.4 billion.[23]

Table 9: Economic Impact of the Genomics-Enabled Industry, 2010 (in Millions, 2010 $)

Impact	Employment (Jobs)	Personal Income	Output	State/Local Tax Revenue	Federal Tax Revenue
Direct Effect	44,372	4,889.0	21,401.3	266.4	924.4
Indirect Impacts	104,126	7,309.2	21,904.2	889.0	1,466.0
Induced Impacts	138,173	6,331.5	20,185.5	1,152.2	1,360.1
Total Impact	**286,672**	**18,529.7**	**63,491.0**	**2,307.6**	**3,750.5**
Impact Multiplier	6.46	3.79	2.97	8.66	4.06

[22] It should also be noted that the Battelle estimate of total pharmaceutical R&D may be considered low. According to recent information from the Pharmaceutical Research and Manufacturers of America (PhRMA), biopharmaceutical companies invested nearly $67.4 billion in R&D in 2010 compared to the Battelle conservative estimate of $45.5 billion. One reason for not using this PhRMA estimate is that it is unclear to what extent companies Battelle includes in the five genomics sectors are included in this figure.

[23] Personal income consists of both employee and proprietor income and includes wages and other financial benefits (e.g., retirement funds).

Through the economic multiplier effects of supplier and personal spending the genomics-enabled industry is responsible for nearly 287,000 U.S. jobs, more than $18.5 billion in personal income, and nearly $63.5 billion in U.S. output in 2010. Furthermore, the employment multiplier of 6.46 indicates that for every job directly connected to the genome-enabled industry activities leads to an additional 5.46 jobs in the U.S. economy.

From a tax revenue perspective the genomics sectors and the genomics-related segment of the pharmaceutical industry directly generate more than $266 million in state/local tax revenue and more than $924 million in federal tax revenue for 2010. **When combined with the spending actions of the industry's suppliers and personnel, the overall human genomics-enabled industry is responsible for generating more than $6 billion in local, state, and federal tax revenue in 2010.**

Capturing the cumulative economic impact of the genomics-enabled industry over the 1993–2010 period leads to even more substantial impact figures (Table 10). **During this period the industry directly created more than 591,000 job-years which led to a total of $64.1 billion in personal income and $252 billion in output.**

Table 10: Cumulative Economic Impact of the Genomics-Enabled Industry, 1993–2010 (in Millions, 2010 $)

Impact	Employment (Job-Years)	Personal Income	Output	State/Local Tax Revenue	Federal Tax Revenue
Direct Effect	591,138	64,137.2	251,957.7	3,315.8	11,841.9
Indirect Impacts	1,241,644	86,204.9	257,160.1	10,453.4	17,360.2
Induced Impacts	1,703,013	78,042.6	248,815.6	14,201.5	16,764.2
Total Impact	**3,535,795**	**228,384.7**	**757,933.5**	**27,970.7**	**45,966.3**
Impact Multiplier	5.98	3.56	3.01	8.44	3.88

As a broadly integrated set of firms and sectors, the indirect and induced impacts stemming from the genomics-enabled industry are also substantial. The industry is responsible for generating an additional 2.9 million job-years (for a total impact of more than 3.5 million job-years) during the 18-year period, or an average of more than 196,000 annual jobs). The industry's cumulative direct, indirect, and induced output or "economic impact" is considerable—reaching $758 billion (in 2010 $).

It is also important to consider the size of the total tax revenue impacts generated by the genomics-enabled industry. Including multiplier effects, this industry is responsible for generating nearly $74 billion in local, state, and federal tax revenue over the past 18 years.

F. Combined Analysis –The Cumulative Economic Impact of Human Genome Sequencing

In 2010 alone, the human genome sequencing projects and associated research and industry activity directly and indirectly generated:

- *$67 billion in U.S. economic output*
- *$20 billion in personal income for Americans*
- *310 thousand jobs.*

Assessing the combined and cumulative impact of human genome sequencing includes a summation of the impacts from the federal HGP and post-HGP investments and from the impacts from the genomics-enabled industry. Table 11 captures this combination of impacts for 2010 and Table 12 covers the total 1988–2010 period.

In 2010 alone, the combined impact of the U.S. human genome sequencing activities generated nearly 52,000 direct jobs and a total employment of more than 310,000 jobs in the U.S. economy. The direct economic output of the research community and genomic-enabled industry reached nearly $23 billion with total economic impact in 2010 exceeding $67 billion.

Table 11: Cumulative Economic Impact of Human Genome Sequencing, 2010 (in Millions, 2010 $)

Impact	Employment (Jobs)	Personal Income	Output	State/Local Tax Revenue	Federal Tax Revenue
Direct Effect	51,655	5,577.2	22,627.5	212.3	952.2
Indirect Impacts	109,520	7,593.1	22,725.9	922.5	1,522.8
Induced Impacts	149,185	6,835.7	21,792.6	1,244.0	1,468.4
Total Impact	**310,360**	**20,006.1**	**67,146.0**	**2,378.8**	**3,943.4**
Impact Multiplier	6.01	3.59	2.97	11.21	4.14

The cumulative, multi-year impact estimates include nearly 711,000 direct job-years, a combined personal income direct impact of more than $71 billion, and direct genomics-driven output of nearly $265 billion. In combined direct tax revenue more than $3.5 billion has been generated directly in state/local taxes and more than $13 billion in federal taxes. **From just the perspective of federal revenue and expenses the direct federal taxes generated to date (in 2010 $) have exceeded the HGP and post-HGP federal investments to date ($12.8 billion) by $166 million.**

As with all economic impact assessments, the combined total impact of the genomics research enterprise and the genomics-enabled industry extends well beyond the direct effects. The differences lie in the extent of their economic reach.

Between 1988 and 2010 the human genome sequencing projects and associated research and industry activity directly and indirectly generated:

- *$796 billion in U.S. economic output*
- *$244 billion in personal income for Americans*
- *3.8 million job-years of employment.*

Table 12: Cumulative Economic Impact of Human Genome Sequencing, 1988–2010 (in Millions, 2010 $)

Impact	Employment (Job-Years)	Personal Income	Output	State/Local Tax Revenue	Federal Tax Revenue
Direct Effect	710,819	71,350.0	264,819.6	3,522.3	13,028.0
Indirect Impacts	1,298,216	89,183.0	265,779.8	10,804.8	17,955.6
Induced Impacts	1,818,459	83,328.8	265,663.4	15,163.8	17,899.8
Total Impact	**3,827,495**	**243,861.8**	**796,262.8**	**29,491.0**	**48,883.4**
Impact Multiplier	5.38	3.42	3.01	8.37	3.75

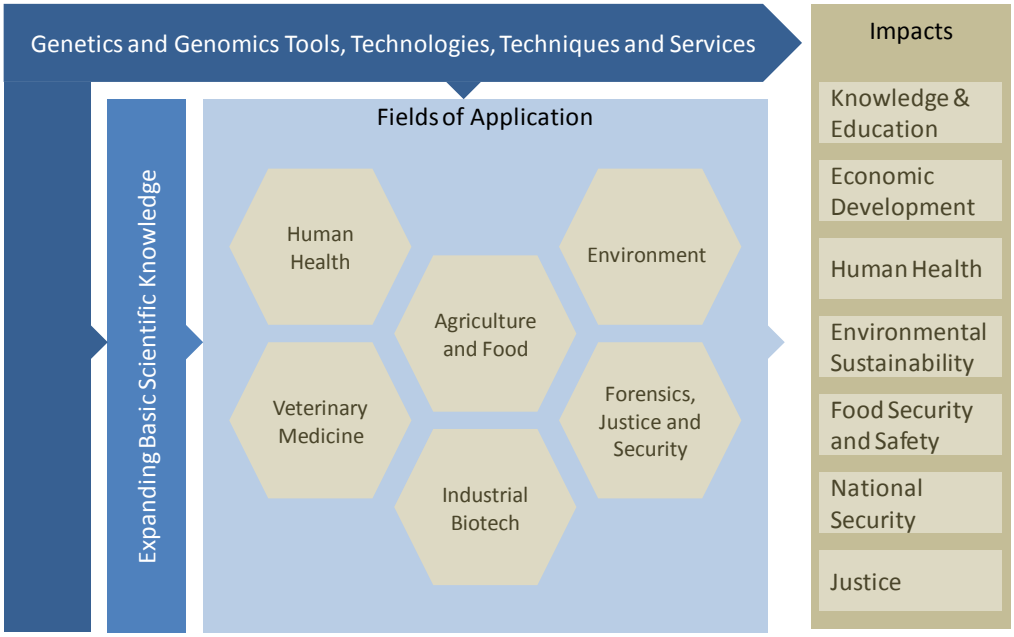
The total employment impact exceeds 3.8 million job-years over the 23-year period. With an employment multiplier of 5.38, genomics research and associated industries generated an additional 4.38 jobs in the U.S. economy for every "genomics" job. This total level of impacted employment generated nearly $244 billion in personal income over the period—in 2010 dollars amounting to an average of $63,700 in personal income per job-year. **The genome sequencing project, the associated research, and the genomics-enabled industry generated a total economic (output) impact of more than $796 billion from 1988–2010.** On a simple calculation basis **the federal investment of $3.8 billion in the Human Genome Project from 1988–2003 ($5.6 billion in 2010 $) enabled the generation of more than $796 billion in economic output for a return on investment (ROI) to the U.S. economy of 141 to 1—every $1 of HGP investment has led to a total of $141 in the economy. Even more important this initial investment will continue to yield both scientific and economic gains for years to come.**

The federal investment of $3.8 billion in the Human Genome Project ($5.6 billion in 2010 $) enabled the generation of more than $796 billion in economic output for a return on investment to the U.S. economy of 141 to 1—every $1 of investment has helped to generate $141 in the economy.

Chapter III: Functional Impacts of the Human Genome Project

The human sequencing programs sought to create benefits for humankind by illuminating the basic molecular processes governing life. It was expected that the resulting advancements in genomics knowledge and technologies would benefit human healthcare, energy, and multiple other fields. In this report we call these benefits *functional impacts.* Figure 4 illustrates the detailed structure of the functional impacts generated by the sequencing of the human genome. There was significant development of genomics tools, technologies and techniques to propel the sequencing efforts forward. These have, in many cases, been commercialized and form the foundation for a highly active and growing commercial genomics-based industry. The application of these genomic tools, technologies and techniques resulted in a truly dramatic expansion of basic biological knowledge.[24] As mentioned in Chapter I, the full human genome sequence unveiled a complex biological system unanticipated by most in science, and has been a paradigm shifting event for biology. Some of the high profile impacts on basic sciences are highlighted in this chapter using illustrative examples and case studies.

Figure 4: The Structure of Functional Impacts Associated with the Human Genome Sequencing.



A. Basic Science and Knowledge Expansion Impacts

The sequencing of the human genome has resulted in a distinct paradigm shift in our understanding of the biology of humans, and indeed all organisms. As such, *ipso facto* the decoding of the human genome stands among the preeminent findings in the history of science.

Until the human genome sequence was published, and its implications understood, the reigning model of thinking in biological science was that biological forms are rendered under a very large but basically

[24] Certain technologies discussed in this report have not received regulatory approval or clearance for clinical uses. In this regard, sequencers are currently intended for research or investigational uses.

Thomas Kuhn first coined the term "paradigm shift" which is, by definition, a rather radical upheaval—a major movement of scientific knowledge and understanding to a new platform. Such movement occurs rarely and research leading to a paradigm shift should, therefore, be viewed as a momentous scientific achievement.

Thomas S. Kuhn. 1962. "The Structure of Scientific Revolutions." University of

linear instruction set coded within DNA. It was expected that life forms and processes result from genes; that each gene generates an RNA template which specifies one protein, and these proteins (greatly simplified) assemble as "life." **The human genome sequence has overturned this one-directional, simple view of the structure of life.** It is now known that:

- The same gene can code for multiple proteins.
- Humans contain far less genes than originally anticipated (about 21,000 instead of upwards of 100,000 to 300,000 once anticipated), yet because of complex processes, such as alternative splicing and post-translational modifications, the result may be over 1 million different protein forms.[25]
- Genes interact with one another and code for multiple RNAs.
- The structural genomic variations between organisms are relatively minor (the human and the chimpanzee, for example, differ from each other by only 1 percent of the DNA sequence), and thus the great variation found across organismal structures is a result of regulatory processes.
- What had been termed "junk DNA" is typically not junk at all and plays important roles in transcription and translational regulation of protein folding
- Most common diseases, such as cancers, heart disease and psychiatric disorders are not homogeneous diseases, but differ dramatically across individual genomes from patient to patient.

The human genome sequence has revealed an immensely more complicated system—one far more complex than previously thought. Rather than the basic linear view of life, what is revealed is a complex network of interactions comprising more players and dialog between players than was thought possible. Professor Steve Jones of University College London described the paradigm shift, noting that:

> *"Now biologists are beginning to face up to the uncomfortable truth that they have only been looking at the nouns in life's lexicon—the crudest and most basic elements of any tongue. Now we are reading the spaces in between—verbs, adverbs, adjectives, pronouns and the rest, and they are complicated indeed. Worse, the genome babbles, stutters and mangles its pronunciation and now and again seems to speak utter nonsense."*[26]

In medicine, genomics has provided the first systematic approaches to discover the genes and cellular pathways underlying disease. Whereas candidate gene studies yielded slow progress, comprehensive approaches have resulted in identification of ~2,850 genes underlying rare Mendelian diseases, ~1,100 loci affecting common polygenic disorders and ~150 new recurrent targets of somatic mutation in cancer. These discoveries are propelling research throughout academia and industry

Eric Lander
Nature. Vol. 470. 2011

The story of the human genome's impact on biology is one of a "revelation of complexity" which will challenge biologists for decades to come. The human genome sequence pushed away a veil and has revealed new pathways to understanding molecular biology and the fundamental processes of life and disease.

The sequencing of the human genome has dramatically altered the process of investigation in biological sciences. Leroy Hood summarized this well:

> *"I saw in the Human Genome Project the introduction of a new type of science in biology—that is, 'discovery driven science.' Discovery-driven science, as compared to hypothesis-driven science, takes an object and enumerates its elements irrespective of any questions. That is, it creates an infrastructure on which hypothesis-driven science can be done far more effectively. …What the Human Genome Project offers us is what I call 'systems biology,' that is, integrating hypothesis-driven and discovery-driven science."*[27]

[25] Swiss Institute of Bioinformatics, Human Proteome Initiative July 2007. Accessed online at http://expasy.org/sprot/hpi/hpi_desc.html

[26] Roger Highfield. 2007. "Life just got a lot more complicated." Daily Telegraph. 19 June, 2007. Steve Jones quoted in this newspaper article was referring to the findings elucidated by the Encode consortium in a follow-on project from the Human Genome project – The Encyclopedia of DNA Elements.

[27] Leroy Hood. 2002. "After the Genome: Where Should We Go?" Chapter in Michael Yudell and Rob DeSalle (editors) "The Genomic Revolution: Unveiling the Unity of Life." 2002. Joseph Henry Press, Washington DC with the American Museum of Natural History.

Sequencing the human genome made clear information sciences, mathematics and biological investigation are now inexorably intertwined. The sequencing of the genome was as much a mathematical and computational achievement as it was a biological one and has helped to give rise to new fields of biological science in "computational biology" and "systems biology". It has been noted that "*The revolutions that have been generated by the first draft of the Human Genome Project have barely been felt, but there is one profound change that has already occurred, and that is the realization that biology is fundamentally an information science*."[28]

Indeed, without the progress made in the latter part of the 20th Century in computer processors, data storage and computational analysis methods, the sequencing of the human genome by the HGP and Celera would not have been possible. Dealing with billions of data points, the efforts of the HGP sequencing centers, for example, were powered by large-scale computing clusters and data storage centers. At the Sanger Centre in the UK, for example, the HGP work engaged a cluster of 250 64-bit Compaq Alpha systems and required terabytes of data storage capacity. High-speed computers were necessary to analyze hundreds of terabytes of raw sequence data and correctly order more than 3 billion pairs of bases. Phil Butcher, head of information technology for the Sanger Centre noted that "*without access to a great deal of the very latest computing resources, we would not even be able to contemplate reaching our targets*."[29] Interestingly, biological sciences are in turn giving back to computational sciences because highly evolved living organisms have developed digital manipulations and strategies that are turning out to be useful within computer science.[30]

For basic science investigation, the human genome sequence has provided a map to guide discovery. Just as a geographic map greatly facilitates directed investigation of a location, so too does the map that is the human genome reference sequence. Eric Lander notes that: "*By providing a comprehensive scaffold, the human sequence has made it possible for scientists to assemble often fragmentary information into landscapes of biological structure and function: maps of evolutionary conservation, gene transcription, chromatin structure, methylation patterns, genetic variation, recombinatorial distance, linkage disequilibrium, association to inherited diseases, genetic alterations in cancer, selective sweeps during human history and three-dimensional organization in the nucleus*."[31]

The fundamental paradigm shifts described above are being felt in the emergence of new scientific sub-disciplines and deep within existing biological and social science disciplines. For example, just some of the centrally impacted science disciplines and sub-disciplines are discussed below.

Genomics Disciplines

The National center for Biotechnology Information at the National Library of Medicine notes that genomics itself has three primary branches, comprising:

- **Structural Genomics** – includes mapping and sequencing
- **Comparative Genomics** – including genetic diversity and evolutionary studies
- **Functional Genomics** – the study of the roles of genes in biological systems.

New and Emerging Disciplines

- **The 'Omics** – At the head of the list, is not one but a series of disciplines that are an outgrowth of genomics and its discoveries, generally grouped by the term 'Omics. They comprise:

[28] Ibid

[29] "Compaq technology enables completion of the human genome." Accessed online at http://www.tgc.com/sponsors/compaq/celera2.html

[30] Leroy Hood. 2002. "After the Genome: Where Should We Go?" Chapter in Michael Yudell and Rob DeSalle (editors) "The Genomic Revolution: Unveiling the Unity of Life." 2002. Joseph Henry Press, Washington DC with the American Museum of Natural History.

[31] Eric S. Lander. 2011. "Initial impact of the sequencing of the human genome." *Nature.* Volume 470, February 10, 2011.

- **Proteomics** – The study of the set of proteins encoded by a genome
- **Metabolomics** – The study of the complete collection of metabolites present in a cell or tissue under a particular set of conditions
- **Transcriptomics** – The study of all RNA molecules, including mRNA, rRNA, tRNA and non-coding RNA in an organism or specific cell.

- **Evolutionary Developmental Biology (EvoDevo)** – Addressing the origins and evolution of embryonic development, this field (an outgrowth of the existing discipline of "embryology") has been empowered by data contained in genome sequencing of human and model organism cells.
- **Computational Biology** – Uses mathematical and computational approaches to address theoretical and experimental questions in biology. Its subdiscipline "computational genomics" performs statistical analysis on the data generated by gene sequencers and microarray technologies to evaluate gene and gene products expressed by various cell types.
- **Bioinformatics** – Is perhaps best viewed as the applied twin of computational biology, and is well defined in the Sanger Institute's glossary as the "science of using computer technology to gather, store, analyze and merge biological data. Expertise in bioinformatics is key to handling the enormous amounts of data produced by the HGP and other sequencing projects, and serving it out to the researchers who use the data."[32]
- **Metagenomics** – Also known as environmental genomics or community genomics, metagenomics investigates the communal genome contained within an environmental sample. It enables the study of the symbiosis and interactions of organismal genomes and genetic products as a biological system.

And, other substantially impacted disciplines, such as

- **Bioarchaeology and Anthropology** – Uses genomics for the study of human evolution and population migrations. It has been noted that "comparative DNA sequence analyses of samples representing distinct modern populations of humans have revolutionized the field of anthropology."[33]
- **Agricultural Sciences** – Whole genome sequencing is being applied across a broad range of crop and livestock species and underpins genetic improvements in input and output traits. Genomics is also being used in diagnostics applications for livestock and zoonotic diseases, plant pathogen identification and in food safety applications. Work is also being performed in developing edible vaccines for incorporation into food products.
- **Environmental Science** – The U.S. Department of Energy and other investigators have sequenced the genomes of multiple nonpathogenic microbes useful for use and development in the environmental waste remediation, energy productions, carbon cycling, and biotechnology applications.

The fundamental advancement of scientific knowledge embodied in the human genome sequencing is also well illustrated by the broad variety of major biological science projects that have developed as follow-on projects or outgrowths of the HGP. Table 13 lists some of these key related initiatives:

[32] http://www.yourgenome.org/glossary/

[33] Judith Fridovich-Keil (primary contributor). "Human Genome Project." Encyclopaedia Britannica Scientific Project.

Table 13: Building Upon the Human Genome Sequence – Related Grand Projects in Biological Sciences

Project Name	Goals
International HapMap Project	The HapMap is a public resource catalog of common genetic variants that occur in human beings. It describes what these variants are, where they occur in human DNA, and how they are distributed among people within populations and among populations in different parts of the world. Using the information in the HapMap, researchers will be able to find genes that affect health, disease, and individual responses to medications and environmental factors.
1000 Genomes Project	Identification of common genetic variation across humans through the sequencing of a large number of individual human genomes. The project aims to find most genetic variants that have frequencies of at least 1 percent in the populations studied.
ENCODE Project	The National Human Genome Research Institute (NHGRI) launched a public research consortium named ENCODE, the Encyclopedia Of DNA Elements, in September 2003, to carry out a project to identify all functional elements in the human genome sequence. The project started with two components: a pilot phase and a technology development phase. The pilot phase tested and compared existing methods to rigorously analyze a defined portion of the human genome sequence. The conclusions from this pilot project were published in June 2007. The findings highlighted the success of the project to identify and characterize functional elements in the human genome. The technology development phase also has been a success with the promotion of several new technologies to generate high throughput data on functional elements.
Microbial Genome Project	The MGP was begun in 1994 as a spinoff from the Human Genome Program. The program sequenced the genomes of a number of nonpathogenic microbes useful in solving DOE's mission challenges in environmental-waste cleanup, energy production, carbon cycling, and biotechnology.
Cancer Genome Anatomy Project	The Cancer Genome Anatomy Project (CGAP), begun in 1996, is an interdisciplinary program established and administered by the U.S. National Cancer Institute to generate the information and technological tools needed to decipher the molecular anatomy of the cancer cell.
Cancer Genome Atlas	The Cancer Genome Atlas (TCGA) is a comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing. TCGA is a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), which are both part of the National Institutes of Health, U.S. Department of Health and Human Services.
Cancer Genome Project	Headquartered at the Wellcome Trust Sanger Institute, the Cancer Genome Project is using the human genome sequence and high throughput mutation detection techniques to identify somatically acquired sequence variants/mutations and hence identify genes critical in the development of human cancers
Human Microbiome Project	Operated by the U.S. NIH Common Fund, the Human Microbiome Project (HMP) aims to characterize the microbial communities found at several different sites on the human body, including nasal passages, oral cavities, skin, gastrointestinal tract, and urogenital tract, and to analyze the role of these microbes in human health and disease.
Genomes to Life Project	Genomes to Life program aims to use microbes and other organisms to address problems in energy production, environmental cleanup, and carbon cycling. The research seeks to understand the chemistry of entire organisms and their interactions with the environment. The project's ten-year goal is to advance systems biology, computation, and technology. These advances will be directed toward increasing biological-based sources of energy, better understanding the earth's carbon cycling, designing novel ways to capture carbon, and developing low-cost methods for cleaning the environment.

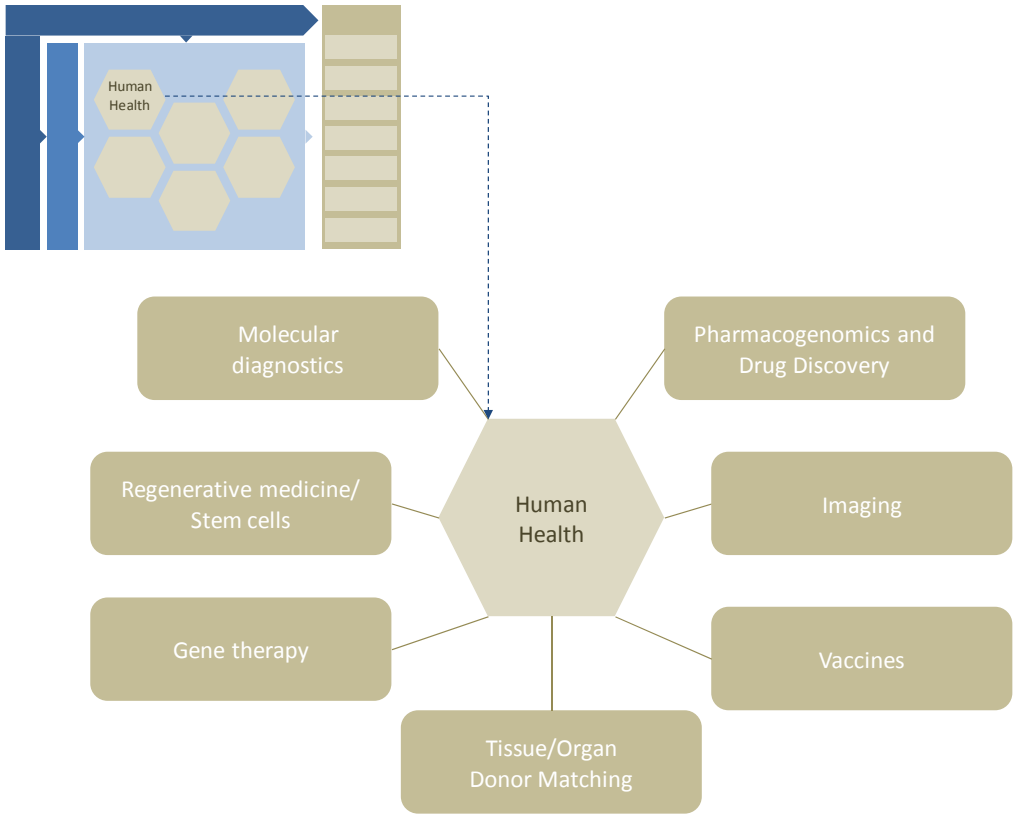
B. Human Health and Medicine

One of the key hopes for the sequencing of the human genome was that, having the entire genome in hand, biomedical scientists would be able to better identify the processes and mechanisms of disease and inherited medical disorders and illuminate new approaches to treating these diseases and disorders.

Modern medicine is built upon a history of fundamental investigations and discoveries across a broad range of life science disciplines: physiology, biochemistry, microbiology, immunology, etc. It is also, however, built upon a foundation of technological innovations enabling these discoveries to be made. Whether it is the invention of the microscope leading to the identification of microorganisms, or the modern development of functional magnetic resonance imaging allowing researchers to watch internal life processes in action, technologies provide new capabilities that propel medical advancement. The sequencing of the human genome actually represents an intense advance on both fronts—both in terms of fundamental knowledge and technology advancement.

As noted above, modern genomics and the human genome sequences have empowered a new view of biological processes. In turn, these new views are being translated into applied biomedical fields, particularly those illustrated in Figure 5.

Figure 5: Functional Impact Areas of Genomics in Human Health



The tools and technologies of modern genomics, in concert with the foundational knowledge in the human genome sequences, are finding application across a broad range of biomedical areas today, and hold extraordinary promise for advancements in the future. Certainly, the complexity of the genome and proteome discussed under basic science implications above have made “instant cures” difficult to

achieve, but the progress being made in applied genomic medicine is quite remarkable. Table 14 illustrates some of these applications:

Table 14: Biomedical Applications of Genomics and Human Genome Sequencing

Potential Application	Genomics Advances Today	Hope for the Future
Diagnosis of single gene, Mendelian diseases and disorders	Specific genes for over 3,000 Mendelian monogenic diseases discovered. Genomic tests are being used to accurately diagnose rare diseases and disorders, many of which were previously misdiagnosed with inappropriate courses of treatment prescribed. Prenatal genetic screening is being performed to inform potential parents of risks for catastrophic inheritable disorders.	Gene therapies will achieve success in repairing genetic abnormalities leading to diseases and disorders. Custom therapeutic products will block or change expressed activity of defective genes.
Knowledge of predisposition towards specific diseases	Multiple genes and biomarkers have been identified for predisposition to multiple diseases such as cancers, neurological diseases, psychiatric disease and cardiovascular disease.	Understanding of risk for disease based upon multi-gene tests will likely lead to appropriate therapeutic interventions and personal behavior/lifestyle modification. Environmental components of disease emergence and progression will be teased-out from genomic factors and addressed appropriately.
Genomics driven drug discovery, known as rational drug development	New drug targets have been identified. Cancer drugs based on the genomics of tumors are on the market, including Gleevec (for chronic myelogenous leukemia), Herceptin (breast cancer), Tarceva (lung cancer) and Avastin (colon, lung and other cancers).	Many new drugs and biologics will be developed to successfully exploit elucidated drug targets.
Therapeutic products custom prescribed based on patient genomics, to maximize effect and reduce or eliminate side effects	Already being applied in the treatment of some forms of cancer and cardiovascular disease. Genetic tests are used for dosage levels in prescription of some drugs such as Coumadin (warfarin).	Routine sequencing of a patient's entire genome wil guide treatment selection and dose for the optimum response. Potential adverse reactions to drugs and treatment regimens, identified via genomic markers, will result in the avoidance of adverse events.
Repurposing or revitalization of some drugs shelved in development because of impact on a "genomic few"	Successful discovery of subpopulations for which previously unapproved drugs are efficacious. Iressa, for example, has been approved with patents testing positive for the EGFR mutation.	There will be a substantial volume of existing drugs found to be efficacious in selected sub-populations, and drug companies will have mined their previously "failed" R&D pipelines to bring forward previously non-marketable drugs to work in selected sub-populations.
Identification of means to combat infectious organisms	Multiple infectious organisms have had their whole genomes sequenced. Public health professionals sequence emerging infectious disease organisms to monitor migrations and mutations. Genetic testing already being applied to direct therapy for HIV/AIDS patients.	Rapid, real-time sequencing of pathogens twill direct public health efforts to combat disease outbreaks in humans and zoonotic disease outbreaks in animals. DNA vaccines will come to the market to impart immunity to infectious diseases. Genetic profiling of patient viruses will assist in customizing treatment regimens.
Gene therapies for inherited genetic diseases and disorders	After publicized setbacks, gene therapies are now achieving success. For example, the fatal brain disorder adrenoleukodystrophy has been treated, with progression stopped, in a sample of children.	Gene therapy may be routinely provided to newborns with identified genomic profiles to correct defective genes, particularly in conditions associated with devastating monogenic disorders.

*"Modern medicine arose when scientists learned to fight some of the worst infectious disease with vaccines and drugs. This strategy has not worked with AIDS, malaria, and a range of other diseases because of their complexity and the way they infiltrate processes in cells. Curing such infectious diseases, cancer, and the health problems that arise from defective genes will require a new type of medicine based on a thorough understanding of how cells work and the development of new methods to manipulate what happens inside them."**

The HGP was and is one of the central projects leading to this "understanding of how cells work" and opening the way for new applications of molecular medicine.

Russ Hodge, "The Future of Genetics: Beyond the Human Genome," 2010.

We are on the leading edge of a true revolution in medicine, one that promises to transform the traditional "one size fits all" approach into a much more powerful strategy that considers each individual as unique and as having special characteristics that should guide an approach to staying healthy.

Francis Collins,
"The Language of Life,"
New York: Harper Collins, 2010

Molecular diagnostics and pharmacogenomics are working hand-in-hand as applications facilitating the growth of "personalized medicine" which uses analysis of patient genomes to assure the right medicine, in the right dose, gets to the right patient.[34] The sequencing of the human genome has enabled the identification of a significant volume of genes and biomarkers to be associated with diseases and health disorders. Biotechnology companies develop molecular diagnostic tools that identify the presence of these biomarkers in patients. The association of a biomarker may indicate susceptibility to a disease, therefore guiding prophylactic treatments or other preventive measures—but biomarkers may also indicate the likely response of a specific patient to a particular therapeutic product. This is the world of pharmacogenomics and pharmacogenetics in which the presence of a single gene, biomarker or collection of genes and biomarkers across the patient's genome can indicate potential efficacy of a particular therapeutic product, dosage or treatment regimen.

The Biotechnology Industry Organization (BIO) notes that the use of genetic and genomic information can help determine:

- Which treatment will likely work best;
- The safest and most effective dosage; and
- Whether a person is predisposed to develop a specific disease later in life.[35]

Likewise, the American Medical Association (AMA) notes that the growth of personalized medicine via pharmacogenomics will "provide tailored drug therapy based on genetically determined variation in effectiveness and side effects."[36] The AMA notes that this will mean:

- "More powerful medicines – Pharmaceutical companies will be able to produce therapies more targeted to specific diseases , maximizing the therapeutic effects while decreasing damage to nearby healthy cells."
- Better, safer drugs the first time – Recovery time will go down and safety will go up as the likelihood of adverse reactions goes down or is eliminated altogether."
- More accurate methods of determining appropriate drug dosages – Current methods of basing dosages on weight and age will be replaced with dosages based on a person's genetics – how well the body processes the medicine and the time it takes to metabolize it.
- Better vaccines – Vaccines made of genetic material, either DNA or RNA, promise the benefits of existing vaccines without all the risks. They theoretically could activate the immune system but be unable to cause infections."[37]

While the AMA sees much hope for the future in pharmacogenomics and personalized medicine, the Association notes that applications are already changing the practice of medicine in selected areas, including cancer, depression and cardiovascular disease.

From an impact standpoint, developments in pharmacogenomics and personalized medicine hold significant promise for more accurate diagnosis of disease, elucidation of preventive measures to avoid disease, more refined disease treatment and avoidance of costs associated with adverse drug reactions.

[34] A more technical definition by the journal Nature: "The use of genetic susceptibility or pharmacogenetic testing to tailor an individual's preventive care or drug therapy. *Nature "*Glossary" accessed online at: http://www.nature.com/nrg/journal/v5/n12/glossary/nrg1495_glossary.html

[35] BIO. 2010. *"Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life."* 2010. The Biotechnology Industry Organization. Accessed online at http://www.valueofbiotech.com/

[36] American Medical Association. "Current Topics: Pharmacogenomics." Accessed online at http://www.ama-assn.org/ama/pub/physician-resources/medical-science/genetics-molecular-medicine/current-topics/pharmacogenomics.page

[37] Ibid

CASE STUDY: Reducing the Economic Impact of Chronic Disease

> *Health is both a personal good and an economic good. Poor health, on the other hand, carries significant personal and economic burdens. Analysis by the Milken Institute indicates that the seven most prevalent chronic diseases in the U.S. in 2003 cost the nation $277 billion in treatment expenditures and $1.05 trillion in lost economic output (largely related to lost personal productivity related to disease), for a combined national cost of $1.3 trillion. Milken projected this cost to rise to over $2 trillion by 2011.*[38]
>
> *The discoveries embodied in human genome sequencing are being directly applied to address progress in the diagnosis and treatment of each of the seven chronic diseases whose impact was quantified by Milken (cancers, diabetes, heart disease, hypertension, stroke, mental disorders and pulmonary conditions). The impact of improved human health is tangible and significant—as the Milken Institute report authors note "good health is an investment in economic growth."*
>
> *Genomic medicine will likely provide cost saving opportunities across multiple fronts in healthcare. Pharmacogenomics (already being applied in cancer treatment for example) allows drugs to be prescribed only to patients with biomarkers indicating the efficacy and safety of a specific drug (avoiding wasted treatments and problem side effects). Patients with a genetic predisposition to develop adult onset diseases or disorders can use this knowledge, with the advice of their physician, to adopt prophylactic treatment regimes or adapt their lifestyle to reduce risk. Drugs that had been shelved due to negative side-effects in a few patients can be rescued for those whom genomics shows may be safely administered the drug—thereby providing new drugs to the pharmacopeia and recapturing the earlier significant development costs in the R&D of those drugs. Over time, advancement in gene therapy will allow disorders and diseases to be completely averted, while already pre-conception genomic tests can provide advice to potential parents or marriage partners regarding risk of devastating illness to their progeny, thus providing actionable information to avoid the heartache and cost of fatal childhood congenital diseases. What is gradually occurring, because of the revolution in part stimulated and facilitated by the sequencing of the human genome, is a move to personalized, data-driven medicine in which the efficacy of a treatment regime is known in advance, rather than the large-scale "prescribe and hope" medicine that dominates clinical practice today. The prevention of disease and the accurate treatment of disease, embodied in the promise of modern genomics, will have a significant impact on the cost of healthcare and the economic burden of disease on society.*

Cancer is one of the areas of human disease in which genomics is playing a fundamental role in the clinic. BIO notes that "new tests have been launched recently that identify patients likely to respond to the following cancer treatments:"[39]

- Tarceva® (Genentech, non-small cell lung and pancreatic cancers);
- Gleevec® (Novartis, chronic myeloid leukemia and gastrointestinal stromal tumor); and
- Campath® (Genzyme, B-cell chronic lymphocytic leukemia).

Mayo Clinic research showed that a recent study showed that hospitalization rates dropped 30 percent when genetic information was used to determine the best dosing for heart patients taking warfarin (the world's most prescribed blood thinner).[40]

---

[38] Ross DeVol and Armen Bedroussian. 2007. "An Unhealthy America: The Economic Burden of Chronic Disease." Milken Institute, October 2007.

[39] Biotechnology Industry Organization. 2010. "Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life." BIO online at www.valueofbiotech.org

[40] Medco and Mayo Clinic. Clinical News Release. March 16, 2010. Accessed online at: http://www.mayomedicallaboratories.com/mediax/articles/features/warfarin/press-release.pdf

CASE STUDY: Tailoring Breast Cancer Treatment

"Some of the most exciting new biotech diagnostic tests help improve breast cancer treatment and care. One of these diagnostic tests determines whether a breast cancer patient has an aggressive form of breast cancer associated with the human epithelial growth factor receptor-2 (HER-2). This form of breast cancer strikes quickly, spreads aggressively, and is often deadly. However, patients treated with Herceptin® (produced by Genentech) experience extremely positive outcomes: two recent clinical trials found that chemotherapy plus Herceptin cut the chance of cancer recurrence in half compared to chemotherapy alone.

Judy Kayse, a breast cancer survivor and patient, knows firsthand the power of the latest breast cancer treatments and diagnostic tests. "I was first diagnosed with breast cancer 15 years ago. At the time there was no testing for HER-2 or a treatment for people with that type of breast cancer, so I was treated with a lumpectomy and radiation and went into remission," recounts Judy. "But at my annual exam in late 2009, we found a tumor. The cancer was back, and it was invasive. But this time my doctors had a new weapon in their arsenal: they were able to test for the HER-2 factor. They determined I was HER-2 positive, meaning I was eligible for Herceptin. It's now part of my treatment program, and doesn't come with the debilitating side effects of other breast cancer treatments. My doctor told me if Herceptin had been available the first time around, I probably would not have had a recurrence."

Herceptin is a tailored treatment that works by blocking the HER-2 receptor, stopping the cancer tumor from growing and eventually killing the cancer cells. Only patients with HER-2 positive breast cancer are eligible for Herceptin—making the HER-2 diagnostic test critical to determining the appropriate breast cancer treatment."

"Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life."
2010. BIO, The Biotechnology Industry Organization (www.valueofbiotech.com)

CASE STUDY: Mayo Clinic Uses Whole-Genome Sequencing as a Diagnostic Tool

The renowned Mayo Clinic has a long track-record in seeing highly challenging cases—the sickest of the sick, patients with unusual illnesses and the most difficult to diagnose cases. Now heavily invested in sequencing technologies and advanced genomics tools, Mayo was early to recognize the future of diagnostics in genomics and is on the front-lines in deploying genome sequencing in patient care. Two recent patient cases illustrate this in action.

Mayo Clinic in Scottsdale, Arizona and the Translational Genomics Research Institute (T-Gen) in Phoenix, Arizona , worked together to sequence Mayo's first whole-patient genome. This involved sequencing both normal cells and tumor cells from a patient with pancreatic cancer (requiring sequencing of over 6 billion base pairs), and, with the sequences in hand, comparative analysis enabled the clinical team to pinpoint "genetic changes that were crucial in tailoring a more effective treatment plan for the patient."[41] Mitesh Borad, a Mayo assistant professor, pointed to the latest generation of gene sequencers and their speed as making such advance possible. He also noted that "utilization of this approach are clearly the way forward, and it is only a matter of time before whole- or targeted-genome sequencing becomes routine in the clinical sphere, ranging from primary care to specialty areas such as oncology."[42]

The Mayo Clinic's team has also sequenced the full genome of an acute myelogenous leukemia patient and found a novel mutation. This led to analysis of additional samples and

[41] Matthew Dublin. 2011. "Researchers Demonstrate Feasibility of Whole-Genome Sequencing in the Clinic." GenomeWeb. April 2011. Accessed online at http://www.genomeweb.com/sequencing/researchers-demonstrate-feasibility-whole-genome-sequencing-clinic

[42] Ibid

identification of 15 additional acute myelogenous leukemia patient cases with the same mutation.

Eric Wieben, Director of Mayo Clinic's Genomic Research Center, and a leader in the Minnesota Partnership for Biotechnology and Medical Genomics, noted in an interview with Battelle that these types of applied genomics advances would not have occurred without the Human Genome Project, because prior to the HGP sequencing technology was stable, but static—it did not have the speed to do what is being done today, and the impetus was not there to rapidly advance the technology. The HGP spurred the advancement of technology to the point that today the speed and cost of whole-genome sequencing is putting it at the doostep of clinical application. Wieben noted that in addition to the cases above, three additional genomes were also recently fully sequenced by Mayo as part of an effort to find genes related to hereditary heart disease. Wieben illustrated the extent of the Mayo genomics commitment and the importance of the HGP to it, noting that Mayo has a staff of about 120 research personnel working in medical genomics, and that "each one of these staff access the reference human genome every day."[43] Mayo Clinic is using medical genomics to develop next generation sequencing-based diagnostics for cancers and mitochondrial diseases, and with sequencing costs continuing to drop whole genome sequencing is expected to be an important clinical tool moving forward. It will become "mainstream" Wieben noted.

One of the key applications of genomics and pharmacogenomics is the ability to "fit the drug and the dose to the patient's genomic profile." As the following case study illustrates, this personalized medicine approach helps to avoid adverse drug reactions and the prescription of drugs unlikely to work for a specific patient genotype.

CASE STUDY: Avoiding the Costs and Physical Harm of Adverse Drug Reactions

The biopharmaceutical sector is one of the leading contributors to the considerable increase in human longevity achieved in the 20th Century. Millions of lives have been saved by the prescription of medicines to combat disease, and the quality of life of almost every American is touched by the biopharmaceutical sector (via vaccines, pain and other symptom relief in chronic disorders, and the effective treatment of disease).

Unfortunately, in a small percentage of cases, the administration of a biopharmaceutical can have serious adverse effects. While we all share the commons characteristic of being homo sapiens, we are also all different—different at a genomic level (with each of us having our own unique genome) and different in the individuality of our daily lives and its impact on our physical health. It is fair to say we "are the same, but different" and the relatively small differences in each of us (many of them genomic differences) may sometimes lead to an adverse drug reaction.

The sequencing of the human genome has enabled scientists to gain a far better understanding of adverse drug reactions as they relate to patient genomics. In some instances, there are now genetic tests on the market that enable physicians to identify the risk of an adverse drug reaction in advance for their patient, and to determine an alternative course of therapy. Genotype-to-drug association studies will increasingly play an important role in personalizing medicine—advancing efficacy and avoiding negative impacts.

While, as noted above, adverse drug reactions represent a small percentage of reactions to administered biopharmaceuticals, in a world with 7 billion people, small percentages quickly add up to large actual numbers of people and significant costs. Genomics is being applied to help alleviate these challenges.

[43] Eric Wieben, Mayo Clinic. Telephone interview with Battelle impact research team. On April 15, 2011.

> *Bond and Raehl studied the issue of adverse drug reactions (ADRs) and report that "based on the literature from 1964-1996, overall incidence of ADRs in hospitalized patients was 6.7 percent (range 1.2–24.1 percent), and of fatal ADRs 0.32 percent (0.1–0.85 percent)."[44] They note that:*
>
> *"These aggregate figures translate to 2,216,000 hospitalized patients each year who experience a serious ADR and 106,000 a year who die from an ADR. Fatal ADRs rank fourth to sixth in leading causes of death. As sobering as these figures appear, ADRs also are one of the more frequent causes of hospitalization (3.7–6.5 percent of patients). Cost estimates for these ADRs are $1.56–$4 billion per year, and as many as half of them may be preventable." Genomics-enabled personalized medicine approaches will help to reduce ADR's in many instances. It should be noted, however, that many ADR's are the result of human errors that genomics would not affect.*
>
> *Clearly, progress being made in genomic tests to avoid ADRs (built upon the backbone of knowledge derived from human genome sequencing) is critically important to avoiding harm and lowering the cost to society of adverse drug reactions.*

J. Craig Venter notes that being able to separate patients into those who will benefit from a drug and those who will not, or those that may experience an adverse drug event and those who will not, allows drugs to be used that otherwise could not be approved. In effect, the power is there to "rescue" drugs that could not progress into the market because of limited negative impacts on a small subset of patients. In 2002 Venter noted that

> *"A type II diabetes drug recently had to be taken off the American market because 1 out of 10,000 people had a severe liver toxicity to it. If we can find simple tests that predict toxicity it will have a huge impact. Not only can we change adverse drug effects in the population, we can also tailor drugs so that they work for more than 30 to 50 percent of the population, the current average."[45]*

Venter's hope for developing genomic tests to refine drug prescription has now come to fruition in a number of cases. Among the important benefits of pharmacogenomics is the paring of a therapeutic with a companion diagnostic. The Food and Drug Administration (FDA) already lists over 70 drugs with pharmacogenomics information labeling (covering aspects such as indications, usage, dosing etc.). The Table 15 indicates examples of key pharmacogenomics-labeled drugs and the major disease/disorder indications with which they are associated:

Table 15: Selected Key Pharmacogenomic Labeled Drugs

Drug	Indications
Abacavir (Ziagen)	HIV
Atorvastatin (Lipitor)	Reduce cholesterol levels
Azathioprine (Imuran)	Prevention of transplanted kidney rejection; osteo-arthritis
Carbamazepine (Tegretol)	An anti-convulsant. Seizure control; trigeminal neuralgia; mania and bi-polar disorder.
Cetuximab (Erbitux)	Head and neck cancer; colon cancer
Clopidogrel (Plavix)	Antiplatelet drug for prevention of stroke and heart attack
Dasatinib (Sprycel)	Leukemia

[44] C.A. Bond and Cynthia Raehl. 2006. "Adverse Drug Reactions in United States Hospitals." Pharmacotherapy. 2006; 26(5):601-608. Accessed online at: http://www.medscape.com/viewarticle/531809

[45] J. Craig Venter. 2002. "Whole-Genome Shotgun Sequencing" Chapter in Michael Yudell and Rob DeSalle (editors) "The Genomic Revolution: Unveiling the Unity of Life." 2002. Joseph Henry Press, Washington DC with the American Museum of Natural History.

Imatinib (Gleevec)	Leukemia and other cancers of the blood cells; gastrointestinal stromal tumors (GIST); dermatofibrosarcoma protuberans (tumors under skin)
Irinotecan (Camptosar)	Colon cancer; rectal cancer
Panitumumab (Vectibix)	Colon cancer; rectal cancer
Rasburicase (Elitek)	High levels of uric acid in chemotherapy patients
Trastuzumab (Herceptin)	Breast cancer
Volproic acid (Stavzor)	An anti-convulsant. Seizures; mania; bi-polar disorder; migraines
Warfarin (Coumadin)	Anticoagulant used in cases of irregular heartbeat; heart valve replacement; heart attack treatment; venous thrombosis; pulmonary embolism.

Source: PubMed Health

CASE STUDY: Pharmacogenomic Drug with Companion Diagnostic for the treatment of HIV/AIDS

> *Pfizer's drug "Selzentry" was introduced in 2007 and represents the first "new-class" of drug for combating HIV in more than a decade. The drug is marketed with a companion diagnostic test called the Trofile assay which detects whether a patient's virus is CCR5-tropic. For patients whose virus is so characterized, Selzentry blocks HIV from attaching to the CCR5 receptor thereby preventing the virus from being able to enter and infect the patients' immune cells.*

CASE STUDY: Rational Drug Design

> *Rational drug design typically relies heavily upon computer modeling and genomics to modify an existing drug or design a new drug that will interact specifically with a selected molecular target important in disease progression. Novartis' drug Gleevec (Imatanib) is an example of an FDA-approved drug on the market today that resulted from a rational drug design approach. Gleevec is a protein-tyrosine kinase inhibitor and works to block the action of an abnormal protein that promotes the proliferation of cancer cells. Simply taken as a tablet once or twice per day, Gleevec works on a range of cancers including: certain types of chronic myelogenous leukemia and other cancers of blood cells; gastrointestinal stromal tumors (GIST);, and a type of tumor that occurs under the top layer of the skin termed dermatofibrosacoma protuberans.*
>
> *The Wellcome Trust notes that "rational drug design involves the design and synthesis of compounds based on the known structure of either a specific target or one of its natural ligands. The results of the HGP and human pathogen genome projects provide many new drug targets."*

For patients with extremely rare disorders, diagnosis has historically presented a serious problem. Primary care physicians and pediatricians may never have encountered a specific type of monogenic disorder that presents in a patient, and unfamiliarity with a rare disorder can lead to an incorrect diagnosis, the wrong treatment, negative side effects from the wrong treatment, and ongoing suffering from the disorder or disease. It is particularly difficult for a physician to diagnose a rare disease when its symptoms may mirror another more common disease. Without definitive genomic tests, physicians are often running practically blind; having to use a one-size fits all approach to their diagnoses and treatments.

In addition to the impact on diagnostics and drugs, the refined understanding of the human genome and related cellular processes has reinvigorated the nascent area of gene therapy. The development of a refined understanding of the genome and the actions of individual genes and groups of genes has paved the way for techniques to change the code in DNA molecules, thereby giving them a new instruction set.

Gene therapy focuses on finding ways to introduce corrective genes into cells in order to:

- Correct a cell malfunction;
- Add a new function to a cell; or,
- In the case of cancer, add a gene to a cancer cell that causes that cell to die.

A variety of different types of gene delivery systems are used to deliver genes into target cells, including:

- Modified viruses, which appear to be very efficient at getting genetic information into cells;
- "Naked" DNA containing the corrected genes; or
- Artificial lipids carrying new DNA.

The goal is to match the appropriate delivery system with the gene, the target cell, and the disease, in order to develop an effective therapy. While suffering early setbacks, gene therapy is now making considerable strides forward and effective treatments are starting to emerge from the development pipeline. Examples include:

- **Adrenoleukodystrophy (ALD)** – is a neurodegenerative disease resulting in progressive brain damage so severe that death typically occurs within 2–5 years post diagnosis. The research team used a modified-HIV viral vector to deliver gene therapy that has resulted in successfully arresting the progress of the disease in children.
- **Leber's congenital amaurosis** – is a congenital eye disease in which a mutated gene results in the lack of production of protein necessary for correct functioning of the retina. Gene therapy in which cells are introduced to create the protein in the retina has resulted in improved vision in the subject patients.
- **Wiskott-Adrich syndrome** – is an immunodeficiency disorder which renders patients open to severe recurrent infections, low blood platelet counts and chronic eczema. Gene therapy trials in children have generated marked improvement in patient condition.

The above disorders are rare diseases for which pharmaceutical development costs are extremely hard to absorb. Gene therapy offers an alternative approach to effective treatment of rare diseases, which total more than 7,000 individual diseases and disorders affecting 25 million Americans and 250 million persons worldwide.[46]

Other early-stage gene therapy clinical research areas include promising results with:

- **HIV/AIDS –** a cell therapy procedure administered in Germany conveyed a gene into the patient that blocks the CCR5 receptor which HIV needs to enter and infect cells.
- **Cancer –** Patients in Scotland have responded positively to trials of gene therapy to address skin cancers (melanoma) with 90 percent of treated cases seeing complete remission from tumors.
- **Adenosine deaminase deficiency –** a severe immunodeficiency disorder akin to the famous "bubble boy" disorder has been treated in successful trials with adenosine deaminase gene transfer.
- **Beta thalassaemia –** a hematological disorder involving a genetic defect that causes defective production of hemoglobin (a disease of over 100,000 infants worldwide annually) has been treated in successful gene therapy trials using lentiviral vectors.

[46] Global Genes Project. http://www.globalgenesproject.org/press/GGPWRDD2011_Final.pdf

CASE STUDY: Sight for the Blind

> *Born with a rare retinal disease called Leber's congenital amaurosis, Corey Haas had lost most of his sight by the time he was 7 and was legally blind. He often clung to his parents when going out, and depended on a teacher's aide, Braille, a large-type computer screen, and cane while at school. Corey expected to eventually lose his vision entirely.*
>
> *But just two years later he enjoys the same activities as any active, healthy 9 year old boy: go-karting, hiking, and playing baseball.*
>
> *Corey's miraculous recovery was due to an experimental gene therapy procedure. In fall 2008, he received an injection in his left eye. The injection contained a virus specially engineered to replace DNA containing a defective version of the gene responsible for his debilitating condition, RPE65, with a normal version of that same gene. In fact, Corey was one of 12 patients in a groundbreaking study on the potential applications of gene therapy for patients suffering from Leber's congenital amaurosis; all patients showed significant improvement in both subjective and objective measures of their vision.*
>
> *"Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life."*
> *2010. BIO, The Biotechnology Industry Organization. (www.valueofbiotech.com)*

Infectious diseases take a terrible toll on human life. According to the World Health Organization (WHO) approximately 15 million people die each year due to infectious diseases.[47] Infectious diseases are caused by bacteria, viruses, fungi and other parasites are major causes of death, disability, and social and economic disruption. The complex nature of even microorganismal genomes has made fighting certain infectious diseases an extremely difficult task—however, now whole genome sequencing has been applied to a very broad range of infectious pathogens, vectors and parasites to help in the global battle against infectious disease.

Viruses are one of the significant causative agents in human disease such as AIDS, encephalitis, hepatitis, influenza, SARS etc. The advances in modern sequencing equipment, built on the research backbone of the human genome sequencing, have been successfully turned to sequencing viruses. Today, the NIH National Center for Biotechnology Information (NCBI) Entrez Genome database contains 3,805 reference sequences for 2,621 viral genomes and 41 reference sequences for viroids.

Likewise, a broad variety of human bacterial pathogens, infectious fungi, protozoa and helminth parasitic worms have also been sequenced. The Sanger Institute (one of the leading participating centers in the HGP), for example, lists just some of the major pathogens sequenced and the diseases associated with them in Table 16.

Table 16: Examples of Pathogen Sequencing Projects and Associated Human Diseases

Bacterium	Associated Diseases
Bordetella	Whooping cough
Chlamydia trachomatosis	Chlamydia
Clostridium botulinum	Botulism (typically food-borne)
Clostridium difficile	Causative pathogen in multiple diseases, typically nosocomial, including severe diarrhea to life threatening colitis
Corynebacterium diphtheiae	Diphtheria
Klebsiella pneumonitae	Pneumonia, liver abscesses, septicemia
Neisseria meningitidis	Bacterial meningitis
Proteus mirabilis	Urinary tract infections

[47] World Health Organization. (2008) WHO global burden of disease: 2004 update. Available from: www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html

Salmonella	Causes broad array of illnesses. Multiple species and subspecies of this bacteria sequenced
Staphylococcus aureus	A nosocomial and community-acquired pathogen. *S. aureus* is the most common cause of hospital acquired infection
Staphylococcus lugdunensis	Invasive endocarditis
Streptococcus pyogenes	A wide variety of illnesses, including streptococcal toxic-shock syndrome, acute rheumatic fever, scarlet fever, and others.
Tropheryma whipplei	Whipple's disease.
Vibrio cholerae	Cholera
Yersinia pestis	Plague
Protozoa	**Associated Diseases**
Bodo saltans	African sleeping sickness, Chagas disease and leishmaniasis
Leishmania (multiple)	Leishmaniasis
Plasmodium (multiple)	Malaria
Trypanosoma brucei	Sleeping sickness
Helminthes	**Associated Diseases**
Ascaris	Intestinal worms
Echinococcus multilocularis	Intestinal tapeworms
Onchocerca volvulus	River blindness
Schistosoma	Blood flukes associated with schistosomiasis and bilharzia

Source: http://www.sanger.ac.uk/resources/downloads/bacteria.

Armed with this information, infectious disease specialists, microbiologists, immunologists, and homeland security professionals are better able to envision and develop approaches to combat infectious disease. Genomic analysis is also being applied to understanding the natural immunity that some individuals have to infectious organisms. Russ Hodge, noted science writer for the Max Delbruck Center for Molecular Medincine, makes note of the central importance of the HGP in allowing these new approaches to disease to develop:

> *"Modern medicine arose when scientists learned to fight some of the worst infectious disease with vaccines and drugs. This strategy has not worked with AIDS, malaria, and a range of other diseases because of their complexity and the way they infiltrate processes in cells. Curing such infectious diseases, cancer, and the health problems that arise from defective genes will require a new type of medicine based on a thorough understanding of how cells work and the development of new methods to manipulate what happens inside them."*[48] *The HGP was and is one of the central projects leading to this understanding of how cells work and opening the way for new applications of molecular medicine.*

Another approach facilitated by the sequencing of the human genome is that of Genome Wide Association Studies—an approach that involves rapidly scanning markers across an individual's genome, to find genetic variations associated with a particular disease. Once new genetic targets are identified, researchers can use the information to develop better strategies to detect, treat and prevent the disease.49 It has been noted that "by discovering large collections of genes that can modulate a phenotype, GWAS has begun to reveal underlying cellular pathways and, in some cases, already pointed to new therapeutic approaches."[50] Eric Lander notes several important examples of key GWAS findings associated with major diseases, including:[51]

[48] Russ Hodge. 2010. "The Future of Genetics: Beyond the Human Genome." Facts on File Publishing, New York.

[49] http://www.hhs.gov/myhealthcare/glossary/glossary.html

[50] Eric S. Lander. "Initial impact of the sequencing of the human genome." Nature. Volume 470, February 10, 2011.

[51] Ibid

- Adult Macular Degeneration
- Crohn's Disease
- Type 2 Diabetes
- Autoimmune Diseases
- Kidney Disease
- Psychiatric Disorders

"*Genome-wide expression analysis has also had a central role in classifying cancers based on their molecular properties, rather than anatomic sites. Studies have revealed distinctive subtypes and shed light on metastatic potential. Expression signatures are already used in the clinic to predict which breast cancer patients will benefit most from adjuvant chemotherapy or surgery.*"[52] Despite the usefulness of GWAS, genomics is an extremely fast moving field and researchers from Duke are finding the latest sequencing equipment enables an even more detailed approach (see case study below).

CASE STUDY: Whole Genome Sequencing for Pinpointing Disease Causing Variants

> *In recently published work, researchers at Duke University are finding that "sequencing may be better than genome-wide association studies at finding causal variants in common diseases."*[53] *"The researchers came to this conclusion after performing whole-genome sequencing studies on 29 individuals and finding that rare variants are significantly more likely than common ones to be functional."*
>
> *"Genome-wide association studies have been used to try to pinpoint the genetic underpinnings of common disease, but have so far only been able to explain a small proportion of the predicted heritability. Rather, the Duke study suggests that rare variants, which GWAS so far have not been designed to find, may be responsible for common diseases.*
>
> *To arrive at their results, researchers in David Goldstein's lab at Duke sequenced 29 individuals to an average 28-fold coverage, and called 5,491,245 single nucleotide variants. They found that for all the functional categories, when the variants are more rare they are more likely to be in the functional region and therefore more likely to cause phenotypic effect.*
>
> *"Sequencing is definitely more powerful [than GWAS] for detecting rare variants. With microarrays the SNPs are already fixed, so they are mostly common variants. But, when you do whole-genome sequencing, you find all the variants that you can find, including both rare and common.*

Other areas of biomedical research and translational science are also using and leveraging the human reference sequence and human genome sequencing technologies to advance. Stem cell researchers, for example, are comparing normal stem cells and cancer stem cells to find pathways to new cancer treatments—for example in leukemia. Understanding the molecular biology of human stem cells and gene regulation in cell differentiation, facilitated by the learning from human genome sequencing, holds promise for the use of these cells for a range of therapeutic applications in regenerative medicine and disease treatment. The NIH reports that "stem cells, directed to differentiate into specific cell types, offer the possibility of a renewable source of replacement cells and tissues to treat diseases including Alzheimer's disease, spinal cord injury, stroke, burns, heart disease, diabetes, osteoarthritis, and rheumatoid arthritis."[54]

---

[52] Ibid

[53] Monica Heger. 2011. "Duke Team Says Sequencing May be Better than GWAS at Pinpointing Disease-Causing Variants." GenomeWeb. April 05, 2011. Accessed online at http://www.genomeweb.com/print/965719

[54] National Institutes of Health. "Stem Cell Information." Accessed online at: http://stemcells.nih.gov/info/basics/basics6.asp

Imaging is also a field both influencing genomics and benefiting from it. On the benefit side of the equation, genomics serves to identify new surrogate markers for therapy monitoring, which may be used as tracers for imaging. Imaging and genomics also go hand-in-hand with the discipline of functional genomics, whereby technologies such as functional magnetic resonance imaging (fMRI) provide the opportunity to explore and evaluate the functional impact of genetic polymorphisms.

As the examples, case studies and narrative above show, the human genome sequencing project has had significant and profound impacts on human medicine and public health. The complexity of the genome and associated regulation activity revealed by the HGP has made progress in therapies and cures move at a pace perhaps slower than some predicted, yet as Nadia Rosenthal notes:

> *"Whether in regard to the early detection and prevention of congenital disease, the promise of a longer and healthier lifespan through a better understanding of the role of gene's in aging, or the creation of new drugs based on a person's unique genetic makeup, the impact of genetics on medicine will continue to grow."*[55]

C. Veterinary Medicine

It should not be surprising that the projects to sequence the human genome, which also incorporated the sequencing of other animal organisms for comparative purposes, has proven to be of benefit to veterinary medicine in ways similar to that of human biomedicine benefits. For example, writing in 2009 M. Breen noted that:

> *"The release of an annotated human genome sequence assembly and the emergence of genomics technologies have led to significant advances in our understanding of many human diseases including cancers. As DNA sequencing technology has become less costly, the field of comparative genomics has progressed rapidly and attention has turned now to generating whole genome assemblies and dedicated genomics resources for veterinary species. Such progress brings a whole new series of opportunities to advance veterinary medicine. Many human and animal diseases share a pathogenetic basis, and although veterinary species need advances in biomedical research in their own right, the consideration of companion animals also as good comparative models for human disease saw the emergence of the "one medicine" concept. The future of many areas of human and veterinary biomedical research is very much interdependent, with one of the closest associations being in oncology. It is inevitable that veterinary oncology will benefit enormously from data derived from genomics and that this era will see a huge shift in the ways in which companion animal cancer patients are evaluated and subsequently treated."*[56]

The College of Veterinary medicine and Biomedical Sciences at Texas A&M University introduces the field of biomedical genomics by noting that:

> *"Genomics is a key and rapidly evolving research discipline that continuously elucidates new dimensions within the dynamic structure of modern biomedicine. The animal genomics research community has witnessed an astonishing expansion and diversification throughout the past decade. Completion of the human and mouse genome sequences sparked enormous interest in the annotation of these genomes through a series of detailed comparative analyses. As a result, genome sequences are now available for nearly all major mammalian companion and livestock species, including cattle, pig, dog, horse, cat, alpaca and chicken, as well as biomedical models such as the rat, rabbit, opossum, shark and Xenopus genomes. Rapid methodological and computational advancements afforded by genome sequence data*

[55] Nadia Rosenthal "Forward" in Russ Hodge. 2010. "The Future of Genetics: Beyond the Human Genome." Facts on File Publishing, New York.

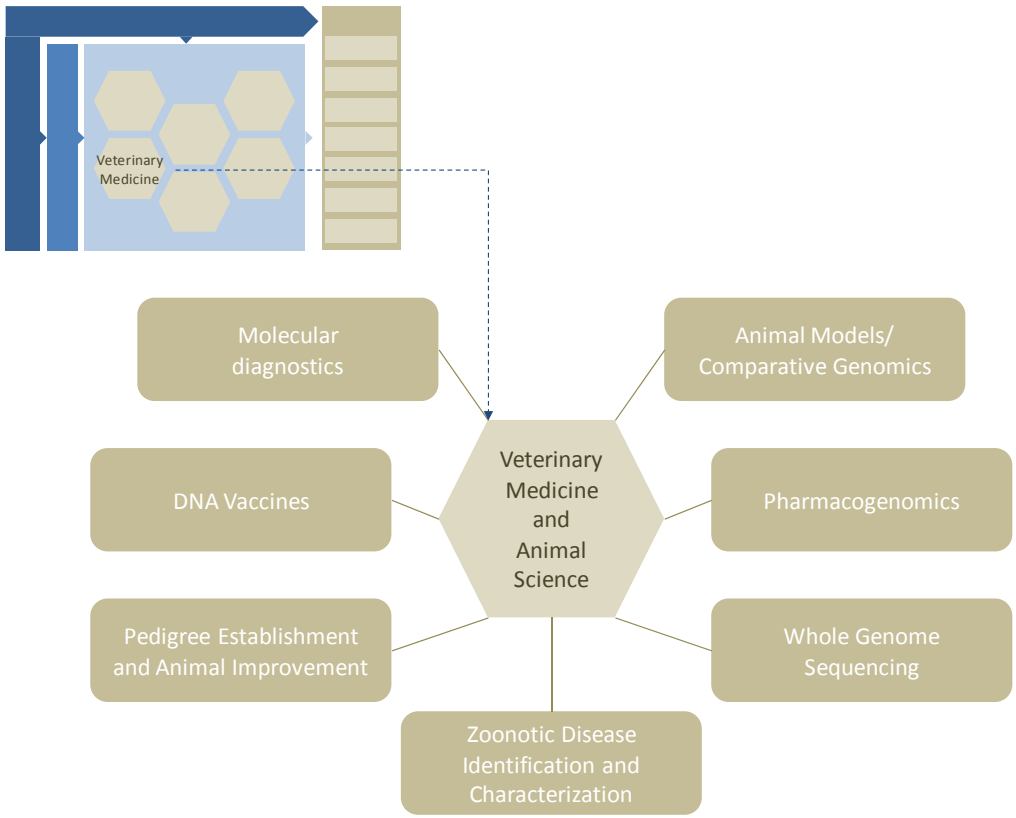
[56] M. Breen. 2009. "Update on genomics in veterinary oncology." Topics in Companion Animal Medicine, 2009. Aug; 24(3):113. Accessed online at: http://www.ncbi.nlm.nih.gov/pubmed/19732729

are providing the means for new areas of research that could barely be imagined just a few years ago. The study and analysis of animal genomes continuously provides new technologies to better understand phenotypes of biomedical and economic importance in these species, while at the same time informing human biology."

Figure 6 illustrates the major areas of veterinary medicine and animal science being impacted by the sequencing of the human genome and the resulting growth in bioscience knowledge and associated genomics technologies. As with human biomedicine, veterinary medicine is benefiting from:

- Genomics-based molecular diagnostics product development
- Development of therapeutics products
- Vaccine development
- Animal improvement.

Figure 6: Functional Impact Areas of Genomics in Veterinary Medicine and Animal Science



Early in the HGP, biomedical scientists realized the value of fully-sequenced mammal and other organism genomes for comparison to the human genome—in part as model organisms for the study of disease. The result has been major sequencing projects directed at animals used in research (the laboratory mouse, chimpanzee, Norwegian brown rat, rhesus macaque, etc.), companion animals (cat, dog—also used in research) , and agricultural animals (such as the horse, pig, cattle, sheep, goats and chickens). In total, the NCBI Entrez Genomes Database lists 1,111 eukaryotic genome sequencing projects, including 40 completed, 408 in assembly and 663 in progress. The list of mammals alone that have been sequenced, are in assembly or have whole or partial genome sequencing is substantial (Table 17).

Table 17: Current Mammalian Organism Sequencing Projects

Completed	Assembly	In Progress
Laboratory mouse	Giant panda	Woolly mammoth
Chimpanzee	Cattle (bos taurus)	Middle-African hedgehog
	Water buffalo	Cattle (bos indicus)
	White tufted-ear marmoset	Bactrian camel
	Dog	Coyote
	Guinea pig	Gray wolf
	Hoffman's two-toed sloth	Goat (domestic)
	Nine-banded armadillo	Sunda flying lemur
	Ord's kangaroo rate	Ring-tailed lemur
	Lesser hedgehog	Crab-eating macaque
	Horse (domestic)	Chinese pangolin
	Western European hedgehog	European polecat
	Cat (domestic)	Bonobo
	Gorilla	Olive baboon
	Alpaca	Yellow baboon
	African elephant	Hamadryas baboon
	Rhesus macaque	Guinea baboon
	Tammar wallaby	Chacma baboon
	Gray mouse lemur	California mouse
	Gray, short-tailed opossum	White-tailed mouse
	Little brown rat	North American deer mouse
	White cheeked gibbon	Oldfield mouse
	American pika	Indus river dolphin
	Duck-billed platypus	Greater horsehoe bat
	Rabbit	Tasmanian devil
	Bush baby	
	Sheep (domestic)	
	Sumatran orangutan	
	Cape rock hyrax	
	Large flying fox	
	Rat	
	European shrew	
	Thirteen-lined ground squirrel	
	Pig (wild)	
	Philippine tarsier	
	Northern tree shrew	
	Bottlenose dolphin	

Source: NCBI Entrez Genome Database

Other categories of animal are also being sequenced, such as reptiles (e.g. the green anole), birds (chicken and zebra finch) and fish (tilapia, salmon, and pufferfish, for example).

These sequencing projects provide data that is highly useful for a broad range of applications: for comparison to the human genome; for the study of evolutionary processes; as model organisms for biological study, and for veterinary medicine. While companion animals, livestock, poultry and economically important aquatic species (such as tilapia, shrimp, salmon etc.) are a key target for sequencing work, there is also application to veterinary medicine in the diverse family of genomes for work with rare and endangered species and invasive species.

It should also be noted that much of the sequencing work performed on pathogenic organisms, parasites and disease vectors has also targeted economically significant diseases infecting livestock and poultry, together with pathogens affecting companion animals. Work in veterinary medicine using genomics and genetics to detect, diagnose and treat disease is also of importance to human health and medicine in the cases where pathogens cause zoonotic diseases (infectious diseases that can be transmitted from an animal to a human).

CASE STUDY: Dog Genomics

> *For pet owners, their pets are part of the family and their health maintenance a legitimate concern. The American Veterinary Medical Association reports that there are 72 million dogs in the U.S. and 82 million cats, with the average annual veterinary expenditure for a household with pets totaling $366 in 2006.[57] Over 79,000 veterinarians are in practice in the U.S.[58]*
>
> *The domestic dog (canis familiaris) is close to full completion and already the impact of genomics is being felt in veterinary medical practice. Multiple DNA mutation tests are on the market for canine-inherited diseases, helping vets diagnose dog disorders ranging from eyesight disorders (such as progressive retinal atrophy) to severe immunodeficiency disorders and hemophilia. A broad variety of veterinary medicine and diagnostics companies have incorporated genomics into their R&D and product development pipelines, such as Optigen, PennGen Laboratories and VetGen.*
>
> *Work in human genomics is also transferring discoveries of relevance to veterinary medicine, and comparative genomics allows the dog to be used as a model for human diseases. In addition, genomics is a tool being applied to purebred animal breeding and the development of breeds with specific desired characteristics.*

The extent to which genomics has penetrated the world of veterinary medicine research can be illustrated by the depth of course offerings in the field at major veterinary medicine colleges in large U.S. land-grant universities. At Texas A&M, for example, courses and research focus areas are available in:[59]

- Mammalian genomics
- Comparative genomics
- Population genomics
- Conservation genomics
- Phylogenomics and genome evolution
- Immunogenomics
- Functional genomics
- Mammalian disease genomics
- Genomics of sex and reproduction
- Computational genomics
- Epigenomics

The human genome field has led the way both in terms of completeness of data and in developing tools and applications. Animal genomics broadly follows a similar route, but there is a major difference in that animals are selected to express (or repress) specific traits, and in fact have been deliberately selected over several thousand years. This means that they also form a unique resource for comparative genomics with other species, including humans. In addition, the use of selection and quantitative genetics in animals is well advanced and this gives the discipline of systems biology at the animal, or population, level a leading edge over the human field. Not only, therefore, does farm animal genomics have the potential to improve sustainable agricultural production, but in key areas it is a tool for developing human genomics research.

Peter Burfening, et al.,"The Future of Livestock Genomics," EC-US Task Force on Biomedical Research.

[57] American Veterinary Medical Association (AVMA). 2007. "U.S. Pet Ownership & Demographics Sourcebook (2007 Edition)."

[58] American Veterinary Medical Association (AVMA). "Veterinary Market Statistics". Accessed online at: http://www.robertsonproperties.com/Veterinary%20Market%20Statistics%20-%20U.S.pdf

[59] Texas A&M University. Accessed online at: http://vetmed.tamu.edu/research/signature-programs/biomedical-genomics

D. Agriculture and Food

Population growth is a major challenge for agriculture because most of the world's usable farmland is already in production. In 1991, 0.81 acres of farmland was available to feed each person. By 2050, only 0.37 acres of farmland will be available for each person. It means the productivity of each unit of land must be increased.

BIO. 2010 "Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life"

Humankind's purposeful manipulation of genetics actually has its roots in thousands of years of agriculture. While certainly early man did not understand the biology behind his actions, the practice of selected breeding of plants and animals was a fundamental genomic process. By gradually improving crops and domesticated animals, humans were able to form location stable societies—generating a food surplus that enabled the development of civilization.[60]

Today, perhaps even more so than in medicine, the impact of modern genomics is seeing widespread application across agriculture. Via transgenics, plants are being improved in terms of input traits (efficient uptake of fertilizer, drought tolerance, pest resistance, etc.) and output traits (nutritional content, food quality, biomass output, etc.). Facilitating this new green-revolution is the fundamentally improved knowledge of genomics and molecular biology stimulated by the projects to sequence the human genome.

The relevance and importance of genomics to agriculture is hard to overstate. Some of the most pressing challenges facing the global community have their solutions rooted in agriculture—issues such as food security, human health, environmental sustainability and economic growth. The scope of these challenges can seem overwhelming. In food security, for example, worldwide population is projected to increase from 7 billion people in 2010 to 9.3 billion by 2030 (an increase of 2.3 billion, equivalent to doubling the entire current population of China and India).[61] As Battelle notes in its recent report for 12 land-grant universities:

> *To meet the rising demand for food (driven both by rising population and increasing income levels), it is anticipated that by 2030 we may actually need to double global food production,[62] yet most cultivatable land is already in production.[63] The inequity of global income levels and access to food across the planet already leads to debilitating levels of malnutrition, undernutrition and associated poor health for over 1 billion people.[64] Meeting the demands of a growing population, for more and better food, and the demand for the fiber, fuels and materials required as inputs for economic growth, is made all the more difficult by the pressing need to do so while reducing environmental impacts and global climate change associated with human economic activity. Against this background of global need and challenges, of threats and opportunities, it becomes clear that the role of agriculture, and associated agbioscience advancements, has come once more to the fore as a critical driver of humankind's future."[65]*

Agriculture has to work to address these daunting challenges using the land already in production, since land not currently cultivated is typically of low soil quality, in forested areas that need to be environmentally protected or in areas with low levels of water availability. Increasing production to the level needed, doing so with lower levels of resource use, and producing more nutritious foods to serve the malnourished are tasks that agricultural genomics directly addresses.

[60] Biotechnology Industry Organization. 2010. "Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life." BIO online at www.valueofbiotech.org

[61] Simon J. Tripp. "Power & Promise: Agbioscience in the North Central United States". Battelle Memorial Institute, April 2011.

[62] Bruce M. Chassey, Wayne A. Parrott and Richard Roush. "Crop Biotechnology and the Future of Food: A Scientific Assessment." CAST Commentary, QTA 2005-2, October 2005.

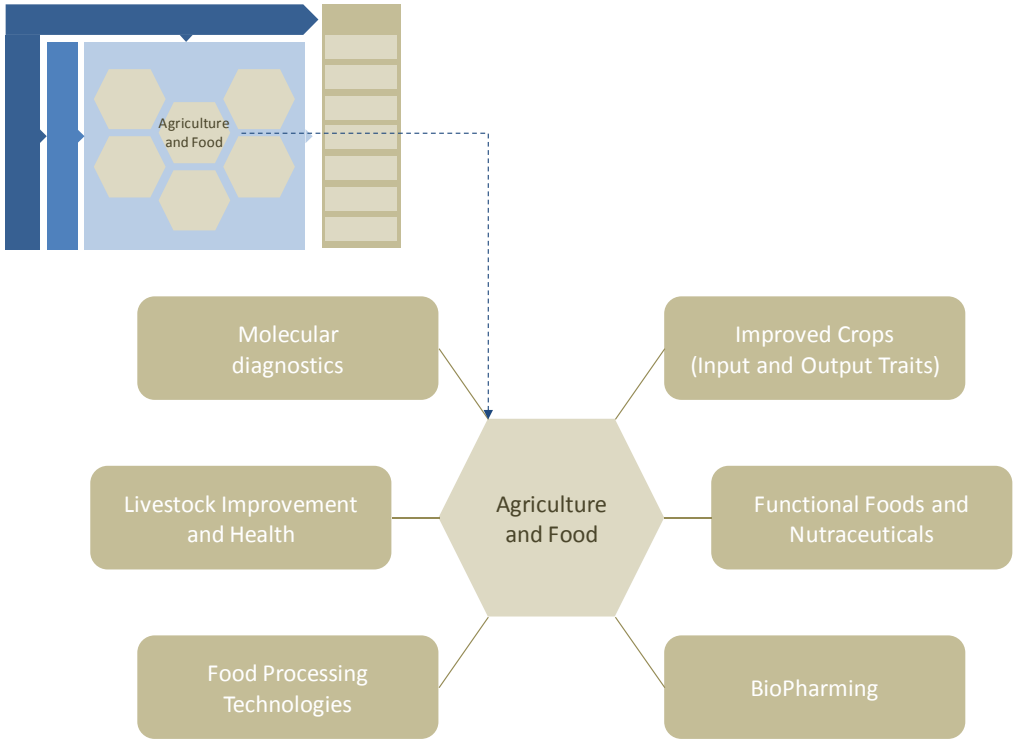
[63] United Nations Food and Agriculture Organization (FAO). "World Agriculture Towards 2015/2030, Summary Report." Accessed online at www.fao.org/documents/show_cdr?url_file/DOCREP

[64] United Nations Food and Agriculture Organization (FAO). "The State of Food Insecurity in the World." Accessed online at www.fao.org/publications/sofi/en/

[65] Simon J. Tripp. "Power &Promise: Agbioscience in the North Central United States." Battelle Memorial Institute, April 2011.

While the application of genetics has a long-history in agriculture, the modern ability to sequence entire plant and animal genomes takes agbioscience to a new level of understanding regarding organismal molecular biology. The tools and technologies of genomics are as applicable to plant and livestock biology as they re to human biomedical biology. Some of the major applications of genomics in agriculture and food are illustrated in Figure 7:

Figure 7: Functional Impact Areas of Genomics in Agriculture and Food.



The U.S. Department of Energy notes that:

> *"Understanding plant and animal genomes will allow us to create stronger, more disease-resistant plants and animals --reducing the costs of agriculture and providing consumers with more nutritious, pesticide-free foods. Already growers are using bioengineered seeds to grow insect- and drought-resistant crops that require little or no pesticide. Farmers have been able to increase outputs and reduce waste because their crops and herds are healthier."*[66]

Indeed, as early as 2002, as the results of the HGP were first being experienced, Robert Bozell noted that "the use of genomics in agricultural biotechnology has already yielded improvements in pest resistance and drought resistance and can be used to preserve biodiversity."[67]

The advancements made in sequencing technology, especially the greatly enhanced speed of sequencing and the dramatically reduced cost of sequencing have made it economically feasible for agbiotech companies, universities and government research groups to deploy the full force of genomic analysis in pursuit of food and agriculture advancement. On the animal agriculture front, sequencing programs in cattle are allowing rapid identification of single nucleotide polymorphisms (SNPs)—with

[66] U.S. Department of Energy Genomic Science Program – Human Genome Project Information. 2009. "Potential Benefits of Human Genome Project Research." Accessed online at: http://www.ornl.gov/sci/techresources/Human_Genome/project/benefits.shtml

[67] Robert Bazell. 2002. "Introduction: Applications of Genomics to Medicine and Agriculture." Chapter in Michael Yudell and Rob DeSalle (editors) "The Genomic Revolution: Unveiling the Unity of Life." 2002. Joseph Henry Press, Washington DC with the American Museum of Natural History.

over 62,000 SNPs identified through the use of massively parallel sequencing equipment to produce short reads across 66 cattle. Next-generation sequencing is also allowing agbioscientists to probe expressed regions of the cattle genome and provide fresh insight on gene regulation impacts on cattle development, meat content, milk production, health and reproduction. Similar applications will occur across other livestock and poultry species with advanced genome sequencing programs, including pigs, chickens, goats and sheep.[68]

CASE STUDY: Unraveling the Amazing Genomic Complexity of Crop Plants

> *The application of genome sequencing to animals is one thing, doing it with some crop plants, however (despite the importance of crop plants to human existence), has been daunting. On the face of it, it seems counterintuitive that a crop plant could have a far larger genome than a complex, thinking, walking, human—yet that is indeed the case. Wheat, for example, has a genome five times the size of the human genome. The wheat genome is 17 gigabases long—a result of the hybridization/breeding process—and the resulting wheat genome actually has three sets of DNA. Crop genomes also tend to have large-scale areas of repetitive sequences making it challenging to reconstruct the sequence of bases on each chromosome (of which wheat has 42).*
>
> *The importance of wheat as a staple crop is hard to overstate, with wheat flour being the basis for such basic foodstuffs as bread and pasta. Indeed, wheat is the leading source of vegetable protein in the human diet. The United Nations reports that 676 million tons of wheat was produced in 2010. With global population and demand increasing, finding ways to increase wheat production, and lessen losses to plant stressors such as disease or drought is critically important. Plant transgenics is proving to be a powerful tool for improving input and output traits in plants, but its promise is best realized if agbioscientists have a complete understanding of the plant's genome. The large-scale human genome sequencing program spurred the development of the high speed/affordable cost sequencing technologies that are today being applied to key crop plants. If the expansion of production challenge is to be met for the world's growing population, it is transgenic crop improvement that is going to get us there—and it is human genome sequencing that propelled the development of the technologies that empower modern crop genomics.*

Genetically modified crops, also known as biotech crops, are now on the frontlines of addressing the agricultural productivity imperative. Matt Ridley, science writer for *The Economist* notes that:

> "*Genetically modified crops are proving to be an unmitigated environmental miracle. Herbicide-tolerant plants are now grown with minimum tillage, which reduces the soil erosion that results from ploughing. Drought-tolerant plants are nearing the market and salt-tolerant ones are not far behind. Within a decade, there may be crops that are no-till, insect-resistant, omega-3-enriched, drought-tolerant, salt-tolerant and nitrogen-efficient. If they boost yields, then the 21st century will see more and more people better and better fed from less and less land.*"[69]

In 2009, more than 330 million acres of biotech crops were grown in 25 countries and 14 million individual farmers chose to grow biotech crops. As BIO reports the findings of researchers that:

> *The adoption of biotech crops has not been limited to farmers in the developed world: 2009 data indicate that almost 93 percent (13 million) of the 14 million farmers growing biotech crops in 2009 were small and resource-poor farmers from developing countries such as China, India, Philippines and South Africa. In terms of total acreage of biotech crops planted, the*

[68] George E. Liu. 2009. "Applications and case Studies of the Next-Generation Sequencing Technologies in Food, Nutrition and Agriculture." Recent Patents on Food, Nutrition and Agriculture, 2009. 1,75-79.

[69] Matt Ridley. 2009. "The NEU thing: nitrogen-use efficiency, the next green revolution." The Economist. 13 November 2009.

developed world (with its larger farms) planted 40 percent of the acreage, while developing countries accounted for 60 percent. Between 1996 and 2009 more than 2.3 billion acres of biotech crops have been successfully grown as a result of approximately 70 million repeat decisions by farmers to grow these crops.[70]

The major categories of benefits from the application of genomics to agriculture are broad, including:

- **Increased agricultural yields and agricultural productivity in crops**—by modifying plant genomes to develop selective herbicide resistance, pest and disease resistance, drought tolerance, faster growth, larger grains, etc. Similar approaches can be taken for forestry production.
- **Increased yield in livestock—**genomics allows agbioscientists to modify livestock, poultry and fish genomes to increase meat content, lower fat content, enhance the uptake of nutrients in the animal, and enhance animal reproduction. Genomics is also powering veterinary medicine to combat pathogens affecting animal health and livestock productivity.
- **Improved food quality, nutrition and health products**—genomics can be applied to develop crops with customized vitamin and nutrient profiles to enhance diets and can be grown to express specific vitamins and even edible vaccines. Plants can also be used for "biopharming" in which plants are used as factories for the production of biopharmaceutical products and the production of nutraceuticals and functional foods. Likewise, transgenic animals may be developed to grow organs for human transplantation.
- **Reducing the environmental impact of agriculture**—genomics allows agbioscientists to develop crops that require the application of lower levels of pesticides or herbicides, that are suited to no-till agronomy, or that require less water for successful growth.
- **Enhanced food safety**—genome sequencing has been applied to several key pathogens that have generated food-borne illness outbreaks, such as E. coli. The ability to genotype strains of infectious organisms allows the FDA to trace the source of outbreaks and directly address the issue.
- **New and expanding biobased industries**—here genomics is being applied to grow a sustainable biobased economy in which biomass is used to derive bioenergy, liquid biofuels, sustainable biobased chemicals, plastics and materials. Industrial biotechnology uses engineered microbes to produce enzymes and other biological agents for the production of valuable products.

The coming decades will be unparalleled and will place plant researchers in the position of being able to modify the nutritional content of major and minor crops to improve many aspects of human and animal health and well-being.

Martina Newell-McGloughlin, Director, University of California Systemwide Biotechnology Research and Education Program (UCBREP) and Co-Director, NIH Training Program in Biomolecular Technology

CASE STUDY: Genomics of Rice, the World's Largest Staple Crop

> *"Two groups of researchers sequenced the genomes of related subspecies of rice. Rice is a food staple for over three billion people. It provides 20 percent of the world's dietary energy supply, while wheat supplies 19 percent and maize 5 percent. Scientists are using the data to create improved varieties of rice that grow in different environments and have desirable traits such as pest resistance and increased nutritional value. Advances made with rice also may be applied to other important crop species such as wheat and corn. Initially, it was estimated that rice had between 43,000 and 63,000 genes. The estimates were later lowered to about 38,000 genes"*[71]

[70] Biotechnology Industry Organization. 2010. "Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life." BIO online at www.valueofbiotech.org

[71] National Human Genome Research Institute. "2002: Rice Genome Sequenced." Understanding the Human genome Project: Dynamic Timeline. Accessed online at: http://www.genome.gov/25520488

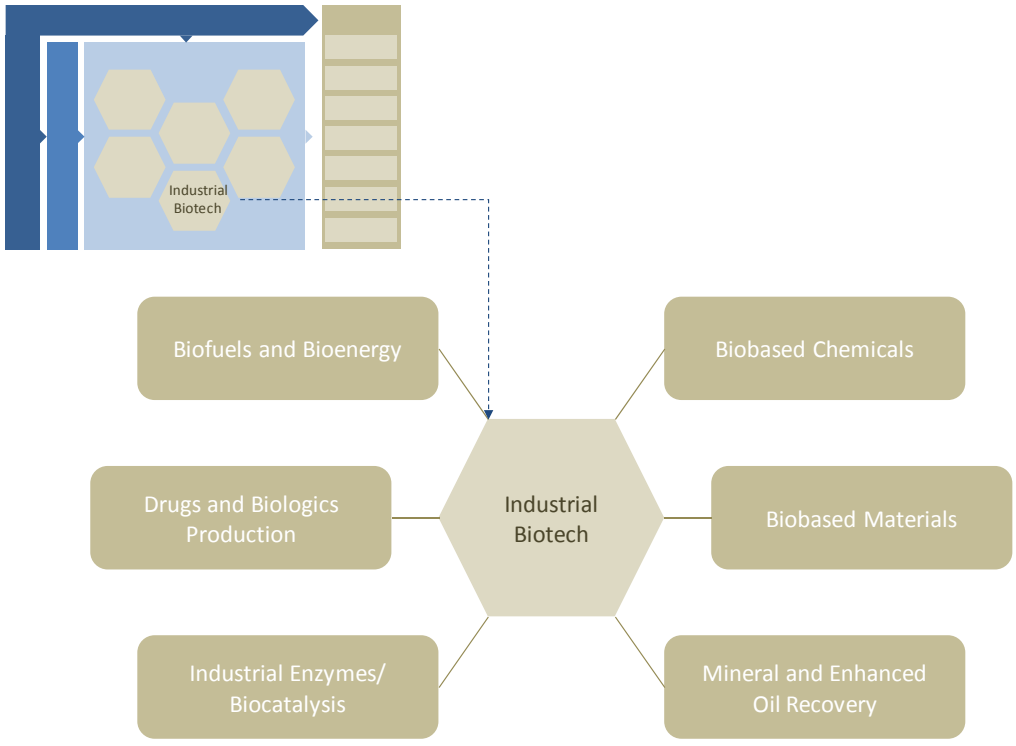
E. Industrial Biotechnology

Today, modern sequencing technologies and genome manipulation techniques are allowing scientists to modify microbes for a tremendous range of uses, and even to synthesize artificial microbial life forms. In doing so, industrial biotechnologists and research scientists are applying the tools of genomics, facilitated by the human genome sequencing development programs, to achieve impacts in many industrial areas.

Biological organisms, especially microbes, have evolved capabilities to thrive in extreme environmental conditions (heat, cold, pressure, and even radioactive environments). Characteristically, these are also environments (particularly the first three) encountered in many industrial processes, especially within the chemicals and energy industries. Recognizing the ability of microorganisms to perform functional activities in these extreme environments, biotechnologists have adapted microbes to use in activities in chemical processing, biofuels production, food processing, pharmaceutical and vitamin production, and the production of sustainable bioproducts, such as plastics and biobased materials.

While science does not know the total number of microbe species on our planet, it is anticipated to number in the millions.[72] These microbes exist in the soil, in water, in air, and on and within other organisms. Providing a range of symbiotic biological services (such as digestion) and even networking functions as they interact within ecosystems (part of the subject of metagenomics), the functional characteristics of microbes have considerable application in industry. At a basic level, humankind has several thousands of years experience in the biotechnology process of fermentation—using yeast to convert sugars to alcohols in brewing. Today, however, modern sequencing technologies and genome manipulation techniques are allowing scientists to modify microbes for a tremendous range of uses, and even to synthesize artificial microbial life forms. In doing so, industrial biotechnologists and research scientists are applying the tools of genomics, facilitated by the human genome sequencing development programs, to achieve impacts in many industrial areas (see Figure 8):

Figure 8: Functional Impact Areas of Genomics in Industrial Biotechnology



[72] Sun-Hee Hong, et al. 2006. "Predicting Microbial Species Richness." Proceedings of the National Academy of Sciences of the United States (PNAS). January 3, 2006. Volume 103, Number 1. Pages 117-122.

The use of finite fossil resources to fuel the U.S. and global economies is now recognized as an unsustainable model. The combustion of oil, natural gas and coal is releasing high levels of carbon dioxide implicated in global climate change. For the U.S., there is also a strategic impetus to avoid overdependence on foreign oil, which represents a security risk. The issues of "peak oil" and high price volatility of oil on global markets also pose threats to U.S. economic stability. As a result, the race is on to find sustainable alternative fuels. While solar, wind, hydro and nuclear power can generate electricity, they do not provide the liquid fuels needed for our existing transportation infrastructure nor the feedstocks needed to supply the chemicals and materials industries. Only biomass offers the opportunity to generate a sustainable stream of such liquid fuels and chemical feedstocks. Because of this, major development programs are relying on genomics to study microbes, algae, plant materials and other biological resources as potential fuel production systems. The U.S. Department of Energy's Joint Genome Institute (JGI) is at the forefront of efforts to sequence genomes for energy, industrial and environmental applications.

Only biomass offers the opportunity to generate a sustainable stream of liquid fuels and chemical feedstocks. Because of this, major development programs are relying on genomics to study microbes, algae, plant materials and other biological resources as potential fuel production systems.

Case Study: Microbial Genomics for Biofuels and Biofeedstocks at the Joint Genome Institute (JGI)

> *The DOE JGI maintains a database of microbial genomes sequenced by the DOE. Already this dataset contains 485 sequenced microbial genomes and 40 microbial community genomes. The rapid advances in genome sequencing speeds, spurred by the HGP and associated projects (of which the DOE was a key leader and funding agency), is allowing DOE and the organizations it funds to search the microbial world for organisms with unique functionality for efficient industrial biotech applications.*
>
> *Considerable sequencing work has been directed towards microbes that digest, or facilitate the digestion of cellulosic biomass. Cellulose is a complex carbohydrate (polysacharine) that makes up the cells walls of plants. Comprising the most abundant organic molecule on Earth, cellulose contains immense amounts of energy locked within its structure. The challenge is that cellulose is difficult to break down into glucose for economical use in industrial feedstock and liquid fuel applications. Numerous microbes have, however, evolved to do this task through the production of enzymes that break-down cellulose (such organisms reside in the gut of ruminants, such as cows, for example, allowing them to digest cellulosic biomass such as grass and straw). The JGI is engaged in multiple sequencing programs aimed at identifying microbes and their expressed enzymes for use in cellulose break-down. The JGI also engaged in plant genome sequencing work to identify biological characteristics that will make cell walls easier to deconstruct and biomass easier to process. The JGI is in the lead in fungi sequencing, again with a goal of finding enzymes to break-down cellulose, as well as metagenomics studies to examine microbial communities and their implications for bioprocessing.*

The application of genomics to industrial feedstocks, fuels and processing technologies has attracted significant industry investment, in addition to government lab and university-based work. Major energy sector companies such as Exxon-Mobil, ConocoPhillips, BP and Shell are actively researching, or sponsoring research, in genomics for biofuels and biobased chemical and plastics feedstocks. Biotechnology and genomics start-up companies have attracted substantial venture capital money within the sector. Plant biotechnology companies are also engaged in producing plant varieties with high biomass content, specialized oil content, or other expressed trait characteristics suiting them to economical industrial bioeconomy applications. Companies such as Monsanto, Land-o-Lakes and Ceres are developing specialized biomass crops, while industrial biotech companies like Novazymes and Genzyme are using genomics in the development of bioprocessing enzymes. Already biofuels are on the market from first-generation starch fermentation technology, and there are multiple examples of sustainable biobased plastics and chemicals successfully competing in the marketplace with petrochemicals (such as Natureworks LLC's Ingeo brand biopolymer produced at a plant in Blair Nebraska with a 300 million pounds annual capacity).

Case Study: The U.S. Economic Impact of Advanced Biofuels Production

> *An analysis by Bio Economic Research Associates (bio-era) calculates the projected economic impacts of building an advanced biofuels economy by 2022. Meeting the U.S. Renewable Fuel Standards requirement for 21 billion annual gallons of production by 2022 would have these projected economic impacts:*
>
> - *29,000 direct jobs created by 2012, 94,000 by 2016, and 190,000 by 2022. Total job impacts in the economy (comprising direct and indirect employment via the employment multiplier effect) could reach 123,000 in 2012, 383,000 in 2016, and 807,000 by 2012.*
> - *Total (direct and indirect) economic output generated by the advanced biofuels industry could reach $20.2 billion by 2012, $64.2 billion by 2016, and $148.7 billion by 2022.*
> - *Anticipated cumulative reduction in petroleum imports over the period 2010–2022 would exceed $350 billion*
>
> *"Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life."*
> *2010. BIO, The Biotechnology Industry Organization (www.valueofbiotech.com)*

Further advancements in genomics and the associated understanding of molecular biology processes is even leading to the development of "synthetic biology"—whereby a synthetic genome can be constructed to achieve specific biological functionality.

Case Study: Synthetic Biology—A New Approach to Engineering Biology

> *Synthetic biology is a new biotechnology tool enabling biotechnologists to go far beyond manipulation of one or two genes, and instead into engineering changes in entire genetic pathways. In effect, scientists can now "write" DNA code, rather than making simple modifications to an existing code. Such biotechnology brings some astounding opportunities to the industrial biotech sphere—most notably an ability to customize engineer microbes, such as bacteria or yeast, to produce specific chemical compounds (compounds that the microbe never previously made).*
>
> *Examples of synthetic biology successfully applied in manufacturing processes already exist in pharmaceuticals manufacturing with production processes for the anti-malarial drug artemisinin and the widely-prescribed cholesterol lowering statin Lipitor® (Pfizer).*
>
> *Codexis is a biotech company active in synthetic biology and novel biocatalyst development from enzymes. Codexis biotechnology is being used to synthesize the chiral side chain of atorvastatin, the active ingredient in Lipitor. The new process has reduced the cost of producing the statin by up to 70 percent and capital expenditures by 35 percent.*
>
> *Synthetic biotechnologies and biocatalysts have the advantage of reducing process steps and operating at lower temperatures and pressures than their chemical-process competition. The net result is less consumption of energy in processing and reduced capital costs for equipment (since the equipment need not withstand high temperatures and pressures).*
>
> *"Healing, Fueling, Feeding: How Biotechnology is Enriching Your Life."*
> *2010. BIO, The Biotechnology Industry Organization (www.valueofbiotech.com)*

The OECD's International Energy Agency states that "*biomass is the most important renewable energy source today*." Biotechnology and genomics tools are, and will continue to be, a key contributor to realizing maximum energy value through biomass feedstocks:

> *Plant-based systems capture solar energy and can be produced in a renewable manner. However, the harvestable parts are not well optimized for energy transfer and this has been a*

significant limitation to the development of economically viable and sustainable biomass energy systems. Biotechnology has provided a new toolset that can be used to design and optimize the capture of solar energy through crops. Further development of biotechnology and genomics tools will enable the development of crops with specific traits that are optimized for biofuels and bioenergy. The implementation of such a system will enable a sustainable platform for centuries to come and should be given a high priority in society.[73]

F. Environmental Applications

Our economic development and industrialization, often comes with an environmental cost. On an annual basis, over 400 million tons of Environmental Protection Agency (EPA)-designated "hazardous wastes" are generated each year, of which fully 256 million tons are generated in the U.S. alone.[74] Included in this waste are substances that are toxic to life, carcinogenic, mutagenic or teratogenic (interfering with embryo or fetus development). Human industrial, agricultural and other daily living activities release air pollutants, water pollutants and contaminate soil. Across the globe there is widespread contamination by heavy metals, pesticides, herbicides, hydrocarbons, radioactive materials and other persistent pollutants. We have areas of oceans and lakes in which pollutant-caused eutrophication has lowered oxygen to the extent that they become dead zones. We have "brownfield sites" across the globe—largely urban legacy industrial sites, which in the U.S. alone have an estimated clean-up cost of $650 billion ($2,100 for every person in the nation).[75] Now we have a preponderance of agreement in science that humankind's activities are even impacting the global climate.

While pollution threatens life on the planet, life itself may hold the solution to pollution. Microorganisms have evolved to thrive in extreme conditions and to absorb and metabolize an incredibly wide suite of chemicals. Organisms thrive in conditions ranging from freezing temperatures to near boiling point water—they thrive in environments of intense pressure in the deepest parts of the ocean, and in extreme arid conditions in deserts and within salt flats. Microorganisms have even been found living in spent nuclear fuel storage ponds. It is now widely recognized that microorganismal diversity has "far-reaching implications for addressing such DOE mission challenges as the remediation of radioactive and hazardous waste sites, sequestering heat-trapping carbon from the atmosphere, and developing renewable energy sources."[76]

As shown in this report, the genome represents a template for governing life processes. As such, by understanding the genomic structure of microbes, scientists are able to discover the genes and regulatory processes that impart useful activity for pollution mitigation, for uptake of carbon and for other environmentally important applications.

The HGP began in 1990, and leveraging the advancements made DOE launched the Microbial Genome Project in 1994 to sequence the genomes of nonpathogenic organisms useful in solving DOE's needs in environmental-waste cleanup, energy production, carbon cycling and biotechnology. Similarly, the Genomes to Life Program (GTL) aimed to use microbes and other organisms to address problems in energy production, environmental cleanup, and carbon cycling. This has evolved into DOE's Genomic Science Program (formerly Genomics:GTL) using "microbial and plant genomic data, high-throughput analytical technologies, and modeling and simulation to develop a predictive understanding of

Recent discoveries from projects funded by DOE's Biological and Environmental Research program highlight the ubiquitous presence and critical importance of microbes in all ecosystems. For example:

Diatoms, ancient and intricately shaped ocean microbes, store an amount of carbon comparable to that in all the earth's rainforests combined. Over geological time, diatoms may have influenced the earth's climate.

More than a million previously undiscovered genes, possibly representing new biochemical functions, were the surprising find in sequencing DNA fragments from the Sargasso Sea—a region heretofore thought to sustain little life. This discovery also was named one of Science magazine's "Breakthroughs of the Year."

Microbes thrive deep within the earth's subsurface and at extremes previously thought to extinguish life. Growing recognition of microbial capabilities and potential applications has made a compelling case for further investigations by DOE and other agencies and institutions.

Source: http://microbialgenomics.energy.gov/benefits.shtml

[73] J.S. McLaren. 2005. "Crop biotechnology provides an opportunity to develop a sustainable future." Trends in Biotechnology. 2005 Jul:23(7):339-42

[74] Microsoft Encarta online encyclopedia, 2009 and the United States Environmental Protection Agency.

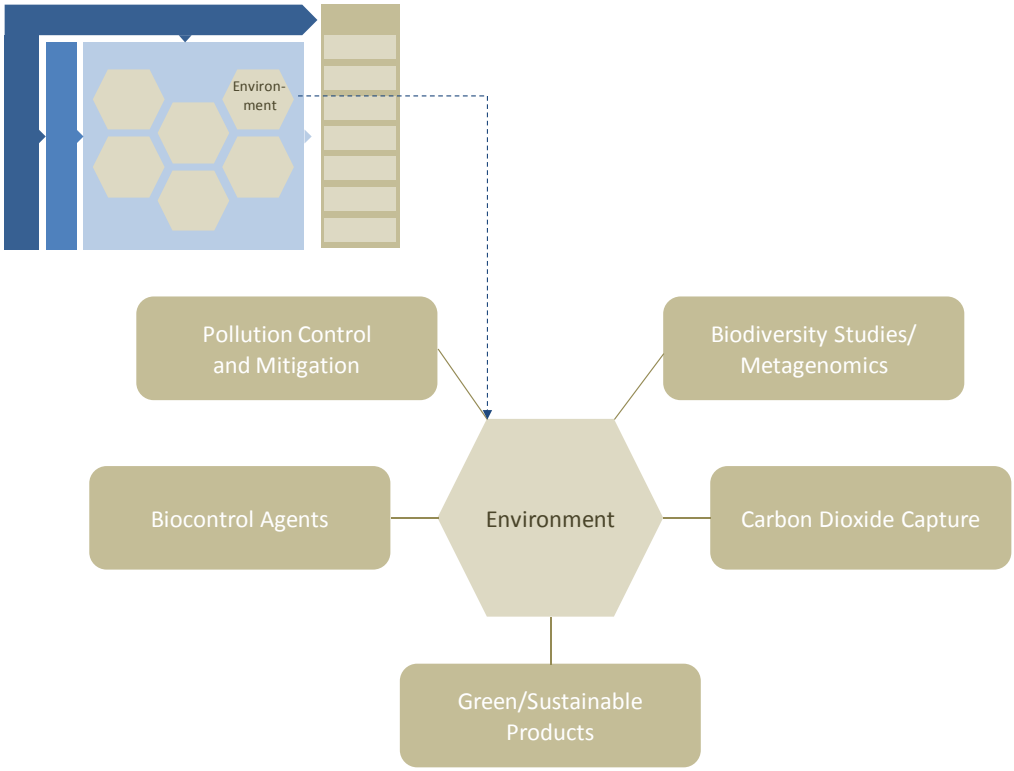
[75] Michigan State University. "EnviroTools Factsheet: Brownfields". Accessed online at: http://www.envirotools.msu.edu/factsheets/brownfields.shtml

[76] U.S. Department of Energy, Joint Genome Institute. "Why Sequence Microbes?" Accessed online at http://www.jgi.doe.gov/sequencing/why/microbes.html

biological systems behavior relevant to solving energy and environmental challenges including bioenergy production, environmental remediation, and climate stabilization."[77]

When applications of genomics are examined across the environmental spectrum it is evident that the five primary macro-categories of application are as shown in Figure 9.

Figure 9: Functional Impact Areas of Genomics in Environmental Science and Sustainability



Pollution control and mitigation microbes play an important function. Indeed it has been noted that "Livestock in the Unites States produce 1.7 billion tons of manure annually—almost all of it is degraded by soil micoorganisms."[78] While manure degradation by micoorganisms occurs naturally in the environment, humans use purpose-developed micoorganisms to degrade and detoxify human sewage, while other engineered organisms are being applied in the treatment of industrial waste materials and hazardous waste. Because microbes have comparatively small genomes (averaging 4-5 million bases versus the 3 billion typical in mammals), they are able to be sequenced cost effectively and rapidly—enabling scientists to understand their life processes and investigate their application to a range of pollution degradation needs. DOE's Joint Genome Institute has, in particular, led programs to sequence a broad range of micoorganisms with application to bioremediation needs as diverse as radioactivity, heavy metals, pesticides and hydrocarbons. Plant genomes are also studied for opportunities to use plants for absorption of pollution from the soil and water (called phytoremediation).

Because communities of microbes of different types can accomplish activities that a single species of microbe alone cannot, there is considerable utility in metagenomics studies that sequence entire populations of microbes from samples of soil or seawater, for example. This is particularly important developing approaches to tackle complex pollutants, such as oil, that have multiple constituent

[77] U.S. Department of Energy. Genomic Science Program. Accessed online at: http://genomicscience.energy.gov/#page=news

[78] Office of Technology Assessment. 2002. "Impacts of Applied Genetics: Micoorganisms, Plants, and Animals." Books for Business, NY.

chemicals that no single organism can degrade. Metagenomics reveals the symbiotic effects of microorganisms working together—in fact, it is found that various microbes may work in sequence, with the waste product of one being the fuel for the next. Metagenomic studies are also enabling researchers to better understand the complexity of ecosystems and the effects of human activities on them.

In addition to bioremediation of pollutants, genomic advancements are enabling scientists to develop biocontrol agents to accomplish activities such as pest control which would otherwise use chemicals for the same function. For example, nematodes (predatory roundworms) can be engineered to hone-in on, intercept and destroy white grubs in lawns, negating the need for chemical white grub pesticide applications which can run-off during rain and negatively impact groundwater and watersheds. Biocontrol genomics is also working to produce environmentally sustainable approaches to the control of plant pathogens. The complete sequence of the biological control organisms *Pseusomonas fluorenscens* Pf-5, for example, has helped researchers understand the genes and regulatory processes involved in this microbe's ability to express antibiotics. By elucidating the genes involved, genes can be transferred to other organisms for similarly useful expression products.

Global climate change concerns are driving efforts to reduce carbon dioxide emissions and to capture carbon dioxide from the atmosphere. Plants metabolize carbon dioxide and produce oxygen; likewise, many species of microbes can metabolize or fix carbon dioxide. Genomic studies are helping to identify opportunities to better leverage plants and microbes as carbon dioxide reservoirs. In addition, of course, plants and algae can produce carbon neutral fuel sources to reduce the release of fossil carbon from coal, oil and natural gas utilization. Genomics is playing a critical role in developing lignocellosic crops that can be more readily processed, microbes and enzymes for breaking-down cellulose to allow cellulosic biofuels to be produced economically, and even specially engineered strains of algae customized for high rates of growth and lipids (oil) production. The development of liquid biofuels is one part of a broader "green chemistry" and sustainable products movement based on biological control via genomics—a movement enabling the development of chemicals, plastics and biobased materials that can be produced in a sustainable fashion and are readily degraded in post-use disposal.

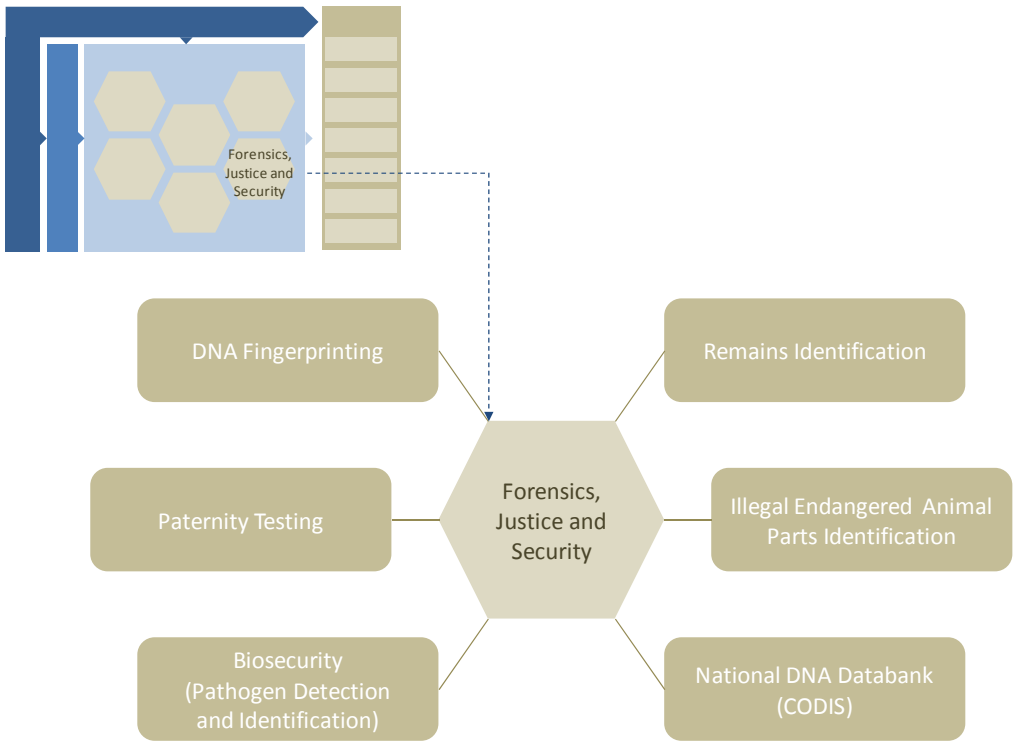
Case Study: The J. Craig Venter Institute and Synthetic Genomics Inc.

> *Recognizing the importance of microbes to environmental sustainability, climate change control and other critically important needs, the non-profit J. Craig Venter Institute has undertaken focused programs to investigate key reservoirs of microbes, such as the earth's oceans, to identify microbes and genes of interest. The Global Ocean Sampling Expedition, for example, uncovered more than six million new genes as a result of its sampling and sequencing program.*
>
> *Synthetic Genomics Inc., (SGI) also founded by J. Craig Venter, is working on commercial application of genes discovered via these and other investigations performed by the J. Craig Venter Institute and other collaborators. SGI is focused first on creating genomic-driven commercial solutions for bioenergy applications and has established important partnerships with major energy companies such as ExxonMobil. Using leading edge genomic technologies, SGI is engineering synthetic algae organisms optimized for biofuel production. SGI is also using genomics in determining pathways to more efficient use of cellulosic materials for bioenergy production, and is investigating microbes for the conversion of coal into fuels and chemicals and other value-added products.*

G. Forensics, Justice and Security

As genomics has opened the study of medicine, biology, and biotechnology to entirely new worlds, so it has done in the field of forensics, criminal and social justice and security. Figure 10 illustrates some of the applications of genomics and genetics in this field currently.

Figure 10: Functional Impact Areas of Genomics in Forensics, Justice and Security



One of the characteristics of DNA is that it varies from person to person. Thus analysis of DNA can be used to specifically link a DNA sample to a specific person –useful in solving crimes, resolving questions of identity in remains identification, and resolving questions in paternity cases. In natural or deliberate disasters (such as the recent Haitian earthquake and the 9/11 terrorist attack) the analysis of DNA samples has proven itself time and time again as a valuable tool for remains identification. Such genetic and genomic analysis is also proving important in the identification of illegally imported organs, tissue, furs and other body parts of endangered species seized by customs or other law enforcement officials.

The increasing and prescient threat of bioterrorism is a particular driver of the use of advanced genomics in security and law enforcement. Indeed, as early as 2001 the field of forensic genomics emerged to tackle a serious threat to domestic security – the 2001 anthrax letter attacks (see case study).

Case Study: Microbial Genomics Solves the Anthrax Attacks

> *After the 2001 anthrax letter attacks, researchers at the Institute for Genome Sciences at the University of Maryland School of Medicine, who worked with researchers at the Federal Bureau of Investigation (FBI), the U.S. Army Medical Research Institute of Infectious Diseases and Northern Arizona University, published a paper recounting the scientific investigation into the anthrax attacks of 2001. The case was groundbreaking because of its use of genomics and microbiology in a criminal investigation.*
>
> *More than 20 people contracted anthrax from Bacillus anthracis spores mailed through the U.S. Postal Service in 2001, and five of those died as a result. The investigation became known as "Amerithrax". The work pioneered the new field known as microbial forensics, a science that will play a key role in the investigations of any future bioterror attacks. The March 7, 2011 Proceedings of the National Academy of Sciences paper describes how Institute for*

Genome Sciences faculty and collaborators from the FBI concluded that the anthrax samples taken from all the attacks were genetically identical. Later, another group of scientists would trace the anthrax spore found in the letters back to a flask of Bacillus anthracis and several samples that were taken from that flask.

"This paper and the Amerithrax investigation really marked the beginning of a new approach for the science we call forensic genomics," says the paper's senior author Jacques Ravel, PhD, who is Associate Professor of Microbiology and Immunology at The University of Maryland School of Medicine and Associate Director for Genomics at the Institute for Genome Sciences. "The science was a critical component of the Amerithrax case. Without genomics, it would have been extremely difficult to narrow the pool of potential suspects."

"Before Amerithrax, no one appreciated the precision, accuracy, and reliability that this type of genomics can offer as a microbial forensic technique," said first author David Rasko, PhD, Assistant Professor for Microbiology and Immunology at the School of Medicine and a research scientist at the Institute for Genome Sciences. As one of the first investigations of its kind, Amerithrax helped to shape the emerging field of microbial forensics. Since the case, Ravel, Rasko, and their colleagues at the Institute for Genome Sciences have also helped to shape the scientific community's effort to develop standards and guidelines for future investigations.[79]

Advanced genomic techniques are also allowing researchers to perform work that was previously impossible. For example, a complex crime scene could contain DNA samples that are mixed—containing, for example, the DNA of a victim, family members, pets, microbes and an assailant. Separating these samples for analysis is a highly complex challenge, but success is being achieved via genomics (see case study below).

Case Study: Separation of an Individual's DNA from Complex Mixed Samples

Researchers at the Translational Genomics Research Institute (TGen) and the University of California – Los Angeles (UCLA) developed a process in 2008 to detect an individual's DNA in a mixture, in proportions which are less than 0.1 percent of the total mixture, with others in a single complex sample.[80] *Led by David Craig at TGen, the team used high-density single nucleotide polymorphism (SNP) genotyping microarrays to demonstrate the ability to accurately determine whether the DNA of separate individuals is in a complex genomic DNA mixture. Then they demonstrated experimentally the identification of the presence of genomic DNA of specific individuals within a series of highly complex genomic mixtures, even where an individual contributes less than 0.1 percent of the total genomic DNA. These findings shift the perceived utility of SNPs for identifying individual trace contributors within a forensics mixture, and suggest future research efforts into assessing the viability of DNA sources thought to be useless due to sample contamination.*

A Virginia based company, Casework Genetics, licensed the technology from TGen while it was a grantee of Catapault Bio of Arizona, a "venture philanthropy" organization that holds an equity position in Casework Genetics.[81]

[79] Rasko, Ravel, "Forensic Genomics Used to Investigate 2001 Anthrax Attacks", March 7, 2011, Proceedings of the National Academy of Sciences.

[80] Homer N, Szelinger S, Redman M, Duggan D, Tembe W, et al. (2008), "Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays." PLoS Genet 4(8): e1000167. doi:10.1371/journal.pgen.1000167

[81] http://www.flinn.org/news/926

Case Study: Identification of Victims of the 9/11 Terror Attacks

> *DOE's Human Genome Project Information Site shows the forensics application of DNA analysis in action on an event that changed America. The DOE website reports that:*
>
> *"Identifying the victims of the September 11, 2001, World Trade Center attack presented a unique forensic challenge because the number and identity of the victims were unknown and many victims were represented only by bone and tissue fragments. At the time of the attack, no systems were in place for rapidly identifying victims in disasters with more than 500 fatalities. The National Institutes of Justice assembled a panel of experts from NIH and other institutions to develop processes to identify victims using DNA collected at the site. Panel members produced forms and kits needed to enable the medical examiner's office to collect reference DNA from victims' previously stored medical specimens. These specimens were collected and entered into a database. The medical examiner's office also received about 20,000 pieces of human remains from the World Trade Center site, and a database of the victims' DNA profiles was created. New information technology infrastructure was developed for data transfer between the state police and medical examiner's office and to interconnect the databases and analytical tools used by panel members. In 2005, the search was declared at an end because many of the unidentified remains were too small or too damaged to be identified by the DNA extraction methods available at that time. Remains of only 1,585 of the 2,792 people known to have died had been identified. In 2007, the medical examiner's office reopened the search after the Bode Technology Group developed a new methodology of DNA extraction that required much less sample material than previously necessary. The victim DNA database and the new methods have allowed more victims to be identified, and further identifications will be possible as forensic DNA technology improves."*[82]

It is likely that infectious disease causing pathogens will be a preferred mode of bioterror attack for the perpetrators. Modern genome sequencers will be on the frontline in characterizing the organisms used in such an attack and tracing the biological agent back to its source. This is highlighted by Slezak et al, who note that:

> *Rapid advances in the genomic sequencing of bacteria and viruses over the past few years have made it possible to consider sequencing the genomes of all pathogens that affect humans and the crops and livestock upon which our lives depend. Recent events make it imperative that full genome sequencing be accomplished as soon as possible for pathogens that could be used as weapons of mass destruction or disruption. This sequence information must be exploited to provide rapid and accurate diagnostics to identify pathogens and distinguish them from harmless near-neighbours and hoaxes.*[83]

Case Study: Genomics on the Frontlines of National Biodefense

> *The U.S. Department of Homeland Security's National Biodefense Analysis and Countermeasures Center (NBACC) is the central national center for the characterization of biological threats to homeland security, and is the core lab, under Presidential Directive, for the bioforensic analysis of evidence from biocrime or terrorism attack. NBACC is managed and operated by the Battelle National Biodefense Institute for the U.S. Department of Homeland Security. Genome sequencing is an integral component of the Center's work and the work of contract labs supporting the Center. NBACC operates multiple genome sequencing platforms, ranging from traditional Sanger sequencers (as used in the HGP) through to instruments from 454 Life Sciences, Pacific Biosciences and the new Ion Torrent*

[82] Human Genome Project Information. "DNA Forensics." Accessed online at: http://www.ornl.gov/sci/techresources/Human_Genome/elsi/forensics.shtml

[83] T. Slezak, et al. 2003. "Comparative genomics tools applied to bioterrorism defense." Briefings in Bioinformatics. June 2003;4(2):133-49

> *desktop machine from Life Technologies. Sequencing is planned as the backbone of a new methodology for supporting biocrime investigators, enabling analysis of samples to be performed without prior development of more limited microarrays. Being able to respond to the case work demands of investigators without needing to pre-develop on-the-shelf biochemistries will be an important step in keeping the nation safe and responding rapidly to biological agent events. The NBACC's research scientists are also utilizing the major reference genome sequences, facilitated by the HGP work, such as the data from the Microbial Genome Project and the Human Microbiome Project.*

Longer term, genomics will also play an important role in the development of vaccines and the rapid development of treatments for infectious agents based on biotechnology using genetically engineered microbial production techniques.

H. Conclusion: Outcomes, Products and Impacts

The sequencing of the human genome has had a profound and paradigm-shifting impact on basic biological science and our understanding of biomolecular life processes. In addition, the large-scale sequencing programs, led by the HGP, spurred the rapid development of advanced sequencing equipment and technology that spawned an entire genomics-based technology sector . Today's sequencing platforms can analyze whole genomes at a speed never before thought possible.

Genome sequence information has great utility across a broad range of scientific and technical disciplines. In human biomedical science a new class of advanced diagnostic tests has been developed because of advances in human genomics, and the field of pharmacogenomics is forming the underpinning of personalized medicine and emerging biomedical applications such as gene therapy and regenerative medicine. In addition, the human reference genome and the technologies of modern genomics have significantly affected science and applied technology deployment in multiple fields outside of human medicine. Disciplines including veterinary medicine, agriculture and food production, forestry, environmental science, industrial biotechnology, biofuels development and biosecurity and forensics are all beneficiaries and users of the knowledge and technological advancements made possible by the HGP and associated programs.

The application of genomics across the above fields and disciplines generates a broad variety of functional economic and social impacts. Based on this detailed review of the literature on genomics applications and associated impacts, Battelle concludes that the primary impact areas include:

- **Knowledge Expansion and Education** – Benjamin Franklin wrote that "*an investment in knowledge pays the best interest*." We live in a modern "knowledge economy" that depends on innovation and technological advancement for economic progress and that equally depends on a knowledgeable and skilled populace to achieve high productivity economic activity. The sequencing of the human genome has generated a startling advance in our knowledge of fundamental biological structures and molecular processes, and that knowledge opens new horizons for future development.
- **Economic Development** – As highlighted in Chapter II, the tools and technologies of genomics, in part empowered by the Human Genome Project and associated programs, are resulting in the growth of commercial enterprise focused on the production of genomic analysis equipment, laboratory supplies and the provision of genomic analysis services. With a distinct concentration of this commercial enterprise in the United States, the HGP and Celera programs have supported the growth of this domestic industry. Furthermore, the application of biological knowledge generated by the HGP is opening new product development pathways in human biomedical products and applications in many other business sectors.

- **Human Health** – Through the sequencing of the human genome, biomedical scientists have a more accurate understanding of the molecular biology of the human. This understanding is being applied to identify targets for new and existing therapeutics, to develop highly refined diagnostic tests, and pave the way for a new paradigm of personalized medicine. With the genome in hand, researchers are also advancing promising new disciplines such as gene therapy and regenerative medicine.
- **Environmental Sustainability** – Human activity has significant impacts on our environment; to the extent that scientific consensus is that we are precipitating climate change. Pollution from human activities has impacted our soils, water and air and, as a result, ourselves and other organisms in the biosphere. The application of genomics advancements, however, is driving the development of technological solutions to environmental challenges. Genetic engineering of microbes to capture, digest and mitigate various environmental pollutants is one impact area, together now with the ability to use genomics-driven synthetic biology to engineer entirely new organisms. The application of genomics to biofuels and environmentally sustainable renewable chemicals and industrial inputs is a "here and now" area of impact, with positive benefits being felt and great promise for the future in advanced biofuels and industrial bioeconomy applications.
- **Food Security and Safety** – Growth in global population and income levels is leading to a dramatic increase in the demand for food. Predictions are that by 2030 we will need to double global food production.[84] Genomics is empowering advancements in crop and livestock improvement, providing us with higher yielding crops, protection of our food supply from disease and pests, and food products with enhanced nutrition content and functional health promoting characteristics. Advanced sequencing technologies, spurred in their development by human genome sequencing, are empowering a new green revolution in agriculture and food production, and also assuring a safer food supply. The positive economic consequences are broad and deeply felt—ranging from new products and opportunities for producers, reduced input costs for farmers, reduced losses, and enhanced health and associated productivity and welfare.
- **National Security** – The application of modern genomic knowledge is also enhancing U.S. national security across multiple fronts. The development of domestic advanced biofuels and biobased industrial inputs is reducing dependence on foreign oil, while genomics is also on the frontlines in addressing the threat of bioterrorism.
- **Justice** – Genomics is being applied to criminal and social justice. As a unique identifier, the genome holds promise for definitive identification of the individual, but beyond that it provides the means to trace biological materials to point of origin (important for investigations). On a social justice front, the sequencing of the human genome also shifts basic thinking on such divisive issues as race, and is providing new paths for more equitable access to food, healthcare and economic opportunity.

Ultimately, we see that the sequencing of the human genome has economic and functional impacts far beyond the obvious human biomedical science impacts. It has provided a fundamental platform for advancements across not only the bioscience of the human, but our understanding and use of all organisms. It has spurred technology development that has impacts felt in disciplines as diverse as industrial biotechnology, environmental science, agriculture, and security.

[84] Food and Agriculture Organization of the United Nations (FAO). "World Agriculture Towards 2015/2030, Summary Report." *www.fao.org/documents/show_cdr?url_file/DOCREP.*

Chapter IV: Into the Future

The sequencing of the human genome was the signature scientific program of the 1990's and early 2000's. When *Science*, in 2010, produced a retrospective on the ten most important insights of the decade, five of the ten most influential scientific insights stem from, or make use of, the human genome sequencing and related genomics advances (Table 18).

Table 18: Science Magazine, Top Insights of the Decade 2010

Insight	Description
Genomic Complexity	**Science writer Elizabeth Pennisi comments:** "Gene regulation has turned out to be a surprisingly complex process governed by various types of regulatory DNA, which may lie deep in the wilderness of so-called junk DNA that lies between genes. Far from being humble messengers, RNAs of all shapes and sizes are actually powerful players in how genomes operate. Finally, there's been increasing recognition of the widespread role of chemical alterations called epigenetic factors that can influence the genome across generations without changing the DNA sequence itself. The scope of this "dark genome" became apparent in 2001, when the human genome sequence was first published." Science 17 December 2010: Vol. 330 no. 6011 p. 1614
The Human Microbiome	**Science writer Elizabeth Pennisi comments:** This past decade has seen a shift in how we see the microbes and viruses in and on our bodies. There is increasing acceptance that they are us, and for good reason. Nine in 10 of the cells in the body are microbial. In the gut alone, as many as 1,000 species bring to the body 100 times as many genes as our own DNA carries. A few microbes make us sick, but most are commensal and just call the human body home. Collectively, they are known as the human microbiome. Likewise, some viruses take up residence in the body, creating a virome whose influence on health and disease is just beginning to be studied. Science 17 December 2010: Vol. 330 no. 6011 p. 1619
Stem Cells and Regenerative Medicine	**Science writer Gretchen Vogel comments:** By prompting a cell to overexpress a few genes, researchers have discovered in the past decade how to turn a skin or blood cell into a pluripotent cell: one that has regained the potential to become any number of cells in the body. Other genes can prompt skin cells to turn directly into neurons or blood cells. Scientists are already using the technique to make cell lines from patients with hard-to-study diseases, and ultimately they hope to grow genetically matched replacement cells and tissues—perhaps even entire organs. Science 17 December 2010: Vol. 330 no. 6011 p. 1618
Ancient DNA	**Science writer Ann Gibbons comments:** In the past decade, powerful new x-ray scans and three-dimensional computer models have transformed the analysis of ancient bones, teeth, and shells. But a new kind of analysis is capable of revealing anatomical adaptations that skeletal evidence can't provide, such as the color of a dinosaur's feathers or how woolly mammoths withstood the cold. The new views of the prehistoric world hinge on the realization that "biomolecules" such as ancient DNA and collagen can survive for tens of thousands of years and give important information about long-dead plants, animals, and humans. Science 17 December 2010: Vol. 330 no. 6011 p. 1616

The Inflammation/ Disease Connection	**Science writer Jennifer Couzin-Frankel comments:** "Not long ago, inflammation had a clear role: It was a sidekick to the body's healers, briefly setting in as immune cells rebuilt tissue damaged by trauma or infection. Today, that's an afterthought. Inflammation has hit the big time. Over the past decade, it has become widely accepted that inflammation is a driving force behind chronic diseases that will kill nearly all of us. Cancer. Diabetes and obesity. Alzheimer's disease. Atherosclerosis. Here, inflammation wears a grim mask, shedding its redeeming features and making sick people sicker."[85] Science 17 December 2010: Vol. 330 no. 6011 p. 1621

The impact on biological science has been profound and paradigm shifting, and the accelerated development of genomics and related technologies has enabled genome sequencing to be applied across a broad variety of important applications in human medicine, veterinary medicine, agriculture, industrial biotechnology, environmental sciences, justice and security.

The world of genomics and genomic technologies would be nowhere near the advancements achieved today without the HGP, its findings, and the development stimulus it provided.

In medicine, genomics is being applied today to quantitatively diagnose diseases and disorders, to develop optimized treatment regimens for patients based on their genomic profile, to develop gene therapies to correct disease-causing genetic variants, and to engineer biological materials to repair and regenerate. New approaches to vaccines are being developed, and infectious disease organisms combated, because of genomics advances. As Battelle interviewed leading scientists and industry representatives, the consistent message was that **the world of genomics and genomic technologies would be nowhere near the advancements achieved today without the HGP, its findings, and the development stimulus it provided.**

The platform of knowledge and the technologies resulting from human genome sequencing have formed the basis of nothing less than a medical revolution. The primary impacts of this revolution in quantitative and personalized medicine may not yet be felt in daily clinical practice, but that day is accelerating towards us. Writing in Nature, Eric Lander notes that:

> "*Medical revolutions require many decades to achieve their full promise. Genomics has only just begun to permeate biomedical research: advances must proceed through fundamental tools, basic discoveries, medical studies, candidate interventions, clinical trials, regulatory approval and widespread adoption. We must be scrupulous not to promise the public a pharmacopoeia of quick pay-offs. At the same time, we should remain unabashed about the ultimate impact of genomic medicine, which will be to transform the health of our children and our children's children.*"[86]
>
> *The ultimate goal is to create a reference catalogue of all genetic variants commons enough to be encountered recurrently in populations, so that they can be examined for association with phenotypes and interpreted in clinical settings. Efforts towards this goal are already well underway. The 100 Genome Project (which plans to study many more than one thousand genomes) aims to find essentially all variants with frequency of greater than 1 percent across the genome and greater than 0.1 percent in protein-coding regions*."[87]

Lander continues:

> "*The ultimate goal is for sequencing to become so simple and inexpensive that it can be routinely deployed as a general-purpose tool throughout biomedicine. Medical applications will eventually include characterizing patients' germline genomes (to detect strongly*

[85] The genomics to inflammation link was highlighted by CW Schmidt in earlier work. Schmidt noted that "Today, genomics defines the cutting edge of inflammation research. Genomic studies, in addition to their proteomic and metabolomic cousins, aim to resolve an age-old mystery: namely, why some patients recover readily from inflammation while others suffer and die from it." See Schmidt CW. 2005. "Critical Care: Applying Genomics to Inflammation Outcomes." Environ Health Perspect 113:A816-A821. doi:10.1289/ehp.113-a816

[86] Eric S. Lander. 2011. "Initial impact of the sequencing of the human genome." *Nature.* Volume 470, February 10, 2011.

[87] Ibid

predictive mutations for presymptomatic counseling where treatments exist, to search for causes of diseases of unknown aetiology, and to detect heterozygous carriers for prenatal counseling); cancer genomes (by identifying somatic mutations to compare tumour and normal DNA); immune repertoires (by reading the patterns of B-cell and T-cell receptors to infer disease exposures and monitor responses to vaccines); and microbiomes (by associating patterns of microbial communities with disease processes). Research applications will include characterizing genomes, epigenomes and transcriptomes of humans and other species, as well as using sequencing as a proxy to probe diverse molecular interactions.

To fulfill this potential, the cost of whole-genome sequencing will need to eventually approach a few-hundred US dollars. With new approaches under development and market-based competition, these goals may be feasible within the next decade."[88]

Writing in 2006 in the journal of the Royal Pharmaceutical Society of Great Britain, Paul Martin and Michael Morrison discussed the future of genomic medicine and summarized the potential horizon for genomic medicine in clinical practice (Table 19). Their conclusions are reasonable in light of experience and what we know about medical development timetables.[89]

Table 19: Potential Timetable for the Advance of Genomic Medicine

Technologies that are well entrenched in the clinic
• Genetic testing for monogenic disorders
• Therapeutic proteins.
The medium term prospect for an expansion of these technologies is very promising and they raise few new social, ethical or practice issues.
Technologies that are starting to become entrenched in the clinic
• Pharmacogenomic drugs
• Pharmacogenetic tests
• Genetic tests for common conditions
• Adult stem cell therapies.
An expansion of these technologies in the medium term is therefore likely. However, each of them has yet to be fully entrenched in a mature market or established set of clinical practices. They still face significant technical difficulties and, with the exception of pharmacogenomic drugs, they will also have to overcome a number of commercial, clinical, ethical and regulatory difficulties.
Technologies that have yet to successfully enter the clinic
• Gene therapy
• Cancer vaccines
• Embryonic stem cell therapies.
Each of these technologies face very significant problems at present and are unlikely to enter the market in anything other than first proof-of-principle products in the near future.

The Future of Genomics in Medicine

Diagnostics*—Appropriate treatment of a disease can only begin once an accurate diagnosis is made. Genomics will increasingly drive advanced, highly accurate diagnostics development by:*

- *Identification of variant genes and their impact on monogenic and multigenic diseases.*
- *Sub classification of diseases by genome analysis to direct personalized medicine approaches (as is already occurring in certain cancers).*
- *Methods for real-time pathogen detection and identification.*

Therapeutics*—Genomics will significantly increase in utility across a range of therapeutic R&D and application areas, including:*

- *Identification of specific disease-relevant targets.*
- *Rational drug design to produce drugs specifically aimed at the characteristics of the identified targets.*
- *Repurposing of existing drugs for application to newly identified genomic targets and signatures*
- *Previously failed drugs "rescued" for application to genomic sub-populations*
- *Clinical trials improved through genomic stratification of participants*
- *Significantly reduced adverse drug events, and associated costs, driven by genotype-guided drug prescription.*
- *Increasing application of gene therapies*
- *Therapeutic strategies developed for genotype-environment interactions and for adaptation to behavioral and lifestyle characteristics.*

Further perspective on many of the above aspects of the future of genomic medicine may be found in: Eric Green and Mark Guyer. 2011. "Charting a course for genomic medicine from base pairs to bedside." Nature. Volume 470. February 10th 2011.Pp:204-213.

[88] Eric S. Lander. 2011. "Initial impact of the sequencing of the human genome." *Nature*. Volume 470, February 10, 2011.

[89] Paul Martin and Michael Morrison. 2006. "Realising the Potential of Genomic Medicine." Royal Pharmaceutical Society of Great Britain. July 2006.

In accepting the Russ Prize, Dr. Leroy Hood summarized the transcendent benefits of the HGP and its current and future impacts on biology and medicine. Hood's comments represent a good synopsis of implications, impacts and direction, and are reproduced here with permission.

> ***Dr. Leroy Hood speech at Russ Prize award, 2011 on the Impact of the Human Genome Project***
>
> *"Let me cite briefly 14 genome project accomplishments that revolutionized both biology and medicine.*
>
> - *First, it democratized genes, that is, it made all genes accessible to all biologists.*
> - *Second, it delineated for the first time all human genes, and by inference, all proteins. This comprehensive "parts list" enabled a new approach to biology that I termed "systems biology", which takes a holistic approach to dealing with biological complexity. This is the focus of my Institute for Systems Biology—the first systems biology organization that was started in 2000.*
> - *Third, it catalyzed the development of high-throughput instrumentation, the very effective generation of biological information in genomics, in proteomics (the study of proteins), in metabolomics (the study of small molecules) and in phenotypic assays.*
> - *Fourth, it pioneered the applications of computer science and mathematics to biology. In fact, it was the genome project that legitimately brought mathematicians and computer scientists and even theoretical physicists into biology to think about acquiring and storing, analyzing, mining, integrating, and ultimately creating predictive and actionable models of complex biological systems.*
> - *Fifth, it was the first biological project whose policy was open source for all data; it demanded the instantaneous release of data to the biological community so everyone could analyze this new information immediately as it was produced.*
> - *Sixth, it created the first rigorous standards for biological data—and project funding depended on meeting these standards.*
> - *Seventh, it gave us access to the genomes of plants, of animals, of microbes, and knowledge of those genomes has transformed many fields of biology.*
> - *Eighth, it revolutionized our understanding of evolution in absolutely magnificent ways.*
> - *Ninth, it transformed how we think about medicine. It created a new field of medical diagnostics using biomarkers in tissues and blood that can actually detect disease early and stratify complex diseases into their different subtypes so physicians can do impedance matches against appropriate therapies.*
> - *Tenth, it also opened up the possibility of using DNA sequencing to identify genes that have actionable behaviors with regard to patients. For example, a gene defect termed Leiden factor 5 leads to an increased tendency for blood clotting. Patients with this defect can drink lots of water, not sit on airplanes for 5 hours without walking and stretching and even use low doses of anti-coagulants. This is particular important for pregnant women with the defective gene.*
> - *Eleventh, physicians are now using DNA sequencing of tumors to look at disease-perturbed biological networks to determine the right drug for the patient.*
> - *Twelfth, the genome project created a scientific environment that in part led to the third paradigm change that I will discuss shortly—P4 Medicine. (**P**redictive, **P**reventive, **P**ersonalized and **P**articipatory).*

- *Thirteenth, the genome project has changed the sociology of biology. For example, it introduced the concept of big science to biology. By big science I mean cross-disciplinary, hypothesis-driven science that integrates different data types to build predictive models to attackhard biological problems such as P4 Medicine. Small science is done by a single investigator and a few co-workers and takes on highly focused and discrete problems.*

  The synergy between big and small science is enormous in that each can take advantage of the strengths of other and they can effectively operate in an integrated manner. Unfortunately, at NIH today, there is a thrust to do away with big science (driven by budget constraints and many scientists who practice small science). To eliminate big science would be a tragic mistake, as it is necessary to attack many of the most complex and challenging problems of biology and society today. We obviously need a mixed portfolio where the strengths of big and small science can be utilized together in an integrated manner.

- *Finally, the genome project also was the first that supported an investigation of the social, ethical, and legal aspects of the genome sequence in a way that **presaged** how we are really going to have to consider medicine of the future.*

One of the key realizations that must be understood regarding the human genome sequencing is that its usefulness is perpetual. While other major big science projects have a life attached to them (the $11 billion Superconducting Super Collider has an estimated life span of 30 years for example, or the $1.5 billion Hubble Space Telescope at 15–20 years), the human genome sequence does not wear out or become obsolete. Rather, the reference human genome is more akin to chemistry's periodic table, a fundamental platform for understanding and advancement of science.

As we have explored in this report, the impact on human medicine and health is profound and important, but the benefits of performing the HGP and related projects extend into areas of human activity and planetary life, far beyond that. Our fundamental understanding of genomics, the sequencing and genomic technologies whose development was spurred by the HGP and Celera efforts, the advanced 'omics disciplines generated, will make contributions on a broad range of fronts. We can expect:

- Agricultural productivity to increase considerably, working towards the challenge of feeding the world's rapidly expanding population in a sustainable manner.
- Not only will food availability increase, but the impact of its production on the global environment will reduce as crops and livestock are developed with traits suited to nitrogen use efficiency, no-till agriculture, water use efficiency and reduced waste production.
- Currently low-value biomass, especially low-value cellulosic biomass, will be converted into higher-value liquid fuels, energy sources, biobased chemicals, plastics and materials. These products will increasingly displace petroleum and other fossil-based inputs, contributing to reduced carbon emissions and associated climate and environmental benefits.
- An increasingly two-way flow of diagnostics, therapeutics and prevention tools will move between human medicine, veterinary medicine and agriculture as the cost of genomic technologies reduces and the applications of discoveries in one area can be applied to another because of comparative genomics and other genomic advancements.
- The legacy of pollution on the planet caused by human activity will increasingly be addressed through the application of genetically-engineered, modified or synthetic organisms designed to perform remediation and mitigation functions.

The above represent just some of the areas in which genomics will be applied for the improvement of health, wellbeing and sustainability.

The Future Cost of Sequencing – Dramatic Cost Reductions Lead to Widespread Application

"It is now possible to order your personal genome sequenced today for a retail cost of under ~$20,000. This cost will likely fall to less than $1,000 by 2012, and to $100 by 2013.

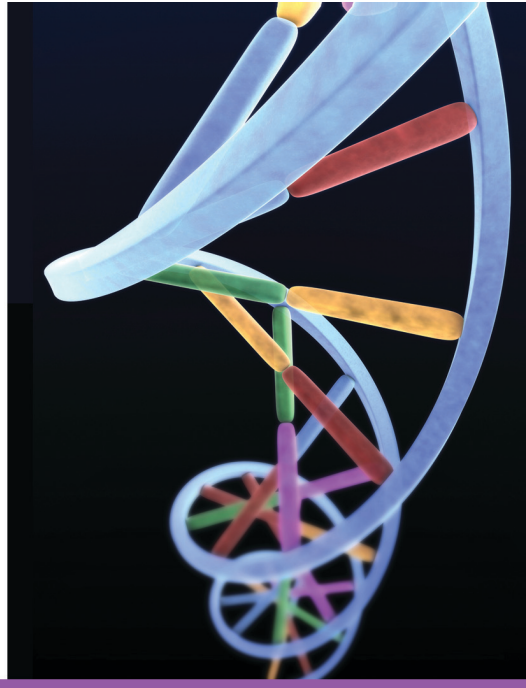
At costs below $1,000 per genome, a number of intriguing applications of DNA sequencing become cost effective. For example, researchers will have access to thousands or even millions of human genomes to seek correlations between genotypes and phenotypes. Medical doctors will be able to order genome sequencing along with standard laboratory tests, and will likely do so if they believe that knowledge of the DNA sequence will facilitate patient diagnosis and/or treatment. Even web-based genetic testing service companies will exploit full genome sequences to gather and dispense medical and ancestry information, and provide genetic counseling."

JASON Program. "The $100 Genome: Implications for the DoD." December 2010. JASON, The MITRE Corporation. McLean VA.

From an economic standpoint, of course, the impact of applied genomics will be equally profound. Modern developed societies are driven forward by innovation—typically technological innovation. The HGP and associated sequencing projects have advanced the state of innovation and technology and resulted in broad economic impacts as highlighted in this report. These positive economic impacts will continue to ripple outwards and expand. Personalized medicine approaches for the prevention and optimized treatment of diseases and disorders holds promise for reducing the future burden of disease and associated healthcare costs.

In the developing world, the promise contained within applied genomics is fundamental to basic economic progress. Without affordable and sustainable food supplies, development will be highly limited as malnutrition impacts developing society productivity and incomes. Without higher productivity crops, more and more marginal lands will be pressed into production and environmental degradation will become increasingly more widespread. Without enhanced abilities to combat current and emerging infectious and endemic diseases, developing societies will again be held back and progress limited. Applied genomics, built upon the foundation of the HGP and associated projects, holds promise for the provision of solutions to each and every one of these challenges and more.

In Battelle's interview with Lee Hood he called the HGP "the single most transformative event in the history of biological science." It would be difficult not to concur with that statement. In advancing basic scientific knowledge—in opening new pathways to enhanced human health—in spurring technologies for economic development—in providing solutions to widespread challenges in global food production and environmental sustainability—**the $3.8 billion spent on the HGP may well represent the best single investment ever made in science.**



Prepared by Battelle Technology Partnership Practice
with support provided by Life Technologies Foundation

Battelle
The Business of Innovation

www.battelle.org

life technologies™
FOUNDATION

www.lifetechnologies.com/foundation