The Human Genome: Unlocking Life's Code

Eric Green, M.D., Ph.D.

Director, NHGRI

NIH National Human Genome Research Institute

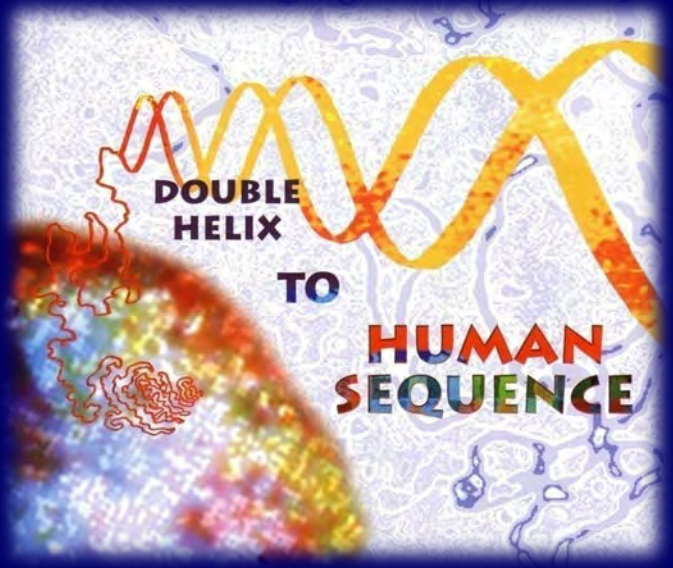
The Relevance of Genomics

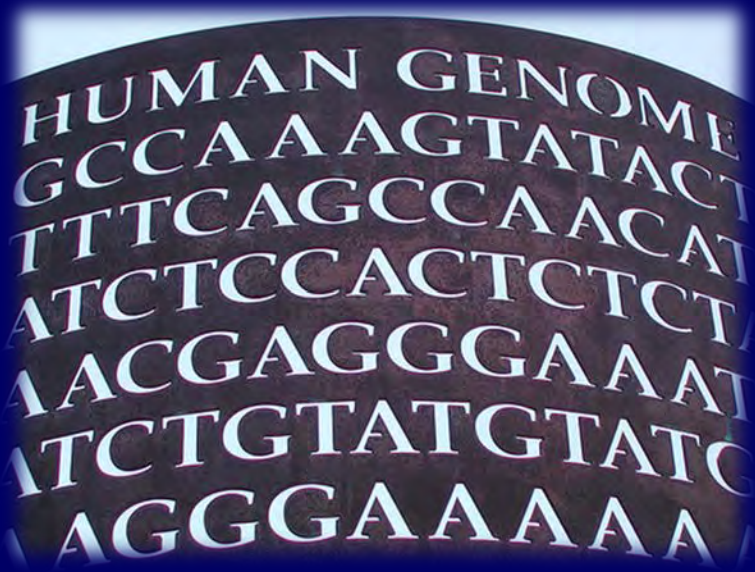
Biomedical Researchers

Healthcare Professionals

Patients (and Friends & Relatives of Patients)

Human Genome Project
1990-2003





NIH

Turning discovery into health

Advancing human health through genomics research

Genomic Medicine

An emerging medical discipline that involves using an individual's genomic information as part of their clinical care

February, 2011


PERSPECTIVE

doi:10.1038/nature09764

Charting a course for genomic medicine from base pairs to bedside

omim and http://www.genome.gov/GWAStudies) and the role of structural variation in disease[8], some of which have already led to new therapies[9–13]. Other advances have already changed medical practice (for example, microarrays are now used for clinical detection of genomic imbalances[14] and pharmacogenomic testing is routinely performed before administration of certain medications[15]). Together, these achievements (see accompanying paper[16]) document that genomics is contributing to a better understanding of human biology and to improving human health.

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

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

The past decade has seen genomics contribute fundamental knowledge about biology and its perturbation in disease. Further deepening this understanding will accelerate the transition to genomic medicine (clinical care based on genomic information). But significant change rarely comes

decade. Similarly, we note three cross-cutting areas that are broadly relevant and fundamental across the entire spectrum of genomics and genomic medicine: bioinformatics and computational biology (Box 3), education and training (Box 4), and genomics and society (Box 5).

Understanding the biology of genomes

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

Comprehensive catalogues of genomic data

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

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

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

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

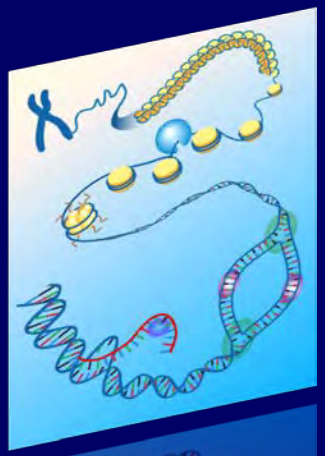
[1]National Human Genome Research Institute, National Institutes of Health, 31 Center Dr., Bethesda, Maryland 20892-2152, USA.
*Lists of participants and their affiliations appear at the end of the paper.

New NHGRI Vision for Genomics Published

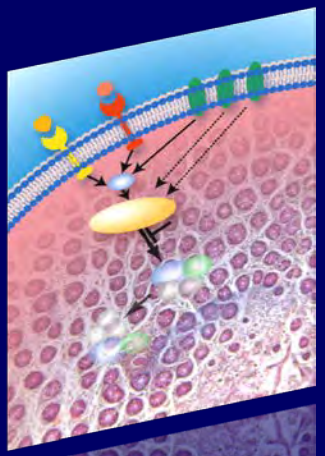
Five Domains of Genomics Research

Understanding the Structure of Genomes

Understanding the Biology of Genomes



Understanding the Biology of Disease

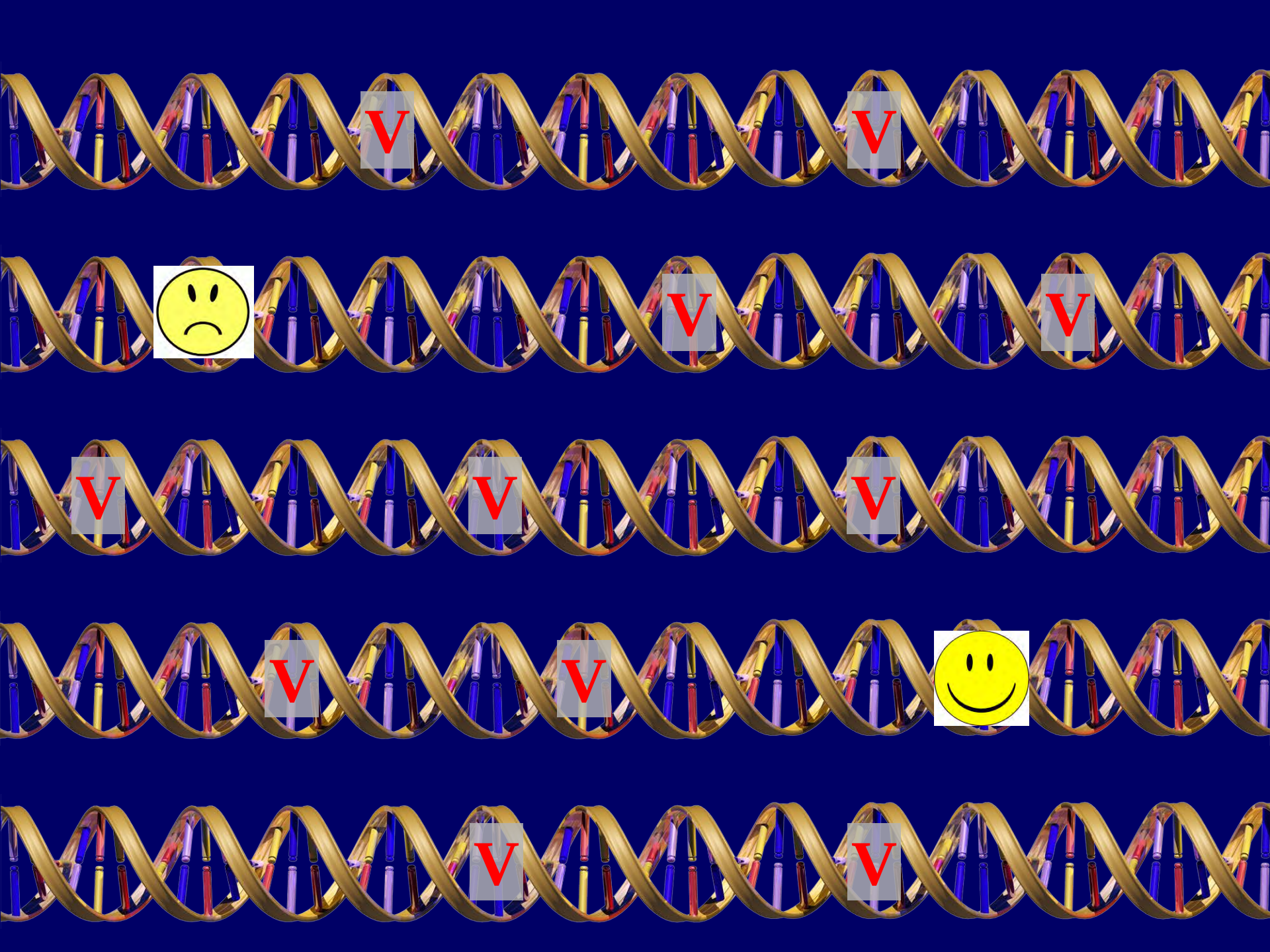


Advancing the Science of Medicine

Improving the Effectiveness of Healthcare

~3,000 bp (0.0001%) of Human Genome Sequence

TGCCGCGGAACTTTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTTCCTTCCCTTTTCAAAATGCACCTTGCAAACGTAACAG
GAACCCGACTAGGATCATCGGGAAAAGGAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGGTGGCAGCGGGTCCTGGGTCTGGCGGACCCTGA
CGCGAAGGAGGGTCTAGGAAGCTCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGTCCTCCAGCGTTGCCAACTGGACCTAAAGAGAGG
CCGCGACTGTCGCCCACCTGCGGGATGGGCCTGGTGCTGGGCGGTAAGGACACGGACCTGGAAGGAGCGCGCGCGAGGGAGGGAGGCTGGGAGTC
AGAATCGGGAAAGGGAGGTGCGGGGCGGCGAGGGAGCGAAGGAGGAGAGGAGGAAGGAGCGGGAGGGGTGCTGGCGGGGGTGCGTAGTGGGTGGA
GAAAGCCGCTAGAGCAAATTTGGGGCCGGACCAGGCAGCACTCGGCTTTTAACCTGGGCAGTGAAGGCGGGGGAAAGAGCAAAAGGAAGGGGTGG
TGTGCGGAGTAGGGGTGGGTGGGGGGAATTGGAAGCAAATGACATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCAGAGTAGTAG
GTCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCCCAGCGCCCGAGAGACCATGCAGAGGTCGCCTCTGGAAAAGGCCAGCGT
TGTCTCCAAACTTTTTTTCAGGTGAGAAGGTGGCCAACCGAGCTTCGGAAAGACACGTGCCCACGAAAGAGGAGGGCGTGTGTATGGGTTGGGTT
TGGGGTAAAGGAATAAGCAGTTTTTAAAAAGATGCGCTATCATTCATTGTTTTGAAAGAAAATGTGGGTATTGTAGAATAAAACAGAAAGCATTA
AGAAGAGATGGAAGAATGAACTGAAGCTGATTGAATAGAGAGCCACATCTACTTGCAACTGAAAAGTTAGAATCTCAAGACTCAAGTACGCTACT
ATGCACTTGTTTTATTTCATTTTTCTAAGAAACTAAAAATACTTGTTAATAAGTACCTAAGTATGGTTTATTGGTTTTCCCCCTTCATGCCTTGG
ACACTTGATTGTCTTCTTGGCACATACAGGTGCCATGCCTGCATATAGTAAGTGCTCAGAAAACATTTCTTGACTGAATTCAGCCAACAAAAATT
TTGGGGTAGGTAGAAAATATATGCTTAAAGTATTTATTGTTATGAGACTGGATATATCTAGTATTTGTCACAGGTAAATGATTCTTCAAAAATTG
AAAGCAAATTTGTTGAAATATTTATTTTGAAAAAAGTTACTTCACAAGCTATAAATTTTAAAAGCCATAGGAATAGATACCGAAGTTATATCCAA
CTGACATTTAATAAATTGTATTCATAGCCTAATGTGATGAGCCACAGAAGCTTGCAAACTTTAATGAGATTTTTTAAAATAGCATCTAAGTTCGG
AATCTTAGGCAAAGTGTTGTTAGATGTAGCACTTCATATTTGAAGTGTTCTTTGGATATTGCATCTACTTTGTTCCTGTTATTATACTGGTGTGA
ATGAATGAATAGGTACTGCTCTCTCTTGGGACATTACTTGACACATAATTACCCAATGAATAAGCATACTGAGGTATCAAAAAAGTCAAATATGT
TATAAATAGCTCATATATGTGTGTAGGGGGGAAGGAATTTAGCTTTCACATCTCTCTTATGTTTAGTTCTCTGCATGTGCAGTTAATCCTGGAAC
TCCGGTGCTAAGGAGAGACTGTTGGCCCTTGAAGGAGAGCTCCTCCCTGTGGATGAGAGAGAAGGACTTTACTCTTTGGAATTATCTTTTTGTGT
TGATGTTATCCACCTTTTGTTACTCCACCTATAAAATCGGCTTATCTATTGATCTGTTTTCCTAGTCCTTATAAAGTCAAAATGTTAATTGGCAT
AAATTATAGACTTTTTTTAGCAGAGAACTTTGAGGAACCTAAATGCCAACCAGTCTAAAAATGCAGTTTTCAGAAGAATGAATATTTCATGGATA
GTTCTAAATACTAATGAACTTTAAAATAGCTTACTATTGATCTGTCAAAGTGGGTTTTTATATAATTTTCTTTTTACAAATCACCTGACACATTT
AATATAGGTTAAAAAATGCTATCAGGCTGGTTTGCAAAGAAAATGTATTACAAAGGCTGCTAAGTGTGTTAAGAGCATACTCATTTCTGTTCTCC
AAAATATTTCATAAGGTGCTTTAAGAATAGGTATGTTTTTAAAAGTTAAGTTCCTACTATTTATAGGAACTGACAATCACCTAAAATACCAATGA
TTACAAACTTCCTTCTGGCCTTCTGGACTGCAATTCTAAAAGTGTAAAAAACATATTTTCTGCATTAAGTTAGGCAGTATTGCTTAGTTTTCAAA
GTGGTAGGCTTTGGAGTCAGATTATTTTGATTCAGATCCTACATCTACTGTTTAGTAGCTCTGTTGCCTGAGGCAGGTCCCTTAACATCTCTGTG
TGTGACTTGACCTTTAAAATTTGGAGACTGTCATAGGGGTTAATCCCTTGAGAAAATGAATGTGAAAAGTTAGCCTAATGTTAACTGCTATTATT
ATGGATTACCATATTTTCACATTCATCACAGTACATGCACCTTGTTAATATAAGATGCTCAATTCATCTTTGAGTATAATTTTGTGACTCTCAAT
CTGGATATGCAATGAGTGGGCCTGTATGAGAATTTAATTTATGAAAAATTGTGTTTCACATGGCCTTACCAGATATACAGGAAACACGTCACATG
TTTCTATTGTATGTTGTTAAATGCCTTAGAATTTAACTTTCTGAATAGGATCCCTTCAGTTTGAGAGTCATAAAAGAGTAAAATTATTATGGTAT

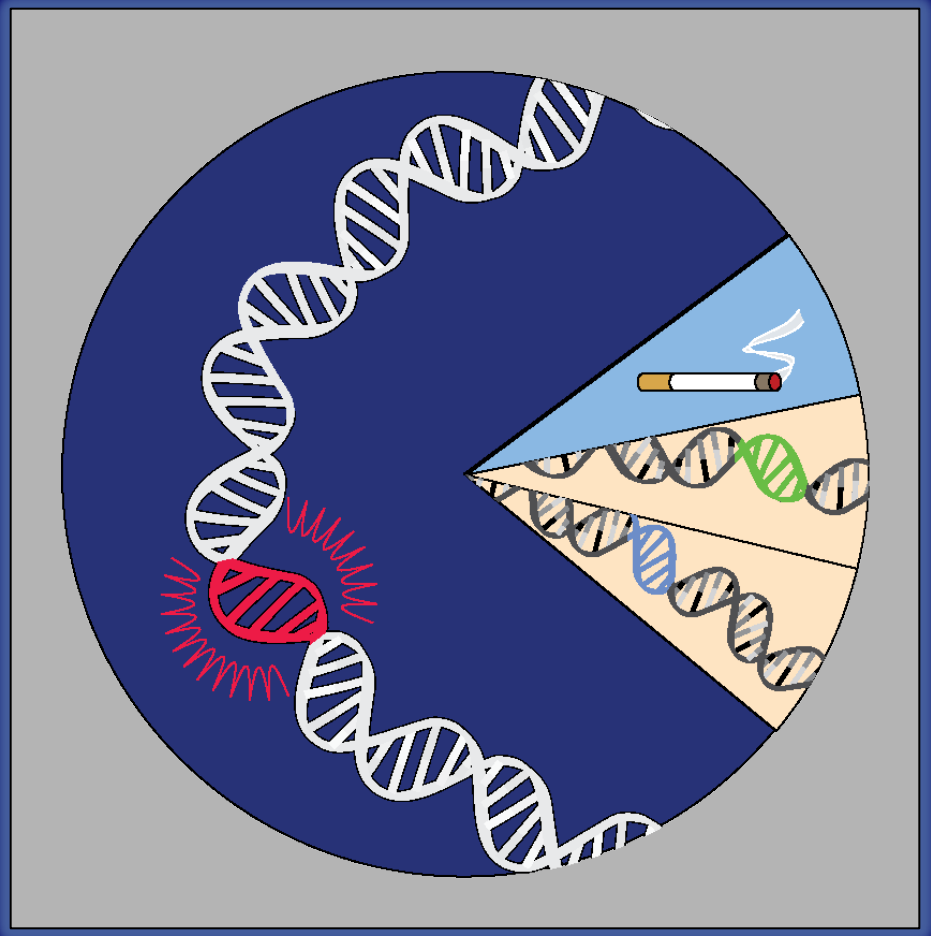

V
V
V
V
V
V
V
V
V
V
V
V
V
V

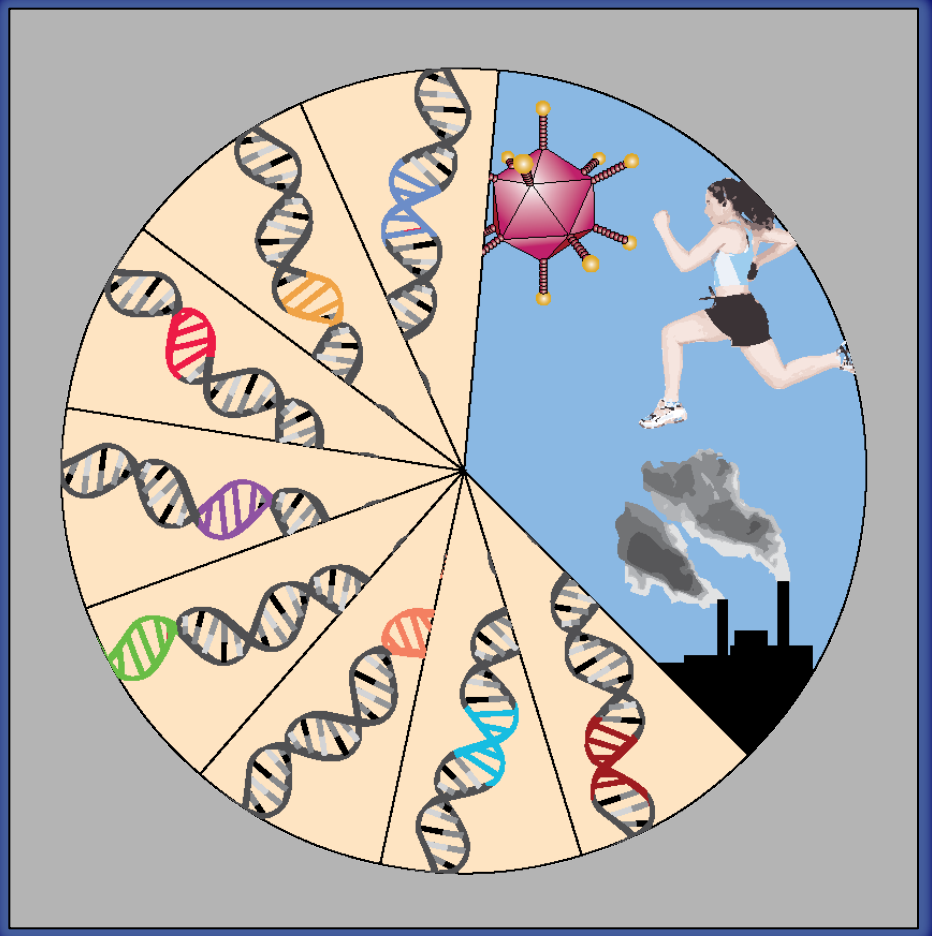
Your Genome: By the Numbers

- ~6B nucleotides
- ~3-5M single-nucleotide variants
  - ~150K not in databases
  - ~60 not in either parent
- ~10-20K structural variants

- ~100 ‘disruptive’ variants in genes
- ~20 completely inactivated genes (both copies)

Genomic Architecture of Genetic Diseases





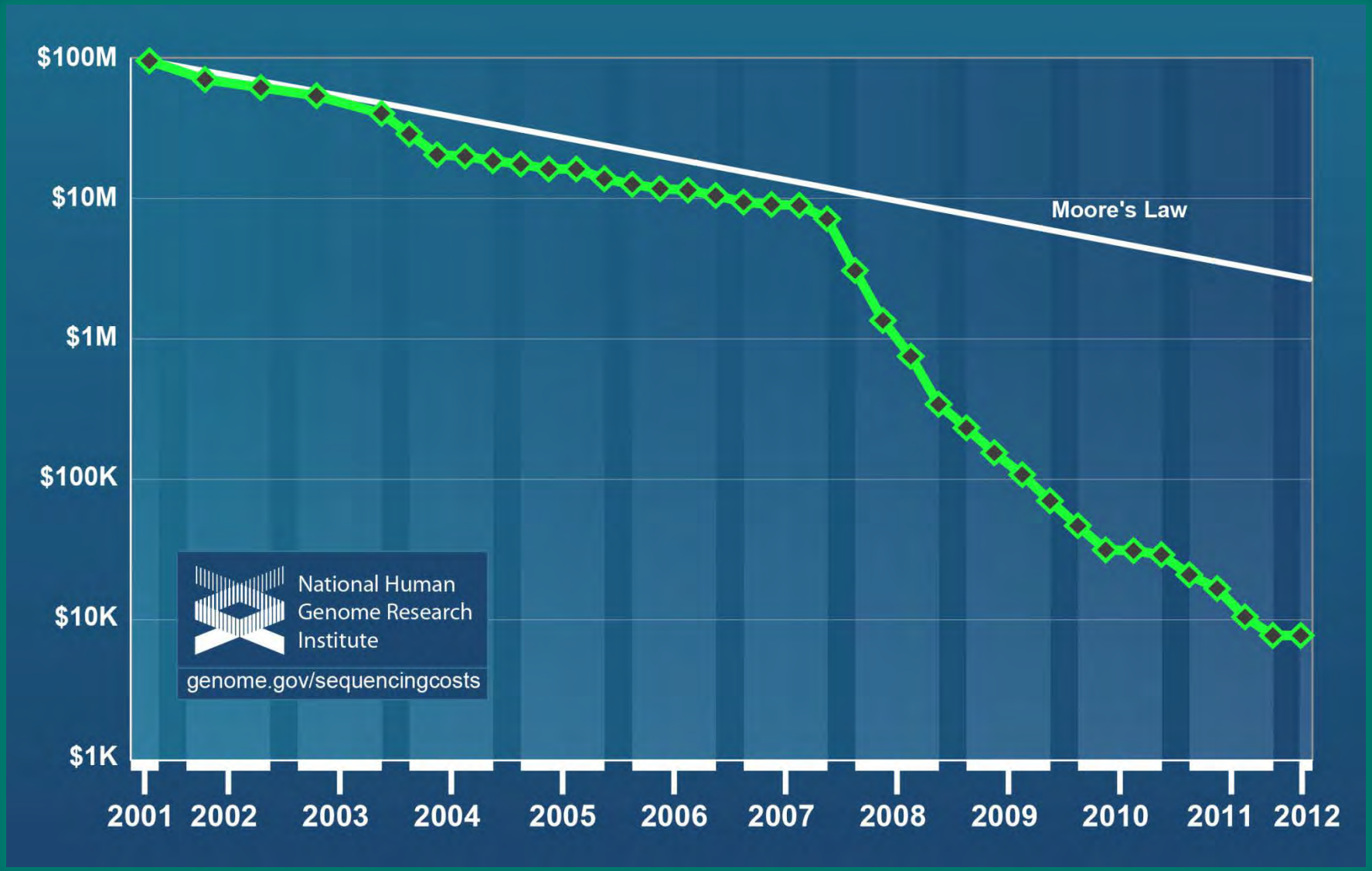
Rare, Simple, Monogenic, Mendelian…

Common, Complex, Multigenic, Non-Mendelian…

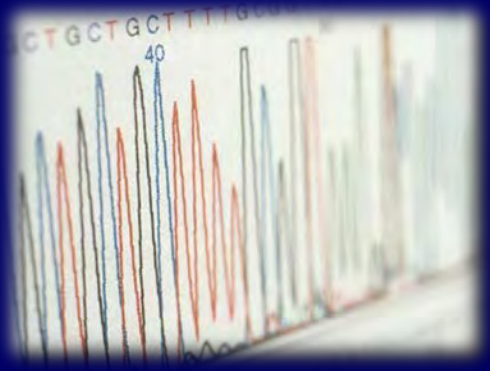
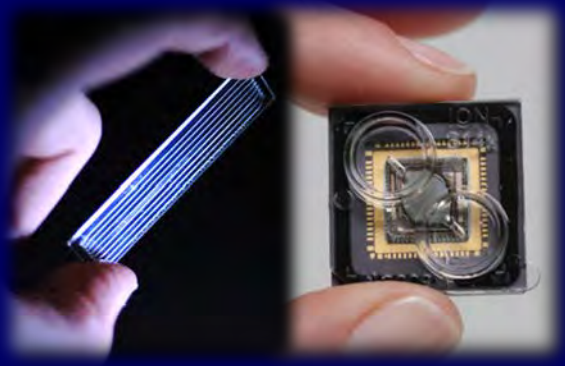
Mostly *Coding* Mutations

Mostly *Non-Coding* Mutations

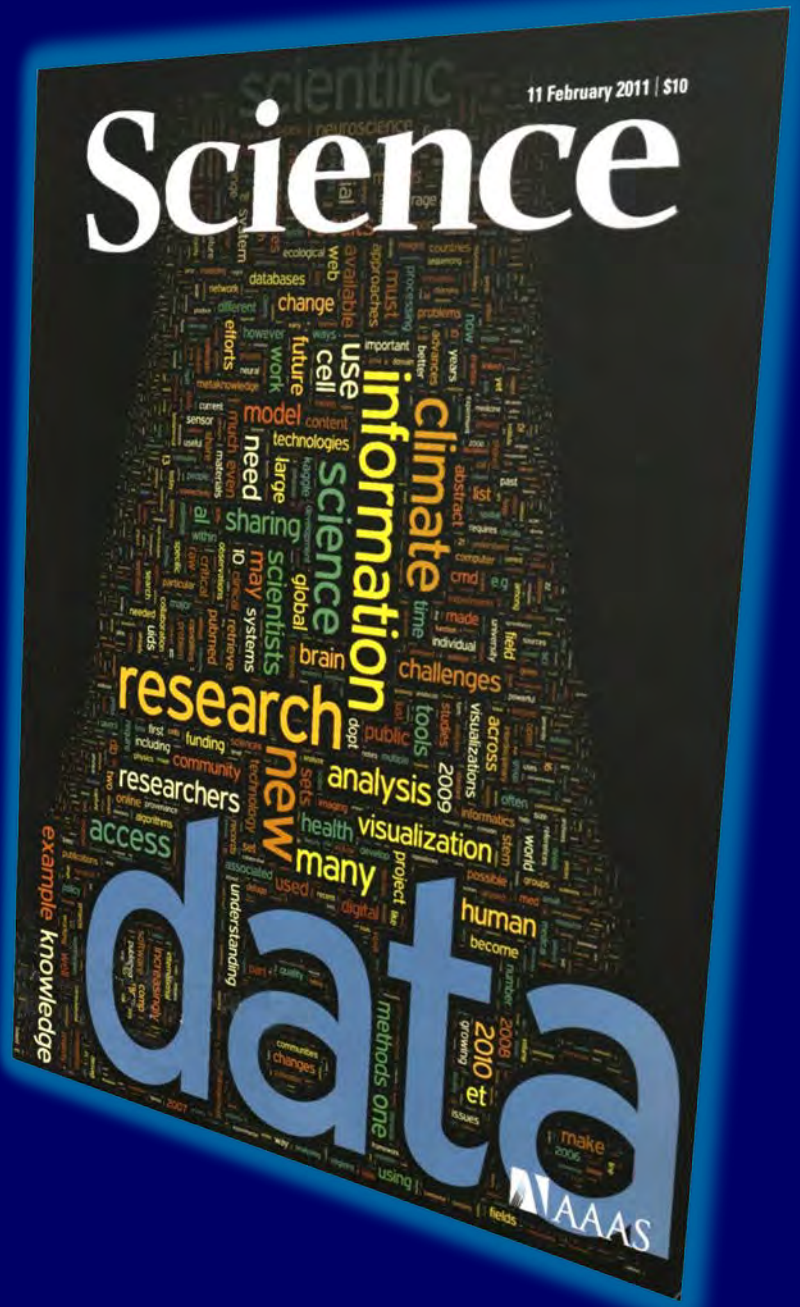
Cost per Sequenced Human Genome



Sequencing a Human Genome

HGP (1st Sequence)	*Immediate Post-HGP*	*Today*
		
~6-8 years	**~3-4 months**	**~2-3 days**
~$1B	**~$10-50M**	**~$4-6K**


4 September 2008 | www.nature.com/nature | $10
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
nature
THE BITER BIT
Viral infections for viruses
TROPICAL CYCLONES
The strong get stronger
BLACK HOLE PHYSICS
A new window on the Galactic Centre
BIG DATA
NATUREJOBS
Minnesota musings
SCIENCE IN THE PETABYTE ERA


11 February 2011 | $10
Science
data
climate
information
science
research
challenges
analysis
visualization
researchers
access
new
many
human
AAAS

Genomics and the Economy

The Impact of Genomics on the U.S. Economy

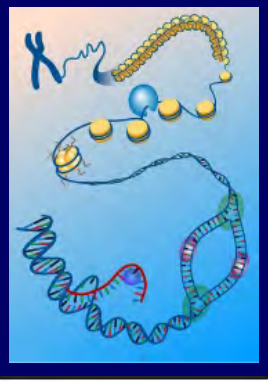
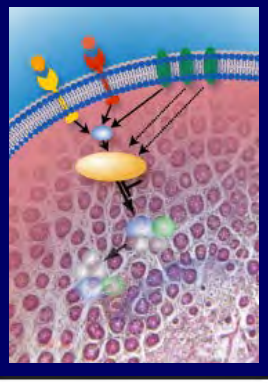
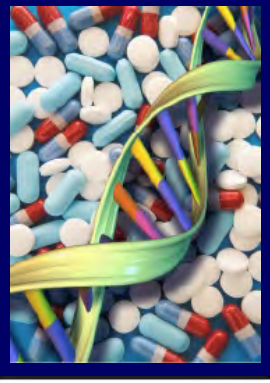
How a federal research investment has changed medical science, improved public health, boosted the U.S. economy and created American jobs

Prepared by Battelle Technology Partnership Practice for United for Medical Research (UMR)

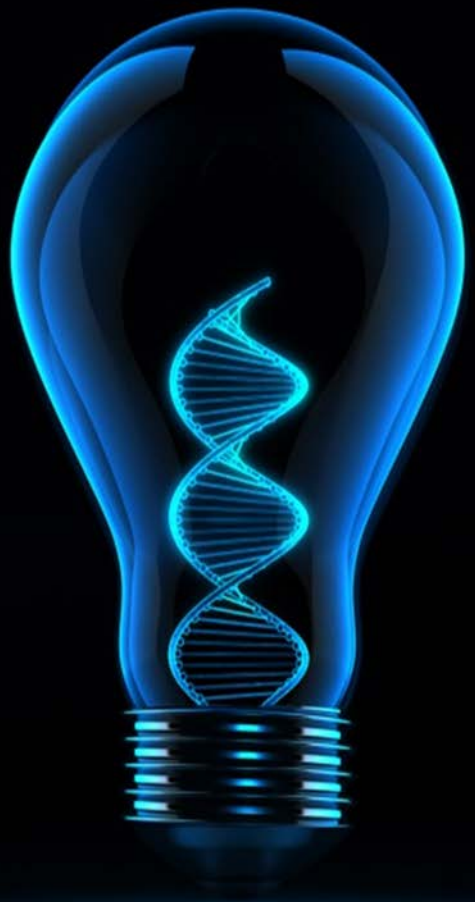
- From 1988-2012, genomic activities generated $965 billion in economic output
- In 2012, genomic activities generated $65 billion in economic output
- Genomics has a remarkable demonstrated return-on-investment


The Future
NEXT EXIT

Genomic Accomplishments Across Domains

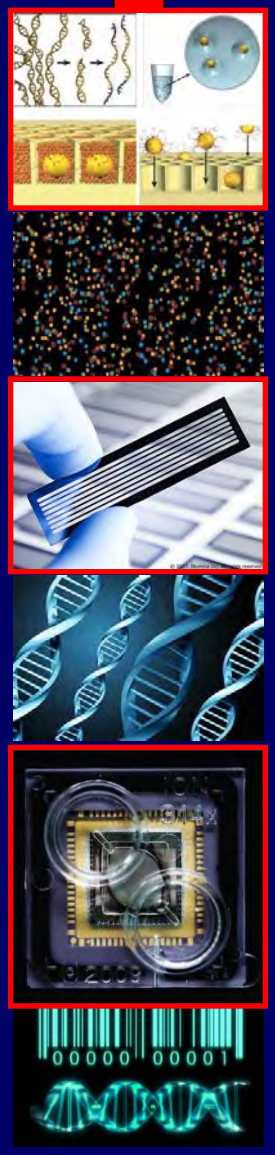
	Understanding the Structure of Genomes	Understanding the Biology of Genomes	Understanding the Biology of Disease	Advancing the Science of Medicine	Improving the Effectiveness of Healthcare
1990-2003 *Human Genome Project*					
2004-2010					
2011-2020					
Beyond 2020					

‘Hot Areas’ in Genomic Medicine



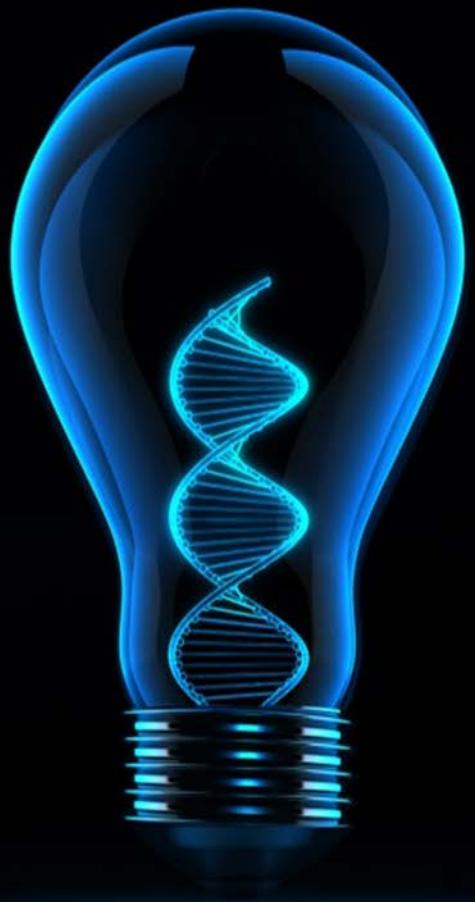
Cancer Genomics

Genome Sequencing: Cancer



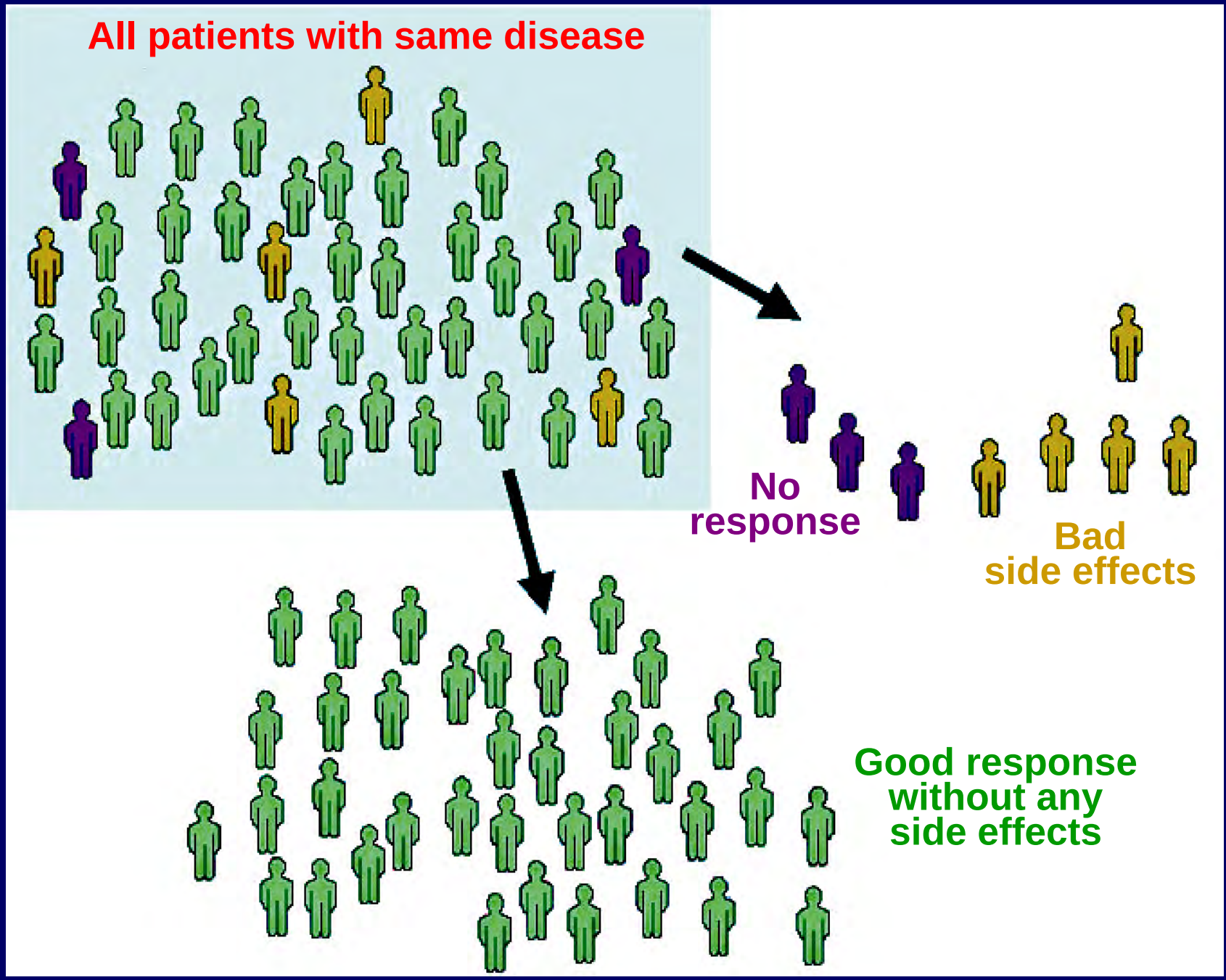
Cancer Genomics

‘Hot Areas’ in Genomic Medicine

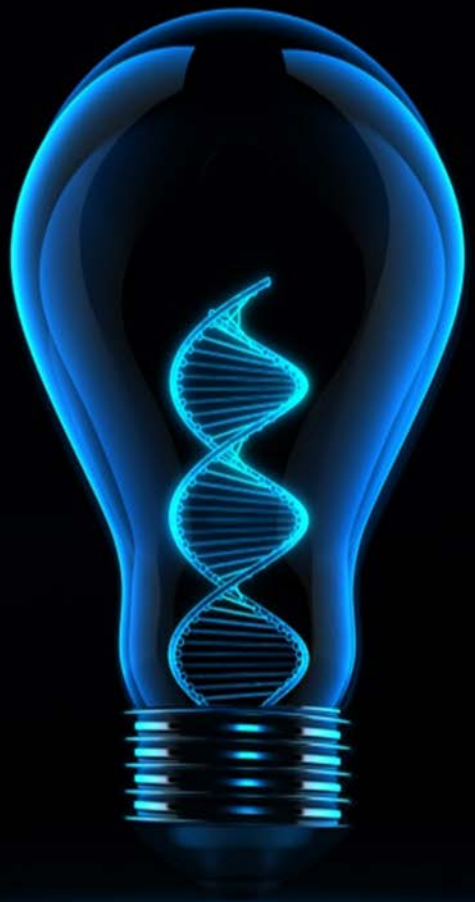


Cancer Genomics

Pharmacogenomics


All patients with same disease
No response
Bad side effects
Good response without any side effects

‘Hot Areas’ in Genomic Medicine

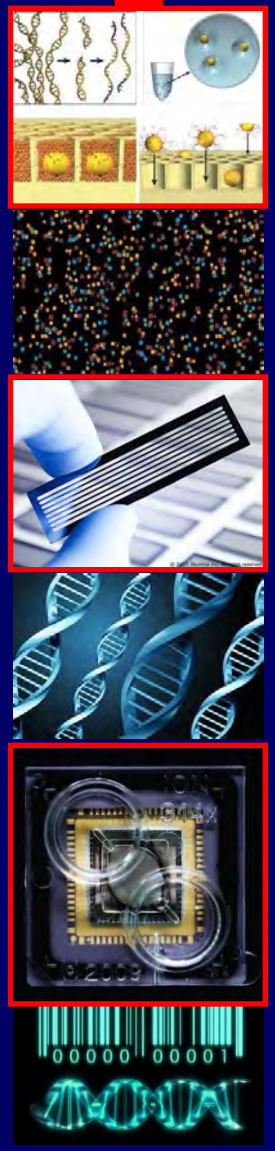


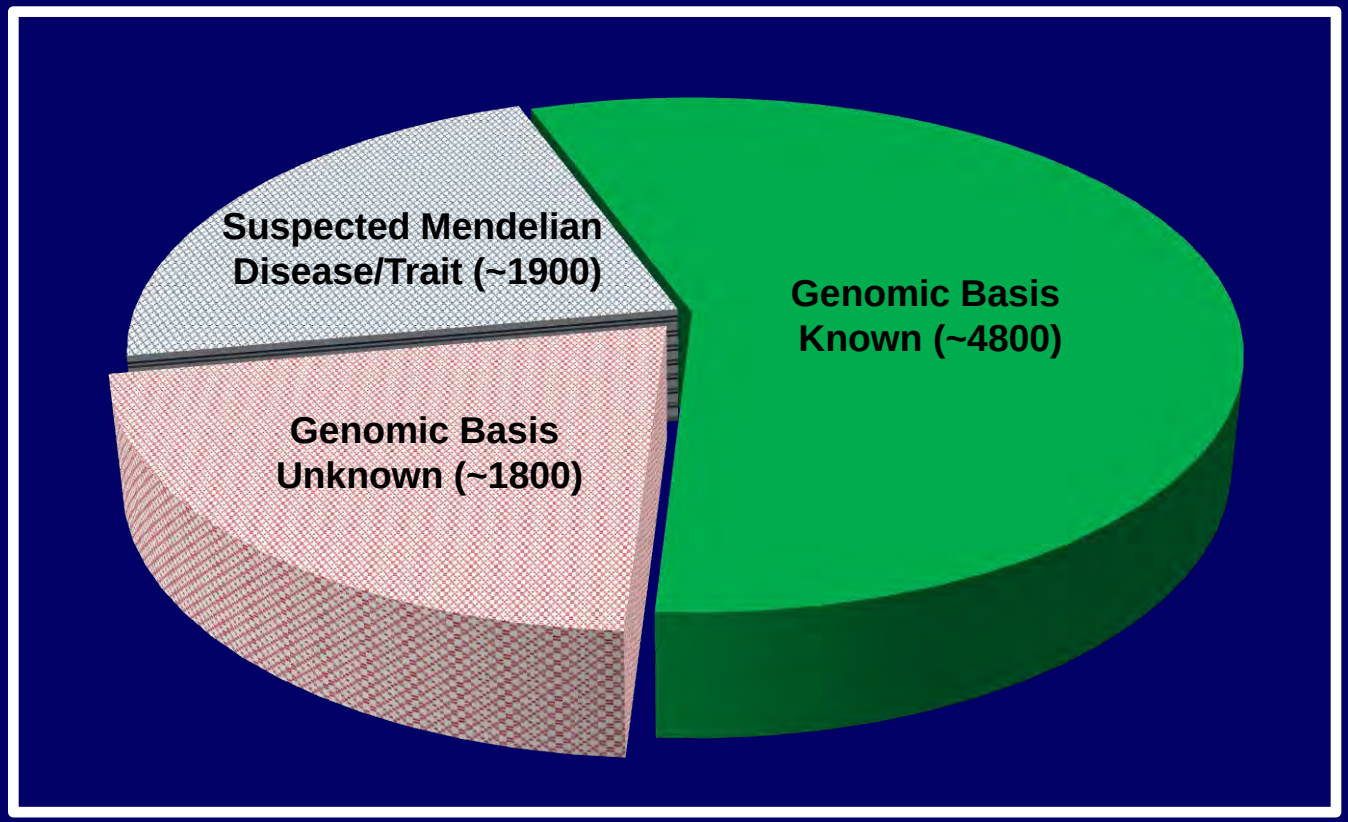
Cancer Genomics

Pharmacogenomics

Rare & Common
Genetic Diseases

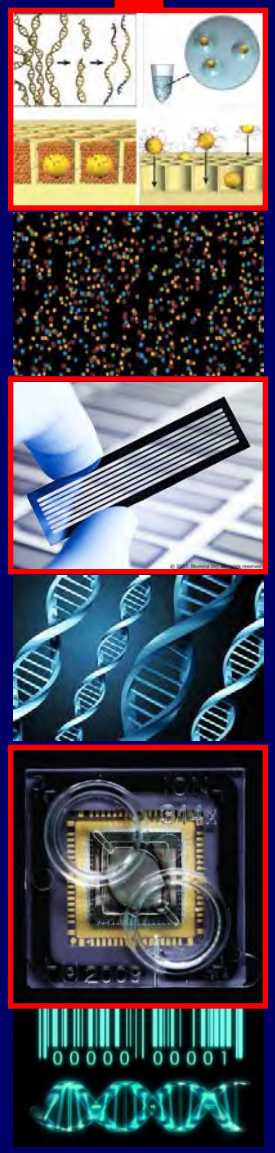
Genome Sequencing: Rare Diseases

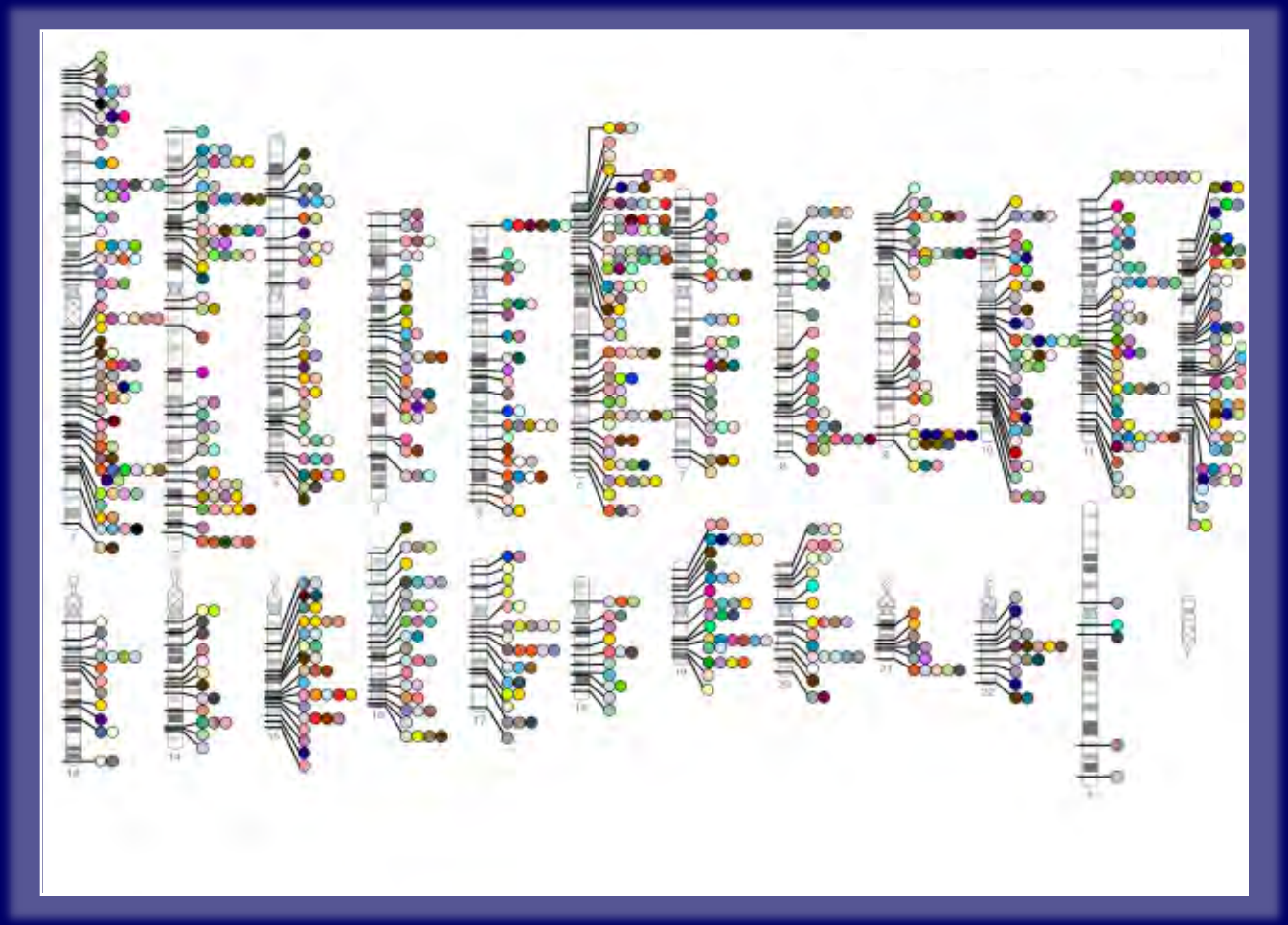




Mendelian Diseases/Traits

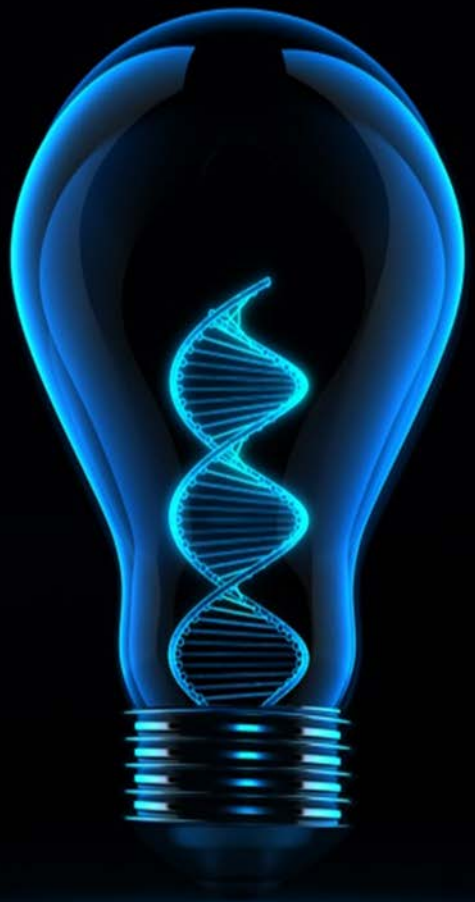
Genome Sequencing: Common Diseases





Complex Diseases/Traits

‘Hot Areas’ in Genomic Medicine

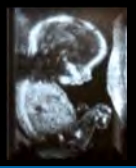


Cancer Genomics

Pharmacogenomics

Rare & Common
Genetic Diseases

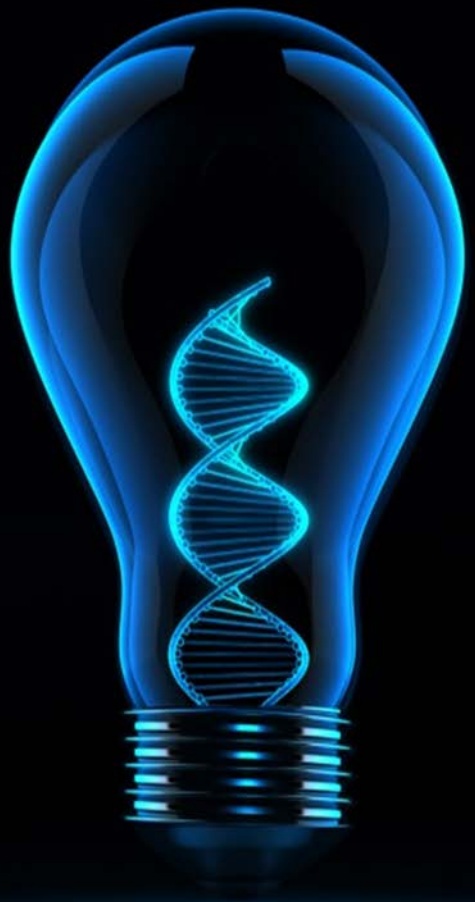
Prenatal & Newborn
Genomic Analysis




Clinical Genomics
Information Systems

Clinical Genomic Information Systems

‘Hot Areas’ in Genomic Medicine

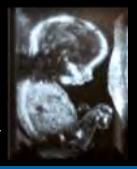


- Cancer Genomics

- Pharmacogenomics

- Rare & Common Genetic Diseases

- Prenatal & Newborn Genomic Analysis



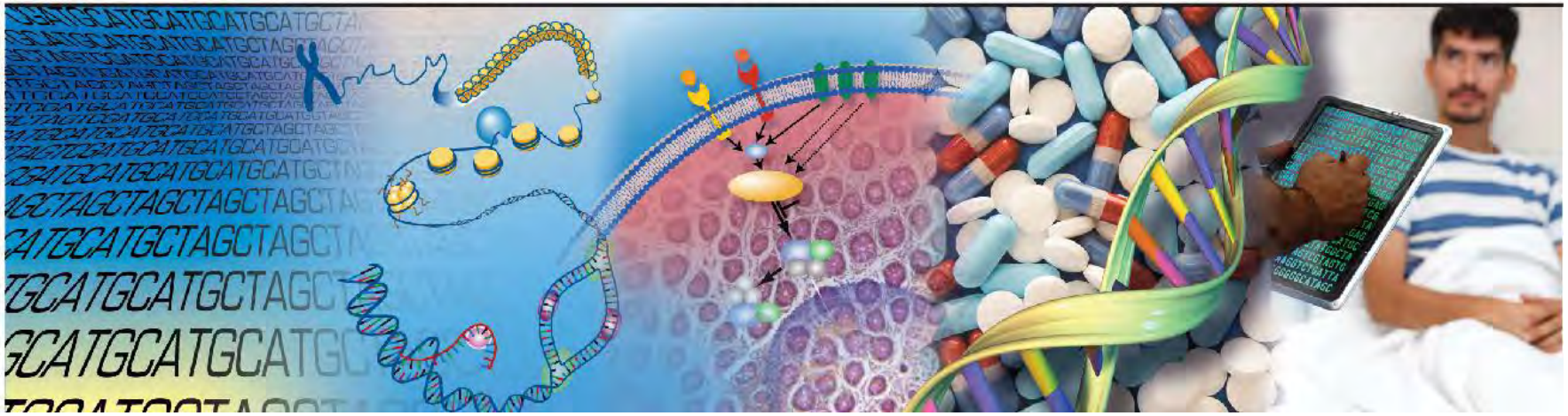

- Clinical Genomics Information Systems

- Genomic Medicine ‘Test Drive’ Programs

Bringing Genomic Medicine into Focus

~1990 ~2003 ~2011 ~2020

Understanding the Structure of Genomes | Understanding the Biology of Genomes | Understanding the Biology of Disease | Advancing the Science of Medicine | Improving the Effectiveness of Healthcare



A pessimist sees the difficulty in every opportunity.

An optimist sees the opportunity in every difficulty.

--Winston Churchill

NATIONAL HUMAN GENOME RESEARCH INSTITUTE

Advancing human health through genomics research