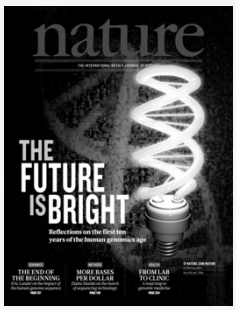


The Genomic Landscape *Circa 2012*

Eric Green, M.D., Ph.D.
Director, NHGRI

JOHNS HOPKINS MEDICINE
CONTINUING MEDICAL EDUCATION

Current Topics in Genome Analysis 2012

Eric Green

No Relevant Financial Relationships with Commercial Interests

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Outline

I. Historical Context for Genomics

II. Major Achievements since the Human Genome Project

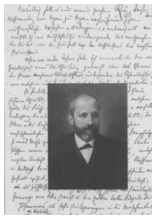
III. The Human Genomics Landscape: 2012 and Beyond

>>> Goal: Place Future Speakers into a Broader Context <<<

Foundational Milestones in Genetics & Genomics

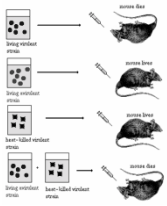
Mendel

1865



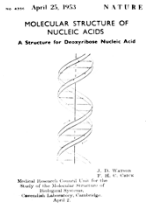
Miescher

1871



Avery

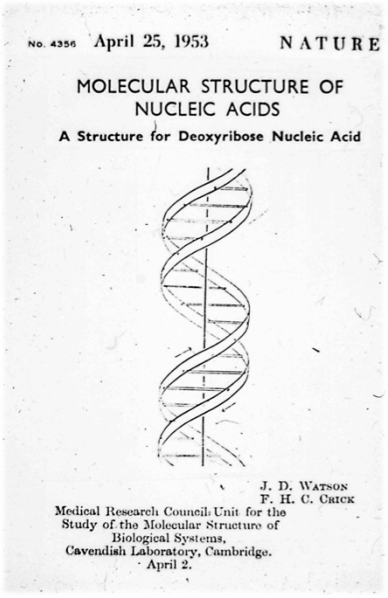
1944



Watson & Crick

1953

April, 1953



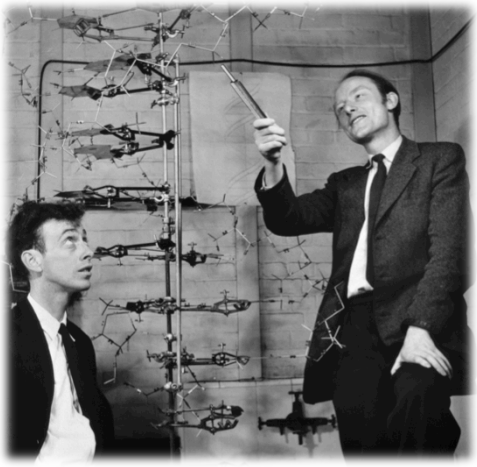
No. 4356 April 25, 1953 NATURE

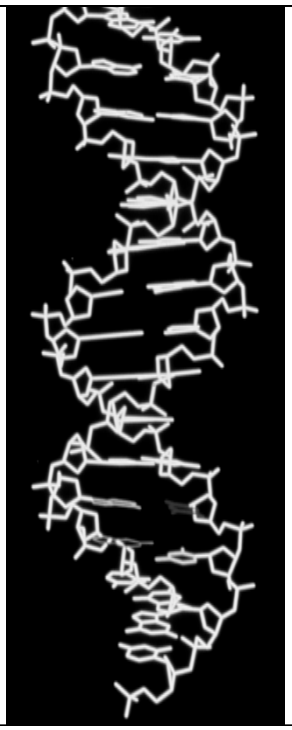
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

J. D. WATSON
F. H. C. CRICK

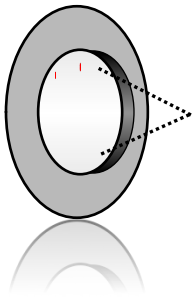
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge.
April 2.





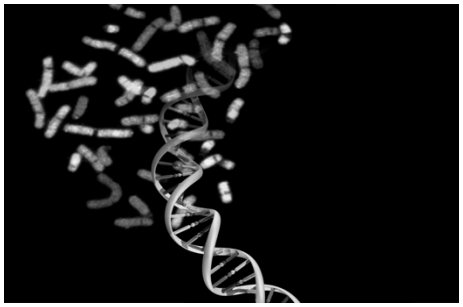
The DNA Alphabet

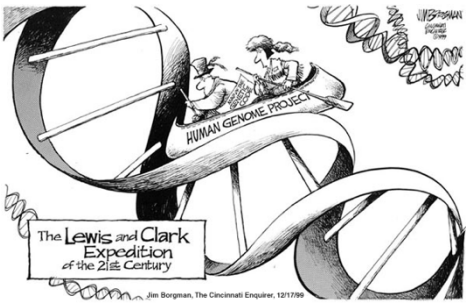
G Guanine
A Adenine
T Thymine
C Cytosine



Human Genome: ~3 billion bases ('letters')

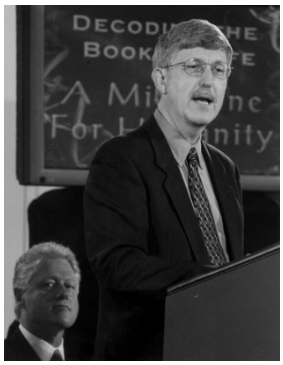
~21 Years Ago

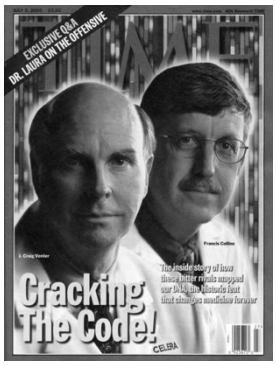




October 1990
Human Genome Project Begins

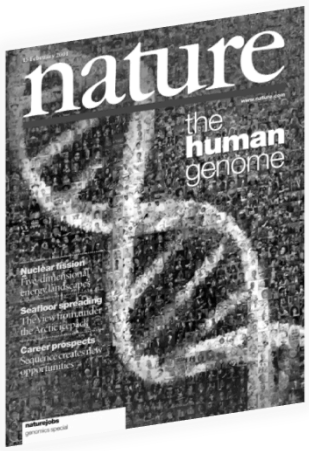
~11 Years Ago





June 2000
Draft Human Genome Sequence Announced

~11 Years Ago



February 2001
Draft Human Genome Sequence Published

~9 Years Ago

April 2003
Human Genome Project Ends

guardian.co.uk

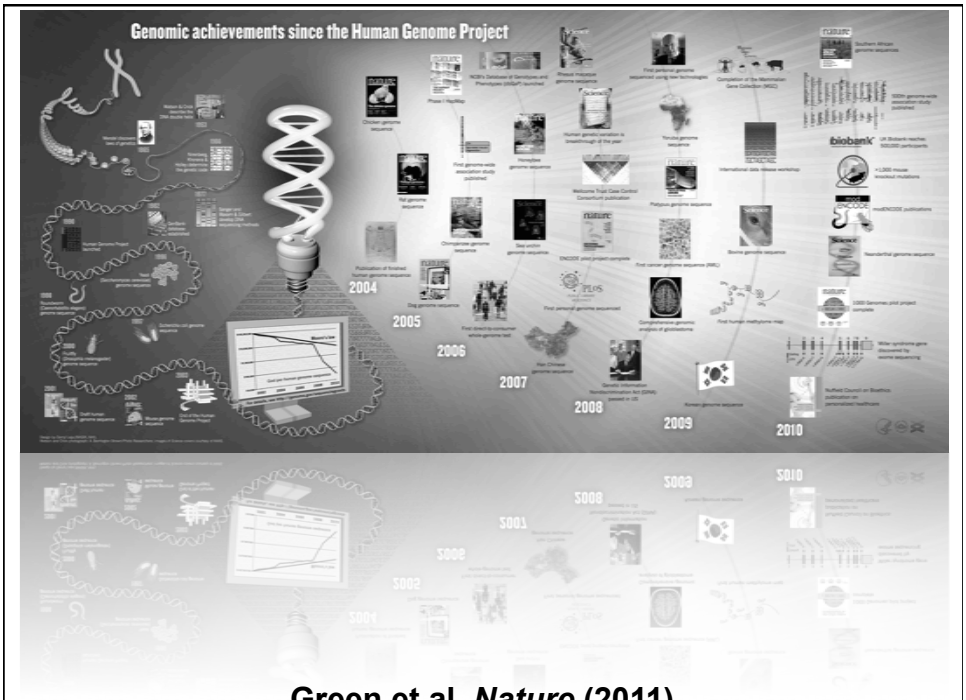
Adam Rutherford
guardian.co.uk, Thursday 21 April 2011 09.59 BST

A larger | smaller

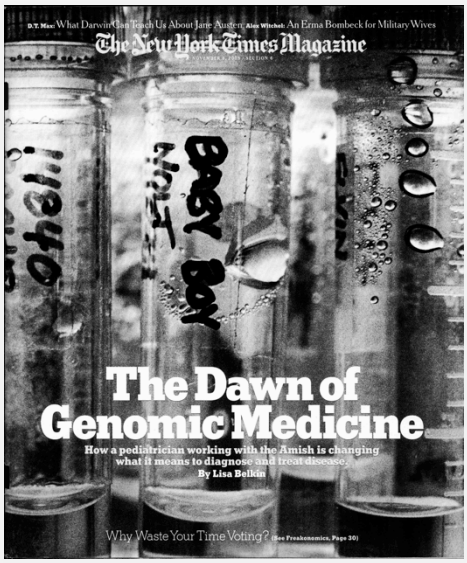
The Human Genome Project was just the starting point

April 2011

"But the mistake that we often make is [saying that the Human Genome Project] was an end point. In fact, the Human Genome Project was a pregnancy… Ten years later, we now have a clue what we don't yet know. The Human Genome Project may be finished, but understanding our genome is only just beginning."



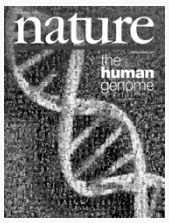
Green et al. *Nature* (2011)



Genomic Medicine

Healthcare tailored to the individual based on genomic information

The Path to Genomic Medicine



?

Human Genome Project

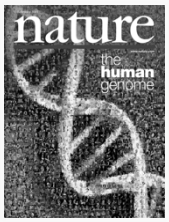
Realization of Genomic Medicine

"Fulfilling the Promise"

✓

?

The Path to Genomic Medicine



Data/Knowledge/Insights

Human Genome Project

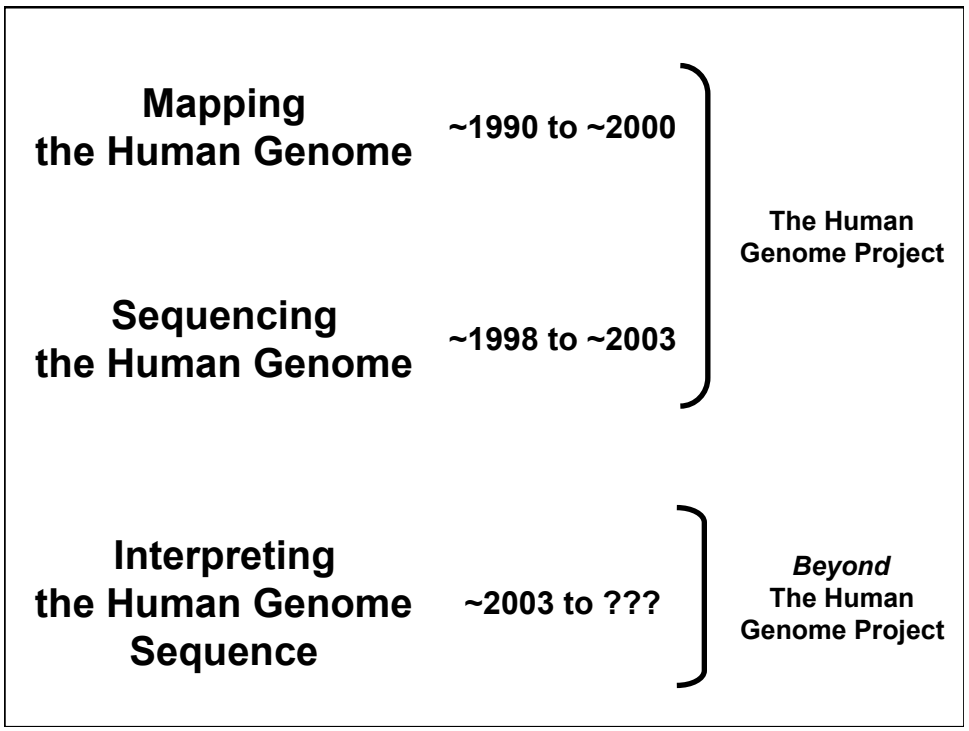
Realization of Genomic Medicine

The Path to Genomic Medicine

Function of the Human Genome Sequence

Human Genome Project

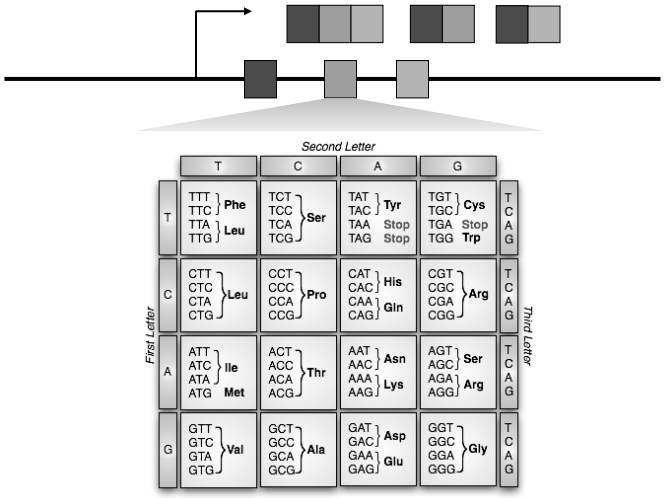
Realization of Genomic Medicine



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

```
TGCCGCGGAACTTTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTTCCTTCCCTTTTCAAAATGCACCTTGCAAACGTAACAG
GAACCCGACTAGGATCATCGGGAAAAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGGTGGCAGCGGGTCCTGGGTCTGGCGGACCCTGA
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
```

Coding Sequences (i.e., Genes)

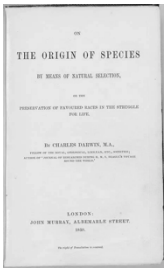


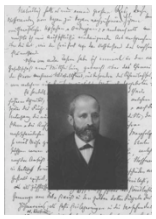
The Genetic Code

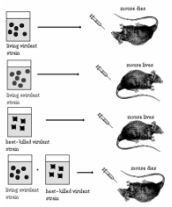
~3,000 bp (0.0001%) of Human Genome Sequence

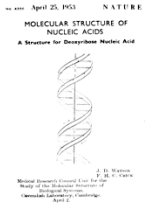
```
TGCCGCGGAACTTTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTTCCTTCCCTTTTCAAAATGCACCTTGCAAACGTAACAG
GAACCCGACTAGGATCATCGGGAAAAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGGTGGCAGCGGGTCCTGGGTCTGGCGGACCCTGA
CGCGAAGGAGGGTCTAGGAAGCTCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGTCCTCCAGCGTTGCCAACTGGACCTAAAGAGAGG
CCGCGACTGTCGCCCACCTGCGGGATGGGCCTGGTGCTGGGCGGTAAGGACACGGACCTGGAAGGAGCGCGCGCGAGGGAGGGAGGCTGGGAGTC
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
```

Foundational Milestones in Genetics & Genomics





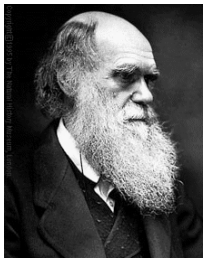




Darwin	Mendel	Miescher	Avery	Watson & Crick
1859	*1865*	*1871*	*1944*	*1953*

"*It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is the most adaptable to change.*"

(Attributed to Darwin)

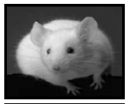
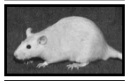


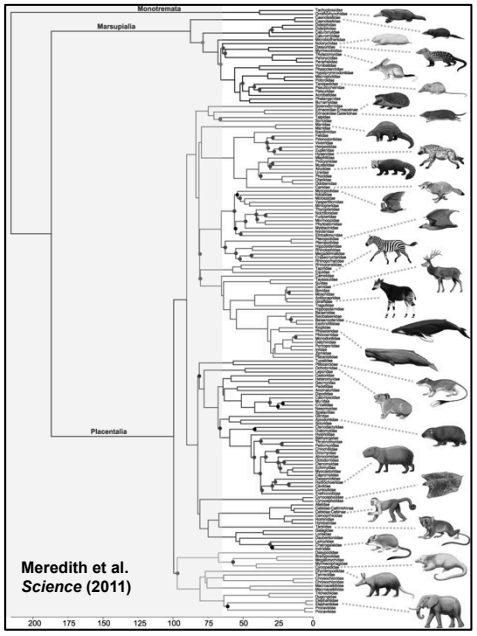
Charles Darwin (1809-1882)

"*For the last three and a half billion years, evolution has been taking notes.*"

–Eric Lander

Comparative Genome Sequencing



Meredith et al.
Science (2011)





Comparative Genome Sequencing

A high-resolution map of human evolutionary constraint using 29 mammals

Kerstin Lindblad-Toh[1,2], Manuel Garber[1]*, Or Zuk[1]*, Michael F. Lin[1,3]*, Brian J. Parker[4]*, Stefan Washietl[3]*, Pouya Kheradpour[1,3]*, Jason Ernst[1,3]*, Gregory Jordan[5]*, Evan Mauceli[1]*, Lucas D. Ward[1,3]*, Craig B. Lowe[6,7,8]*, Alisha K. Holloway[9]*, Michele Clamp[1,10]*, Sante Gnerre[1]*, Jessica Alföldi[1], Kathryn Beal[5], Jean Chang[1], Hiram Clawson[6], James Cuff[11], Federica Di Palma[1], Stephen Fitzgerald[5], Paul Flicek[5], Mitchell Guttman[1], Melissa J. Hubisz[12], David B. Jaffe[1], Irwin Jungreis[3], W. James Kent[9], Dennis Kostka[9], Marcia Lara[1], Andre L. Martins[12], Tim Massingham[5], Ida Moltke[4], Brian J. Raney[6], Matthew D. Rasmussen[3], Jim Robinson[1], Alexander Stark[13], Albert J. Vilella[5], Jiayu Wen[4], Xiaohui Xie[1], Michael C. Zody[1], Broad Institute Sequencing Platform and Whole Genome Assembly Team†, Kim C. Worley[14], Christie L. Kovar[14], Donna M. Muzny[14], Richard A. Gibbs[14], Baylor College of Medicine Human Genome Sequencing Center Sequencing Team†, Wesley C. Warren[15], Elaine R. Mardis[15], George M. Weinstock[14,15], Richard K. Wilson[15], Genome Institute at Washington University†, Ewan Birney[5], Elliott H. Margulies[16], Javier Herrero[5], Eric D. Green[17], David Haussler[6,8], Adam Siepel[12], Nick Goldman[5], Katherine S. Pollard[9,18], Jakob S. Pedersen[4,19], Eric S. Lander[1] & Manolis Kellis[1,3]

***Nature* (2011)**

The Human Genome… by the Numbers

~5% of Human Genome Sequence is Constrained Across Mammals (and Presumed Functional)

5% of 3B Bases = ~150M Bases

Lower Bound for the Amount that is Functional

~1.5% Encodes for Protein (Genes)

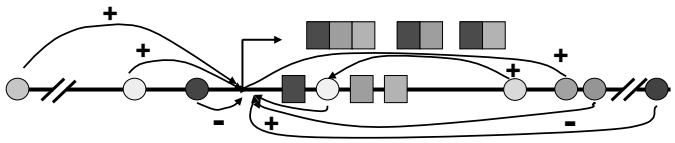
Corresponds to ~18-22K Genes

Many More than ~22K Different Proteins

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

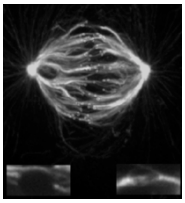
TGCCGCGGAACTTTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTTCCTTCCCTTTTCAAAATGCACCTTGCAAACGTAACAG
GAACCCGACTAGGATCATCGGGAAAAGGAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGGTGGCAGCGGGTCCTGGGTCTGGCGGACCCTGA
CGCGAAGGAGGGTCTAGGAAGCTCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGTCCTCCAGCGTTGCCAACTGGACCTAAAGAGAGG
CCGCGACTGTCGCCCACCTGCGGGATGGGCCTGGTGCTGGGCGGTAAGGACACGGACCTGGAAGGAGCGCGCGCGAGGGAGGGAGGCTGGGAGTC
AGAATCGGGAAAGGGAGGTGCGGGGCGGCGAGGGAGCGAAGGAGGAGAGGAGGAAGGAGCGGGAGGGGTGCTGGCGGGGGTGCGTAGTGGGTGGA
GAAAGCCGCTAGAGCAAATTTGGGGCCGGACCAGGCAGCACTCGGCTTTTAACCTGGGCAGTGAAGGCGGGGGAAAAGAGCAAAAGGAAGGGGTGG
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

Non-Coding Functional Sequences

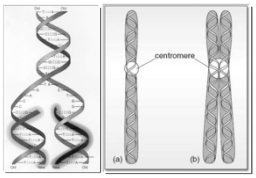


Gene Regulation

Chromosome Packaging



Chromosome Segregation



Chromosome Replication

Non-Coding RNAs

?

The Human Genome… by the Numbers

~5% of Human Genome Sequence is Constrained Across Mammals (and Presumed Functional)

- 5% of 3B Bases = ~150M Bases
- Do NOT Yet Know the Position of these ~150M Functional Bases
- Lower Bound for the Amount that is Functional

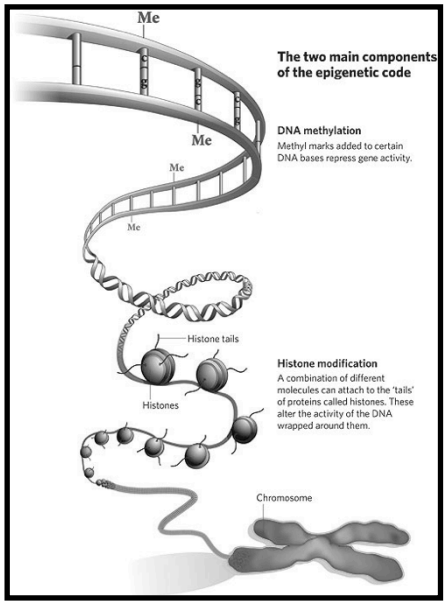
~1.5% Encodes for Protein (Genes)

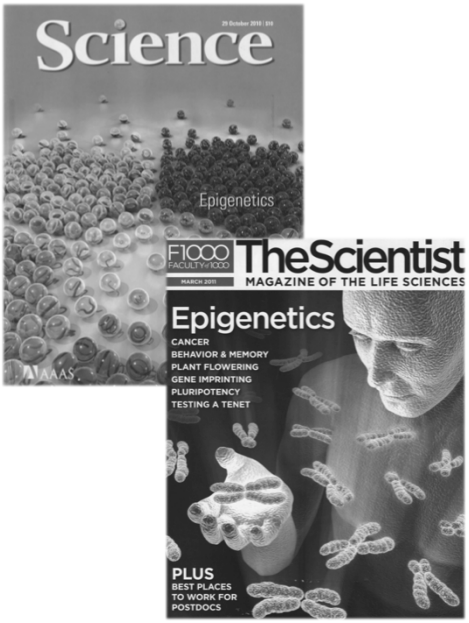
- Corresponds to ~18-22K Genes
- Many More than ~22K Different Proteins

~3.5% Functional But Non-Coding

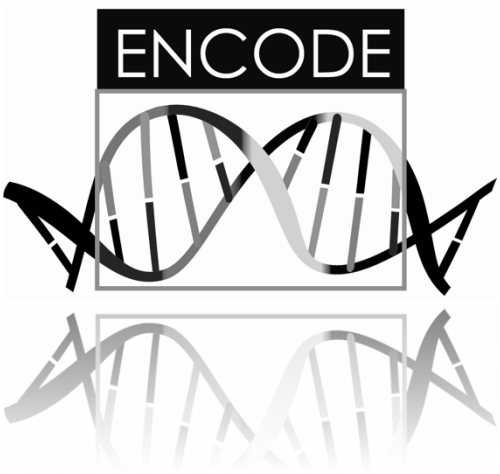
- Gene Regulatory Elements
- Chromosomal Functional Elements
- Undiscovered Functional Elements (NOT Yet in Textbooks!)

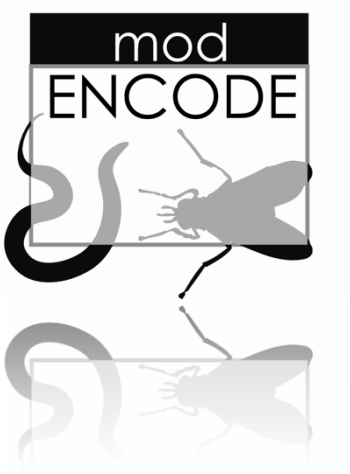
The Epigenomic Landscape

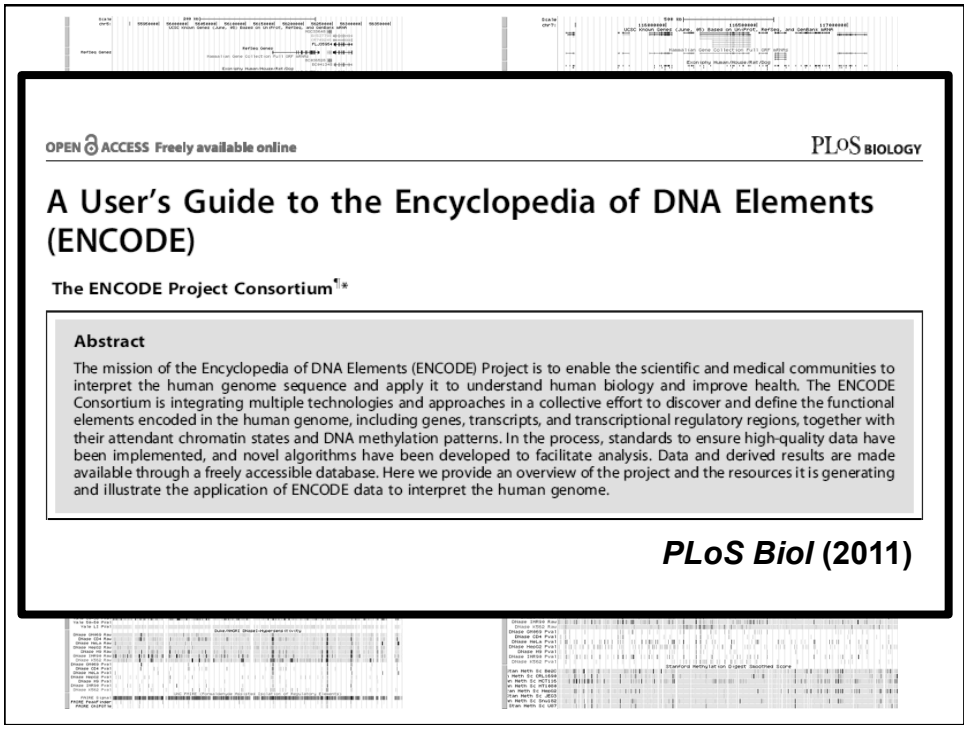




The ENCODE Portfolio







OPEN ACCESS Freely available online

PLoS BIOLOGY

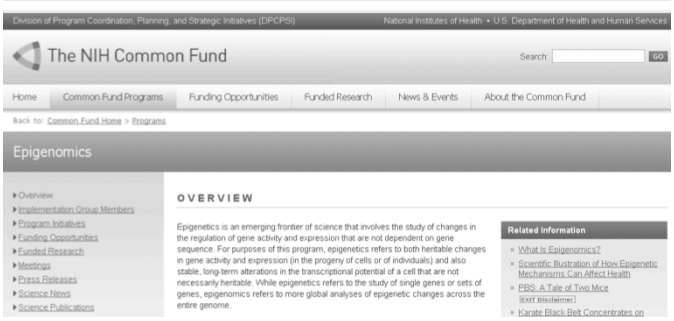
A User's Guide to the Encyclopedia of DNA Elements (ENCODE)

The ENCODE Project Consortium¶*

Abstract

The mission of the Encyclopedia of DNA Elements (ENCODE) Project is to enable the scientific and medical communities to interpret the human genome sequence and apply it to understand human biology and improve health. The ENCODE Consortium is integrating multiple technologies and approaches in a collective effort to discover and define the functional elements encoded in the human genome, including genes, transcripts, and transcriptional regulatory regions, together with their attendant chromatin states and DNA methylation patterns. In the process, standards to ensure high-quality data have been implemented, and novel algorithms have been developed to facilitate analysis. Data and derived results are made available through a freely accessible database. Here we provide an overview of the project and the resources it is generating and illustrate the application of ENCODE data to interpret the human genome.

***PLoS Biol* (2011)**

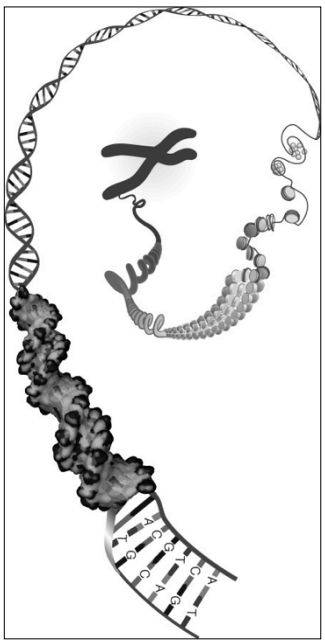


nihroadmap.nih.gov/epigenomics

The NIH Roadmap Epigenomics Mapping Consortium

Bradley E Bernstein, John A Stamatoyannopoulos, Joseph F Costello, Bing Ren, Aleksandar Milosavljevic, Alexander Meissner, Manolis Kellis, Marco A Marra, Arthur L Beaudet, Joseph R Ecker, Peggy J Farnham, Martin Hirst, Eric S Lander, Tarjei S Mikkelsen & James A Thomson

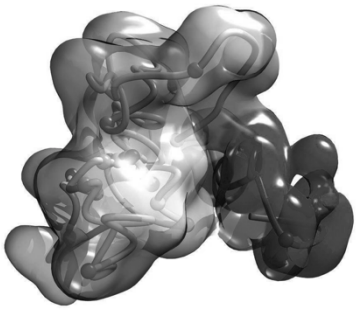
***Nature Biotechnology* (2010)**



TECHNOLOGY FEATURE

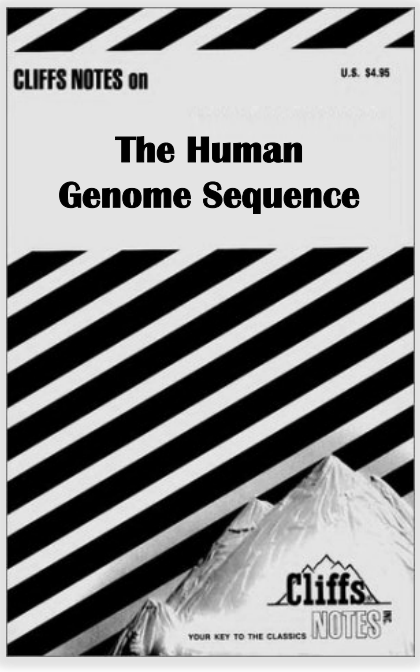
GENOMES IN THREE DIMENSIONS

A DNA sequence isn't enough; to understand the workings of the genome, we must study chromosome structure.

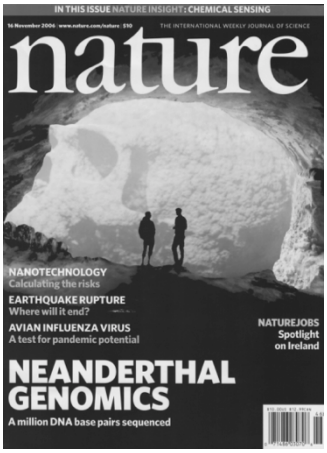


***Nature* (2011)**

```
TGCCGCGGAACTTTTCGGCTCTCTAAGGCTGTATTTTGATAT
ACGAAAGGCACATTTTCCTTCCCTTTTCAAAATGCACCTTGC
AAACGTAACAGGAACCCGACTAGGATCATCGGGAAAAGGAGG
AGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGGTGGCAGCGGG
TCCTGGGTCTGGCGGACCCTGACGCGAAGGAGGGTCTAGGAA
GCTCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGT
CCTCCAGCGTTGCCAACTGGACCTAAAGAGAGGCCGCGACTG
TCGCCCACCTGCGGGATGGGCCTGGTGCTGGGCGGTAAGGAC
ACGGACCTGGAAGGAGCGCGCGCGAGGGAGGGAGGCTGGGAG
TCAGAATCGGGAAAGGGAGGTGCGGGGCGGCGAGGGAGCGAA
GGAGGAGAGGAGGAAGGAGCGGGAGGGGTGCTGGCGGGGGTG
CGTAGTGGGTGGAGAAAGCCGCTAGAGCAAATTTGGGGCCGG
ACCAGGCAGCACTCGGCTTTTAACCTGGGCAGTGAAGGCGGG
GGAAAGAGCAAAAGGAAGGGGTGGTGTGCGGAGTAGGGGTGG
GTGGGGGGAATTGGAAGCAAATGACATCACAGCAGGTCAGAG
AAAAAGGGTTGAGCGGCAGGCACCCAGAGTAGTAGGTCTTTG
GCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCCCA
GCGCCCGAGAGACCATGCAGAGGTCGCCTCTGGAAAAGGCCA
GCGTTGTCTCCAAACTTTTTTTCAGGTGAGAAGGTGGCCAAC
CGAGCTTCGGAAAGACACGTGCCCACGAAAGAGGAGGGCGTG
TGTATGGGTTGGGTTTGGGGTAAAGGAATAAGCAGTTTTTAA
AAAGATGCGCTATCATTCATTGTTTTGAAAGAAAATGTGGGT
ATTGTAGAATAAAACAGAAAGCATTAAGAAGAGATGGAAGAA
TGAACTGAAGCTGATTGAATAGAGAGCCACATCTACTTGCAA
CTGAAAAGTTAGAATCTCAAGACTCAAGTACGCTACTATGCA
CTTGTTTTATTTCATTTTTCTAAGAAACTAAAAATACTTGTT
AATAAGTACCTAAGTATGGTTTATTGGTTTTCCCCCTTCATG
CCTTGGACACTTGATTGTCTTCTTGGCACATACAGGTGCCAT
GCCTGCATATAGTAAGTGCTCAGAAAACATTTCTTGACTGAA
TTCAGCCAACAAAAATTTTGGGGTAGGTAGAAAATATATGCT
TAAAGTATTTATTGTTATGAGACTGGATATATCTAGTATTTG
TCACAGGTAAATGATTCTTCAAAAATTGAAAGCAAATTTGTT
GAAATATTTATTTTGAAAAAAGTTACTTCACAAGCTATAAAT
TTTAAAAGCCATAGGAATAGATACCGAAGTTATATCCAACTG
ACATTTA
```



The Genomics of Human Evolution



GENOME 10K©

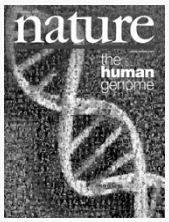
Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species

GENOME 10K COMMUNITY OF SCIENTISTS*

J. Heredity **(2009)**

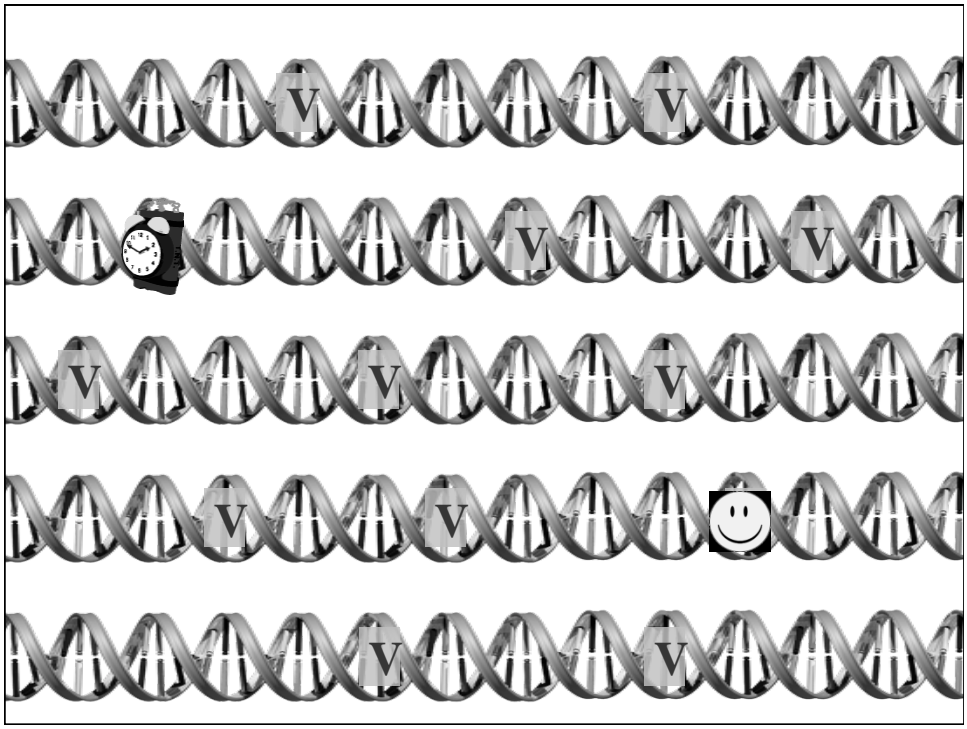
The Path to Genomic Medicine

Human Genomic Variation



Human Genome Project

Realization of Genomic Medicine



A haplotype map of the human genome

The International HapMap Consortium*

Inherited genetic variation has a critical but as yet largely uncharacterized role in human disease. Here we report a public database of common variation in the human genome: more than one million single nucleotide polymorphisms (SNPs) for which accurate and complete genotypes have been obtained in 269 DNA samples from four populations, including ten 500-kilobase regions in which essentially all information about common DNA variation has been extracted. These data document the generality of recombination hotspots, a block-like structure of linkage disequilibrium and low haplotype diversity, leading to substantial correlations of SNPs with many of their neighbours. We show how the HapMap resource can guide the design and analysis of genetic association studies, shed light on structural variation and recombination, and identify loci that may have been subject to natural selection during human evolution.

2005

A second generation human haplotype map of over 3.1 million SNPs

The International HapMap Consortium*

We describe the Phase II HapMap, which characterizes over 3.1 million human single nucleotide polymorphisms (SNPs) genotyped in 270 individuals from four geographically diverse populations and includes 25–35% of common SNP variation in the populations surveyed. The map is estimated to capture untyped common variation with an average maximum r^2 of between 0.9 and 0.96 depending on population. We demonstrate that the current generation of commercial genome-wide genotyping products captures common Phase II SNPs with an average maximum r^2 of up to 0.8 in African and up to 0.95 in non-African populations, and that potential gains in power in association studies can be obtained through imputation. These data also reveal novel aspects of the structure of linkage disequilibrium. We show that 10–30% of pairs of individuals within a population share at least one region of extended genetic identity arising from recent ancestry and that up to 1% of all common variants are untaggable, primarily because they lie within recombination hotspots. We show that recombination rates vary systematically around genes and between genes of different function. Finally, we demonstrate increased differentiation at non-synonymous, compared to synonymous, SNPs, resulting from systematic differences in the strength or efficacy of natural selection between populations.

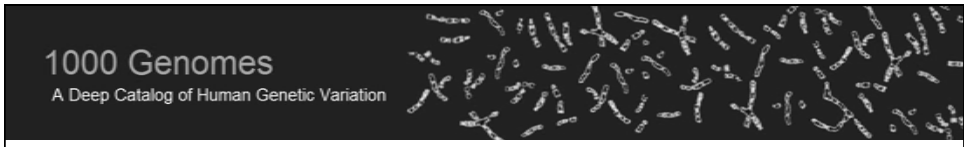
2007

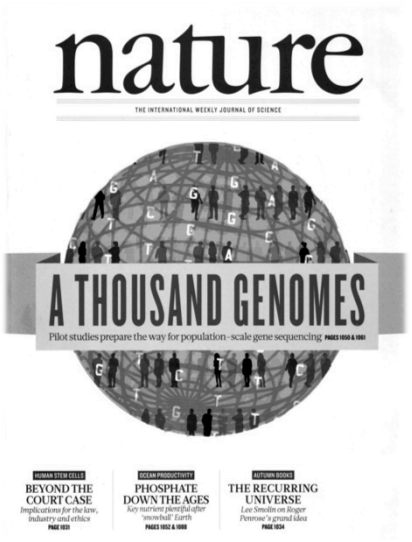
Integrating common and rare genetic variation in diverse human populations

The International HapMap 3 Consortium*

Despite great progress in identifying genetic variants that influence human disease, most inherited risk remains unexplained. A more complete understanding requires genome-wide studies that fully examine less common alleles in populations with a wide range of ancestry. To inform the design and interpretation of such studies, we genotyped 1.6 million common single nucleotide polymorphisms (SNPs) in 1,184 reference individuals from 11 global populations, and sequenced ten 100-kilobase regions in 692 of these individuals. This integrated data set of common and rare alleles, called 'HapMap 3', includes both SNPs and copy number polymorphisms (CNPs). We characterized population-specific differences among low-frequency variants, measured the improvement in imputation accuracy afforded by the larger reference panel, especially in imputing SNPs with a minor allele frequency of ≥5%, and demonstrated the feasibility of imputing newly discovered CNPs and SNPs. This expanded public resource of genome variants in global populations supports deeper interrogation of genomic variation and its role in human disease, and serves as a step towards a high-resolution map of the landscape of human genetic variation.

2010





ARTICLE

doi:10.1038/nature09534

A map of human genome variation from population-scale sequencing

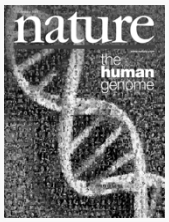
The 1000 Genomes Project Consortium*

The 1000 Genomes Project aims to provide a deep characterization of human genome sequence variation as a foundation for investigating the relationship between genotype and phenotype. Here we present results of the pilot phase of the project, designed to develop and compare different strategies for genome-wide sequencing with high-throughput platforms. We undertook three projects: low-coverage whole-genome sequencing of 179 individuals from four populations; high-coverage sequencing of two mother–father–child trios; and exon-targeted sequencing of 697 individuals from seven populations. We describe the location, allele frequency and local haplotype structure of approximately 15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants, most of which were previously undescribed. We show that, because we have catalogued the vast majority of common variation, over 95% of the currently accessible variants found in any individual are present in this data set. On average, each person is found to carry approximately 250 to 300 loss-of-function variants in annotated genes and 50 to 100 variants previously implicated in inherited disorders. We demonstrate how these results can be used to inform association and functional studies. From the two trios, we directly estimate the rate of de novo germline base substitution mutations to be approximately 10^{-8} per base pair per generation. We explore the data with regard to signatures of natural selection, and identify a marked reduction of genetic variation in the neighbourhood of genes, due to selection at linked sites. These methods and public data will support the next phase of human genetic research.

***Nature* 2010**

The Path to Genomic Medicine

Genomic Basis for Human Diseases

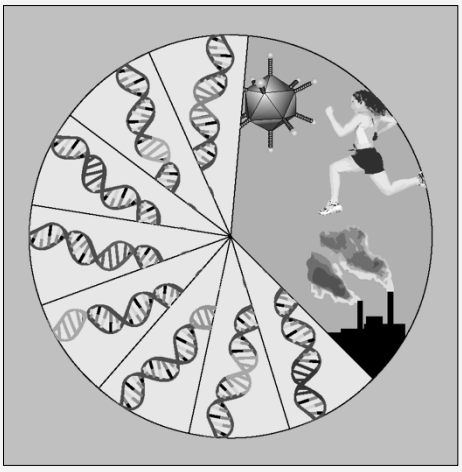


Human Genome Project

Realization of Genomic Medicine

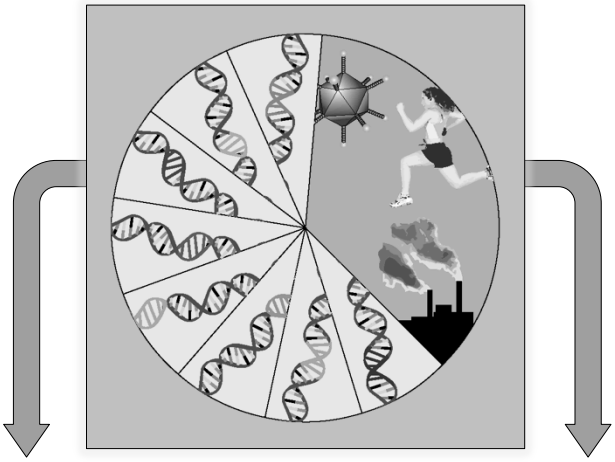
Genomic Architecture of Genetic Diseases

Rare, Simple, Monogenic, Mendelian…

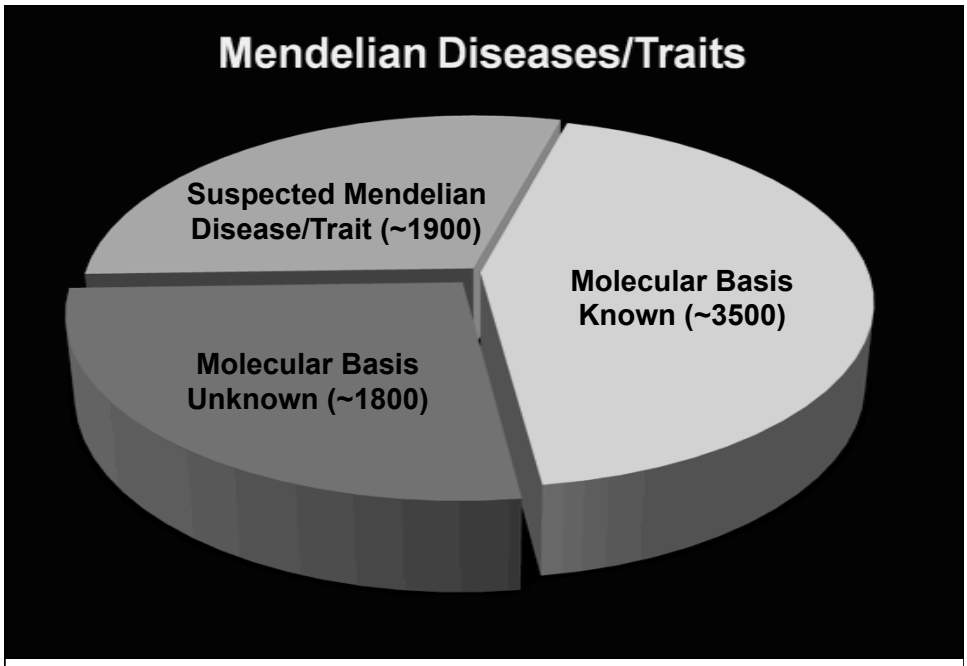


Common, Complex, Multigenic, Non-Mendelian…

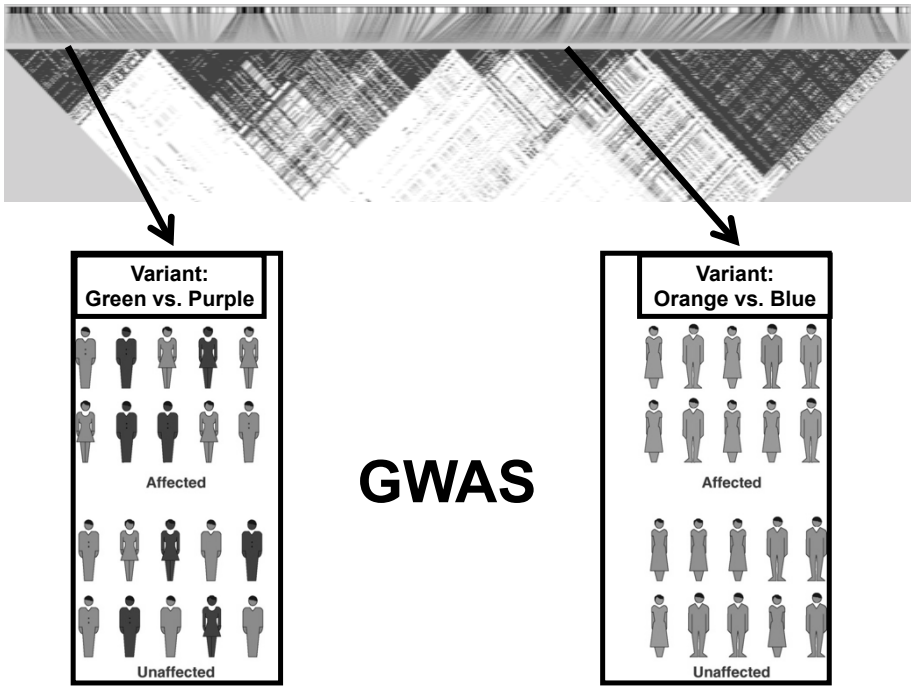
The Role of Both the Genome and the Environment in Human Disease



Genome Analysis Technologies **>>>** **Environmental Monitoring Technologies**



Source: ***Online Mendelian Inheritance in Man*** **(www.ncbi.nlm.nih.gov/omim)**

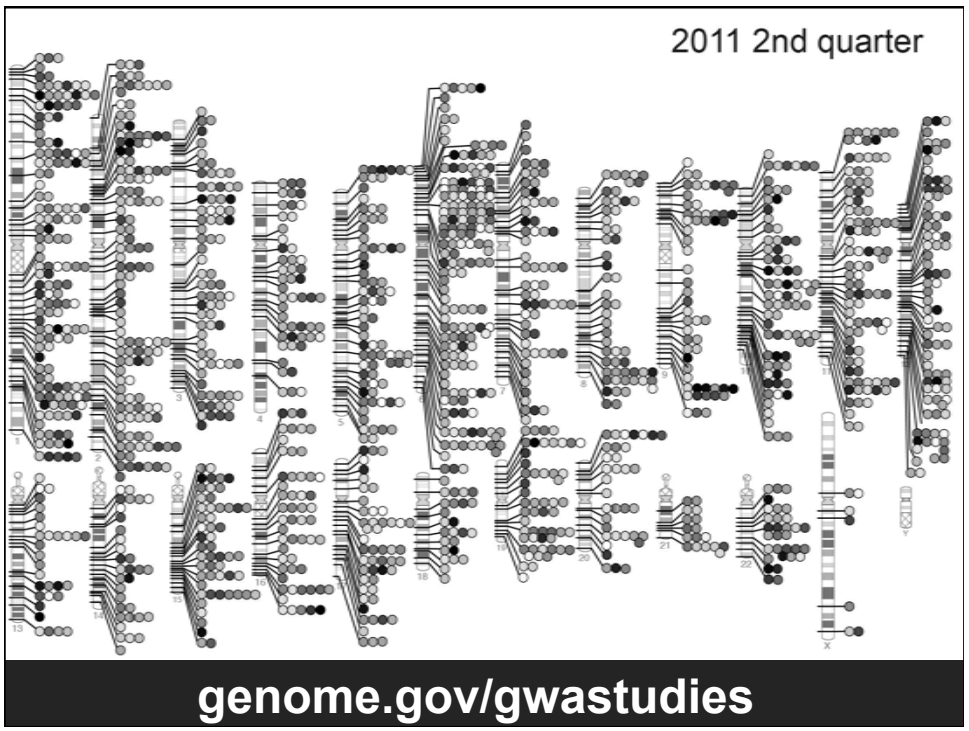


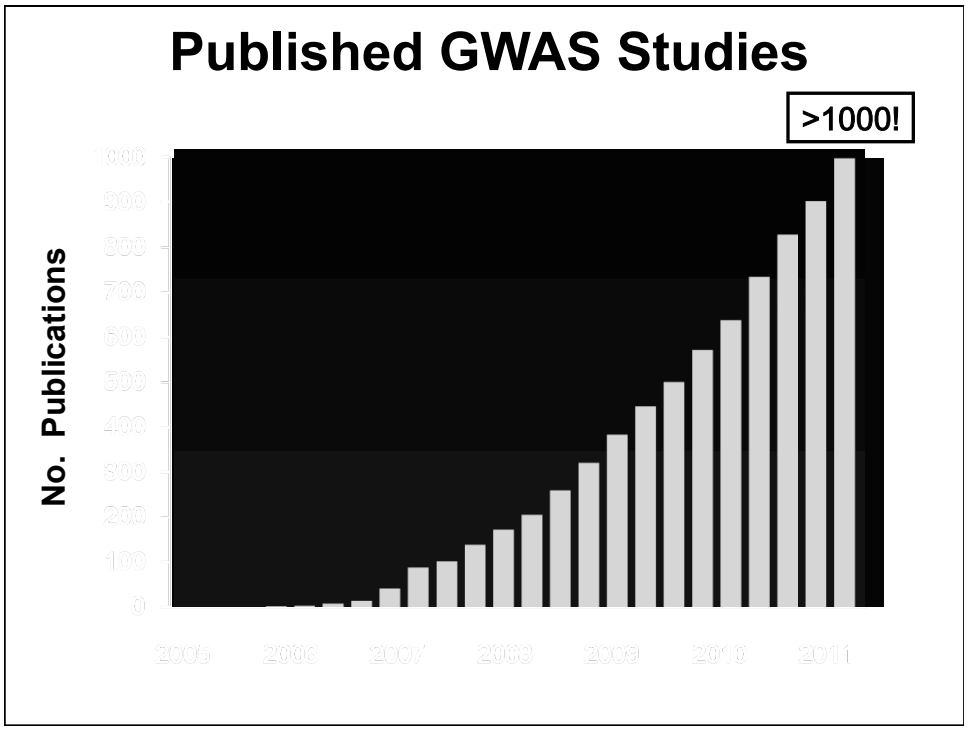
The First GWAS Success Story: Age-Related Macular Degeneration

Complement Factor H Polymorphism in Age-Related Macular Degeneration

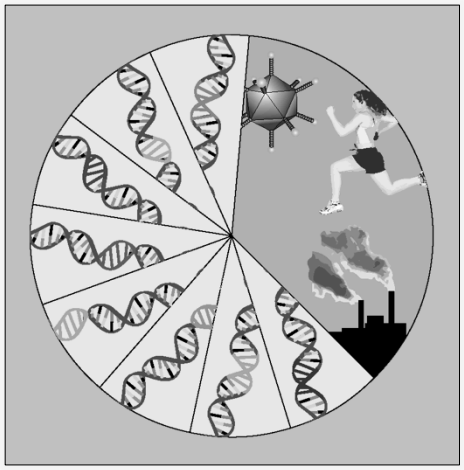
Robert J. Klein,[1] Caroline Zeiss,[2*] Emily Y. Chew,[3*] Jen-Yue Tsai,[4*] Richard S. Sackler,[1] Chad Haynes,[1] Alice K. Henning,[5] John Paul SanGiovanni,[3] Shrikant M. Mane,[6] Susan T. Mayne,[7] Michael B. Bracken,[7] Frederick L. Ferris,[3] Jurg Ott,[1] Colin Barnstable,[2] Josephine Hoh[7†]

***Science* (2005)**


2011 2nd quarter
genome.gov/gwastudies


Published GWAS Studies
>1000!
No. Publications
1000
900
800
700
600
500
400
300
200
100
0
2005
2006
2007
2008
2009
2010
2011

Genomic Architecture of Genetic Diseases



Rare, Simple, Monogenic, Mendelian…

Mostly *Coding* Mutations

Common, Complex, Multigenic, Non-Mendelian…

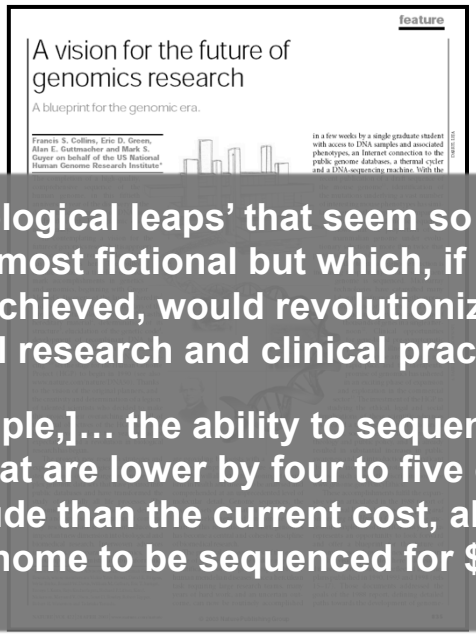
Mostly *Non-Coding* Mutations

The Path to Genomic Medicine

Routine Whole-Genome Sequencing

Human Genome Project

Realization of Genomic Medicine


feature

A vision for the future of genomics research

A blueprint for the genomic era.

Francis S. Collins, Eric D. Green, Alan E. Guttmacher and Mark S. Guyer on behalf of the US National Human Genome Research Institute*

in a few weeks by a single graduate student with access to DNA samples and associated phenotypes, an Internet connection to the public genome databases, a thermal cycler and a DNA-sequencing machine. With the

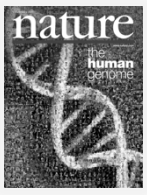
"…'technological leaps' that seem so far off as to be almost fictional but which, if they could be achieved, would revolutionize biomedical research and clinical practice.

[For example,]… the ability to sequence DNA at costs that are lower by four to five orders of magnitude than the current cost, allowing a human genome to be sequenced for $1,000 or less."

***Nature*, April 2003**

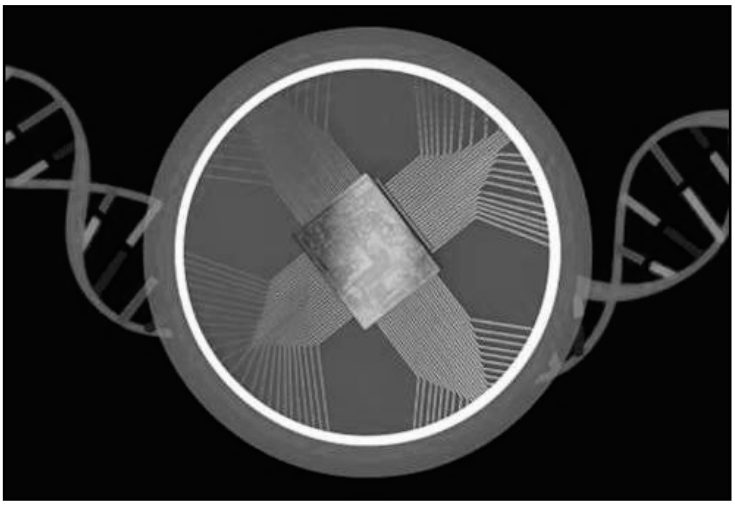
Human Genome Sequence

~$1,000,000,000

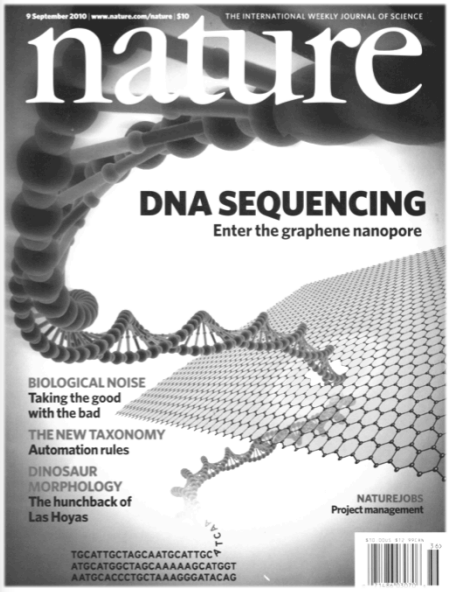


~$1,000

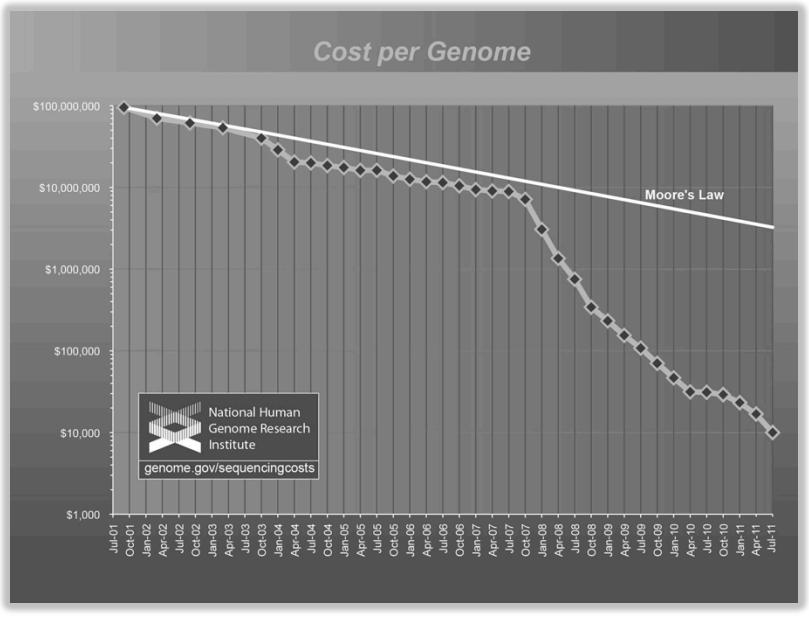
"The $1000 Genome"



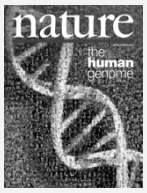

POLONATOR G.007



Cost per Sequenced Human Genome



Human Genome Sequence

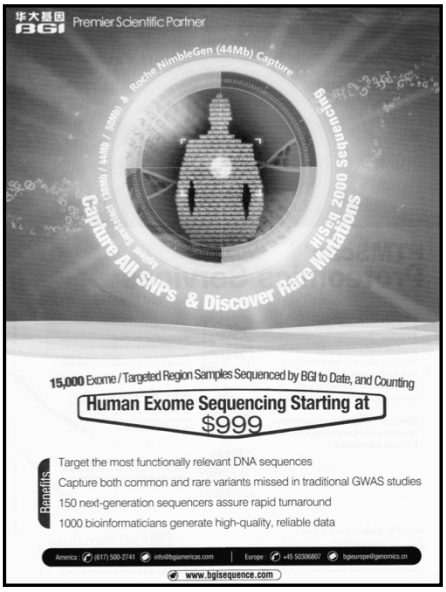


~$1,000,000,000

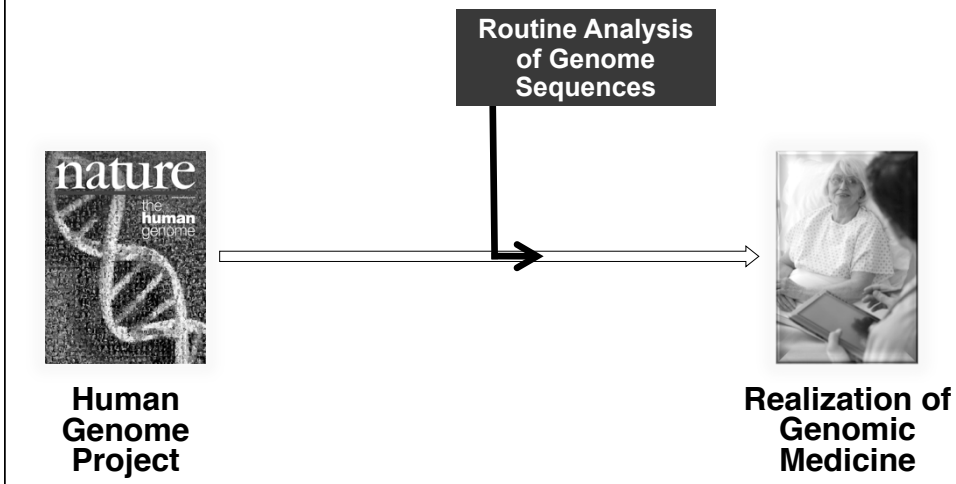
Current Cost

~$1,000 ***"The $1000 Genome"***

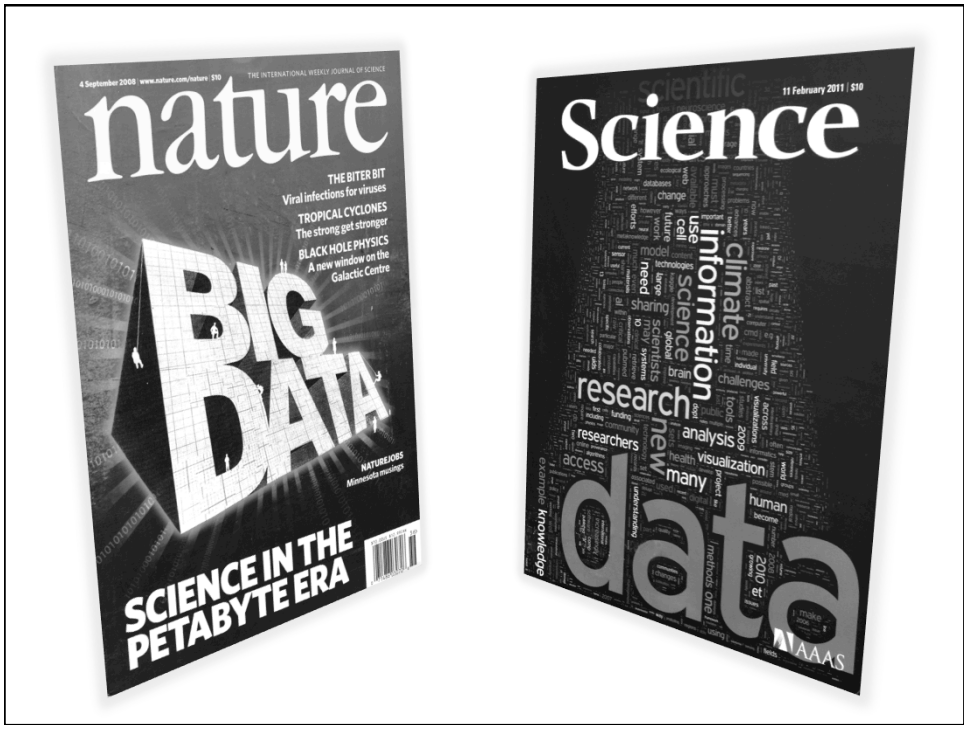
Genome Sequencing as a 'Commodity'



The Path to Genomic Medicine

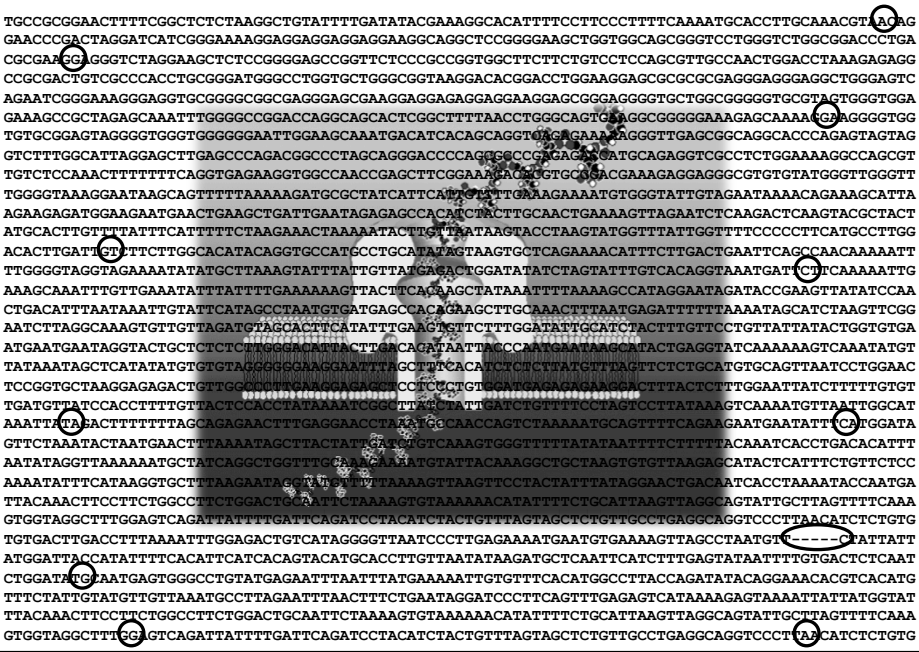


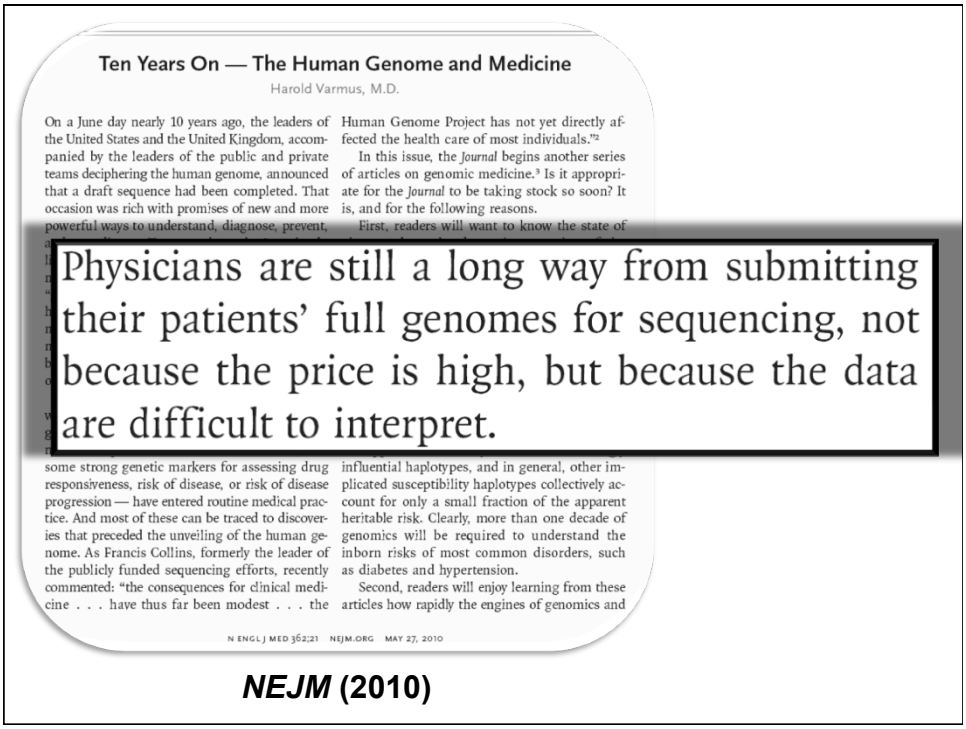
The Largest Current Bottleneck in Genomics…



The Computational Bottleneck

The Informational Bottleneck





Ten Years On — The Human Genome and Medicine

Harold Varmus, M.D.

On a June day nearly 10 years ago, the leaders of the United States and the United Kingdom, accompanied by the leaders of the public and private teams deciphering the human genome, announced that a draft sequence had been completed. That occasion was rich with promises of new and more powerful ways to understand, diagnose, prevent,

Human Genome Project has not yet directly affected the health care of most individuals."[2]

In this issue, the *Journal* begins another series of articles on genomic medicine.[3] Is it appropriate for the *Journal* to be taking stock so soon? It is, and for the following reasons.

First, readers will want to know the state of

Physicians are still a long way from submitting their patients' full genomes for sequencing, not because the price is high, but because the data are difficult to interpret.

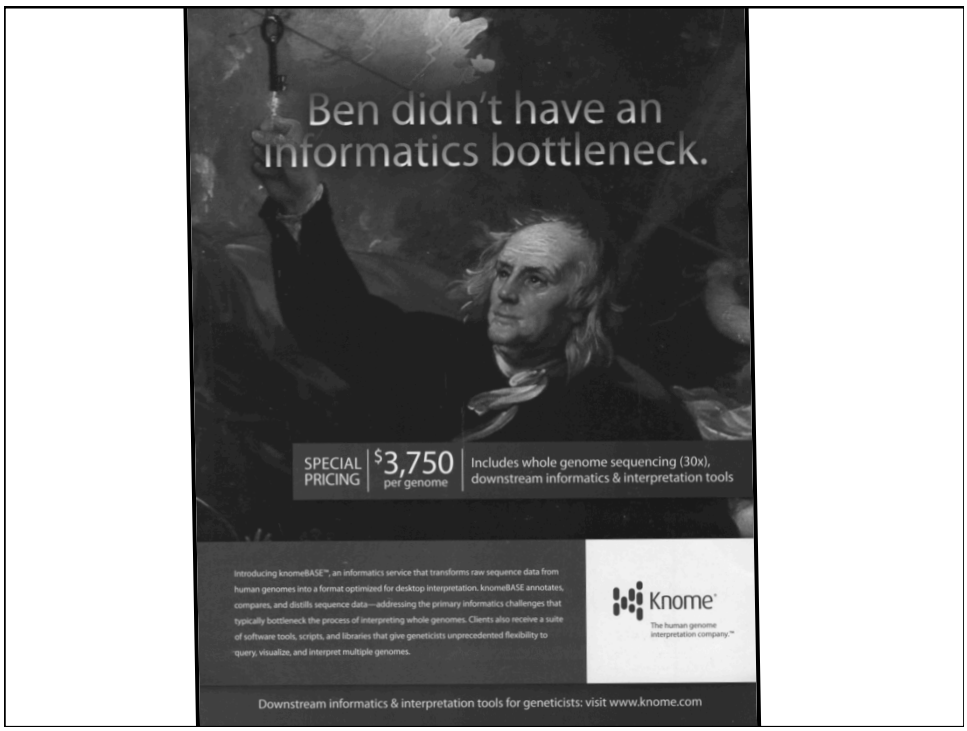
some strong genetic markers for assessing drug responsiveness, risk of disease, or risk of disease progression — have entered routine medical practice. And most of these can be traced to discoveries that preceded the unveiling of the human genome. As Francis Collins, formerly the leader of the publicly funded sequencing efforts, recently commented: "the consequences for clinical medicine . . . have thus far been modest . . . the

influential haplotypes, and in general, other implicated susceptibility haplotypes collectively account for only a small fraction of the apparent heritable risk. Clearly, more than one decade of genomics will be required to understand the inborn risks of most common disorders, such as diabetes and hypertension.

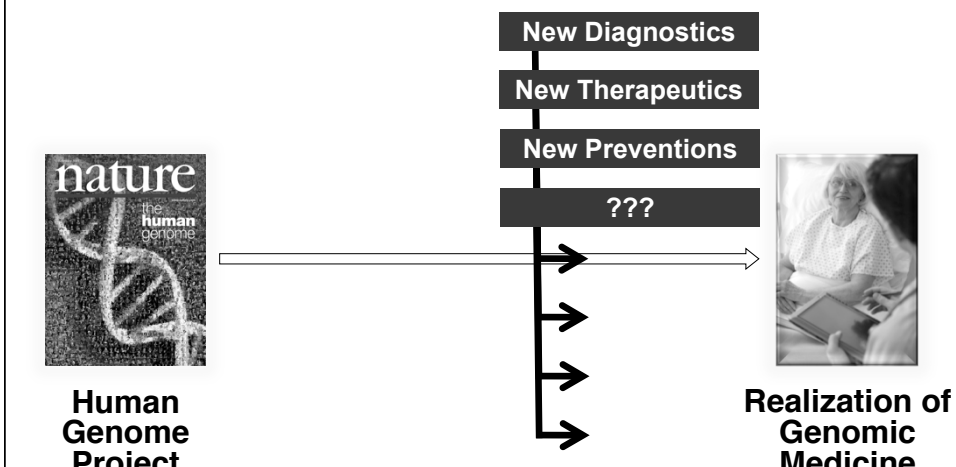
Second, readers will enjoy learning from these articles how rapidly the engines of genomics and

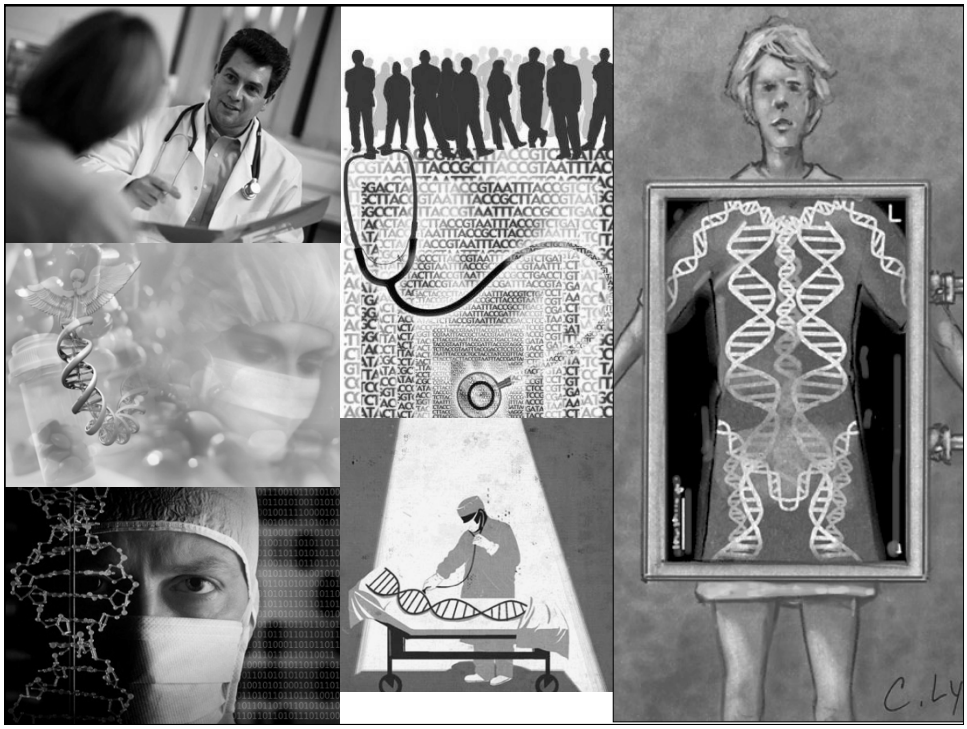
N ENGL J MED 362;21 NEJM.ORG MAY 27, 2010

***NEJM* (2010)**



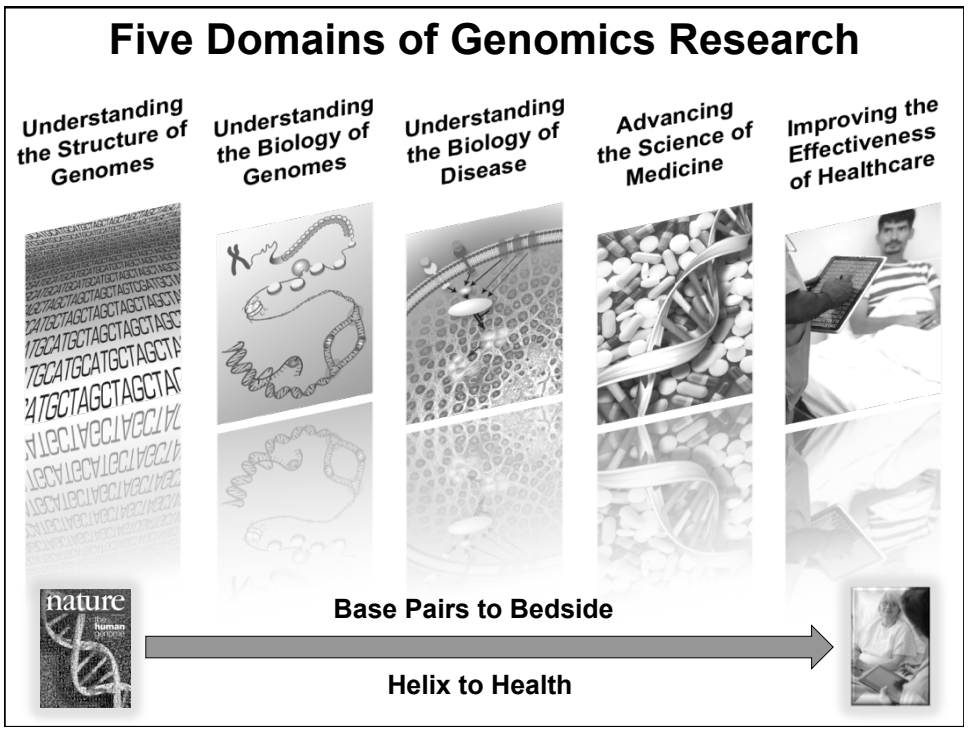
The Path to Genomic Medicine

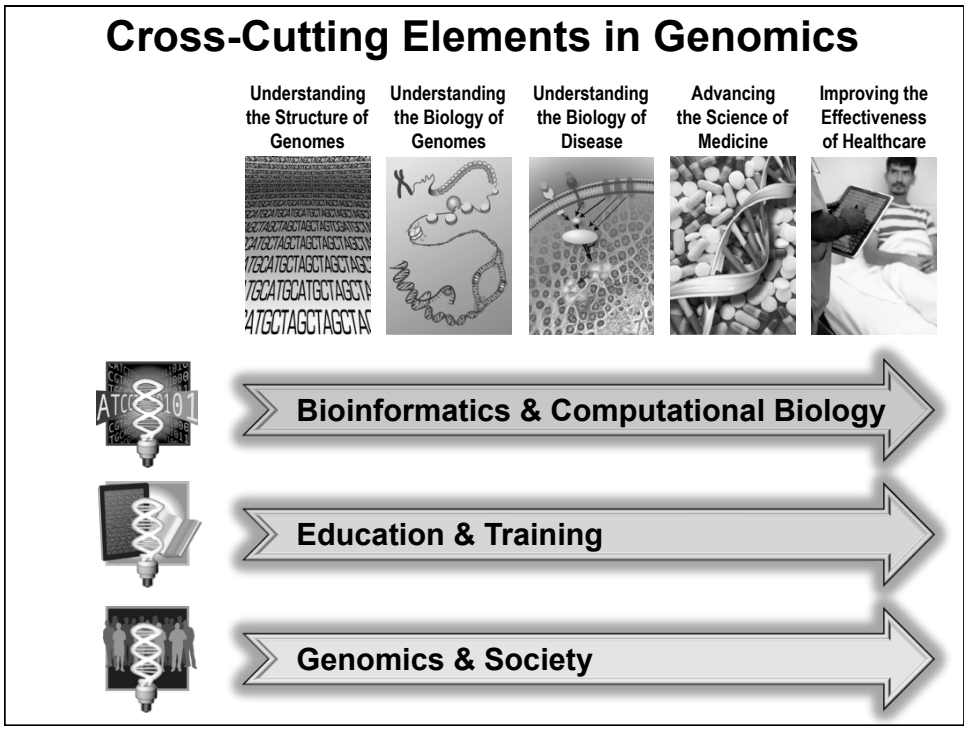


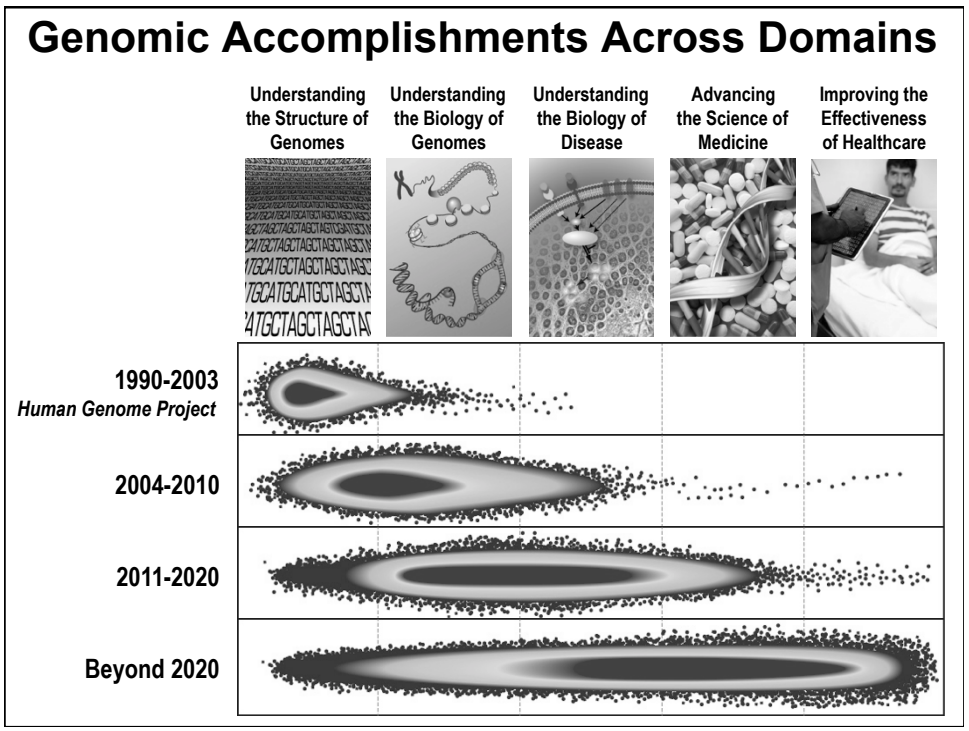


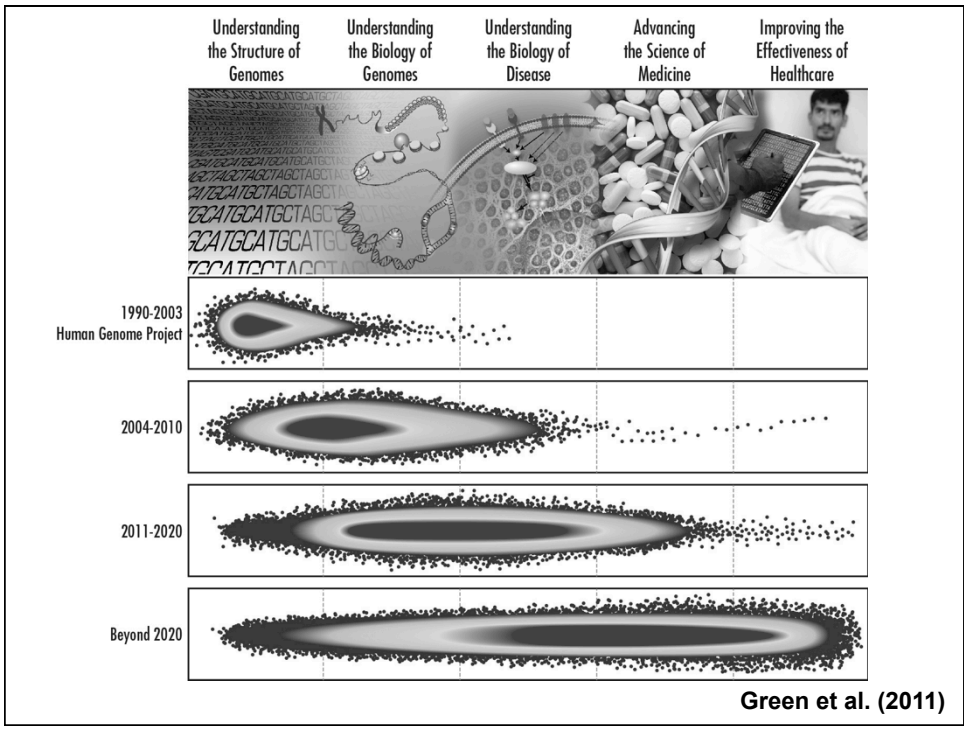
~11 Months Ago

February 2011
NHGRI Published New Vision for Genomics


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
nature
Base Pairs to Bedside
Helix to Health

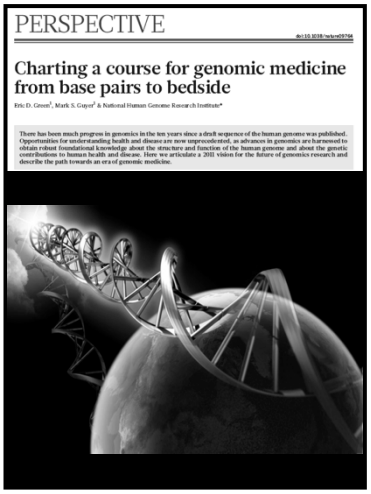

Cross-Cutting Elements in Genomics
Understanding the Structure of Genomes
Understanding the Biology of Genomes
Understanding the Biology of Disease
Advancing the Science of Medicine
Improving the Effectiveness of Healthcare
Bioinformatics & Computational Biology
Education & Training
Genomics & Society





Green et al. (2011)

2011 NHGRI Strategic Plan for Genomics



PERSPECTIVE

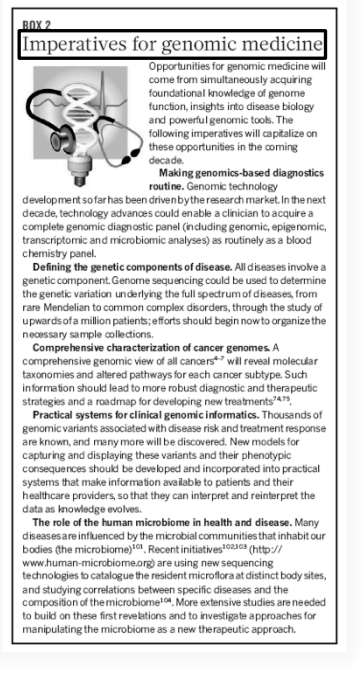
Charting a course for genomic medicine from base pairs to bedside

Eric D. Green[1], Mark S. Guyer[2] & National Human Genome Research Institute*

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

>>>

NHGRI



BOX 2

Imperatives for genomic medicine

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

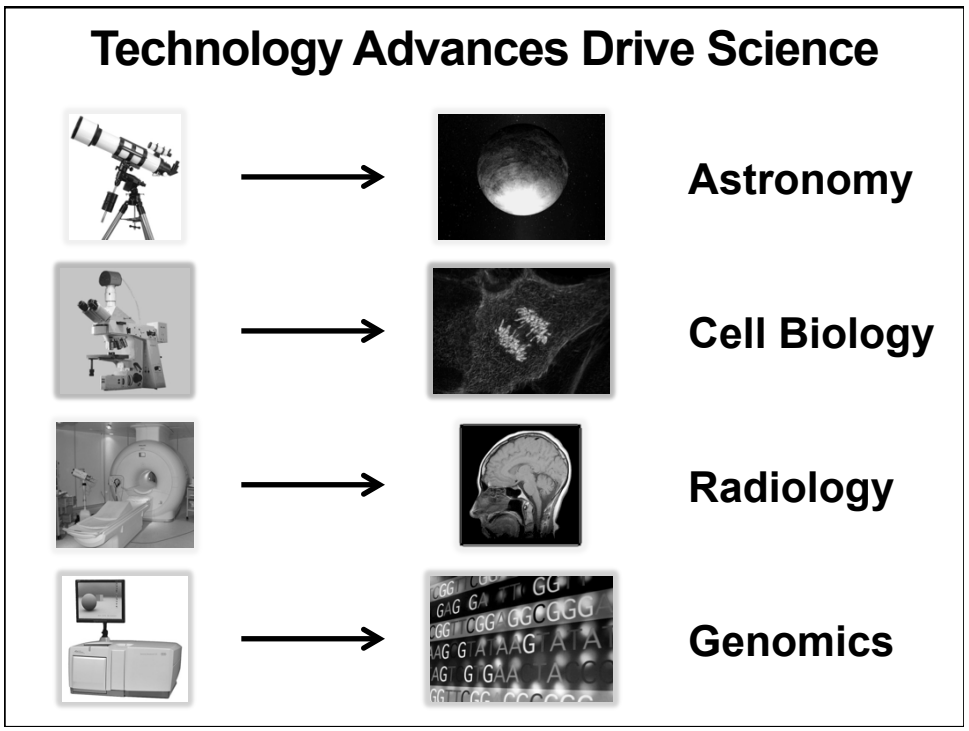
Making genomics-based diagnostics routine. Genomic technology development so far has been driven by the research market. In the next decade, technology advances could enable a clinician to acquire a complete genomic diagnostic panel (including genomic, epigenomic, transcriptomic and microbiomic analyses)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

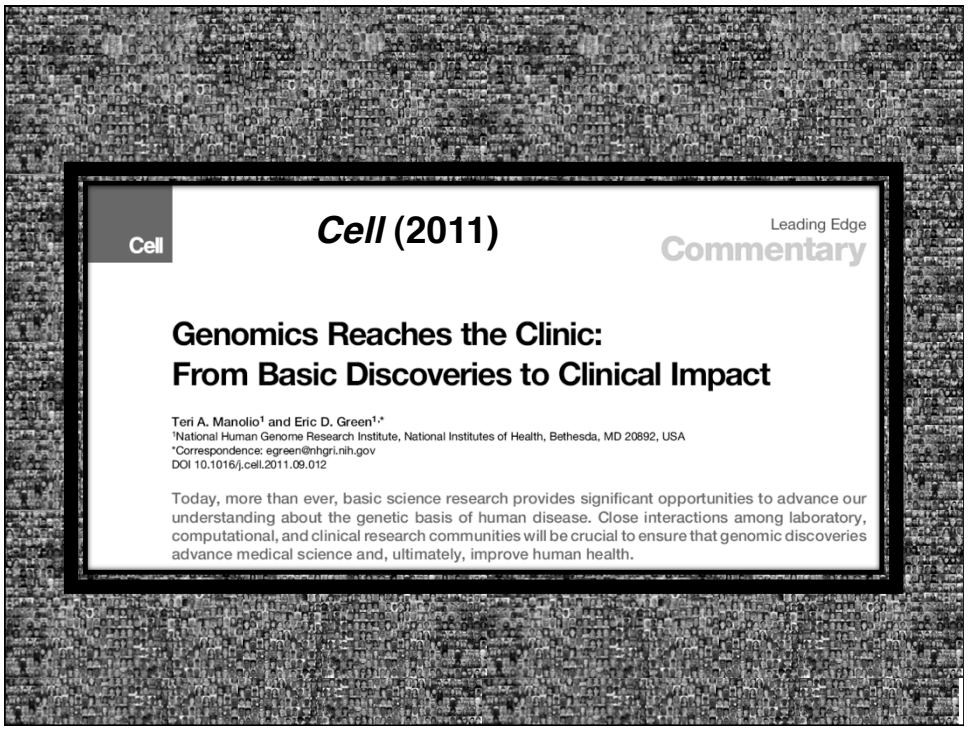
Defining the genetic components of disease. All diseases involve a genetic component. Genome sequencing could be used to determine the genetic variation underlying the full spectrum of diseases, from rare Mendelian to common complex disorders, through the study of upwards of a million patients; efforts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

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

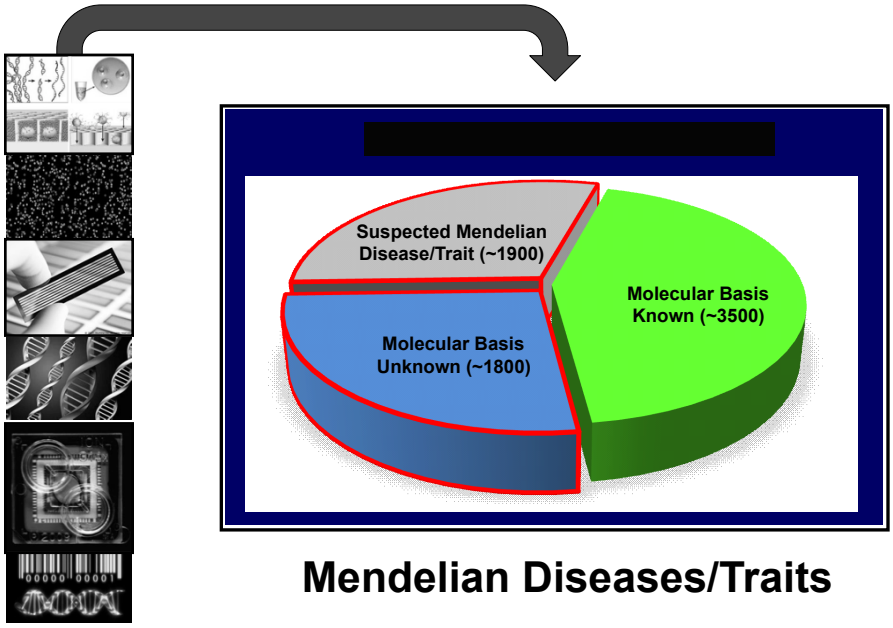
Practical systems for clinical genomic informatics. Thousands of genomic variants associated with disease risk and treatment response are known, and many more will be discovered. New models for capturing and displaying these variants and their phenotypic consequences should be developed and incorporated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

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


Technology Advances Drive Science
Astronomy
Cell Biology
Radiology
Genomics

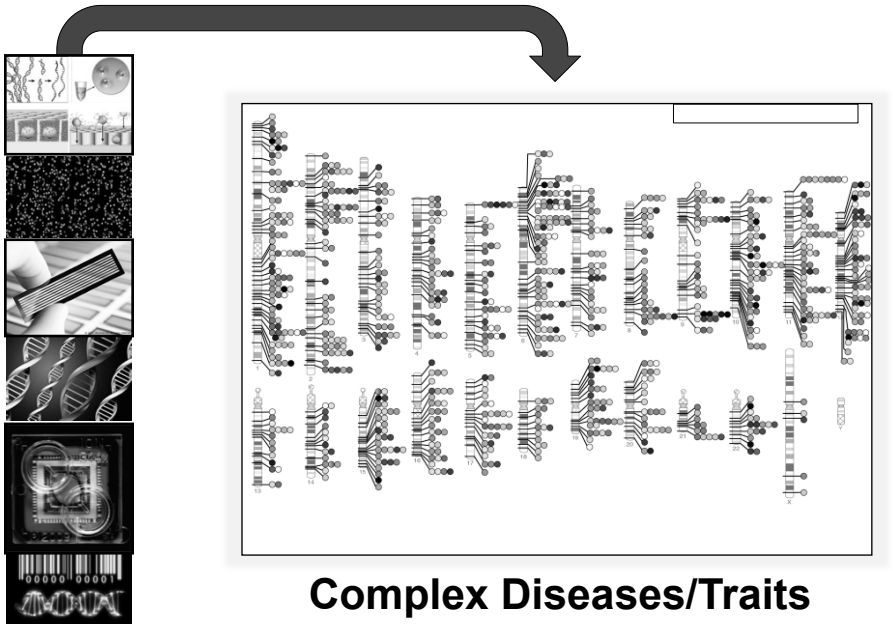

Cell
Cell (2011)
Leading Edge
Commentary
Genomics Reaches the Clinic:
From Basic Discoveries to Clinical Impact
Teri A. Manolio[1] and Eric D. Green[1,*]
[1]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
*Correspondence: egreen@nhgri.nih.gov
DOI 10.1016/j.cell.2011.09.012
Today, more than ever, basic science research provides significant opportunities to advance our understanding about the genetic basis of human disease. Close interactions among laboratory, computational, and clinical research communities will be crucial to ensure that genomic discoveries advance medical science and, ultimately, improve human health.

The Future: Genome Sequencing



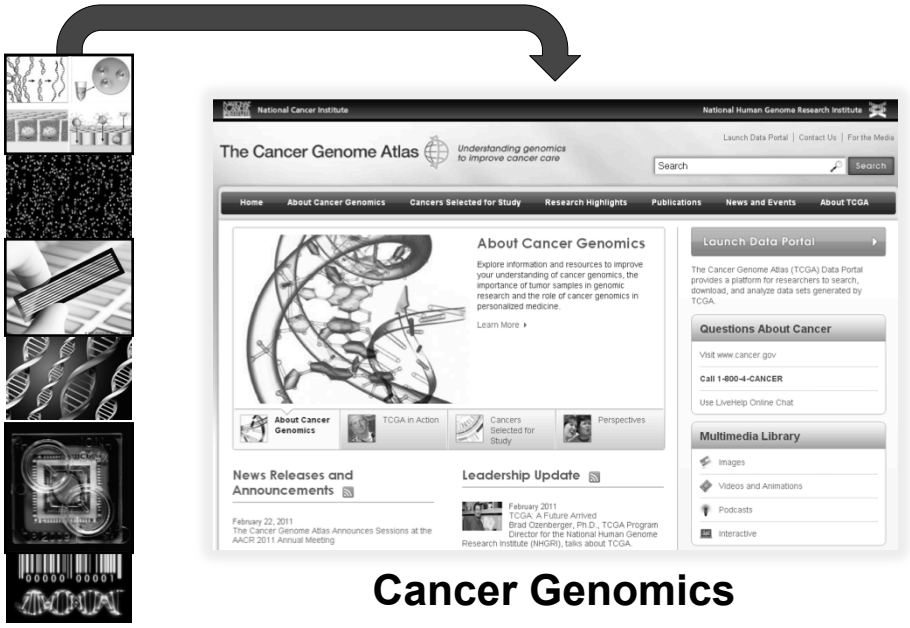
Mendelian Diseases/Traits

The Future: Genome Sequencing

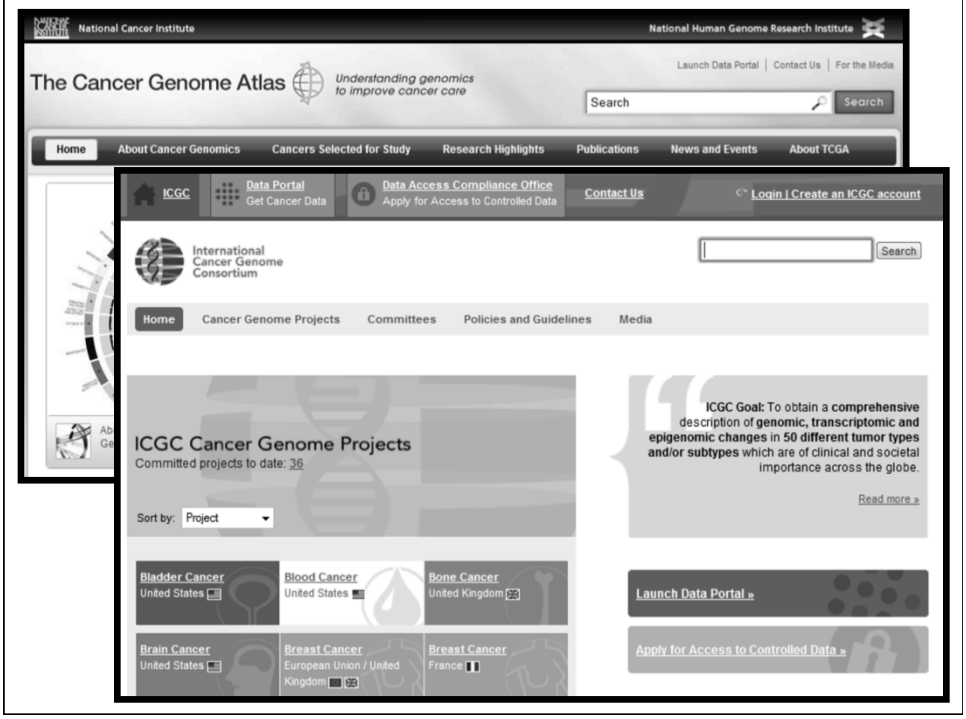


Complex Diseases/Traits

The Future: Genome Sequencing

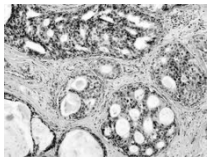


Cancer Genomics

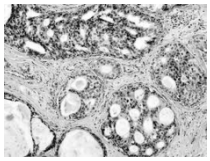


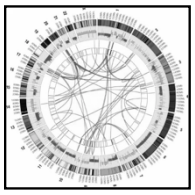
Genomic Medicine: Cancer Diagnostics

Now

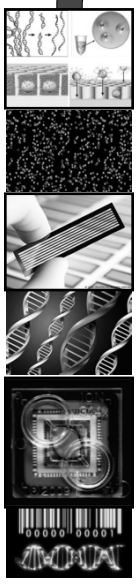


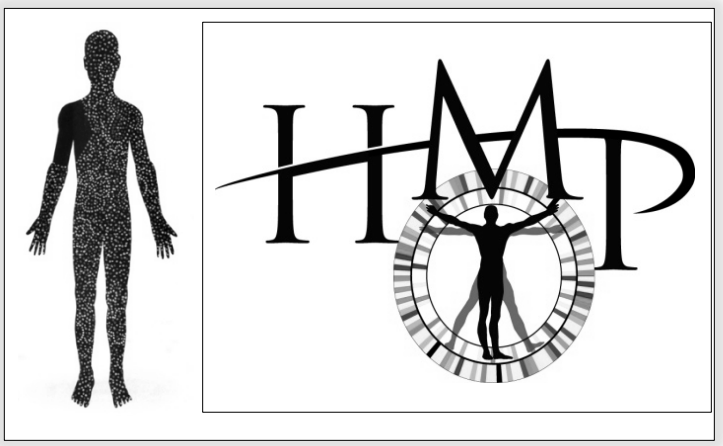
Future



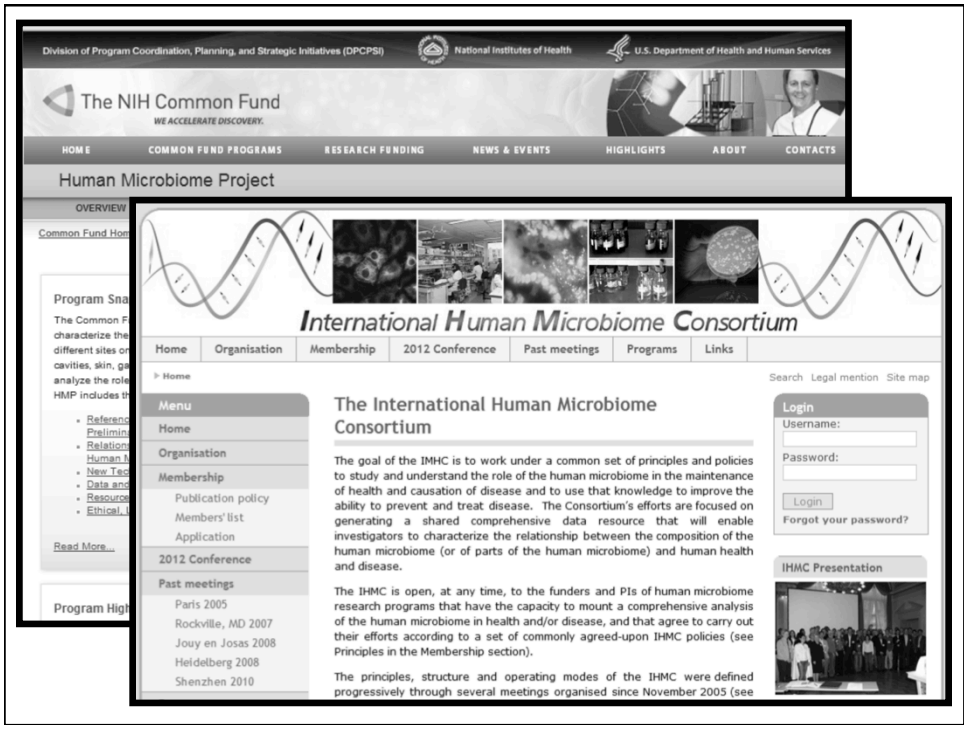


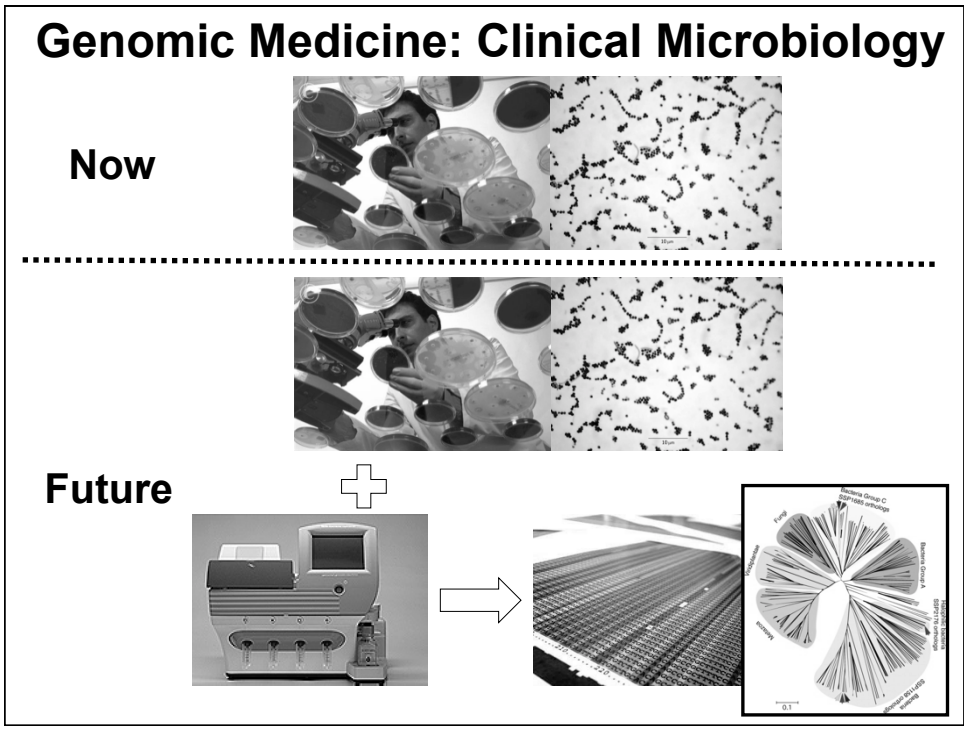
The Future: Genome Sequencing

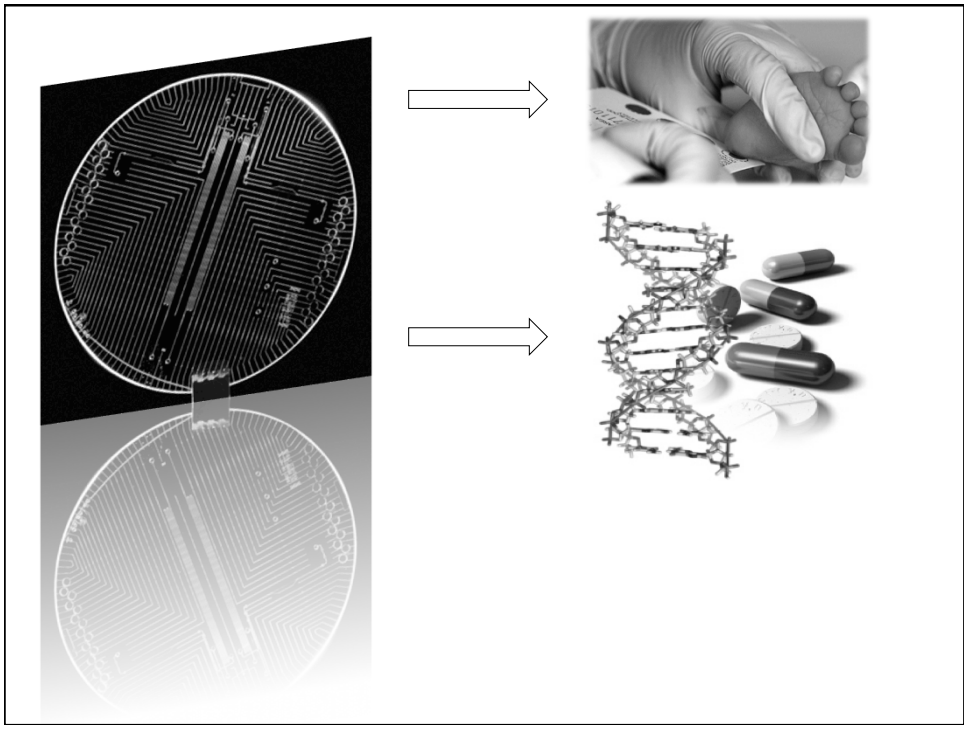


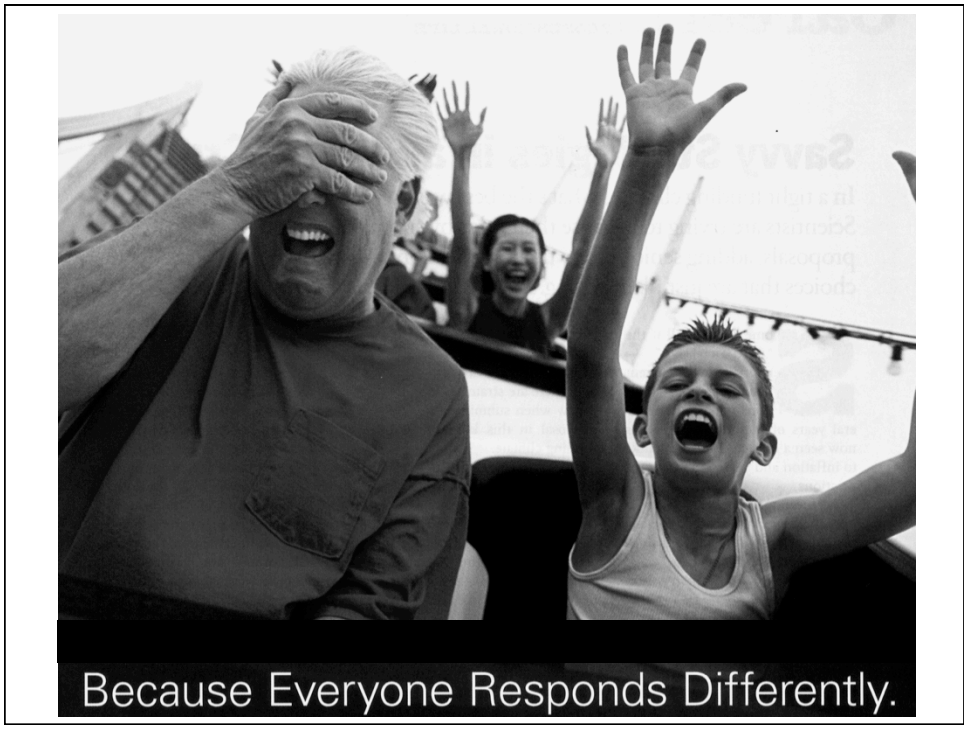


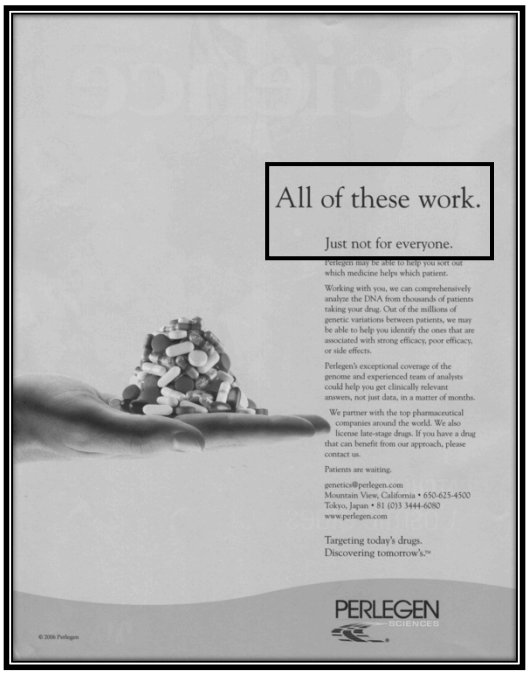
Microbiome

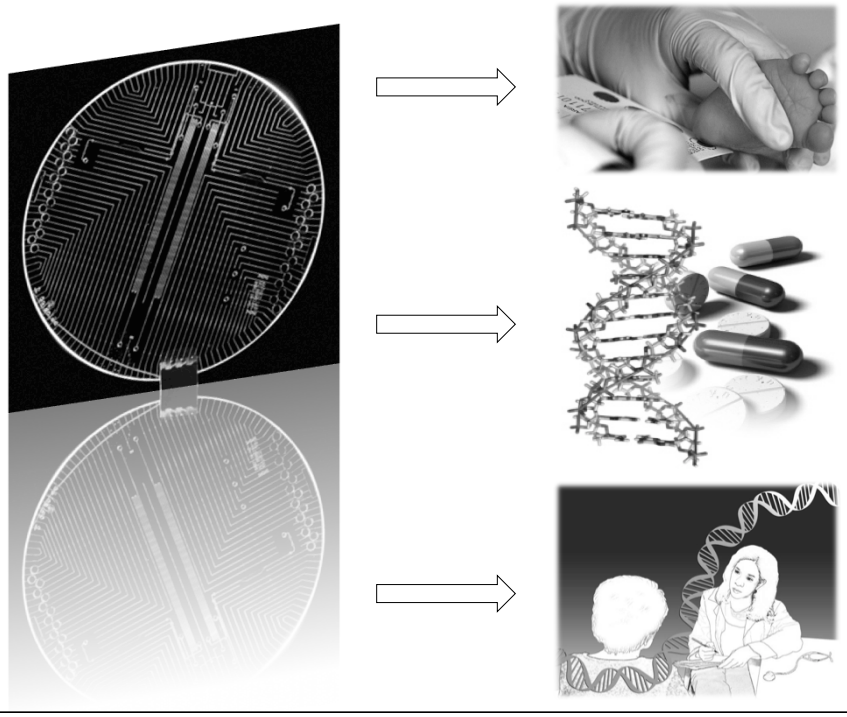

Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI)
National Institutes of Health
U.S. Department of Health and Human Services
The NIH Common Fund
WE ACCELERATE DISCOVERY.
HOME
COMMON FUND PROGRAMS
RESEARCH FUNDING
NEWS & EVENTS
HIGHLIGHTS
ABOUT
CONTACTS
Human Microbiome Project
OVERVIEW
International Human Microbiome Consortium
Home
Organisation
Membership
2012 Conference
Past meetings
Programs
Links
Search Legal mention Site map
Menu
Home
Organisation
Membership
Publication policy
Members' list
Application
2012 Conference
Past meetings
Paris 2005
Rockville, MD 2007
Jouy en Josas 2008
Heidelberg 2008
Shenzhen 2010
The International Human Microbiome Consortium
The goal of the IMHC is to work under a common set of principles and policies to study and understand the role of the human microbiome in the maintenance of health and causation of disease and to use that knowledge to improve the ability to prevent and treat disease. The Consortium's efforts are focused on generating a shared comprehensive data resource that will enable investigators to characterize the relationship between the composition of the human microbiome (or of parts of the human microbiome) and human health and disease.
The IHMC is open, at any time, to the funders and PIs of human microbiome research programs that have the capacity to mount a comprehensive analysis of the human microbiome in health and/or disease, and that agree to carry out their efforts according to a set of commonly agreed-upon IHMC policies (see Principles in the Membership section).
The principles, structure and operating modes of the IHMC were defined progressively through several meetings organised since November 2005 (see
Login
Username:
Password:
Login
Forgot your password?
IHMC Presentation


Genomic Medicine: Clinical Microbiology
Now
Future

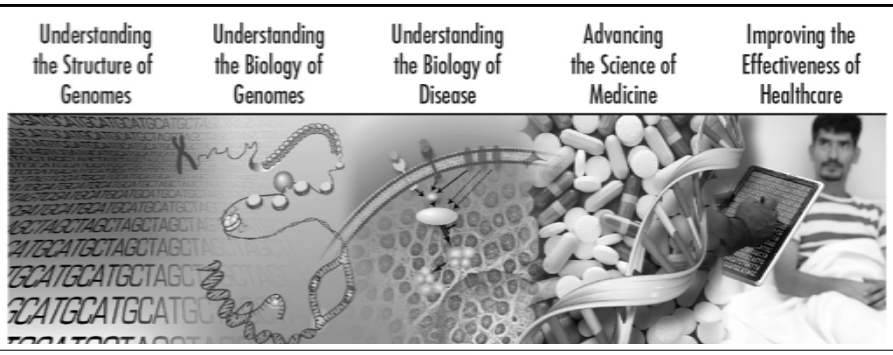



Because Everyone Responds Differently.


All of these work.
Just not for everyone.
Perlegen may be able to help you sort out which medicine helps which patient.
Working with you, we can comprehensively analyze the DNA from thousands of patients taking your drug. Out of the millions of genetic variations between patients, we may be able to help you identify the ones that are associated with strong efficacy, poor efficacy, or side effects.
Perlegen's exceptional coverage of the genome and experienced team of analysts could help you get clinically relevant answers, not just data, in a matter of months.
We partner with the top pharmaceutical companies around the world. We also license late-stage drugs. If you have a drug that can benefit from our approach, please contact us.
Patients are waiting.
genetics@perlegen.com
Mountain View, California • 650-625-4500
Tokyo, Japan • 81 (03) 3444-6080
www.perlegen.com
Targeting today's drugs.
Discovering tomorrow's.™
PERLEGEN
SCIENCES
© 2006 Perlegen



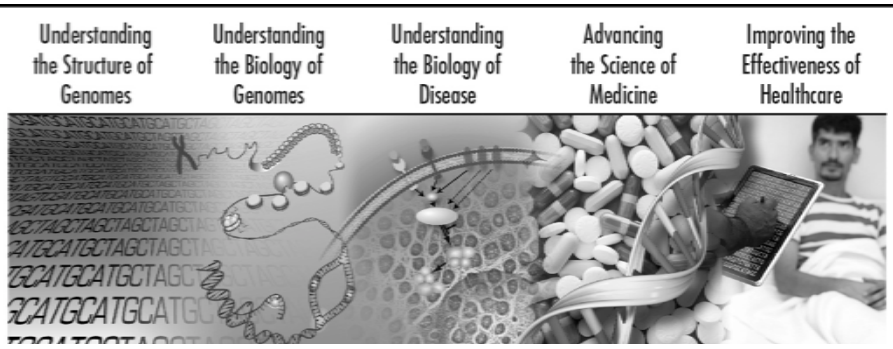
Clinical Genomic Information Systems



Basic Science

Translational Science

Implementation Science



A pessimist sees the difficulty in every opportunity.
An optimist sees the opportunity in every difficulty.

--Winston Churchill