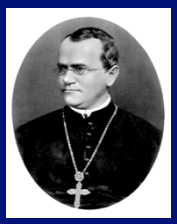
National Human Genome Research Institute Division of Intramural Research

The Genomic Landscape: *circa* 2010

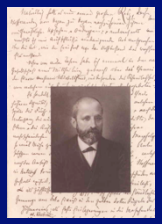
Eric Green, M.D., Ph.D.
Director, NHGRI

Foundational Milestones in Genetics & Genomics



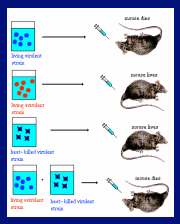
Mendel

1865



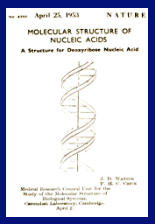
Miescher

1871



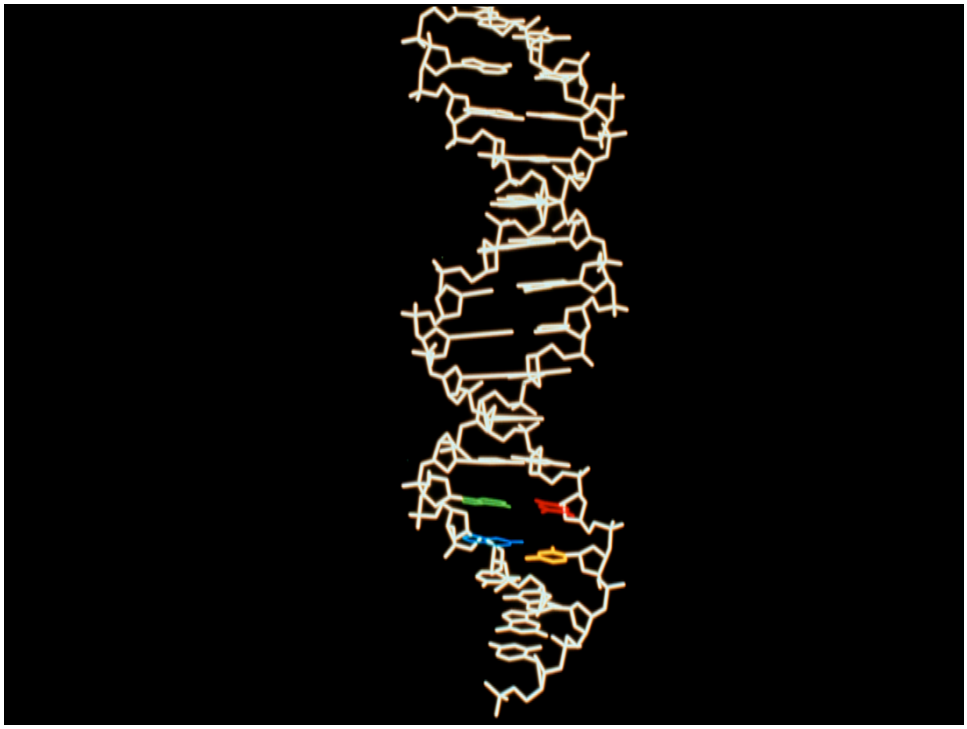
Avery

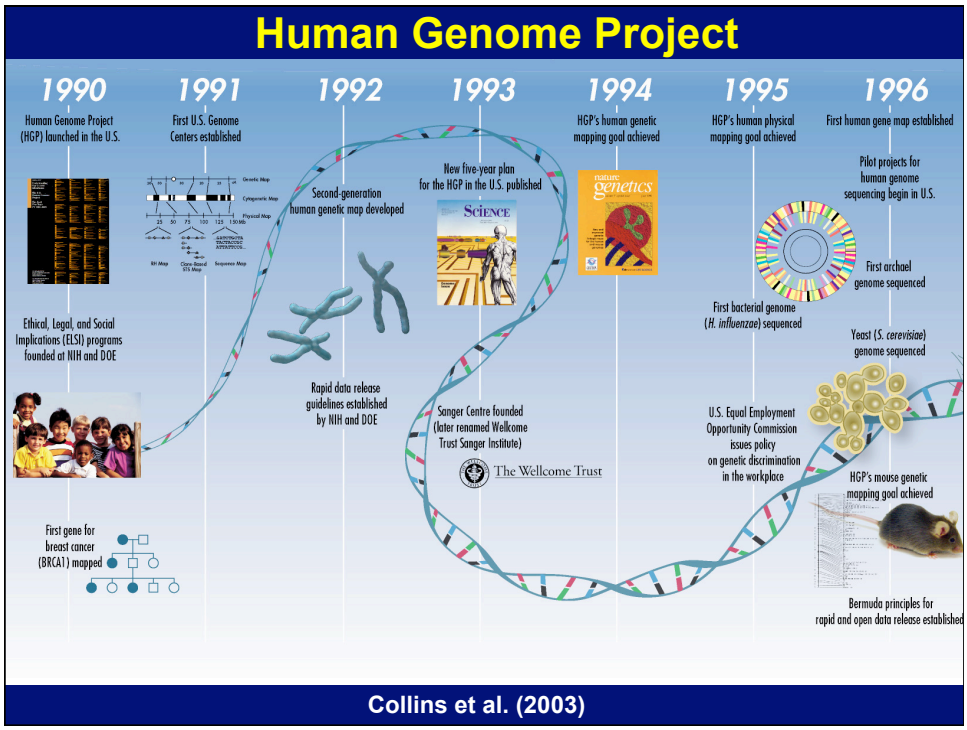
1944

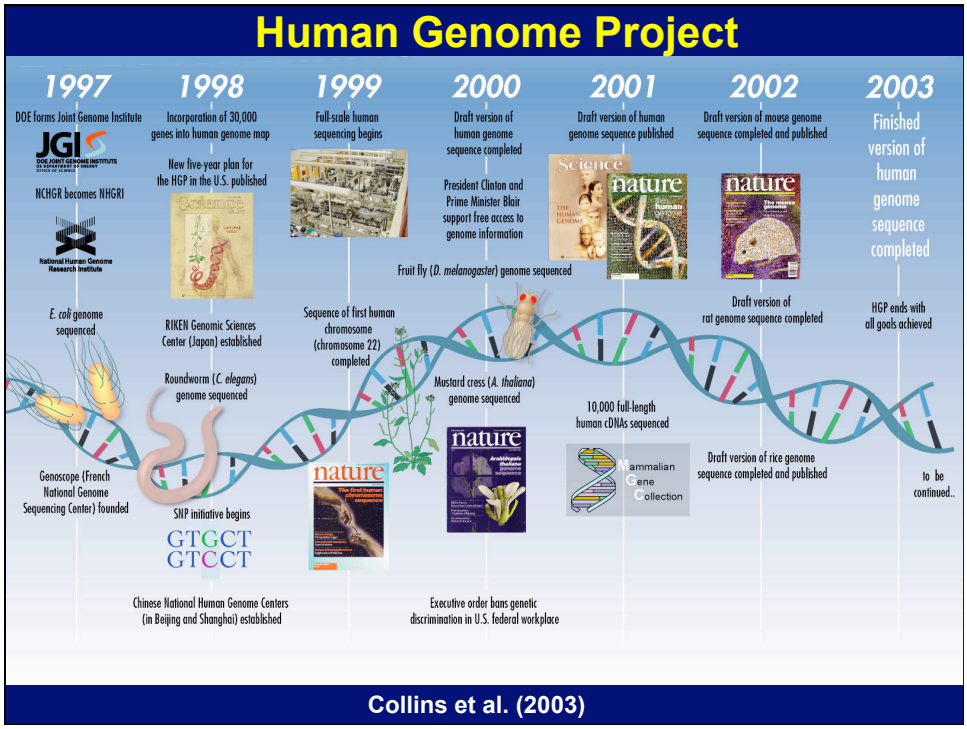


Watson & Crick

1953




Human Genome Project
1990
Human Genome Project (HGP) launched in the U.S.
1991
First U.S. Genome Centers established
1992
Second-generation human genetic map developed
1993
New five-year plan for the HGP in the U.S. published
1994
HGP's human genetic mapping goal achieved
1995
HGP's human physical mapping goal achieved
1996
First human gene map established
Pilot projects for human genome sequencing begin in U.S.
First archael genome sequenced
First bacterial genome (H. influenzae) sequenced
Yeast (S. cerevisiae) genome sequenced
HGP's mouse genetic mapping goal achieved
Bermuda principles for rapid and open data release established
Ethical, Legal, and Social Implications (ELSI) programs founded at NIH and DOE
Rapid data release guidelines established by NIH and DOE
Sanger Centre founded (later renamed Wellcome Trust Sanger Institute)
The Wellcome Trust
U.S. Equal Employment Opportunity Commission issues policy on genetic discrimination in the workplace
First gene for breast cancer (BRCA1) mapped
Collins et al. (2003)



Outline

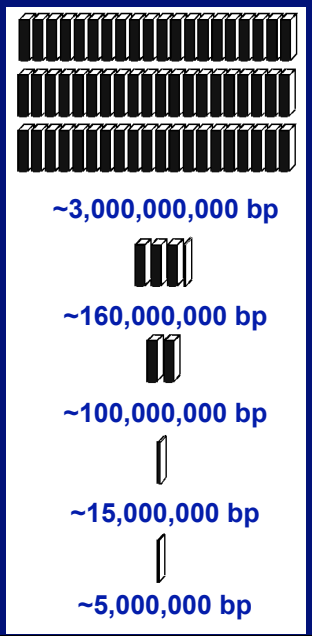
I. Fundamentals of Genome Mapping & Sequencing

II. Mapping & Sequencing in the Human Genome Project

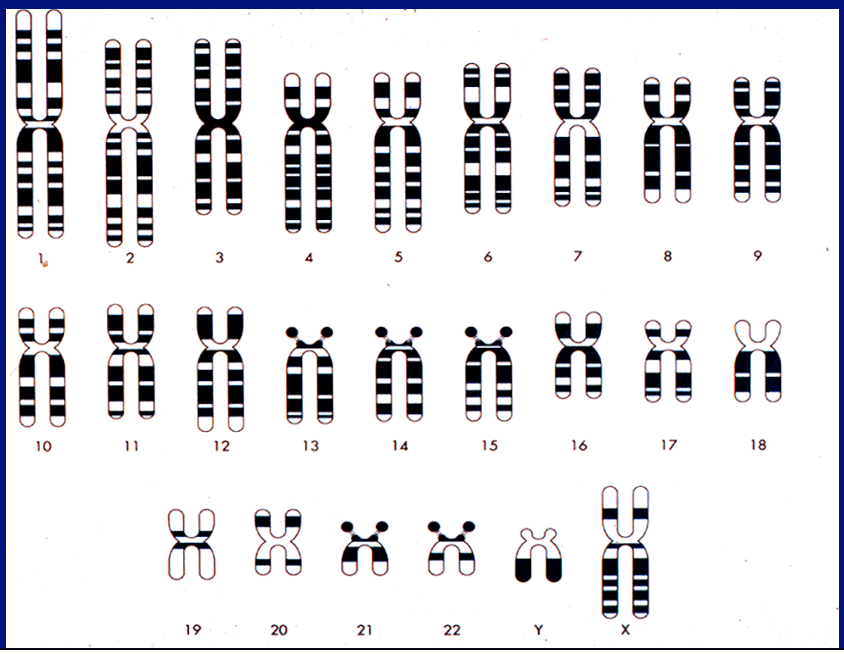
III. Comparative Sequencing

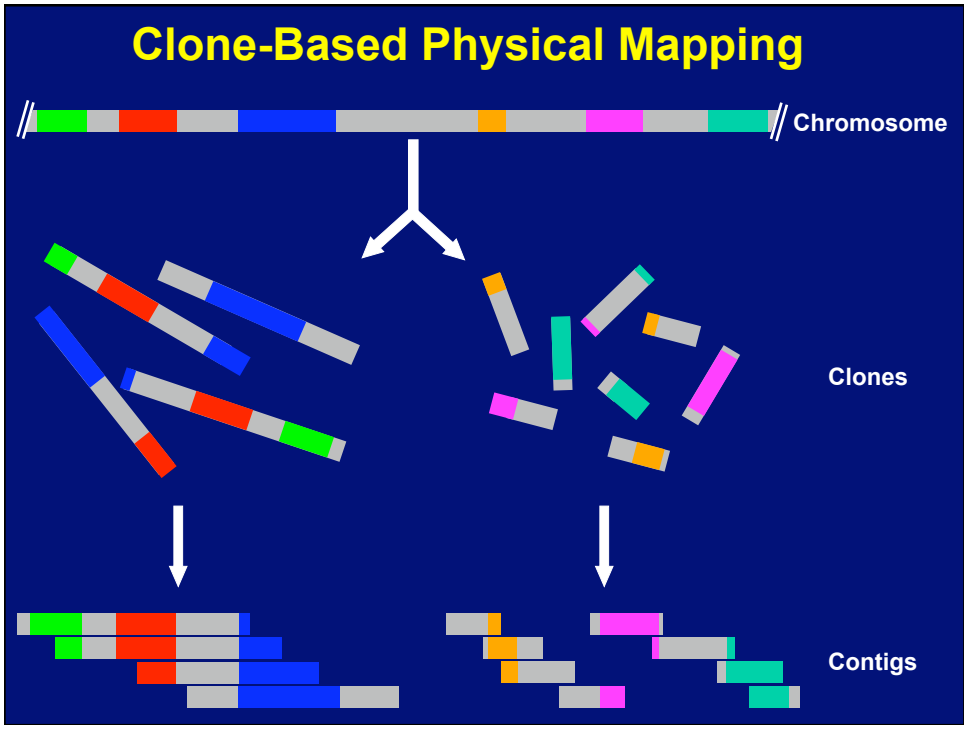
IV. New Frontiers in Genomics

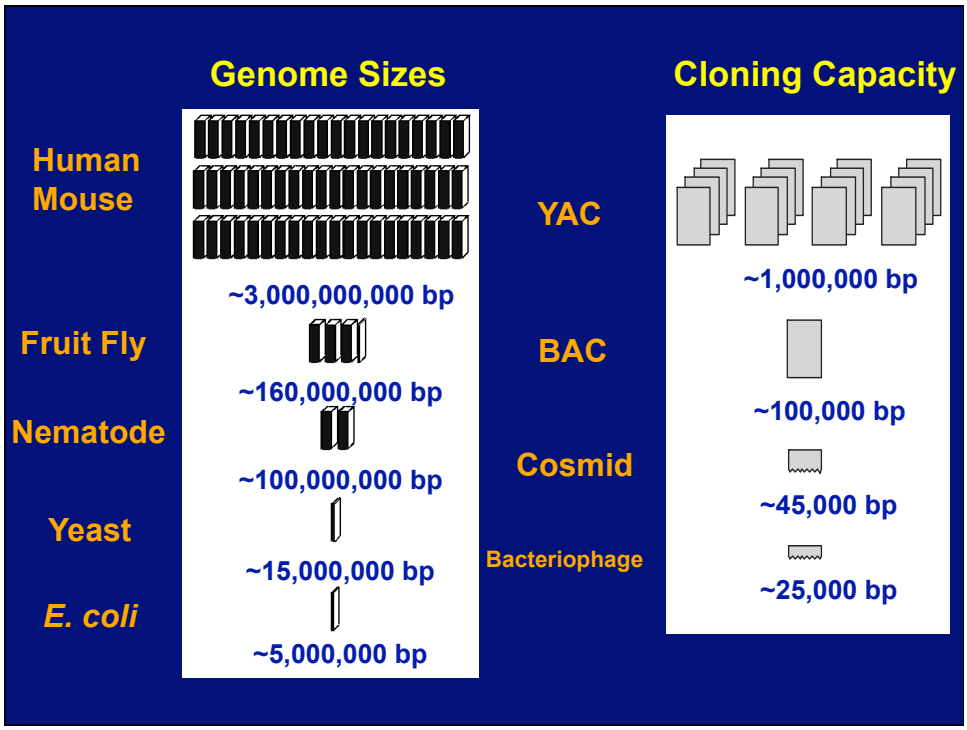
Genome Sizes

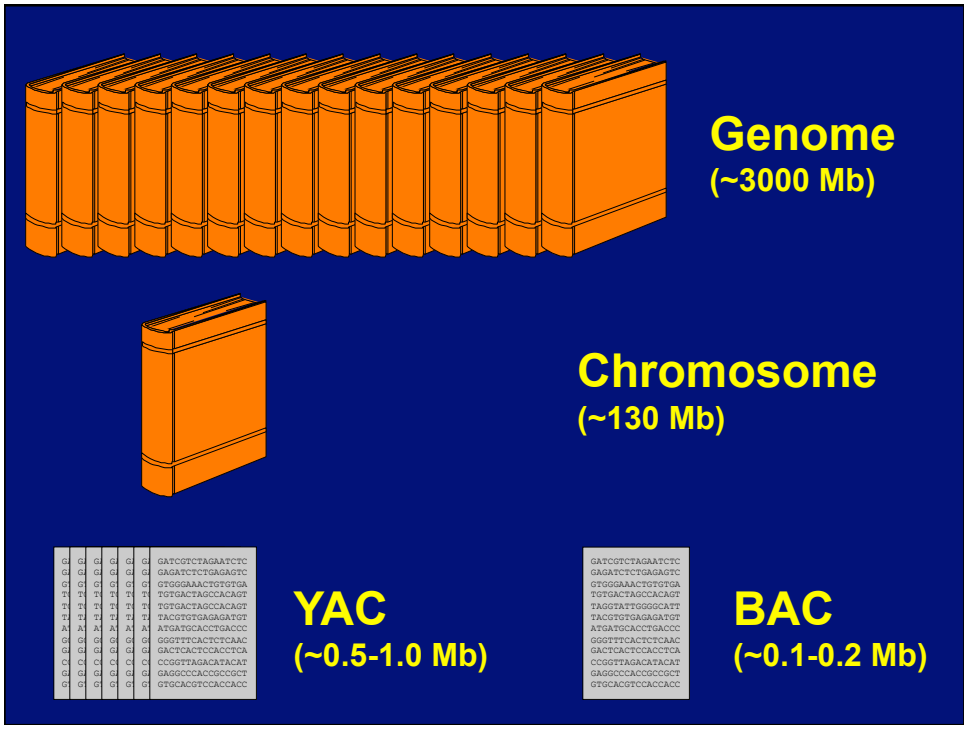


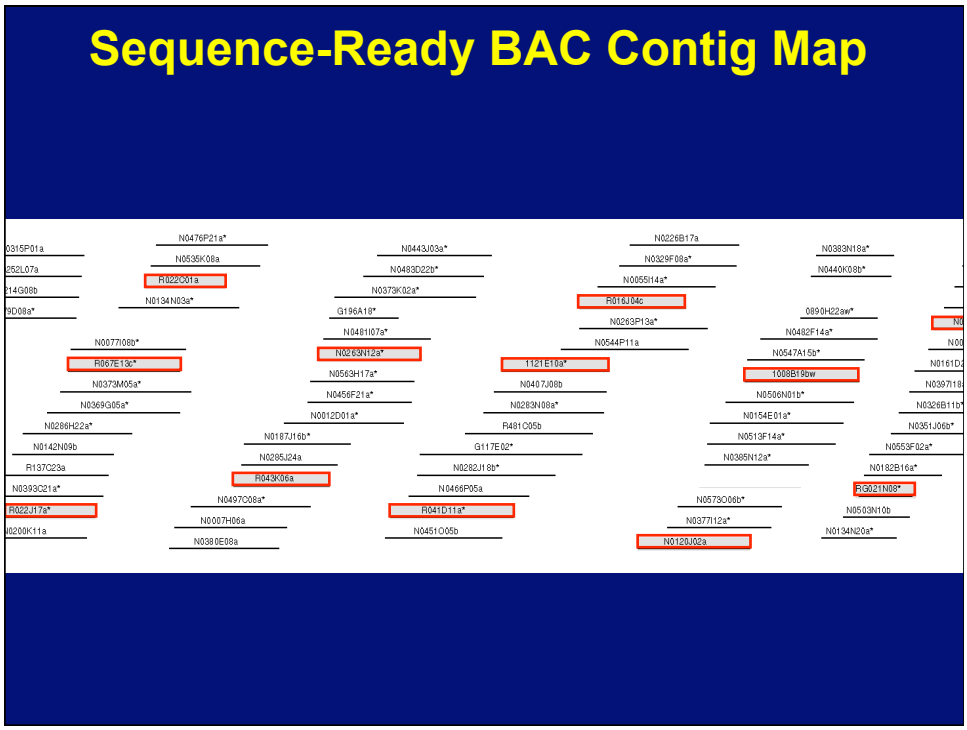
The Human Cytogenetic Map




Clone-Based Physical Mapping
Chromosome
Clones
Contigs

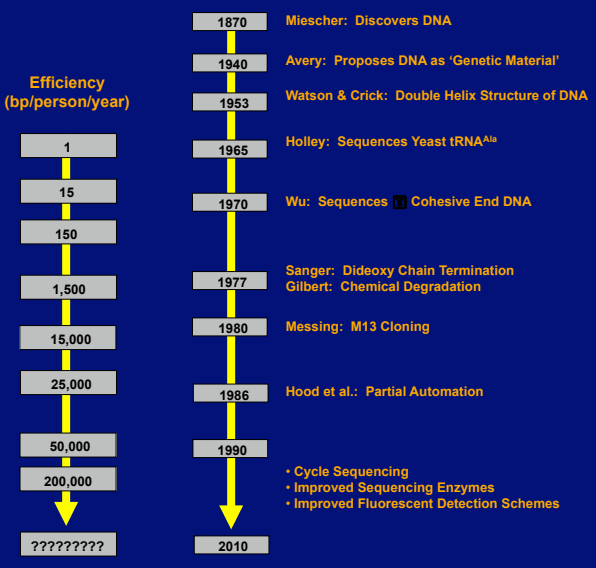

Genome Sizes
Cloning Capacity
Human
Mouse
~3,000,000,000 bp
Fruit Fly
~160,000,000 bp
Nematode
~100,000,000 bp
Yeast
~15,000,000 bp
E. coli
~5,000,000 bp
YAC
~1,000,000 bp
BAC
~100,000 bp
Cosmid
~45,000 bp
Bacteriophage
~25,000 bp

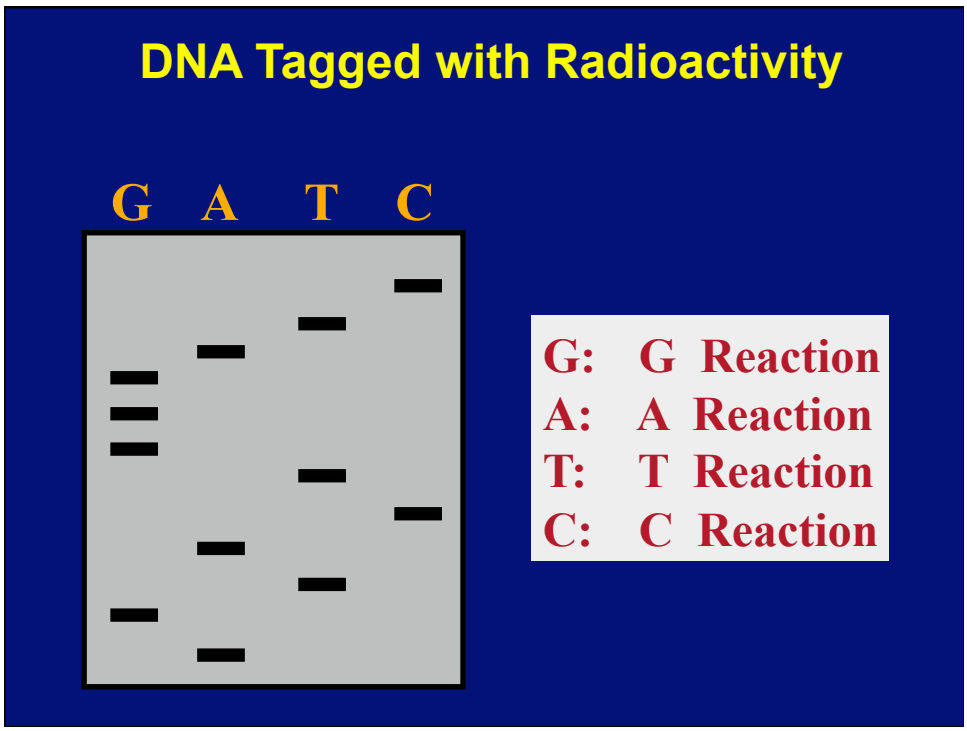

Genome
(~3000 Mb)
Chromosome
(~130 Mb)
YAC
(~0.5-1.0 Mb)
BAC
(~0.1-0.2 Mb)

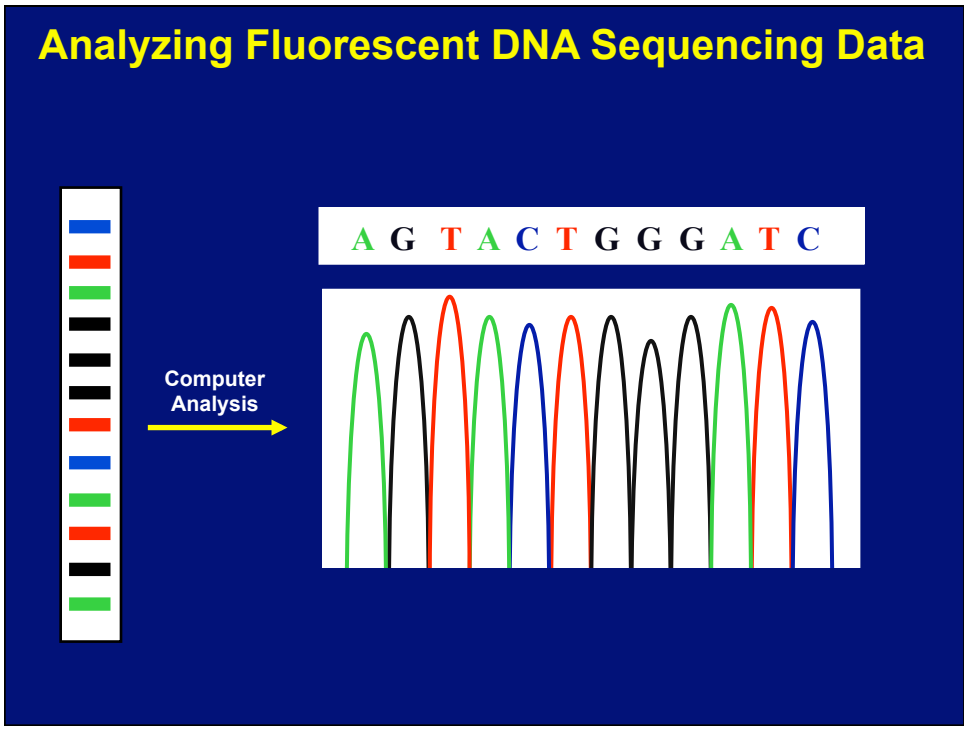

Sequence-Ready BAC Contig Map

DNA Sequencing

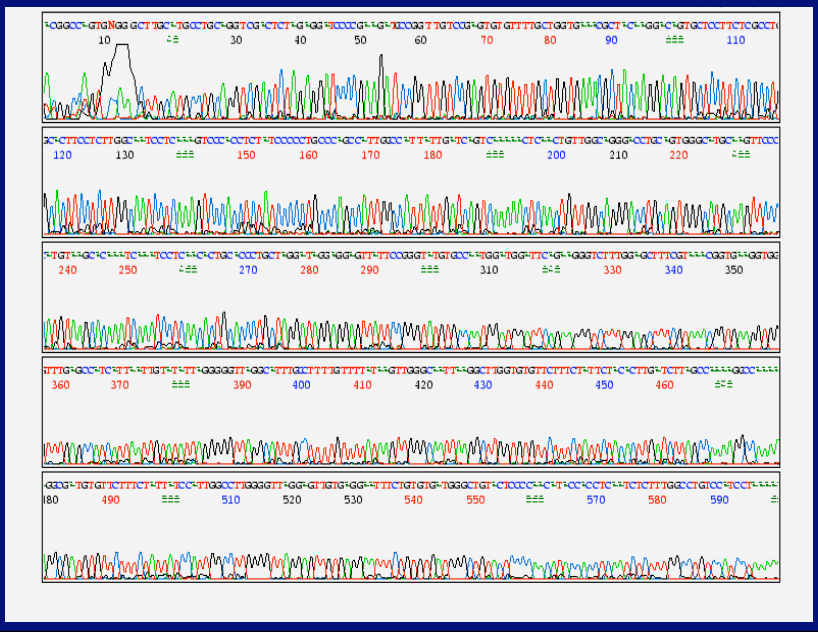
History of DNA Sequencing




DNA Tagged with Radioactivity
G A T C
G: G Reaction
A: A Reaction
T: T Reaction
C: C Reaction


Analyzing Fluorescent DNA Sequencing Data
Computer Analysis
A G T A C T G G G A T C

Fluorescent DNA Sequencing Results

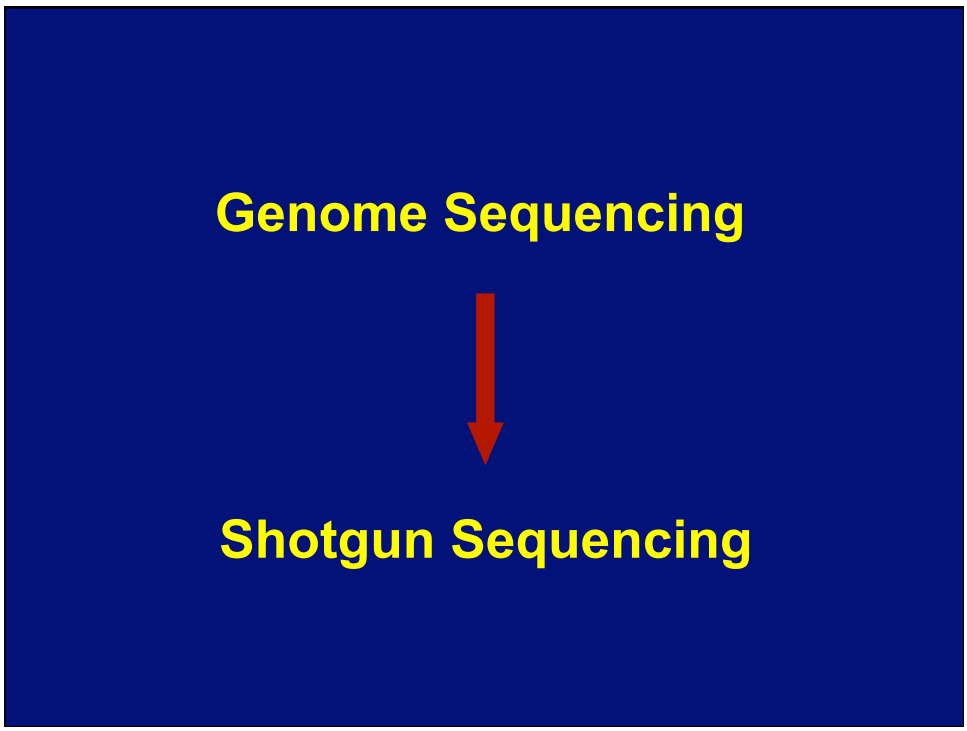


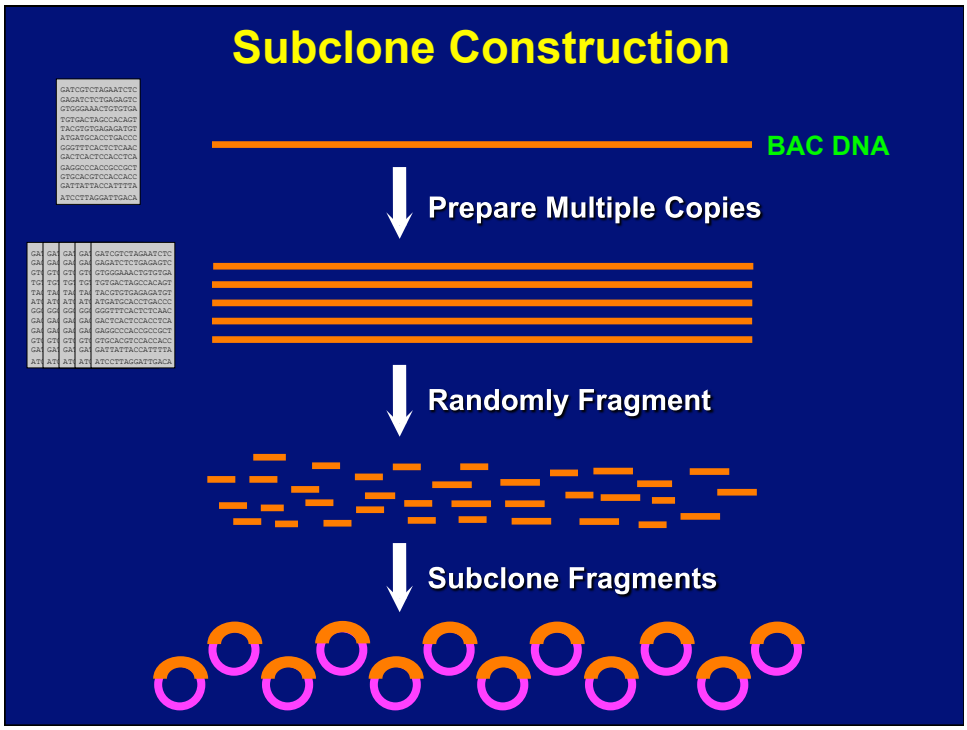
Analysis of Gene Expression

- **ESTs:** Expressed-Sequence Tags
- **SAGE:** Serial Analysis of Gene Expression
- Full-Insert (Full-Length) cDNA Sequencing

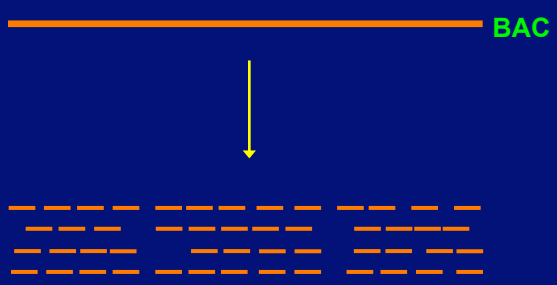
Mammalian Gene Collection

mgc.nci.nih.gov


Genome Sequencing
Shotgun Sequencing


Subclone Construction
BAC DNA
Prepare Multiple Copies
Randomly Fragment
Subclone Fragments

Shotgun Sequencing Strategy



Poisson Calculations

The sequencing strategy for the shotgun approach follows the Lander and Waterman application of the Poisson distribution

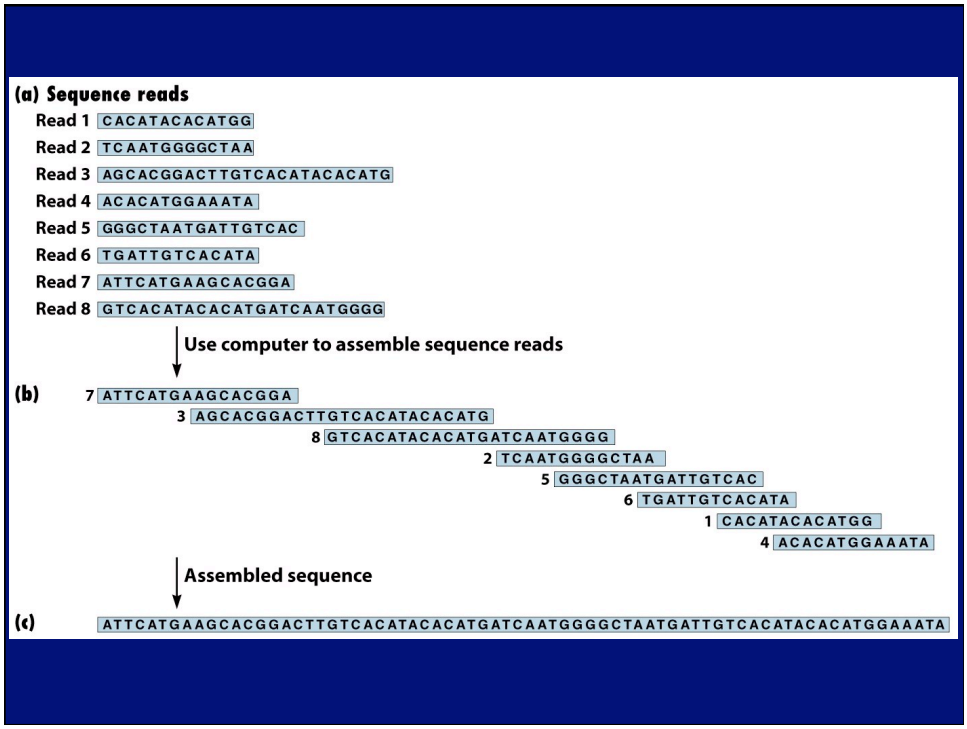
The probability a base is not sequenced is given by:

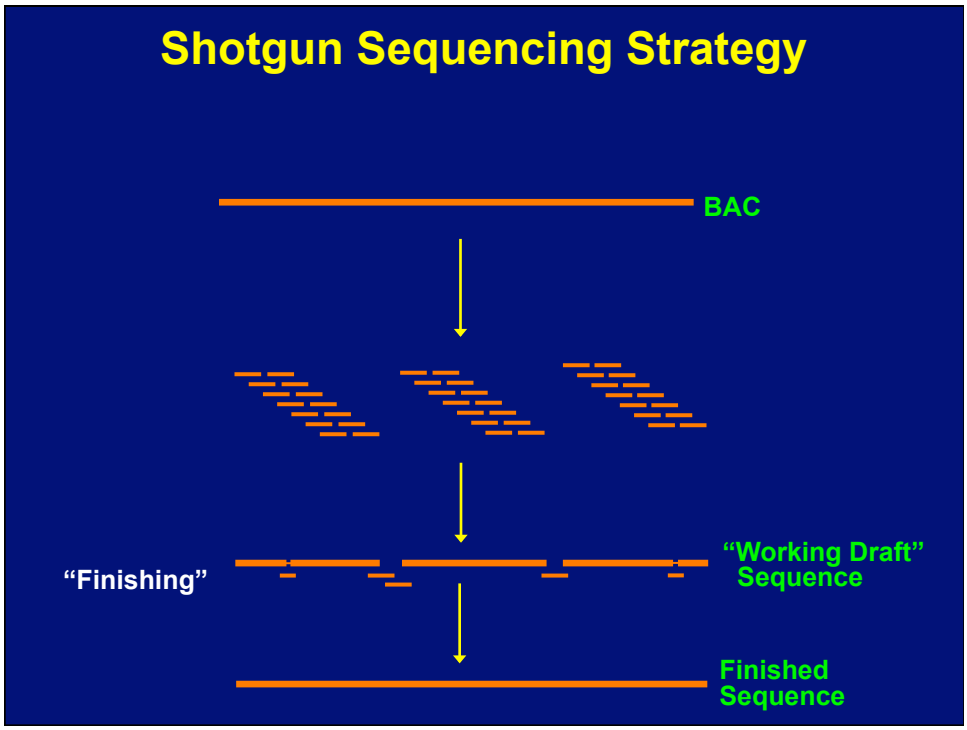
$$P_0 = e^{-c}$$

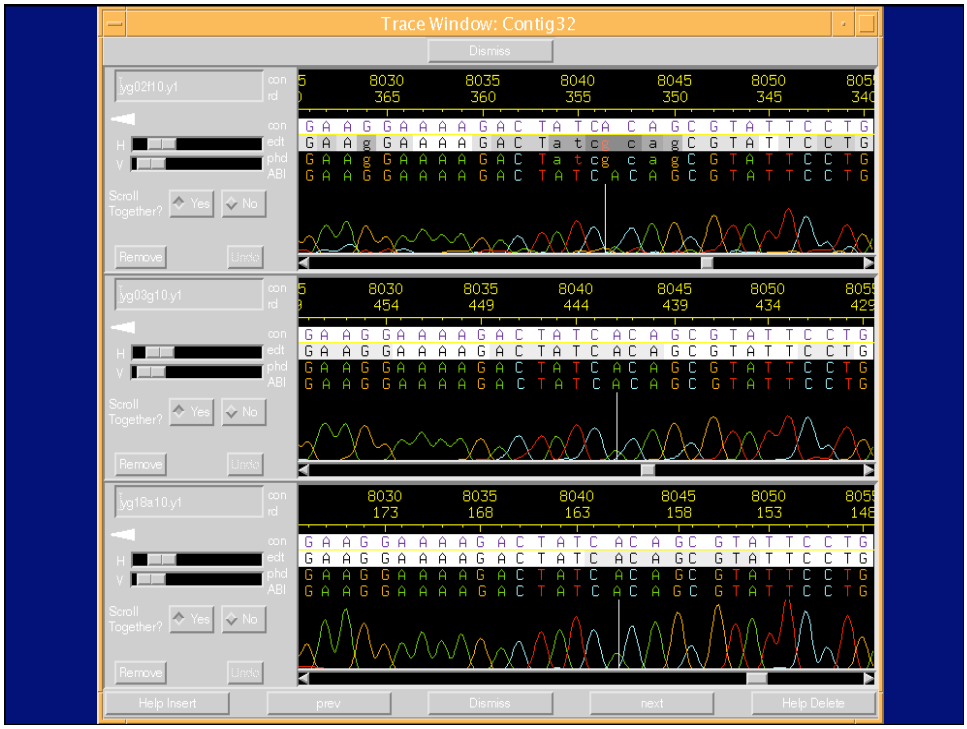
Where:

- c = fold sequence coverage (c=LN/G),
- LN = # bases sequenced, i.e. L = average sequencing read length and N = # reads
- G = target sequence length
- e = 2.718 (e=2.718281828459)

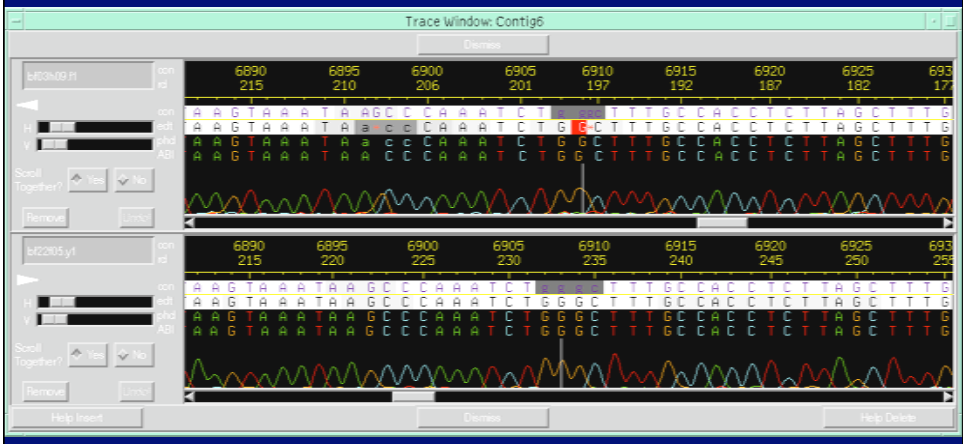
Fold Coverage	$P_0=e^{-c}$	% not sequenced	% sequenced
1	0.37	37%	63%
2	0.135	13.5%	87.5%
3	0.05	5%	95%
4	0.018	1.8%	98.2%
5	0.0067	0.6%	99.4%
6	0.0025	0.25%	99.75%
7	0.0009	0.09%	99.91%
8	0.0003	0.03%	99.97
9	0.0001	0.01%	99.99%
10	0.000045	0.005%	99.995%


(a) Sequence reads
Read 1 CACATACACATGG
Read 2 TCAATGGGGCTAA
Read 3 AGCACGGACTTGTCACATACACATG
Read 4 ACACATGGAAATA
Read 5 GGGCTAATGATTGTCAC
Read 6 TGATTGTCACATA
Read 7 ATTCATGAAGCACGGA
Read 8 GTCACATACACATGATCAATGGGG
Use computer to assemble sequence reads
(b)
7 ATTCATGAAGCACGGA
3 AGCACGGACTTGTCACATACACATG
8 GTCACATACACATGATCAATGGGG
2 TCAATGGGGCTAA
5 GGGCTAATGATTGTCAC
6 TGATTGTCACATA
1 CACATACACATGG
4 ACACATGGAAATA
Assembled sequence
(c)
ATTCATGAAGCACGGACTTGTCACATACACATGATCAATGGGGCTAATGATTGTCACATACACATGGAAATA


Shotgun Sequencing Strategy
BAC
"Finishing"
"Working Draft" Sequence
Finished Sequence



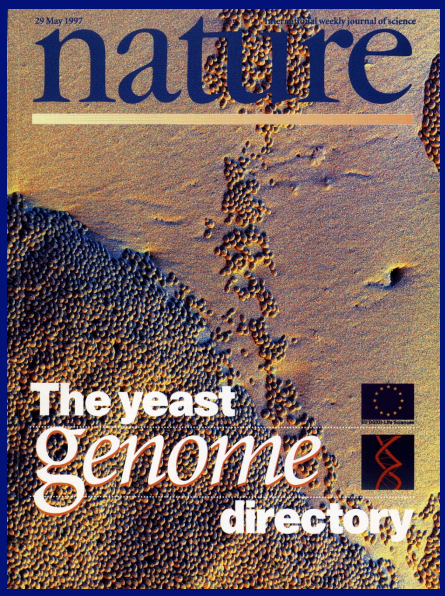
Sequence Finishing: Resolving Ambiguities



*** Sequence Finishing: Remains Relatively Expensive ***

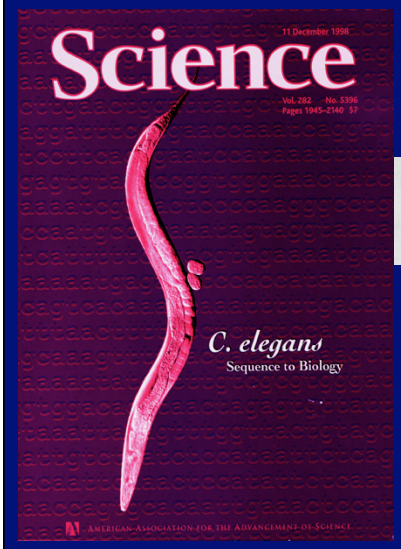
Historically Significant Genome Sequencing Projects

First Eukaryotic Genome Sequence



Goffeau et al. (1997)

First Animal Genome Sequence

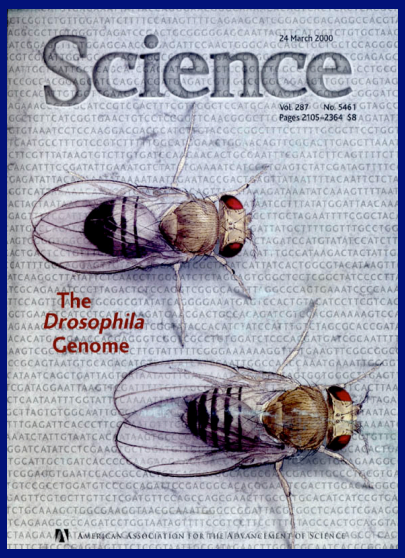


Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology

The *C. elegans* Sequencing Consortium*

***C. elegans* Sequencing Consortium (1998)**

Second Animal Genome Sequence



THE *DROSOPHILA* GENOME

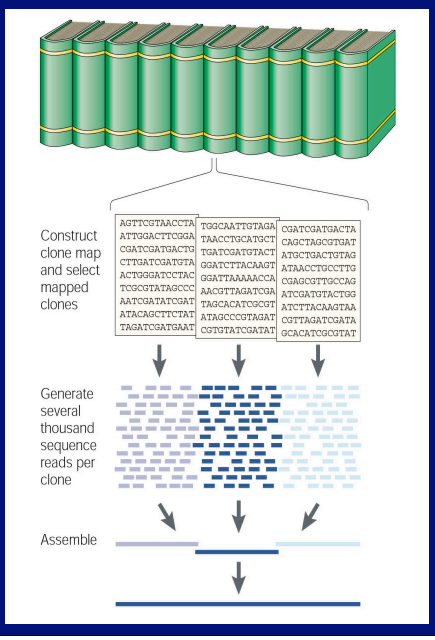
REVIEW

The Genome Sequence of *Drosophila melanogaster*

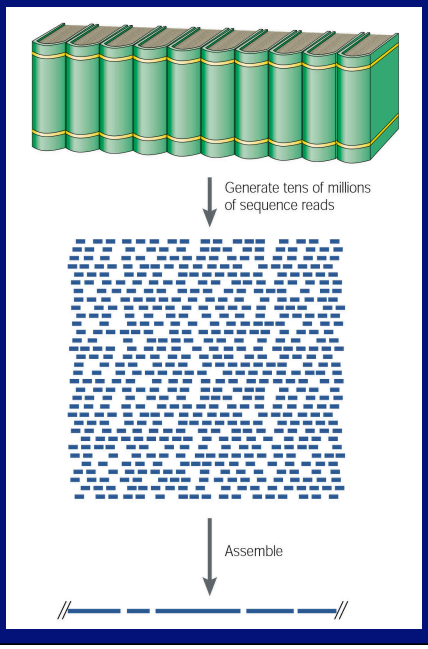
Mark D. Adams,[1*] Susan E. Celniker,[2] Robert A. Holt,[1] Cheryl A. Evans,[1] Jeannine D. Gocayne,[1] Peter G. Amanatides,[1] Steven E. Scherer,[3] Peter W. Li,[1] Roger A. Hoskins,[2] Richard F. Galle,[2] Reed A. George,[2] Suzanna E. Lewis,[4] Stephen Richards,[2] Michael Ashburner,[5] Scott N. Henderson,[1] Granger G. Sutton,[1] Jennifer R. Wortman,[1] Mark D. Yandell,[1] Qing Zhang,[1] Lin X. Chen,[1] Rhonda C. Brandon,[1] Yu-Hui C. Rogers,[1] Robert G. Blazej,[3] Mark Champe,[2] Barret D. Pfeiffer,[2] Kenneth H. Wan,[2] Clare Doyle,[2] Evan G. Baxter,[2] Gregg Helt,[6] Catherine R. Nelson,[6] George L. Gabor Miklos,[7] Josep F. Abril,[8] Anna Agbayani,[2] Hui-Jin An,[1] Cynthia Andrews-Pfannkoch,[1] Danita Baldwin,[1] Richard M. Ballew,[1] Anand Basu,[1] James Baxendale,[1] Leyla Bayraktaroglu,[9] Ellen M. Beasley,[1] Karen Y. Beeson,[1] P. V. Benos,[10] Benjamin P. Berman,[2] Deepali Bhandari,[1] Slava Bolshakov,[11] Dana Borkova,[12] Michael R. Botchan,[13] John Bouck,[3] Peter Brokstein,[6] Phillipe Brottier,[14] Kenneth C. Burtis,[15] Dana A. Busam,[1] Heather Butler,[16] Edouard Cadieu,[17] Angela Center,[1] Ishwar Chandra,[1] J. Michael Cherry,[18] Simon Cawley,[19] Carl Dahlke,[1] Lionel B. Davenport,[1] Peter Davies,[1] Beatriz de Pablos,[20] Arthur Delcher,[1] Zuoming Deng,[1] Anne Deslattes Mays,[1] Ian Dew,[1] Suzanne M. Dietz,[1] Kristina Dodson,[1] Lisa E. Doup,[1] Michael Downes,[21] Shannon Dugan-Rocha,[3] Boris C. Dunkov,[22] Patrick Dunn,[1] Kenneth J. Durbin,[1] Carlos C. Evangelista,[1] Concepcion Ferraz,[23] Steven Ferriera,[1] Wolfgang Fleischmann,[5] Carl Fosler,[1] Andrei E. Gabrielian,[1] Neha S. Garg,[1] William M. Gelbart,[9] Ken Glasser,[1] Anna Glodek,[1] Fangcheng Gong,[1] J. Harley Gorrell,[9] Zhiping Gu,[1] Ping Guan,[1] Michael Harris,[1] Nomi L. Harris,[2] Damon Harvey,[1] Thomas J. Heiman,[1] Judith R. Hernandez,[3] Jarrett Houck,[1] Damon Hostin,[1] Kathryn A. Houston,[2] Timothy J. Howland,[1] Ming-Hui Wei,[1] Chinyere Ibegwam,[1] Mena Jalali,[1] Francis Kalush,[1] Gary H. Karpen,[21] Zhaoxi Ke,[1] James A. Kennison,[24] Karen A. Ketchum,[1] Bruce E. Kimmel,[2] Chinnappa D. Kodira,[1] Cheryl Kraft,[1] Saul Kravitz,[1] David Kulp,[6] Zhongwu Lai,[1] Paul Lasko,[25] Yiding Lei,[1] Alexander A. Levitsky,[1] Jieyin Li,[1] Zhenya Li,[1] Yong Liang,[1] Xiaoying Lin,[26] Xiangjun Liu,[1] Bettina Mattei,[1] Tina C. McIntosh,[1] Michael P. McLeod,[3] Duncan McPherson,[1] Gennady Merkulov,[1] Natalia V. Milshina,[1] Clark Mobarry,[1] Joe Morris,[6] Ali Moshrefi,[2] Stephen M. Mount,[27] Mee Moy,[1] Brian Murphy,[1] Lee Murphy,[28] Donna M. Muzny,[3] David L. Nelson,[3] David R. Nelson,[29] Keith A. Nelson,[1] Katherine Nixon,[2] Deborah R. Nusskern,[1] Joanne M. Pacleb,[2] Michael Palazzolo,[2] Gjange S. Pittman,[1] Sue Pan,[1] John Pollard,[1] Vinita Puri,[1] Martin G. Reese,[6] Knut Reinert,[1] Karin Remington,[1] Robert D. C. Saunders,[30] Frederick Scheeler,[1] Hua Shen,[2] Bixiang Christopher Shue,[1] Inga Sidén-Kiamos,[11] Michael Simpson,[1] Marian P. Skupski,[1] Tom Smith,[1] Eugene Spier,[1] Allan C. Spradling,[31] Mark Stapleton,[2] Renee Strong,[1] Eric Sun,[1] Robert Svirskas,[32] Cyndee Tector,[1] Russell Turner,[1] Eli Venter,[1] Aihui H. Wang,[1] Xin Wang,[1] Zhen-Yuan Wang,[1] David A. Wassarman,[33] George M. Weinstock,[3] Jean Weissenbach,[14] Sherita M. Williams,[1] Trevor Woodage,[1] Kim C. Worley,[3] David Wu,[1] Song Yang,[2] Q. Alison Yao,[1] Jane Ye,[1] Ru-Fang Yeh,[10] Jayshree S. Zaveri,[1] Ming Zhan,[1] Guangren Zhang,[1] Qi Zhao,[1] Lianzheng Zheng,[1] Xianggun H. Zheng,[1] Fei N. Zhong,[1] Wenyan Zhong,[1] Xiaojun Zhou,[3] Shiaoping Zhu,[1] Xiaohong Zhu,[1] Hamilton O. Smith,[1] Richard A. Gibbs,[3] Eugene W. Myers,[1] Gerald M. Rubin,[34] J. Craig Venter[1]

Adams et al. (2000)

Clone-Based Shotgun Sequencing



Whole-Genome Shotgun Sequencing

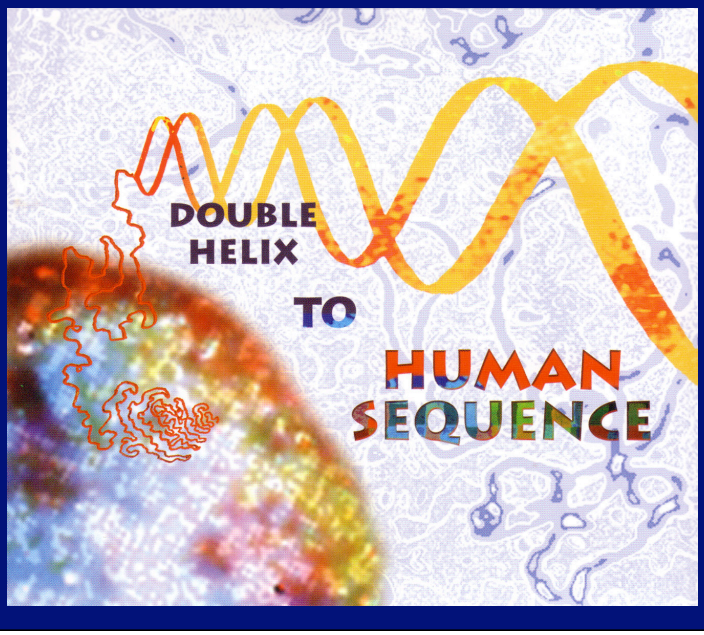


February, 2001 Draft Sequence

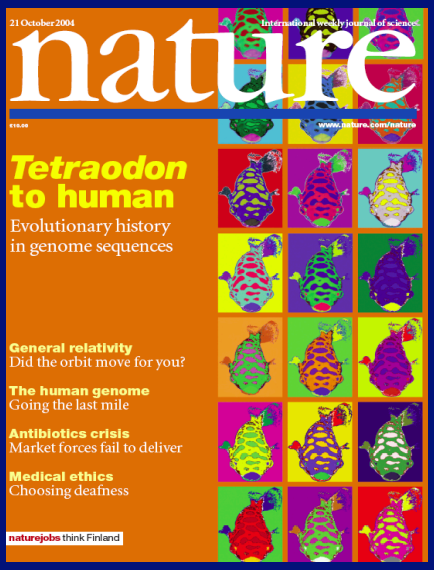
International Human Genome Sequencing Consortium (2001)

Venter et al. (2001)

April, 2003 Completion



October, 2004 Publication



articles

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium*

* A list of authors and their affiliations appears in the Supplementary Information

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human enome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

The Human Genome Project (HGP) was launched in 1990 with the goal of obtaining a highly accurate sequence of the vast majority of the euchromatic portion of the human genome. The initial work followed a two-pronged approach: (1) the mapping of the human and mouse genomes[1–6] to allow the study of inherited disease and provide a crucial scaffold for genome assembly; and (2) the sequencing of organisms with smaller, simpler genomes[7,8] to serve as a testbed for method development and assist in interpreting the human genome. With success along both paths, the sequencing of the human genome itself eventually became feasible. The International Human Genome Sequencing Consortium (IHGSC), an open collaboration involving twenty centres in six countries, was formed to carry out this component of the HGP.

In February 2001, the IHGSC[9] and Celera Genomics[10] each reported draft sequences providing a first overall view of the human genome. These sequences allowed systematic study of the human genome itself, including identification of genes, combinatorial architecture of proteins, regional differences in genome composition, distribution and history of transposable elements, distribution of polymorphism and relationship between genetic recombination and physical distance. Moreover, systematic knowledge of the human genome has enabled new tools and approaches that have markedly accelerated biomedical research.

Both draft sequences, however, had important shortcomings. The IHGSC sequence, for example, omitted ~10% of the euchromatic genome; it was interrupted by ~150,000 gaps and the order and orientation of many segments within local regions had not been established. The IHGSC thus turned to the challenge of completing the sequence of the euchromatic genome. Operationally, a finished sequence was defined as having an error rate of, at most, one event per 10^4 bases, and the goal for completion was coverage in finished sequence of at least 95% of the euchromatic genome, with the only gaps being those refractory to all available techniques[11] (see http://www.genome.gov/10000925). The goal was challenging because the human genome is replete with such features as dispersed repeats and large segmental duplications, which greatly complicate the determination of genome structure and sequence. In fact, near-complete sequences have been obtained so far only for three multicellular organisms: the nematode[12], mustard weed[13] and the fruitfly[14]. These genomes are all roughly 30-fold smaller than the human genome and have much simpler structure.

We describe here the results of a multiyear effort by the IHGSC towards the goal of a complete human sequence. The number of gaps has been reduced 403-fold to only 341, most of which are associated with segmental duplications and will require new methods for resolution. The assembled near-complete genome sequence has an error rate of only ~1 event per 100,000 bases; it contains 2.85 billion nucleotides and covers ~99% of the euchromatic genome. This paper describes the current genome sequence and the process used to produce it; examines the accuracy and completeness of the sequence; and illustrates biological analyses made possible by the sequence. We do not attempt here a comprehensive analysis of the contents of the human genome. An initial analysis was previously reported[15] and a series of papers is being written describing the individual chromosomes[16–36], including annotation of genes and other features.

Current genome sequence

Finishing process

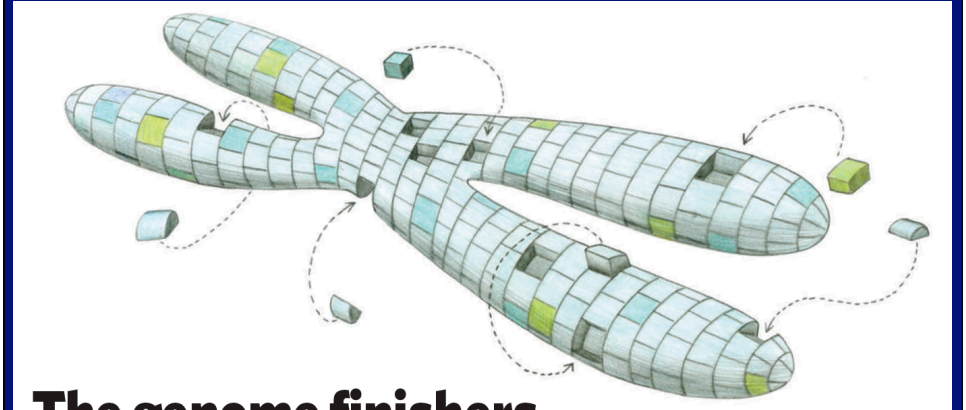
The process of converting the initial draft sequence into a near-complete sequence is referred to as 'finishing'. It is a complex iterative process that proceeds simultaneously at multiple scales, ranging from single nucleotides to the integrity of whole chromosomes. The fundamental challenge is that genomic regions that are not well represented or readily resolved through random shotgun sequencing tend to be highly enriched in problematic sequences. Resolving such regions required the development of special approaches, which evolved substantially over time and varied among centres.

Broadly, the finishing process involved two distinct components: (1) producing finished maps, consisting of continuous and accurate paths of overlapping large-insert clones spanning the euchromatic region of each chromosome arm; and (2) producing finished clones, consisting of continuous and accurate nucleotide sequence across each large-insert clone. In practice, these two components were tightly intertwined in that progress in each often depended on results from the other. The components are described in Boxes 1 and 2. Further information about the finishing process and finishing standards can be found in the Supplementary Information (Note 1) and at http://www.genome.gov/10000925.

In total, we generated a shotgun sequence from 59,208 large-insert clones (total length ~5.84 gigabases (Gb)) and finished the sequence from 45,742 of these clones (total length ~3.67 Gb). The clones consisted primarily of bacterial artificial chromosomes

NATURE | VOL 431 | 21 OCTOBER 2004 | www.nature.com/nature ©2004 Nature Publishing Group 931

International Human Genome Sequencing Consortium (2004)

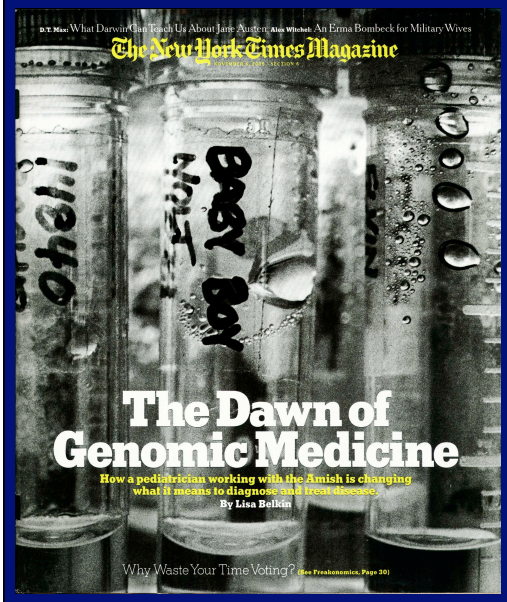


The genome finishers

Dedicated scientists are working hard to close the gaps, fix the errors and finally complete the human genome sequence. **Elie Dolgin** looks at how close they are.

***Nature* (2009)**

Human Genome Project



The Path to Genomic Medicine

HGP

Realization of Genomic Medicine

Genomic Medicine

Healthcare tailored to the individual based on genomic information

The Path to Genomic Medicine

HGP

→→ ? →→

Realization of Genomic Medicine

✓ *"Fulfilling the Promise"*

?

Mapping the Human Genome	~1990 to ~2000	The Human Genome Project
Sequencing the Human Genome	~1998 to ~2003	
Interpreting the Human Genome Sequence	~2003 to ???	*Beyond* The Human Genome Project

~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
AAATTATAGACTTTTTTTTAGCAGAGAACTTTGAGGAACCTAAATGCCAACCAGTCTAAAAATGCAGTTTTCAGAAGAATGAATATTTCATGGATA
GTTCTAAATACTAATGAACTTTAAAATAGCTTACTATTGATCTGTCAAAGTGGGTTTTTATATAATTTTCTTTTTACAAATCACCTGACACATTT
AATATAGGTTAAAAAATGCTATCAGGCTGGTTTGCAAAGAAAATGTATTACAAAGGCTGCTAAGTGTGTTAAGAGCATACTCATTTCTGTTCTCC
AAAATATTTCATAAGGTGCTTTAAGAATAGGTATGTTTTTAAAAGTTAAGTTCCTACTATTTATAGGAACTGACAATCACCTAAAATACCAATGA
TTACAAACTTCCTTCTGGCCTTCTGGACTGCAATTCTAAAAGTGTAAAAAACATATTTTCTGCATTAAGTTAGGCAGTATTGCTTAGTTTTCAAA
GTGGTAGGCTTTGGAGTCAGATTATTTTGATTCAGATCCTACATCTACTGTTTAGTAGCTCTGTTGCCTGAGGCAGGTCCCTTAACATCTCTGTG
TGTGACTTGACCTTTAAAATTTGGAGACTGTCATAGGGGTTAATCCCTTGAGAAAATGAATGTGAAAAGTTAGCCTAATGTTAACTGCTATTATT
ATGGATTACCATATTTTCACATTCATCACAGTACATGCACCTTGTTAATATAAGATGCTCAATTCATCTTTGAGTATAATTTTGTGACTCTCAAT
CTGGATATGCAATGAGTGGGCCTGTATGAGAATTTAATTTATGAAAAATTGTGTTTCACATGGCCTTACCAGATATACAGGAAACACGTCACATG
TTTCTATTGTATGTTGTTAAATGCCTTAGAATTTAACTTTCTGAATAGGATCCCTTCAGTTTGAGAGTCATAAAAGAGTAAAATTATTATGGTAT

The Human Genome… by the Numbers

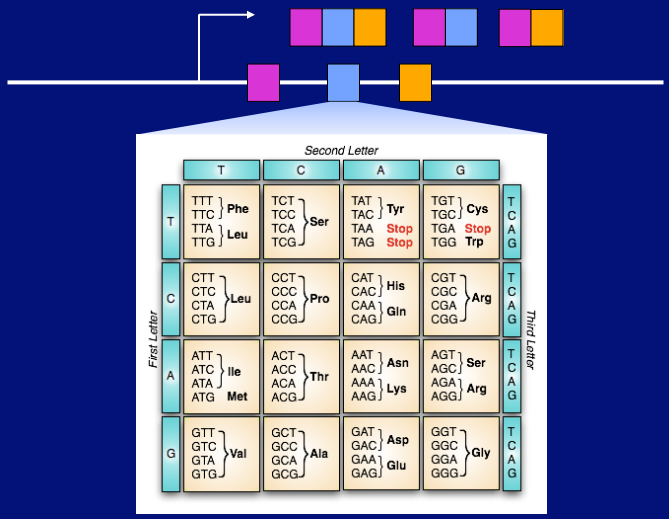
~5% of Human Genome Sequence is Constrained Across Mammals (and Presumed Functional)

- 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

~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
AAATTATAGACTTTTTTTTAGCAGAGAACTTTGAGGAACCTAAATGCCAACCAGTCTAAAAATGCAGTTTTCAGAAGAATGAATATTTCATGGATA
GTTCTAAATACTAATGAACTTTAAAATAGCTTACTATTGATCTGTCAAAGTGGGTTTTTATATAATTTTCTTTTTACAAATCACCTGACACATTT
AATATAGGTTAAAAAATGCTATCAGGCTGGTTTGCAAAGAAAATGTATTACAAAGGCTGCTAAGTGTGTTAAGAGCATACTCATTTCTGTTCTCC
AAAATATTTCATAAGGTGCTTTAAGAATAGGTATGTTTTTAAAAGTTAAGTTCCTACTATTTATAGGAACTGACAATCACCTAAAATACCAATGA
TTACAAACTTCCTTCTGGCCTTCTGGACTGCAATTCTAAAAGTGTAAAAAACATATTTTCTGCATTAAGTTAGGCAGTATTGCTTAGTTTTCAAA
GTGGTAGGCTTTGGAGTCAGATTATTTTGATTCAGATCCTACATCTACTGTTTAGTAGCTCTGTTGCCTGAGGCAGGTCCCTTAACATCTCTGTG
TGTGACTTGACCTTTAAAATTTGGAGACTGTCATAGGGGTTAATCCCTTGAGAAAATGAATGTGAAAAGTTAGCCTAATGTTAACTGCTATTATT
ATGGATTACCATATTTTCACATTCATCACAGTACATGCACCTTGTTAATATAAGATGCTCAATTCATCTTTGAGTATAATTTTGTGACTCTCAAT
CTGGATATGCAATGAGTGGGCCTGTATGAGAATTTAATTTATGAAAAATTGTGTTTCACATGGCCTTACCAGATATACAGGAAACACGTCACATG
TTTCTATTGTATGTTGTTAAATGCCTTAGAATTTAACTTTCTGAATAGGATCCCTTCAGTTTGAGAGTCATAAAAGAGTAAAATTATTATGGTAT

Coding Sequences (i.e., Genes)



The Genetic Code

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

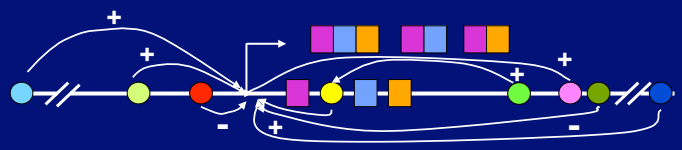
~1.5% Encodes for Protein (Genes)

- Corresponds to ~18-22K Genes
- Many More than ~22K Different Proteins
- Good Inventory at Present

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

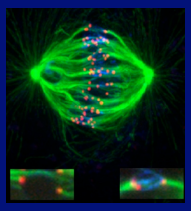
TGCCGCGGAACTTTTCGGCTCTCTAAGGCTGTATTTTGATATACGAAAGGCACATTTTCCTTCCCTTTTCAAAATGCACCTTGCAAACGTAACAG
GAACCCGACTAGGATCATCGGGAAAAGGAGGAGGAGGAGGAAGGCAGGCTCCGGGGAAGCTGGTGGCAGCGGGTCCTGGGTCTGGCGGACCCTGA
CGCGAAGGAGGGTCTAGGAAGCTCTCCGGGGAGCCGGTTCTCCCGCCGGTGGCTTCTTCTGTCCTCCAGCGTTGCCAACTGGACCTAAAGAGAGG
CCGCGACTGTCGCCCACCTGCGGGATGGGCCTGGTGCTGGGCGGTAAGGACACGGACCTGGAAGGAGCGCGCGCGA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

Non-Coding Functional Sequences

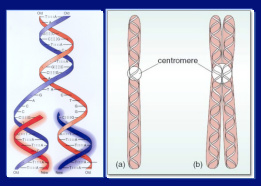


Gene Regulation

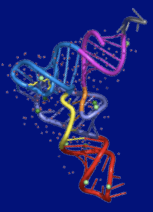
Chromosome Packaging



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

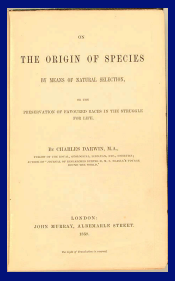
~1.5% Encodes for Protein (Genes)

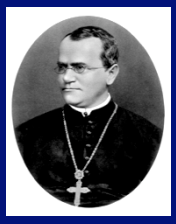
- Corresponds to ~18-22K Genes
- Many More than ~22K Different Proteins
- Good Inventory at Present

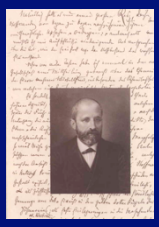
~3.5% Functional But Non-Coding

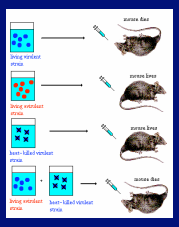
- Gene Regulatory Elements
- Chromosomal Functional Elements
- Undiscovered Functional Elements (NOT Yet in Textbooks!)
- Poor Inventory at Present

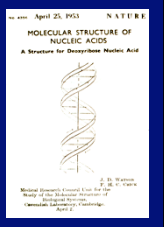
Foundational Milestones in Genetics & Genomics



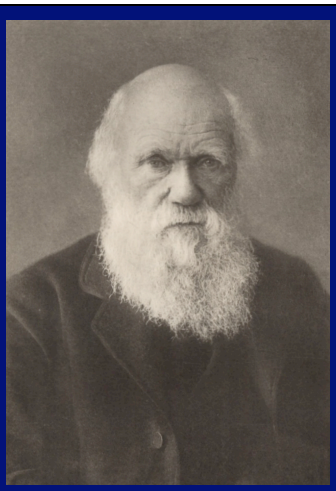








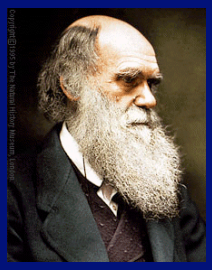
Darwin	Mendel	Miescher	Avery	Watson & Crick
1859	*1865*	*1871*	*1944*	*1953*



Charles Darwin
Born February 12, 1809

"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."

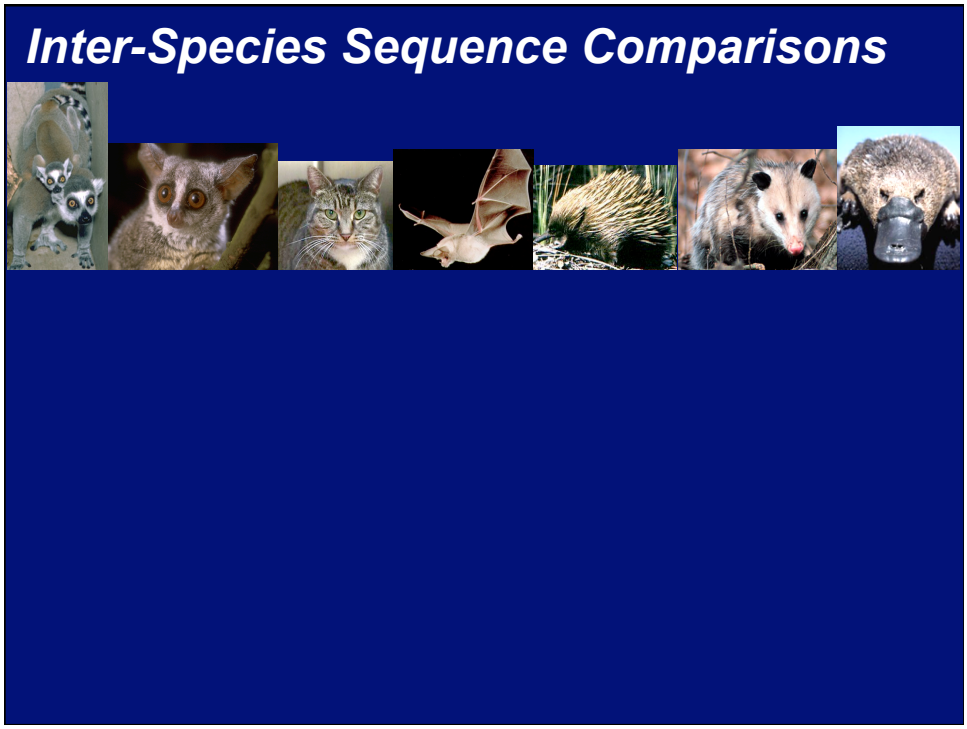
(Attributed to Darwin)

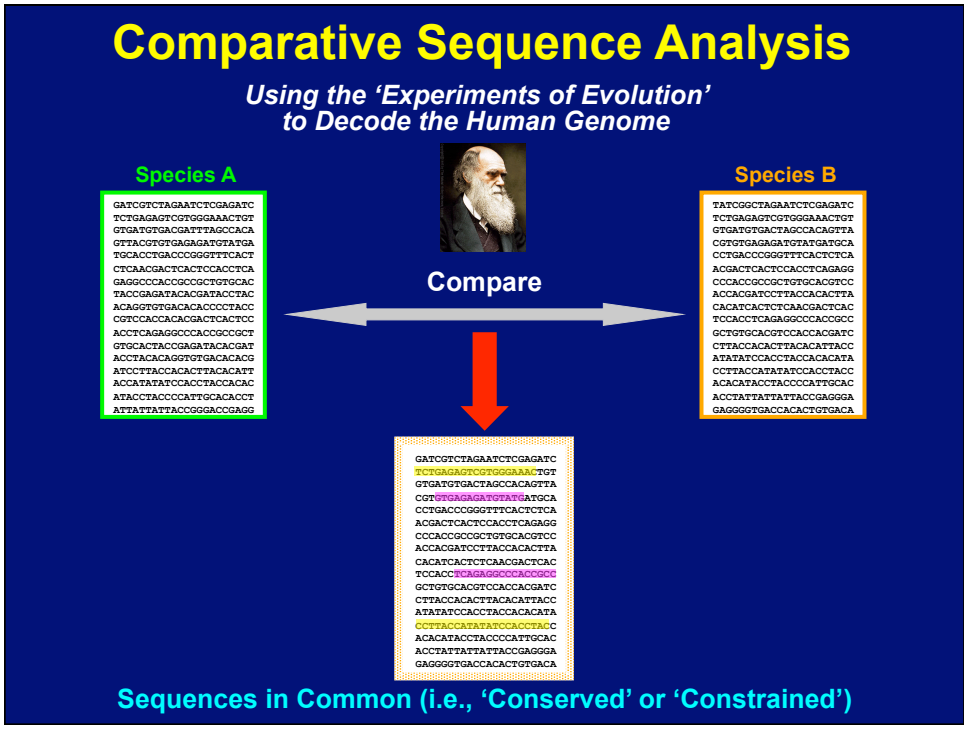


Charles Darwin (1809-1882)

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

–Eric Lander


Inter-Species Sequence Comparisons


Comparative Sequence Analysis
Using the 'Experiments of Evolution'
to Decode the Human Genome
Species A
GATCGTCTAGAATCTCGAGATC
TCTGAGAGTCGTGGGAAACTGT
GTGATGTGACGATTTAGCCACA
GTTACGTGTGAGAGATGTATGA
TGCACCTGACCCGGGTTTCACT
CTCAACGACTCACTCCACCTCA
GAGGCCCACCGCCGCTGTGCAC
TACCGAGATACACGATACCTAC
ACAGGTGTGACACACCCCTACC
CGTCCACCACACGACTCACTCC
ACCTCAGAGGCCCACCGCCGCT
GTGCACTACCGAGATACACGAT
ACCTACACAGGTGTGACACACG
ATCCTTACCACACTTACACATT
ACCATATATCCACCTACCACAC
ATACCTACCCCATTGCACACCT
ATTATTATTACCGGGACCGAGG
Compare
Species B
TATCGGCTAGAATCTCGAGATC
TCTGAGAGTCGTGGGAAACTGT
GTGATGTGACTAGCCACAGTTA
CGTGTGAGAGATGTATGATGCA
CCTGACCCGGGTTTCACTCTCA
ACGACTCACTCCACCTCAGAGG
CCCACCGCCGCTGTGCACGTCC
ACCACGATCCTTACCACACTTA
CACATCACTCTCAACGACTCAC
TCCACCTCAGAGGCCCACCGCC
GCTGTGCACGTCCACCACGATC
CTTACCACACTTACACATTACC
ATATATCCACCTACCACACATA
CCTTACCATATATCCACCTACC
ACACATACCTACCCCATTGCAC
ACCTATTATTATTACCGAGGGA
GAGGGGTGACCACACTGTGACA
GATCGTCTAGAATCTCGAGATC
TCTGAGAGTCGTGGGAAACTGT
GTGATGTGACTAGCCACAGTTA
CGTGTGAGAGATGTATGATGCA
CCTGACCCGGGTTTCACTCTCA
ACGACTCACTCCACCTCAGAGG
CCCACCGCCGCTGTGCACGTCC
ACCACGATCCTTACCACACTTA
CACATCACTCTCAACGACTCAC
TCCACCTCAGAGGCCCACCGCC
GCTGTGCACGTCCACCACGATC
CTTACCACACTTACACATTACC
ATATATCCACCTACCACACATA
CCTTACCATATATCCACCTACC
ACACATACCTACCCCATTGCAC
ACCTATTATTATTACCGAGGGA
GAGGGGTGACCACACTGTGACA
Sequences in Common (i.e., 'Conserved' or 'Constrained')

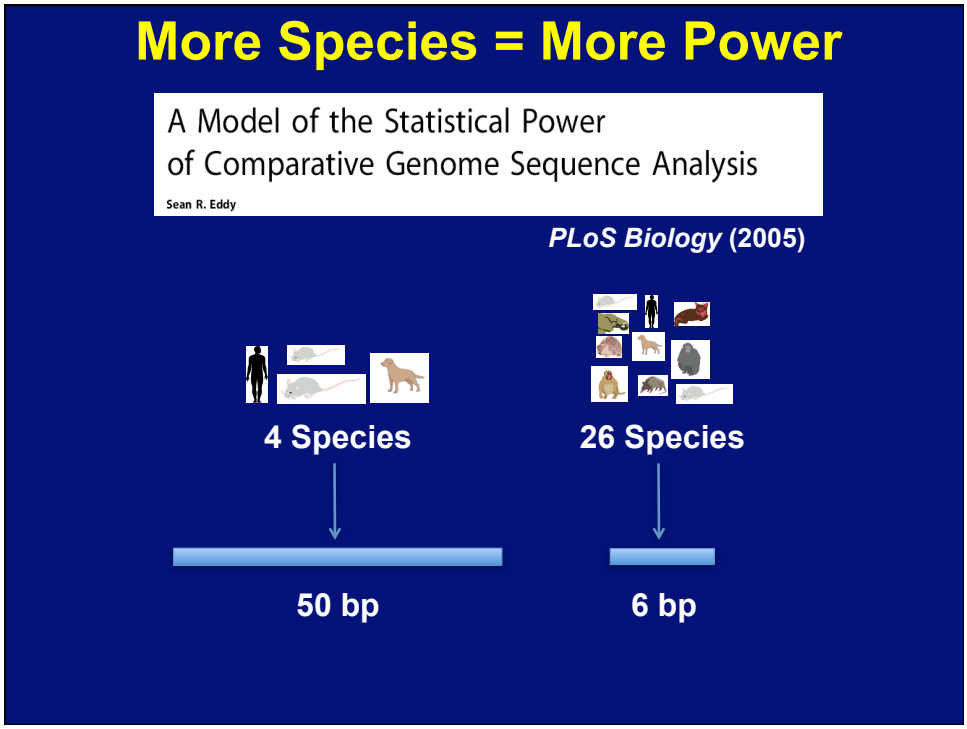
Vertebrate Genome Sequences

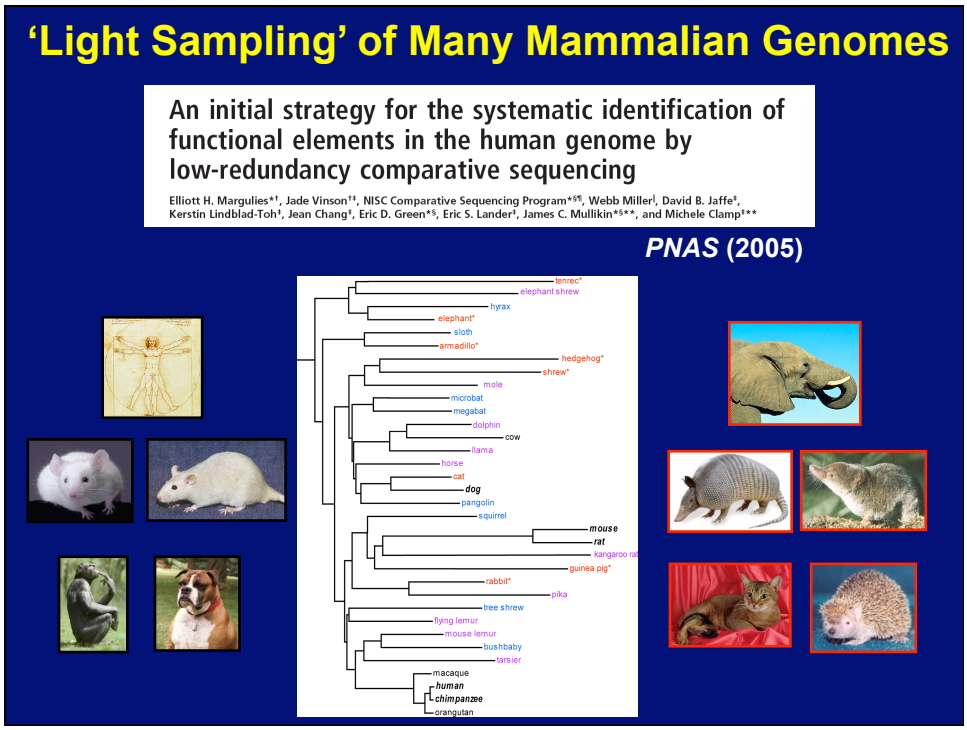
Mouse Rat Chicken Chimpanzee Dog Macaque

Monodelphis Platypus Cow Pufferfish

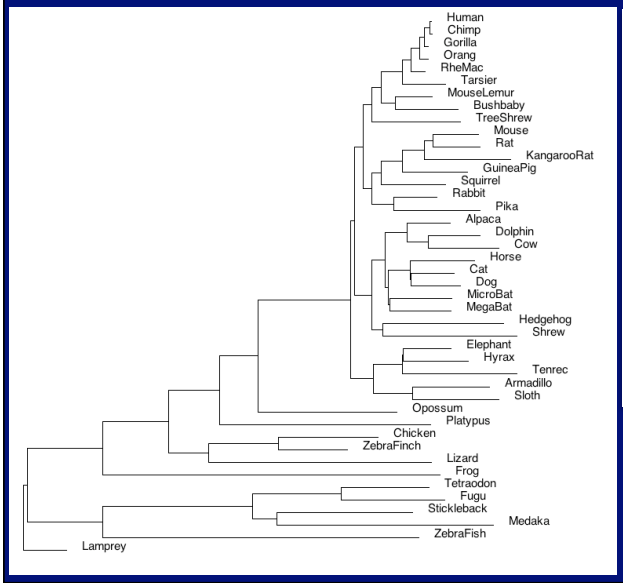
Diverse Landscape of Genome Sequencing

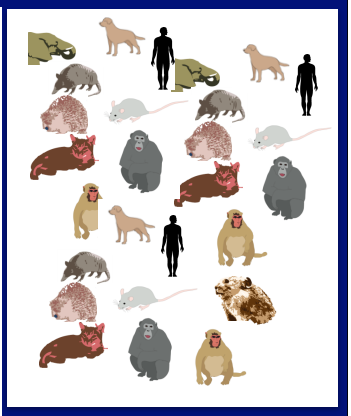
Human
Mouse
Rat
Pufferfish
Zebrafish
Chicken
Chimpanzee
Dog
Cow
Xenopus
Monodelphis
Macaque
Platypus
Marmoset
etc….


More Species = More Power
A Model of the Statistical Power of Comparative Genome Sequence Analysis
Sean R. Eddy
PLoS Biology (2005)
4 Species
26 Species
50 bp
6 bp

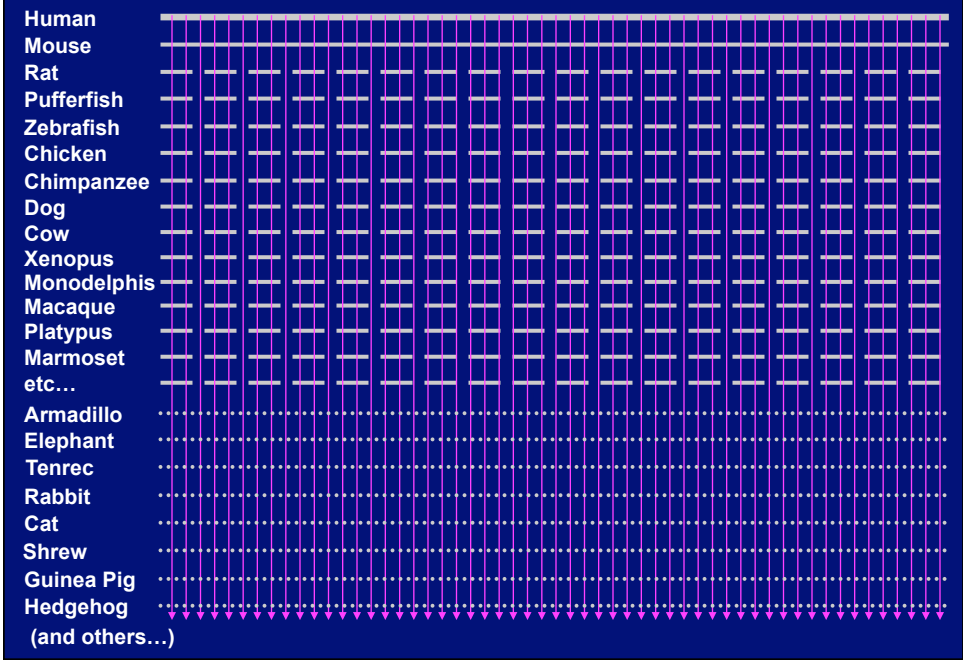

'Light Sampling' of Many Mammalian Genomes
An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing
Elliott H. Margulies*†, Jade Vinson†‡, NISC Comparative Sequencing Program*§¶, Webb Miller||, David B. Jaffe‡, Kerstin Lindblad-Toh‡, Jean Chang‡, Eric D. Green*§, Eric S. Lander‡, James C. Mullikin*§**, and Michele Clamp‡**
PNAS (2005)
tenrec*
elephant shrew
hyrax
elephant*
sloth
armadillo*
hedgehog*
shrew*
mole
microbat
megabat
dolphin
cow
llama
horse
cat
dog
pangolin
squirrel
mouse
rat
kangaroo rat
guinea pig*
rabbit*
pika
tree shrew
flying lemur
mouse lemur
bushbaby
tarsier
macaque
human
chimpanzee
orangutan

22 Additional Mammalian Genome Sequences (@ Low Redundancy)

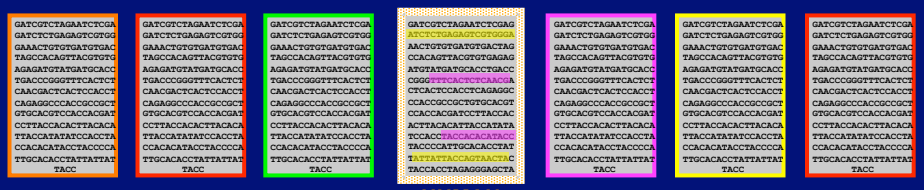




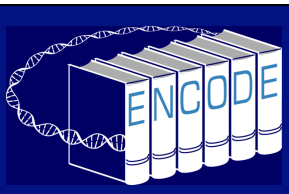
Diverse Landscape of Genome Sequencing



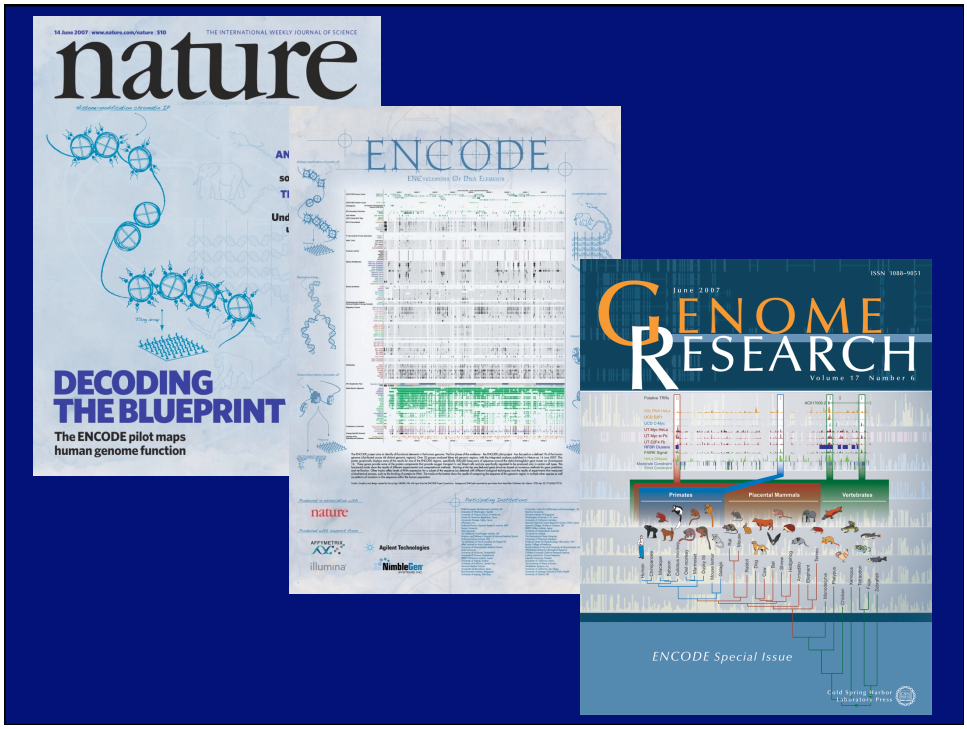
Multi-Species Sequence Comparisons



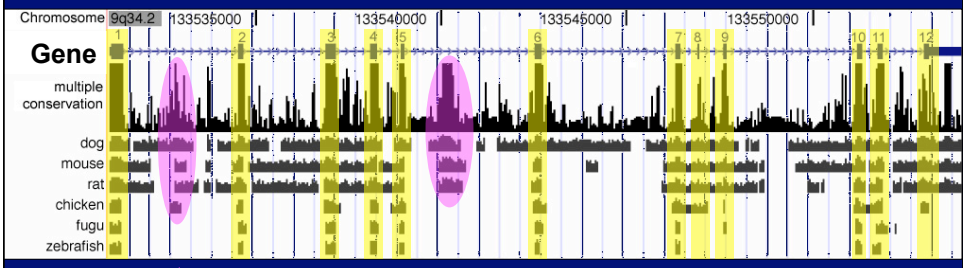
ENCODE Project



- ENCODE: ENCyclopedia Of DNA Elements
- Goal: Compile a *Comprehensive Encyclopedia* of All Functional Elements in the Human Genome
- Initial Pilot Project: 1% of Human Genome
- Apply Multiple, Diverse Approaches to Study and Analyze that 1% in a Consortium Fashion

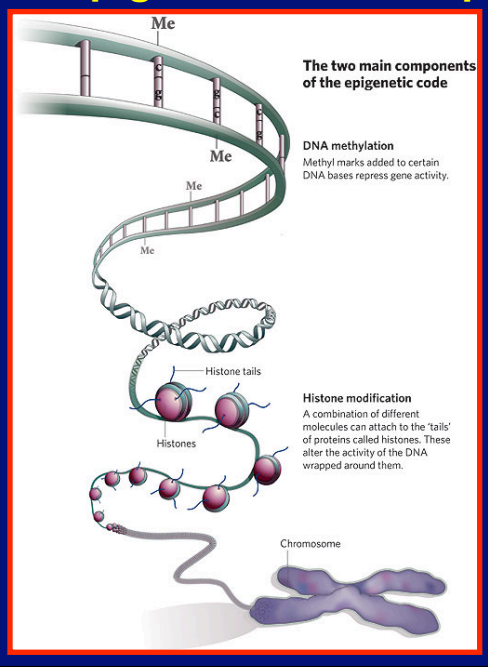


Conserved Coding and Non-Coding Sequences



Augment with 'Functional Annotations'

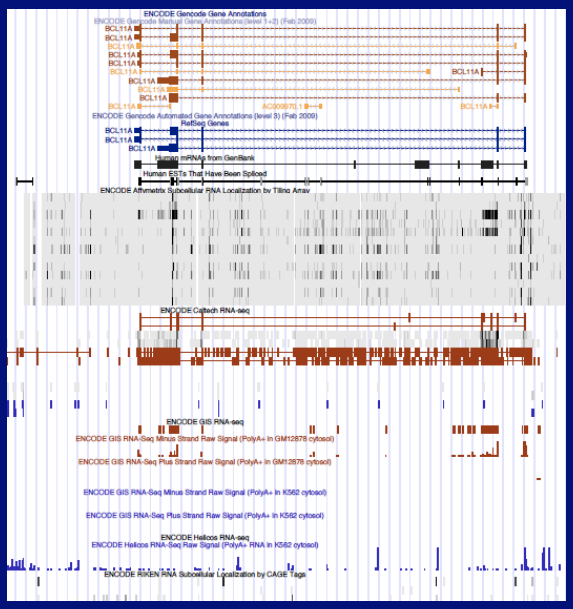
The Epigenetic Landscape

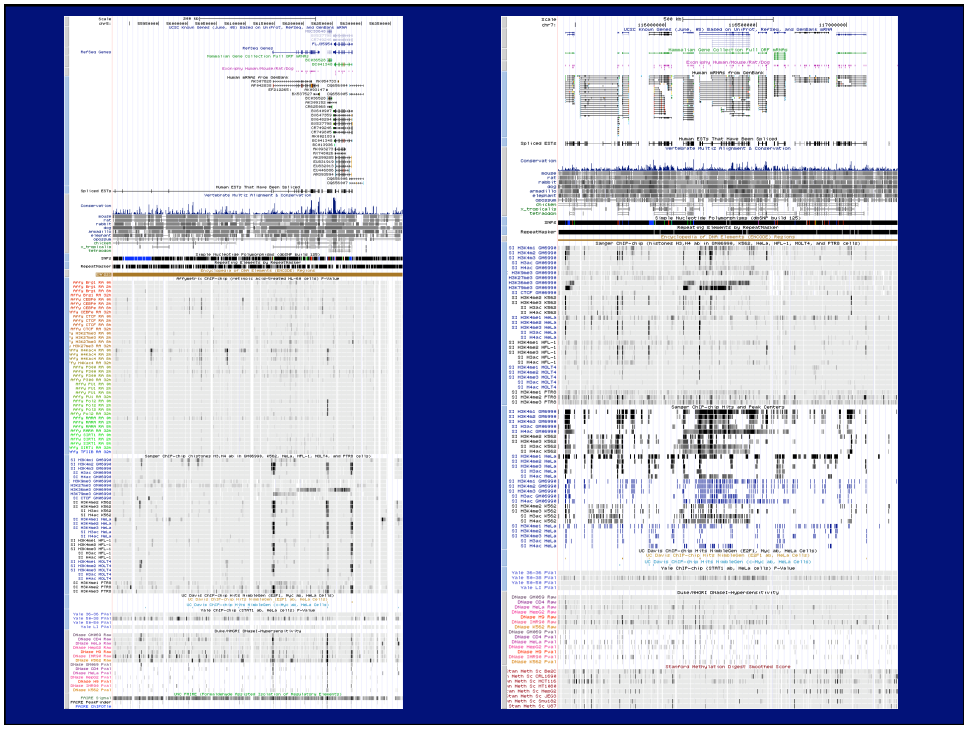


ENCODE: Lots of Data and Data Types

Generated by:

RNA-Seq
RNA-array
TF ChIP-Seq
Histone modif ChIP-Seq
DNaseHS-Seq
FAIRE-Seq
Methyl-Seq
Methyl27-bisulfite
etc.
etc.
etc.

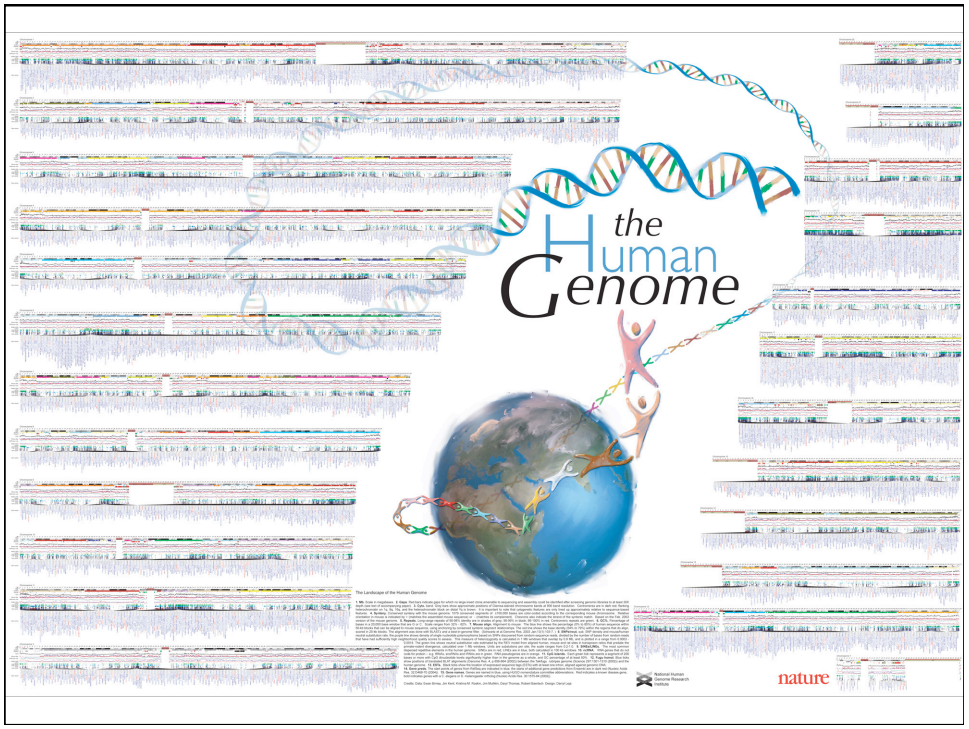




Expanding ENCODE Portfolio

ENCODE

mod ENCODE


the
Human
Genome
nature


Inter-Species Sequence Comparisons

The Genomics of Human Evolution

GENOME 10K©

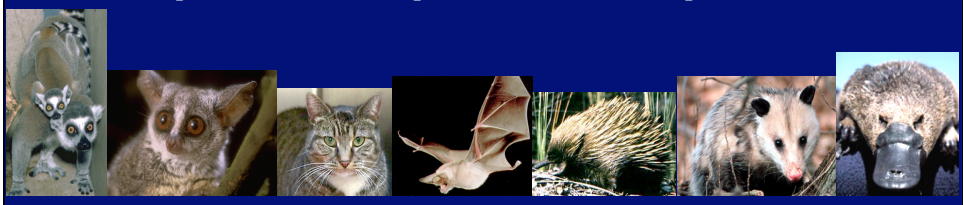
(David Haussler, Stephen O'Brien, & Oliver Ryder)

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

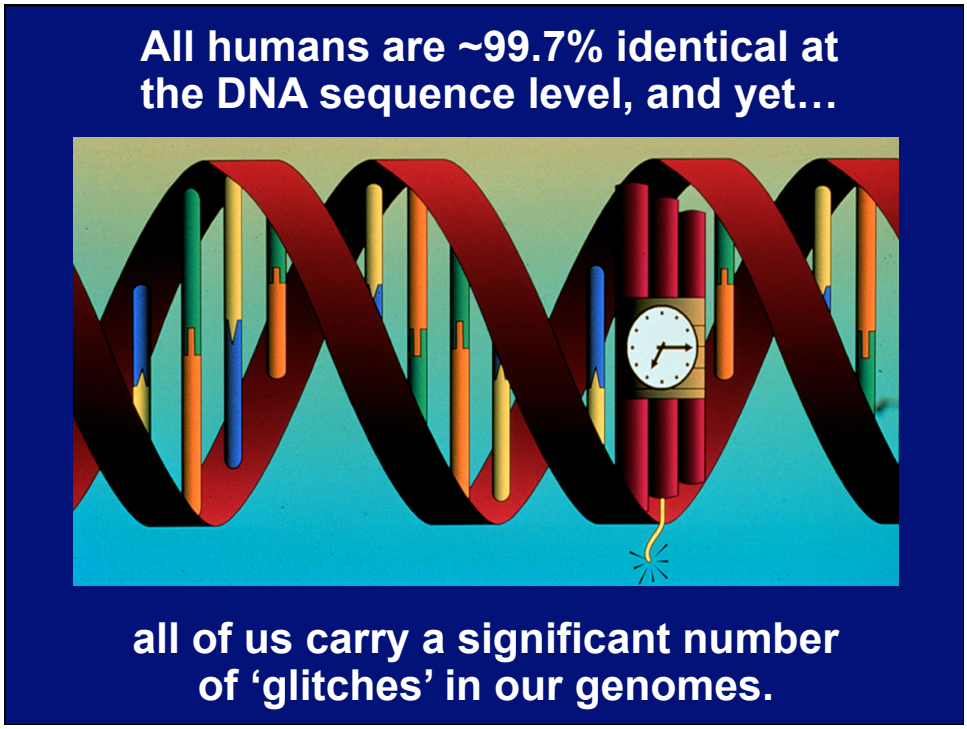
Genome 10K Community of Scientists*

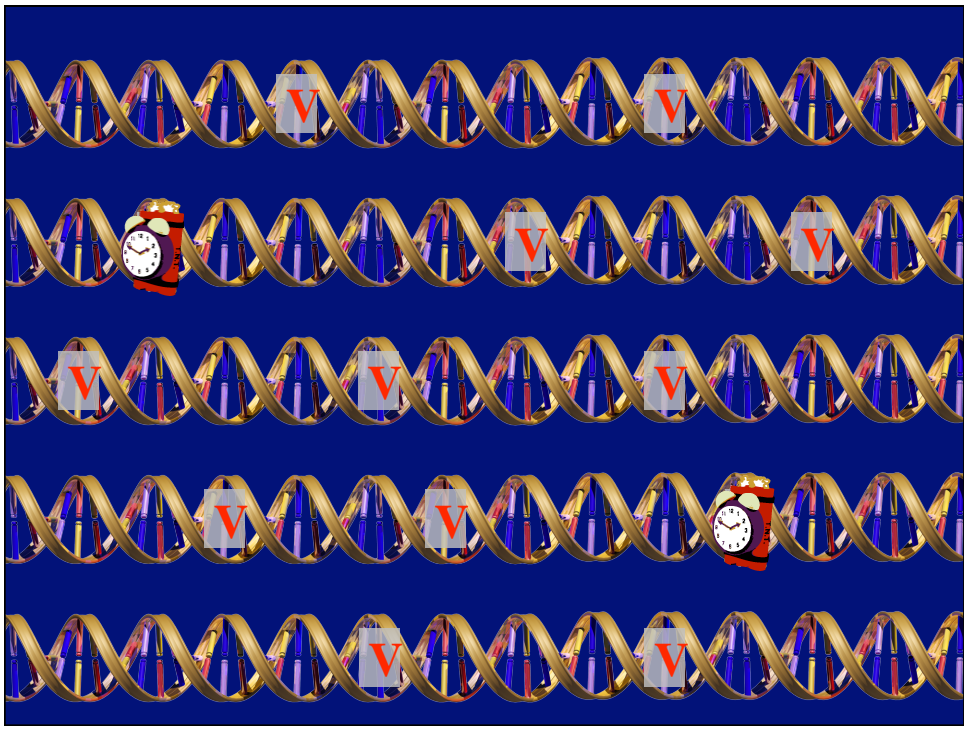
J. Heredity (2009)

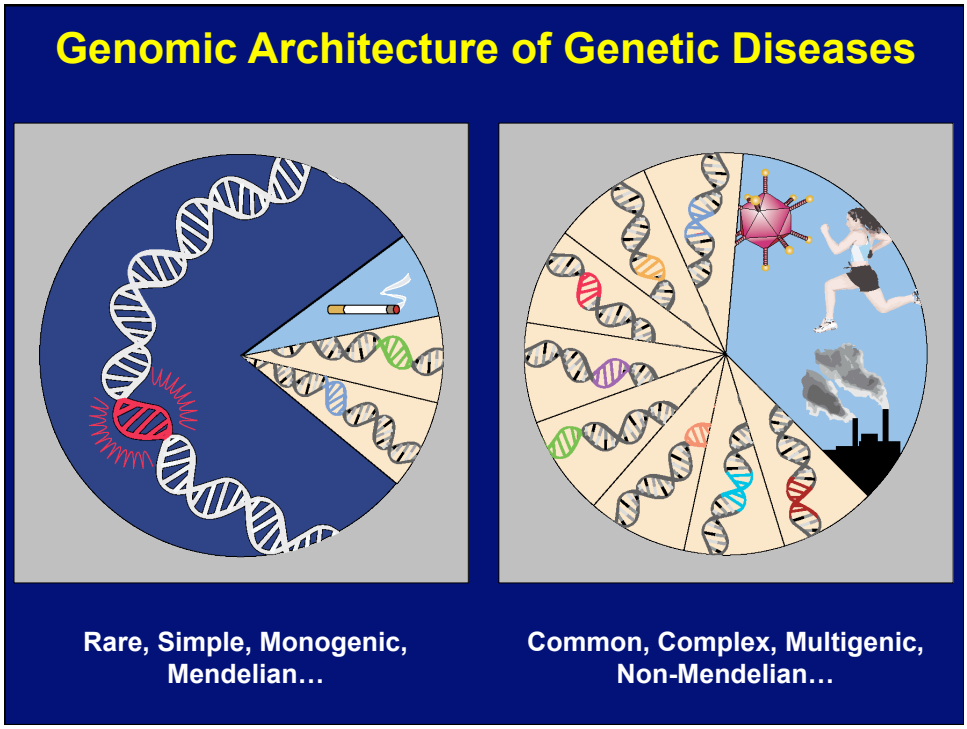
Inter-Species Sequence Comparisons

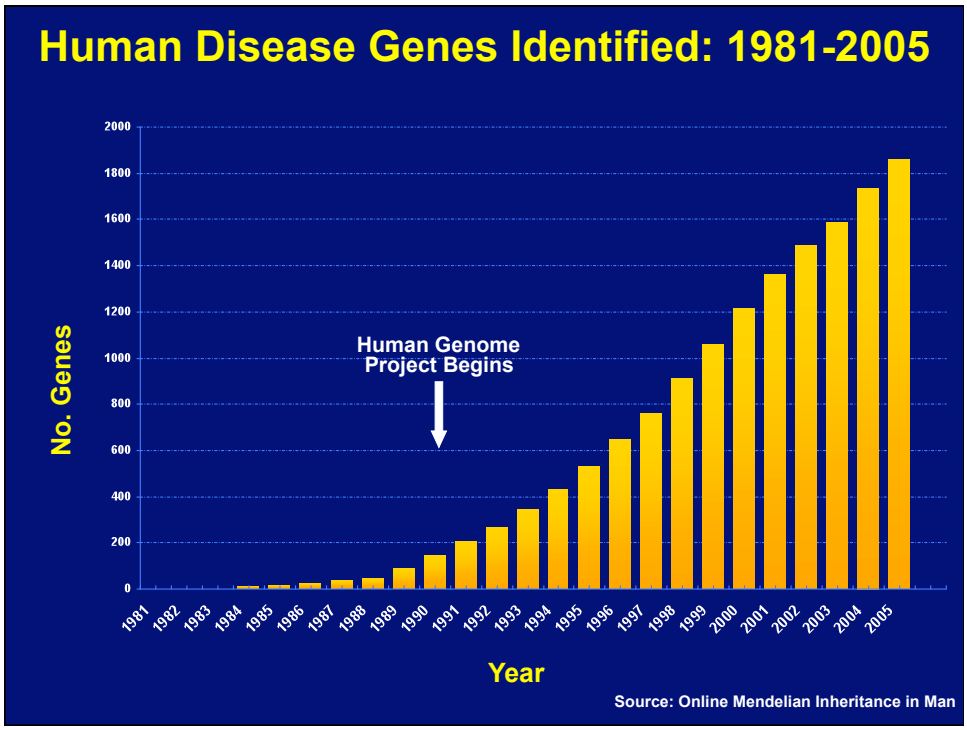


Intra-Species Sequence Comparisons


All humans are ~99.7% identical at the DNA sequence level, and yet…
all of us carry a significant number of 'glitches' in our genomes.


V

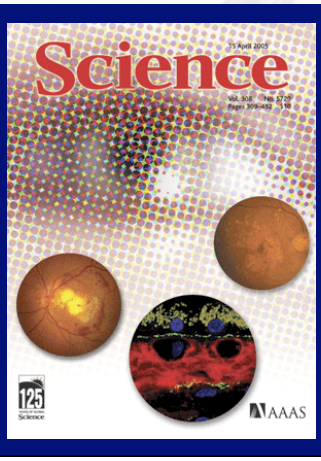

Genomic Architecture of Genetic Diseases
Rare, Simple, Monogenic, Mendelian…
Common, Complex, Multigenic, Non-Mendelian…

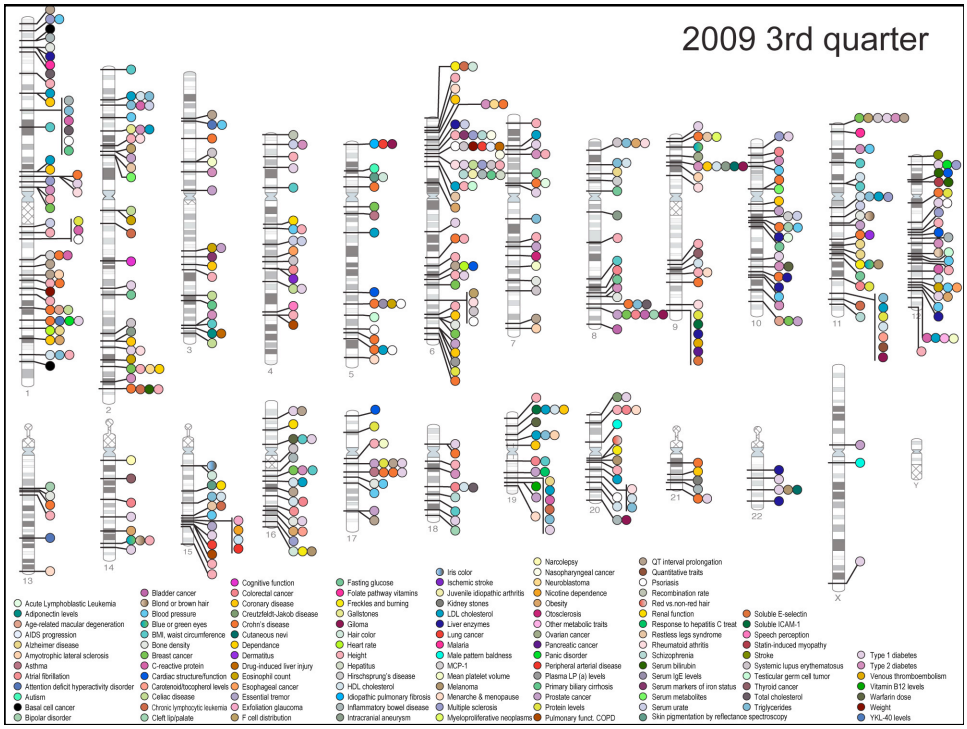

Human Disease Genes Identified: 1981-2005
No. Genes
Year
Human Genome Project Begins
Source: Online Mendelian Inheritance in Man

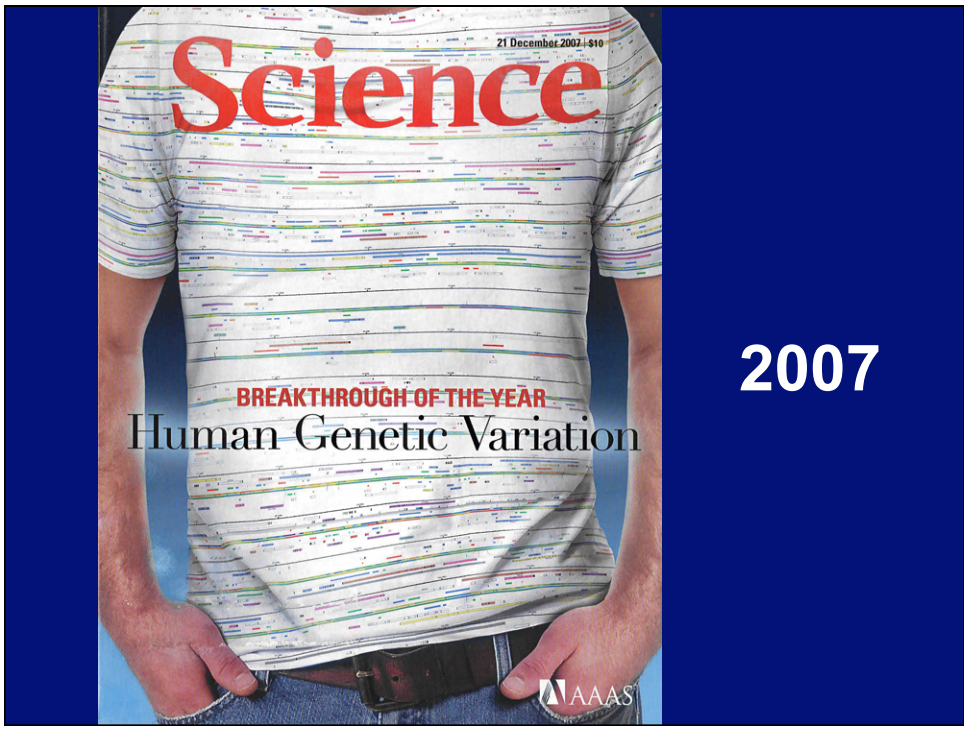
The First HapMap Success Story: Age-Related Macular Degeneration

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†]




2009 3rd quarter
Acute Lymphoblastic Leukemia
Adiponectin levels
Age-related macular degeneration
AIDS progression
Alzheimer disease
Amyotrophic lateral sclerosis
Asthma
Atrial fibrillation
Attention deficit hyperactivity disorder
Autism
Basal cell cancer
Bipolar disorder
Bladder cancer
Blond or brown hair
Blood pressure
Blue or green eyes
BMI, waist circumference
Bone density
Breast cancer
C-reactive protein
Cardiac structure/function
Carotenoid/tocopherol levels
Celiac disease
Chronic lymphocytic leukemia
Cleft lip/palate
Cognitive function
Colorectal cancer
Coronary disease
Creutzfeldt-Jakob disease
Crohn's disease
Cutaneous nevi
Dependance
Dermatitis
Drug-induced liver injury
Eosinophil count
Esophageal cancer
Essential tremor
Exfoliation glaucoma
F cell distribution
Fasting glucose
Folate pathway vitamins
Freckles and burning
Gallstones
Glioma
Hair color
Heart rate
Height
Hepatitis
Hirschsprung's disease
HDL cholesterol
Idiopathic pulmonary fibrosis
Inflammatory bowel disease
Intracranial aneurysm
Iris color
Ischemic stroke
Juvenile idiopathic arthritis
Kidney stones
LDL cholesterol
Liver enzymes
Lung cancer
Malaria
Male pattern baldness
MCP-1
Mean platelet volume
Melanoma
Menarche & menopause
Multiple sclerosis
Myeloproliferative neoplasms
Narcolepsy
Nasopharyngeal cancer
Neuroblastoma
Nicotine dependence
Obesity
Otosclerosis
Other metabolic traits
Ovarian cancer
Pancreatic cancer
Panic disorder
Peripheral arterial disease
Plasma LP (a) levels
Primary biliary cirrhosis
Prostate cancer
Protein levels
Pulmonary funct., COPD
QT interval prolongation
Quantitative traits
Psoriasis
Recombination rate
Red vs.non-red hair
Renal function
Response to hepatitis C treat
Restless legs syndrome
Rheumatoid athritis
Schizophrenia
Serum bilirubin
Serum IgE levels
Serum markers of iron status
Serum metabolites
Serum urate
Skin pigmentation by reflectance spectroscopy
Soluble E-selectin
Soluble ICAM-1
Speech perception
Statin-induced myopathy
Stroke
Systemic lupus erythematosus
Testicular germ cell tumor
Thyroid cancer
Total cholesterol
Triglycerides
Type 1 diabetes
Type 2 diabetes
Venous thromboembolism
Vitamin B12 levels
Warfarin dose
Weight
YKL-40 levels

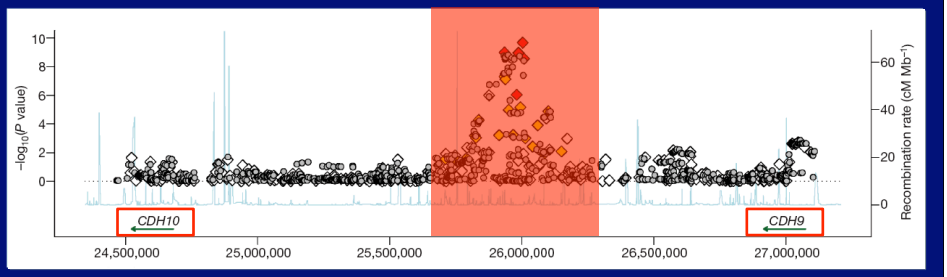

Science
21 December 2007 | $10
BREAKTHROUGH OF THE YEAR
Human Genetic Variation
AAAS
2007

Genetic Association within Intergenic Region

Common genetic variants on 5p14.1 associate with autism spectrum disorders

Kai Wang[1]*, Haitao Zhang[1]*, Deqiong Ma[2]*, Maja Bucan[3], Joseph T. Glessner[1], Brett S. Abrahams[4], Daria Salyakina[2], Marcin Imielinski[1], Jonathan P. Bradfield[1], Patrick M. A. Sleiman[1], Cecilia E. Kim[1], Cuiping Hou[1], Edward Frackelton[1], Rosetta Chiavacci[1], Nagahide Takahashi[5], Takeshi Sakurai[5], Eric Rappaport[6], Clara M. Lajonchere[7], Jeffrey Munson[8], Annette Estes[9], Olena Korvatska[8], Joseph Piven[9], Lisa I. Sonnenblick[4], Ana I. Alvarez Retuerto[4], Edward I. Herman[4], Hongmei Dong[4], Ted Hutman[4], Marian Sigman[4], Sally Ozonoff[10], Ami Klin[11], Thomas Owley[12], John A. Sweeney[12], Camille W. Brune[12], Rita M. Cantor[13], Raphael Bernier[8], John R. Gilbert[2], Michael L. Cuccaro[2], William M. McMahon[14], Judith Miller[14], Matthew W. State[11], Thomas H. Wassink[15], Hilary Coon[14], Susan E. Levy[6], Robert T. Schultz[6], John I. Nurnberger Jr[16], Jonathan L. Haines[17], James S. Sutcliffe[18], Edwin H. Cook[12], Nancy J. Minshew[19], Joseph D. Buxbaum[5,20], Geraldine Dawson[8], Struan F. A. Grant[1,6], Daniel H. Geschwind[4], Margaret A. Pericak-Vance[2], Gerard D. Schellenberg[21] & Hakon Hakonarson[1,6]

Nature (2009)

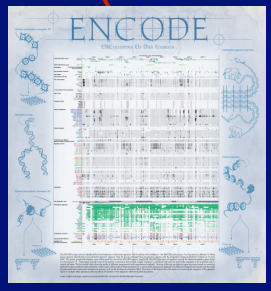


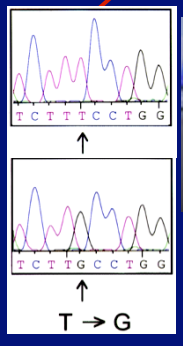
The Pathway to Genomic Medicine

Interpreting the Human Genome Sequence

Implicating Genetic Variants with Human Disease

HGP

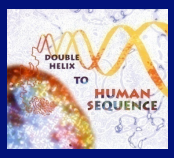




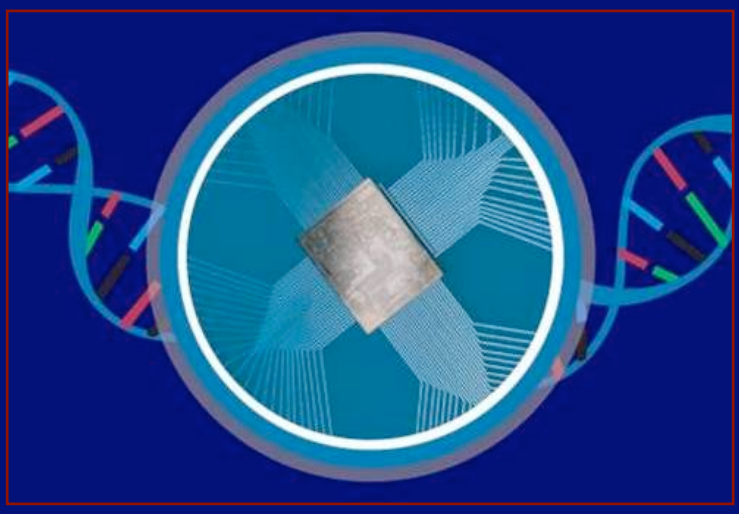
Realization of Genomic Medicine

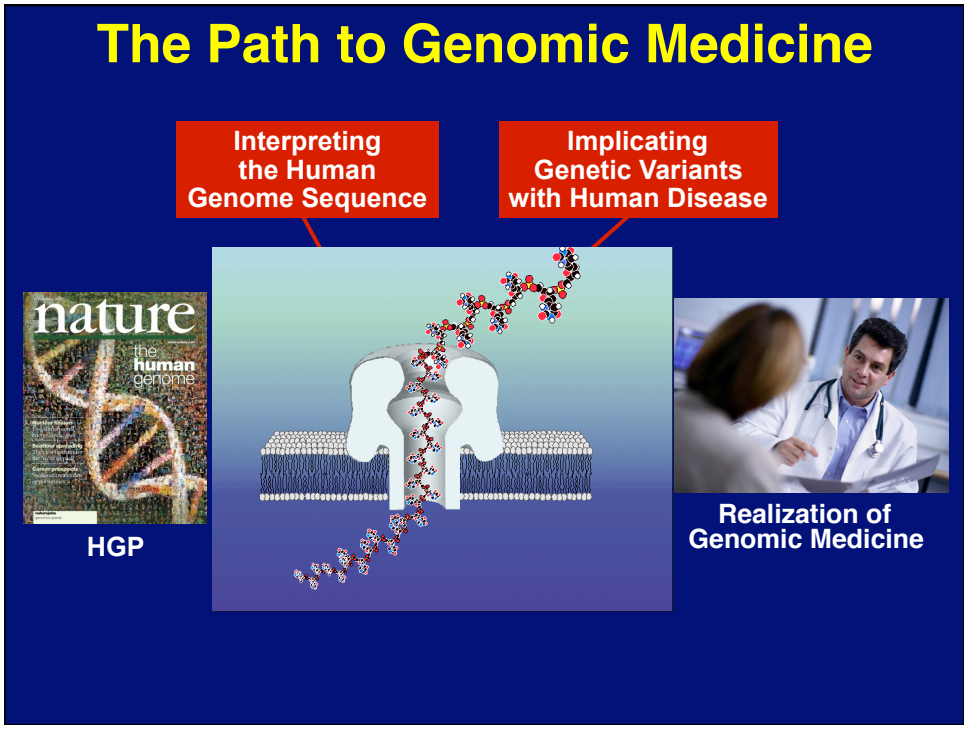
Human Genome Sequence

>$1,000,000,000



~$1,000




The Path to Genomic Medicine
Interpreting the Human Genome Sequence
Implicating Genetic Variants with Human Disease
nature
the human genome
HGP
Realization of Genomic Medicine

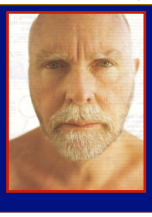

6 November 2008 | www.nature.com/nature | $10
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
nature
Individual genomes from Africa and China
Acute myeloid leukaemia genome
Designer nucleases for gene therapy
Tracing gene flow across Europe
BREAK GLASS WHEN READY
YOUR LIFE IN YOUR HANDS
Instructions for the personal genome age
NATUREJOBS
Mentor awards

Individual Genome Sequences

The Diploid Genome Sequence of an Individual Human

Samuel Levy[17], Granger Sutton[1], Pauline C. Ng[1], Lars Feuk[2], Aaron L. Halpern[1], Brian P. Walenz[1], Nelson Axelrod[1], Jiaqi Huang[1], Ewen F. Kirkness[1], Gennady Denisov[3], Yuan Lin[1], Jeffrey R. MacDonald[2], Andy Wing Chun Pang[2], Mary Shago[2], Timothy B. Stockwell[1], Alexia Tsiamouri[1], Vineet Bafna[3], Vikas Bansal[3], Saul A. Kravitz[1], Dana A. Busam[1], Karen Y. Beeson[1], Tina C. McIntosh[1], Karin A. Remington[1], Josep F. Abril[4], John Gill[1], Jon Borman[1], Yu-Hui Rogers[1], Marvin E. Frazier[1], Stephen W. Scherer[2], Robert L. Strausberg[1], J. Craig Venter[1]

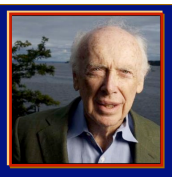
PLoS Biol (2007)



The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler[1]*, Maithreyan Srinivasan[2]*, Michael Egholm[2]*, Yufeng Shen[1]*, Lei Chen[1], Amy McGuire[3], Wen He[2], Yi-Ju Chen[2], Vinod Makhijani[2], G. Thomas Roth[2], Xavier Gomes[2], Karrie Tartaro[2]†, Faheem Niazi[2], Cynthia L. Turcotte[2], Gerard P. Irzyk[2], James R. Lupski[4,5,6], Craig Chinault[4], Xing-zhi Song[1], Yue Liu[1], Ye Yuan[1], Lynne Nazareth[1], Xiang Qin[1], Donna M. Muzny[1], Marcel Margulies[2], George M. Weinstock[1,4], Richard A. Gibbs[1,4] & Jonathan M. Rothberg[2]†

Nature (2008)



Accurate whole human genome sequencing using reversible terminator chemistry

A list of authors and their affiliations appears at the end of the paper

Nature (2008)

The diploid genome sequence of an Asian individual

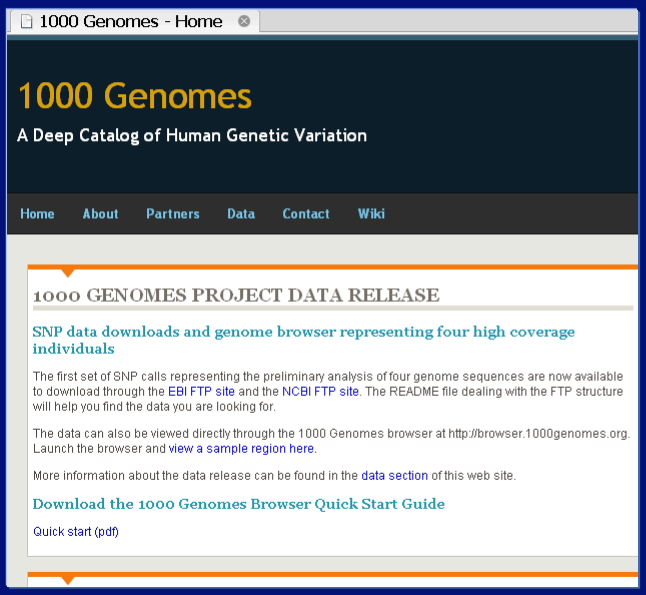
Jun Wang[1,2,3,4]*, Wei Wang[1,3]*, Ruiqiang Li[1,3,4]*, Yingrui Li[1,3,4]*, Geng Tian[1,2], Laurie Goodman[1], Wei Fan[1], Junqing Zhang[1], Jun Li[1], Juanbin Zhang[1], Yiran Guo[1,2], Binxiao Feng[1], Heng Li[1,4], Yao Lu[1], Xiaodong Fang[1], Huiqing Liang[1], Zhenglin Du[1], Dong Li[1], Yiqing Zhao[1,2], Yujie Hu[1,2], Zhenzhen Yang[1], Hancheng Zheng[1], Ines Hellmann[5], Michael Inouye[6], John Pool[6], Xin Yi[1,2], Jing Zhao[1], Jinjie Duan[1], Yan Zhou[1], Junjie Qin[1,2], Lijia Ma[1,2], Guoqing Li[1], Zhentao Yang[1], Guojie Zhang[1,7], Bin Yang[1], Chang Yu[1], Fang Liang[1,2], Wenjie Li[1], Shaochuan Li[1], Dawei Li[1], Peixiang Ni[1], Jue Ruan[1,2], Qibin Li[1,2], Hongmei Zhu[1], Dongyuan Liu[1], Zhike Lu[1], Ning Li[1,2], Guangwu Guo[1,2], Jiangao Zhang[1], Jia Ye[1], Lin Fang[1], Qin Hao[1,2], Quan Chen[1,2], Yu Liang[1,2], Yeyang Su[1,2], A. san[1,2], Cuo Ping[1,2], Shuang Yang[1], Fang Chen[1,2], Li Li[1], Ke Zhou[1], Hongkun Zheng[1,4], Yuanyuan Ren[1], Ling Yang[1], Yang Gao[1,8], Guohua Yang[1,2], Zhuo Li[1], Xiaoli Feng[1], Karsten Kristiansen[9], Gane Ka-Shu Wong[1,10], Rasmus Nielsen[9], Richard Durbin[6], Lars Bolund[1,11], Xiuqing Zhang[1,8], Songgang Li[1,2,3], Huanming Yang[1,2,3] & Jian Wang[1,2,3]

Nature (2008)

A highly annotated whole-genome sequence of a Korean individual

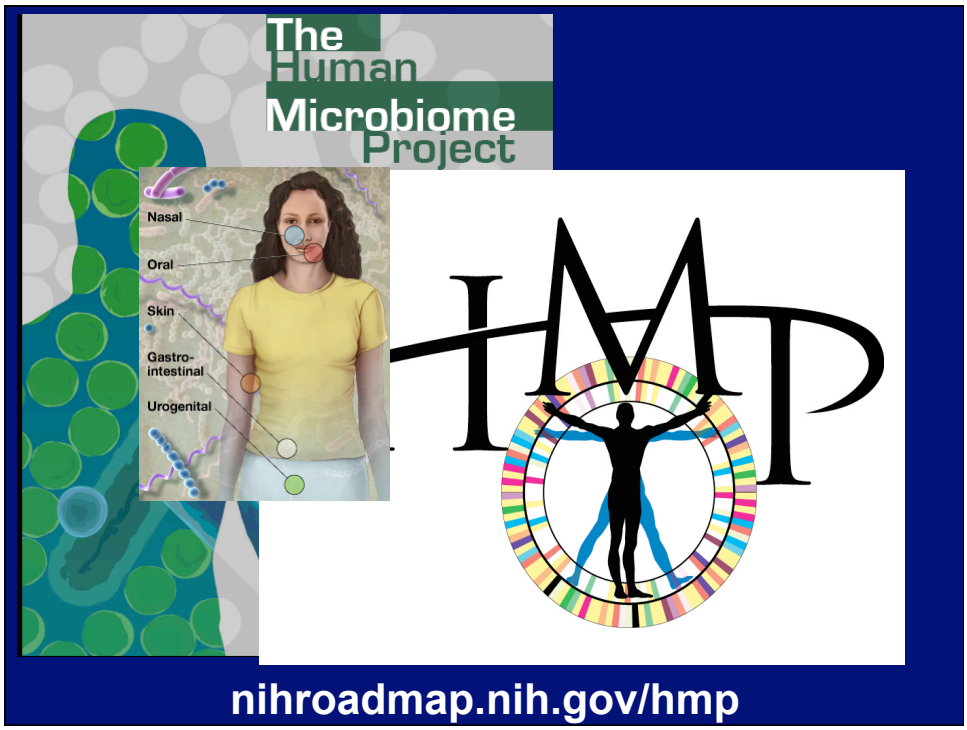
Jong-Il Kim[1,2,4,5]*, Young Seok Ju[1,2]*, Hansoo Park[1,3], Sheehyun Kim[4], Seonwook Lee[4], Jae-Hyuk Yi[1], Joann Mudge[6], Neil A. Miller[6], Dongwan Hong[1], Callum J. Bell[6], Hye-Sun Kim[3], In-Soon Chung[4], Woo-Chung Lee[4], Ji-Sun Lee[4], Seung-Hyun Seo[3], Ji-Young Yun[3], Hyun Nyun Woo[4], Heewook Lee[4], Dongwhan Suh[1,2,3], Seungbok Lee[1,2,3], Hyun-Jin Kim[1,3], Maryam Yavartanoo[1,2], Minhye Kwak[1,2], Ying Zheng[1,2], Mi Kyeong Lee[3], Hyunjun Park[1], Jeong Yeon Kim[3], Omer Gokcumen[2], Ryan E. Mills[2], Alexander Wait Zaranek[6], Joseph Thakuria[6], Xiaodi Wu[6], Ryan W. Kim[6], Jim J. Huntley[6], Shujun Luo[6], Gary P. Schroth[6], Thomas D. Wu[6], HyeRan Kim[3], Kap-Seok Yang[4], Woong-Yang Park[1,2,3], Hyungtae Kim[4], George M. Church[6], Charles Lee[7], Stephen F. Kingsmore[8] & Jeong-Sun Seo[1,2,3,4,5]

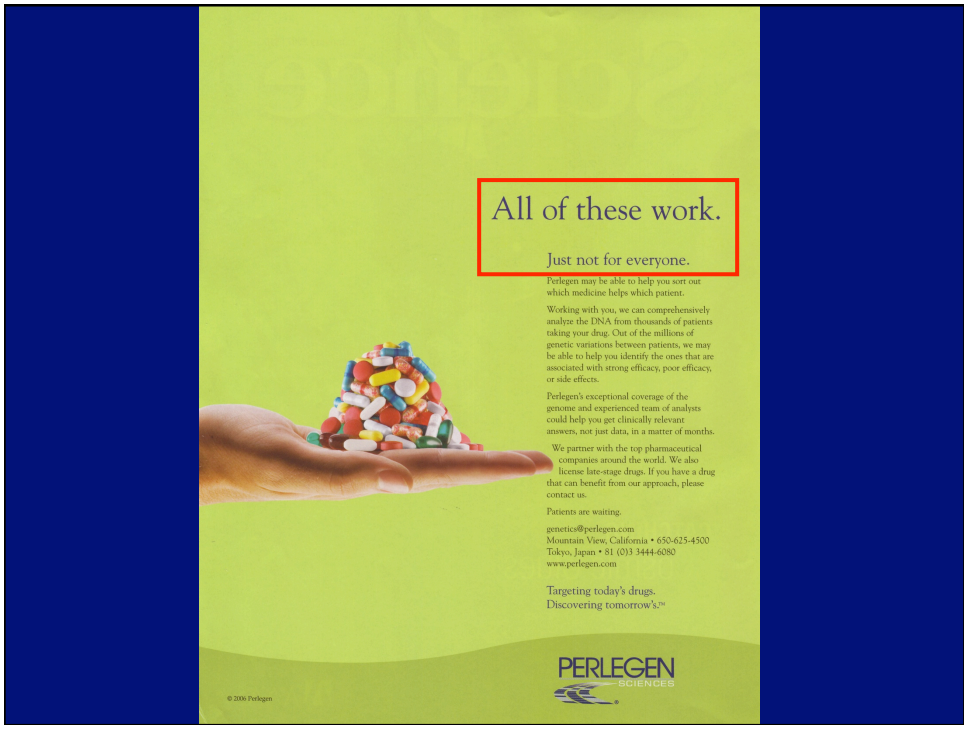
Nature (2009)



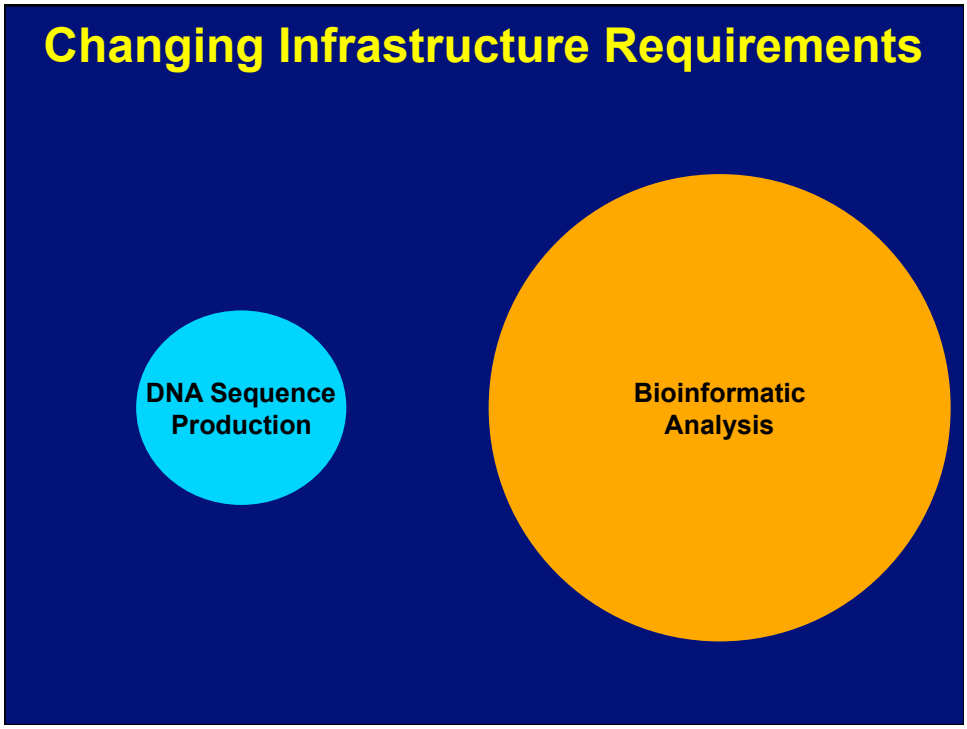
1000genomes.org

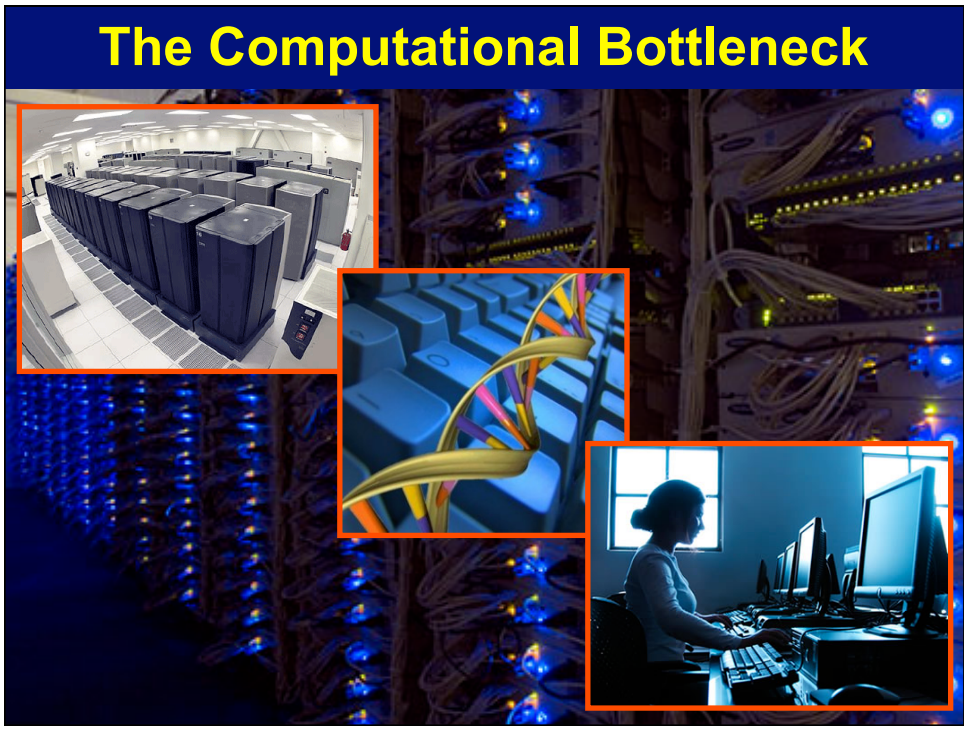

National Cancer Institute
National Human Genome Research Institute
THE CANCER GENOME ATLAS
Search
GO
About TCGA
Program Components
Policies
Media Center
Launch Data Portal
Mission and Goal
The Cancer Genome Atlas (TCGA) is a comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing.
Learn more >>
News from the Pilot Project
NEW* TCGA Network Identifies More Than 6,000 Targets for Sequencing
The Cancer Genome Atlas (TCGA) network has selected more than 6,000 gene and miRNA targets for sequencing that represent both protein-coding genes and microRNAs (miRNAs). While not exhaustive, this list represents genes and
TCGA Data Portal
Access TCGA Data Portal
New* View the list of target genes and miRNAs selected by the TCGA network for sequencing.
NEW Data Available* View Molecular Characterization Data for Ovarian Cancer.
TCGA: How Will It Work?
Click here for more information
cancergenome.nih.gov


The Human Microbiome Project
Nasal
Oral
Skin
Gastro-intestinal
Urogenital
HMP
nihroadmap.nih.gov/hmp


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 (0)3 3444-6080
www.perlegen.com
Targeting today's drugs.
Discovering tomorrow's.™
PERLEGEN
SCIENCES
© 2006 Perlegen


Realities of New DNA Sequencing Technologies…


Changing Infrastructure Requirements
DNA Sequence Production
Bioinformatic Analysis


The Computational Bottleneck

The Genomic Era: *circa* 2010

The Genomic Revolution Continues

The Top 10 Medical Advances of the Decade

By PEGGY PECK and LAUREN COX
ABC News Medical Unit in Collaboration with MedPage Today
Dec. 17, 2009

ABC NEWS
6 comments

Share this story with friends
E-mail RSS Print Font Size A A
Digg submit Facebook Twitter Reddit StumbleUpon More

The first decade of the 21st century brought a number of discoveries, mistakes and medical advances that influenced medicine from the patient's bedside to the medicine cabinet.

Dr. John Sulston, Director of the Sanger Centre near Cambridge takes part in the Human Genome Project. (Ho New/Reuters)

In some cases, these advances changed deeply rooted beliefs in medicine. In others, they opened up possibilities beyond what doctors thought was possible years ago.

ABC News, in collaboration with MedPage Today, reached out to more than 800 specialists for their suggestions. More than 125 experts in various fields and specialties responded. Their suggestions were then sent to the American Association for the History of Medicine, which narrowed the pool down to an authoritative list of 10 medical advances this decade that have had the most impact.

1. Human Genome Discoveries Reach the Bedside

In 2000, scientists in California released a rough draft of the human genome to the public on the Internet. For the first time, the world could download and read the complete set of human genetic information and begin to discover what our roughly 23,000 genes do.

Mapping the human genome was a race involving time and money in the 1990s, with two competitors at the lead -- the government-funded Human Genome Project, which completed its task in 15 years using more than $3 billion in taxpayer money, and a private company, Celera Genomics, which used $100 million and took less than a decade.

Both groups announced drafts of the human genome at a June 26, 2000 press conference with then president Bill Clinton and former British Prime Minister Tony Blair.