

DNA@60: THE BIG DATA DECADE

FEBRUARY 28, 1953

JAMES WATSON AND FRANCIS CRICK DEDUCE THE STRUCTURE OF DNA, RECOGNIZING THE POTENTIAL OF ITS DOUBLE HELICAL STRUCTURE FOR STORING GENETIC INFORMATION—THE BLUEPRINT OF LIFE



APRIL 14, 2003

THE HUMAN GENOME REFERENCE SEQUENCE IS “COMPLETED”

ENCODE (ENCyclopedia Of DNA Elements) project to identify and catalog all functional elements in human DNA is launched by NIH’s National Human Genome Research Institute

2004
FIRST DIAGNOSTIC MICROARRAY system (from Roche) approved for clinical use

2005
NEXT-GENERATION SEQUENCING bursts onto the scene: 454 Life Sciences (now a Roche company) rolls out massively parallel pyrosequencing technology

FIRST HAPLOTYPE MAP OF HUMAN GENOME (PHASE I HAPMAP) locates >1 million single nucleotide polymorphisms (SNPs)

FIRST GENOME-WIDE ASSOCIATION STUDY (GWAS) PUBLISHED

THE CANCER GENOME ATLAS LAUNCHED

2006
REFINED HUMAN GENOME REFERENCE PUBLISHED

CANCER-RELATED MUTATIONS IDENTIFIED in coding genes of breast and colon cancers

SEQUENCING BY SYNTHESIS (SBS) NEXT-GEN PLATFORM launched by Solexa, soon to be Illumina

2007
FIRST COMPLETE PERSONAL GENOME PUBLISHED is that of J. Craig Venter

PHASE II HAPMAP published, characterizing over 3.1 million SNPs from 4 geographically diverse populations

2008
MOORE’S LAW GETS LICKED: Cost of sequencing a complete human genome plummets by a factor of 10

JAMES WATSON’S PERSONAL GENOME PUBLISHED, the first to be sequenced by next-generation technology

SOLID LIGATION-BASED NEXT-GEN SEQUENCING introduced by Life Technologies

FIRST COMPLETE CANCER GENOME PUBLISHED

COMMERCIAL WHOLE-GENOME SEQUENCING OFFERED FOR \$350,000 by Knome

2009
GENOME 10K PROJECT ANNOUNCED: Sequencing of 10,000 vertebrate species, roughly one per genus, to shed light on evolutionary relationships

FIRST MAP OF HUMAN METHYLOME in embryonic stem cells and fibroblasts hints at roles methylation plays in differentiation and gene expression

2010
NEANDERTHAL GENOME SEQUENCED

ROADMAP EPIGENOMICS PROJECT (NIH) issues first major data release, “more than 300 maps of epigenetic changes in 56 cell and tissue types. . . a significant step toward the complete epigenome” (*Nature*)

178 STRAINS OF THE HUMAN MICRO-BIOME have reference genomes cataloged

modENCODE publishes maps of the functional regions in the genomes of *C. elegans* and *D. melanogaster*

NEXT NEXT GEN—LIGHT-FREE SEQUENCING debuts in Life Technologies’ semiconductor-based Ion Torrent

2011
1000TH GWAS PUBLISHED

NEXT NEXT GEN—SINGLE MOLECULE SEQUENCING: SMRT technology from Pacific Biosciences skips the amplification step to detect fluorescently labeled nucleotides as they are added to a single, long strand of DNA

i5K PROJECT LAUNCHED to sequence insect and arthropod genomes

2012
THE ENCODE DATA CORNUCOPIA: 30+ open-access papers, distilled from 15 trillion bytes of raw data, reveal that more than 80 percent of the human genome is biochemically active, and a majority can be transcribed. More than 4 million apparent regulatory regions or “switches” are identified

THE 1000 GENOMES PROJECT documents geographic and functional variation in the genomes of 1,092 people from 14 populations, including a haplotype map of 38 million SNPs

COMMERCIAL WHOLE-GENOME SEQUENCING FOR \$5,000 offered by Complete Genomics

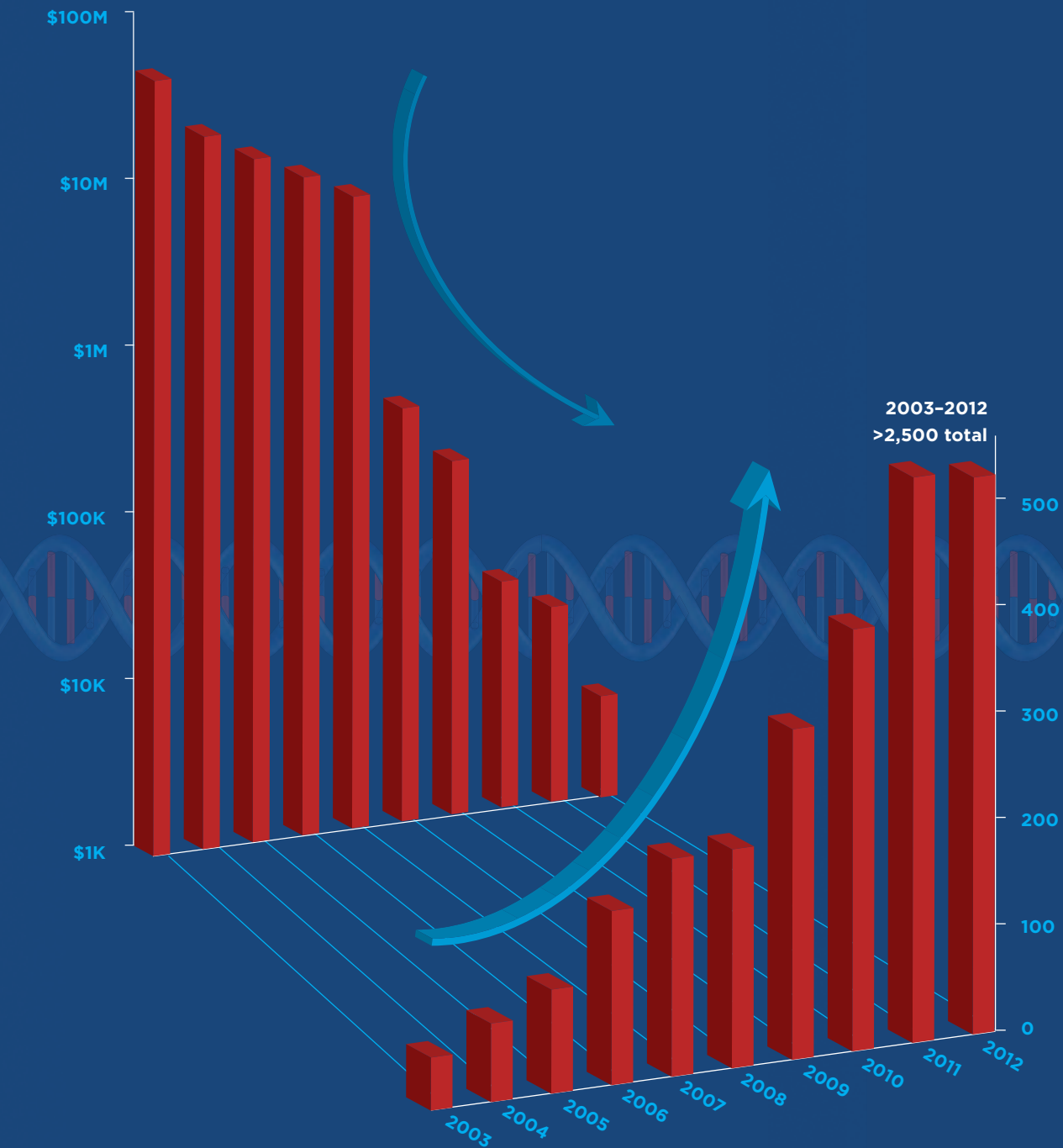
2013. . .

Sponsored by



COST OF SEQUENCING 1 HUMAN GENOME

(Source: www.genome.gov)



COMPLETED GENOMES
(Archaea, Bacteria, and Eukaryotes by NCBI release date)

CELEBRATING 60 YEARS OF THE DOUBLE HELIX