N GENOME

n

o

11

Science Genome Maps

1999
Comparative
Genomics

10

1998
Arabidopsis
thaliana

9

1997
Building
Gene Familles

8

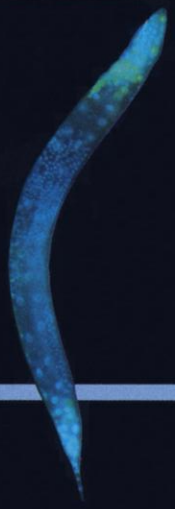
8

1996
The Human
Transcript Map

7

1995
Caenorhabditis
elegans

6

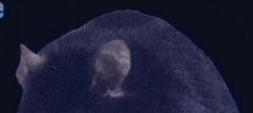


1994
Human
Genetic Map

5

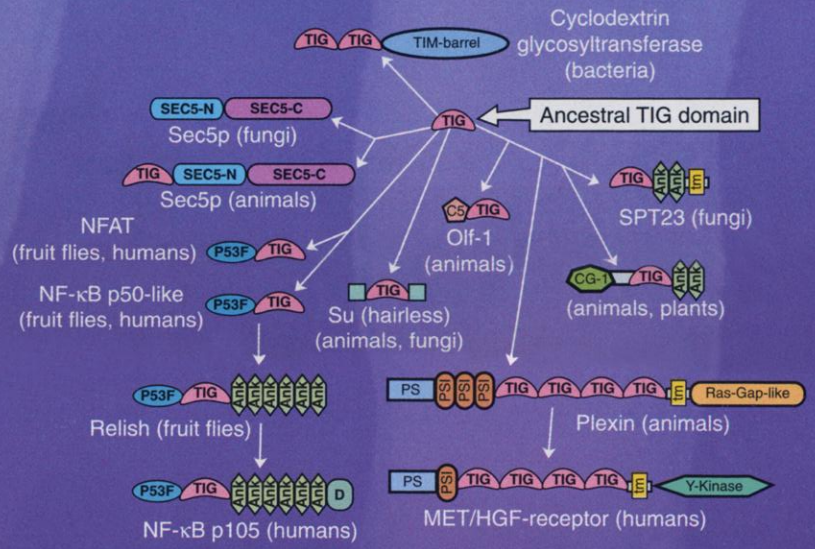
1993
The Mouse

4



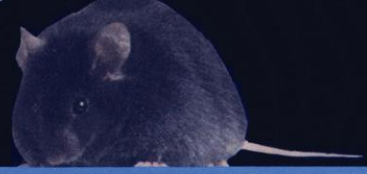
Comparative Genomics

With the ability to determine and compare complete genomic sequences, we will be able to reach the ultimate level of resolution of "comparative anatomy." The extent of similarity among humans, flies, worms, and even bacteria has provided evidence for the commonality of life, yet the relatively small differences between humans and nonhuman primates hold some of the keys to what distinguishes us as human beings. Initial examinations of the predicted protein complement of the human genome indicates that vertebrates have not evolved primarily by addition of new protein domains but through novel ways of putting these modules together to make proteins. It is mostly the architecture rather than the building blocks that distinguishes us from other organisms.



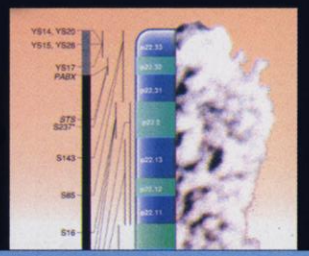
The evolutionary history of the transcription-associated immunoglobulin (TIG) domain, which is present both in transcription regulators such as NF-κB and in the extracellular portions of receptors such as MET/HGF.

4



1992
The X Chromosome

3

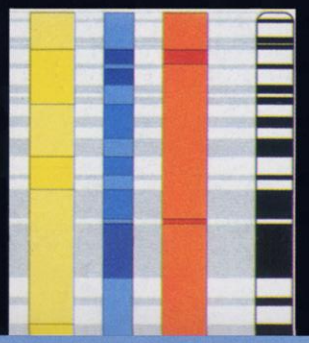


1991
Drosophila melanogaster

2

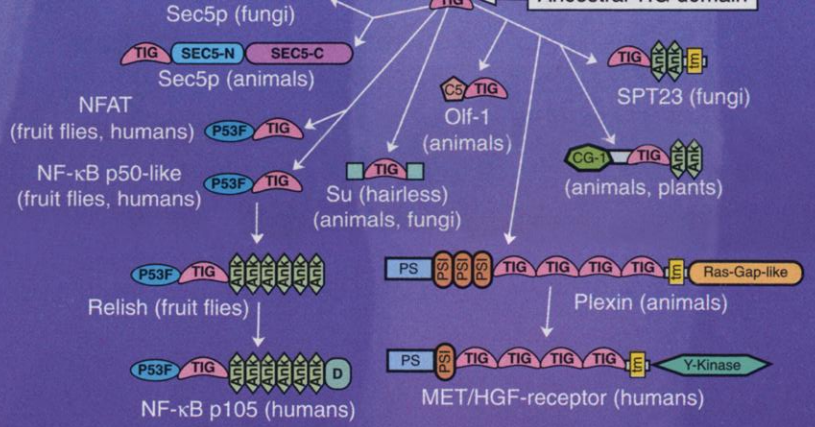
1990
The Human Genome Map

1



Science

addition of new protein domains but through novel ways of putting these modules together to make proteins. It is mostly the architecture rather than the building blocks that distinguishes us from other organisms.



The evolutionary history of the transcription-associated immunoglobulin (TIG) domain, which is present both in transcription regulators such as NF-κB and in the extracellular portions of receptors such as MET/HGF.

Web Resources

Genome Central
http://www.ncbi.nlm.nih.gov/genome/central
http://www.ensembl.org/genome/central

Celera Genomics
http://www.celera.com

National Human Genome Research Institute
http://www.nhgri.nih.gov

ELSI (Ethical, Legal, and Social Implications of Human Genetics Research)
http://www.nhgri.nih.gov/ELSI

Department of Energy Human Genome Program
http://www.ornl.gov/hgmis

Virtual Library: Genetics
http://www.ornl.gov/hgmis/genetics.html

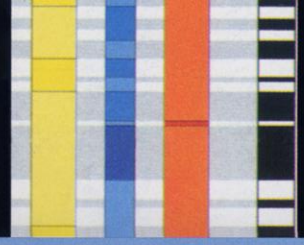
National Center for Biotechnology Information
http://www.ncbi.nlm.nih.gov

European Bioinformatics Institute
http://www.ebi.ac.uk

DNA Data Bank of Japan
http://www.ddbj.nig.ac.jp

1990
The Human Genome Map

1



Science

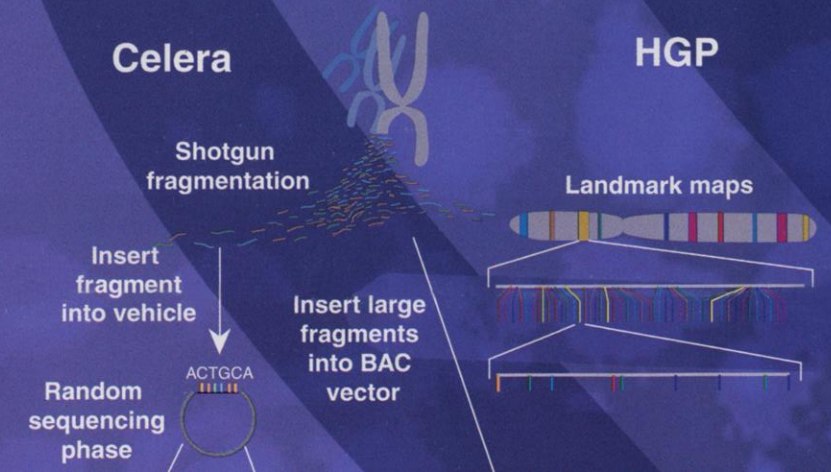
Coordinator: Barbara R. Jasny
Design: Tracy Keaton Drew; C. Faber Smith
Art Direction: C. Faber Smith
Illustration: Cameron Slayden
Production Assistance: Debra Morgenegg
Copyeditor: Harry Jach
Contributors: Mark Adams (Celera sequence strategy flow chart; sequence variation), Celera Genomics, Rockville, MD, USA; Evan Eichler and Julie Horvath (transchromosomal duplication of genomic segments), Case Western Reserve School of Medicine and University Hospitals of Cleveland, Ohio, USA; Todd Golub (cancer microarray), Harvard University, MA, USA; L. Aravind and Eugene Koonin (evolutionary history of a TIG domain), NCBI, NIH, Bethesda, MD, USA; US Department of Energy Human Genome Program, Robert Moyzis (telomere staining), University of California, Irvine, USA.
Reviewers: David Cox, Stanford University, Stanford, CA, USA; Bert Vogelstein, Johns Hopkins University, Baltimore, MD, USA.

The Human G

Two groups have been using different strategies to complete the sequencing of the human genome. Both have now reached goals that were set—the publicly funded effort to produce a working draft of the human genome by a map-based strategy and Celera, to sequence the human genome by the whole-genome shotgun approach. The availability of sequence material obtained through different approaches greatly facilitates the ability of the entire scientific community to interpret the data. This chart provides an introduction to these efforts and some of the revolutionary questions that can be approached with the human genome sequence as a tool.

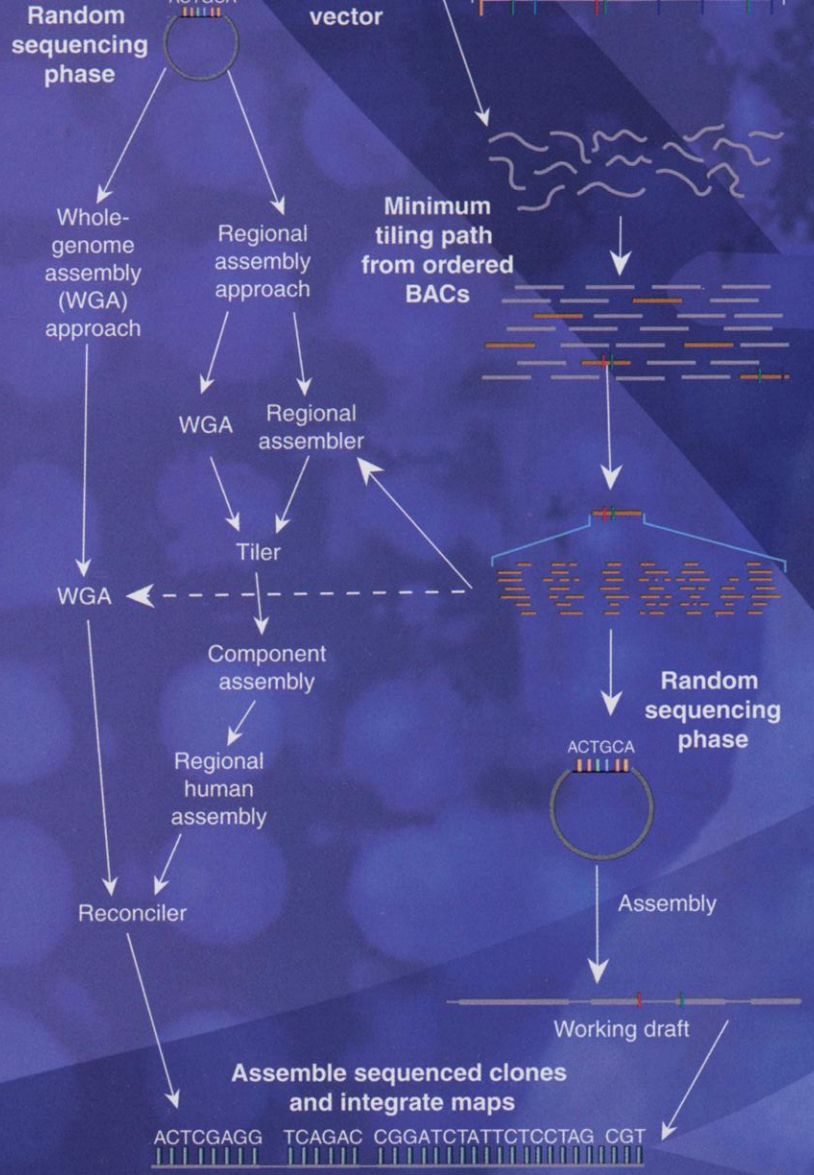
Strategies for Sequencing the Human Genome

The strategy originally established by the publicly funded effort (HGP) was based on localizing bacterial artificial chromosomes (BACs) containing large fragments of human DNA within the framework of a landmark-based physical map. Ideally, sequencing would have been done on a clone-by-clone basis, with



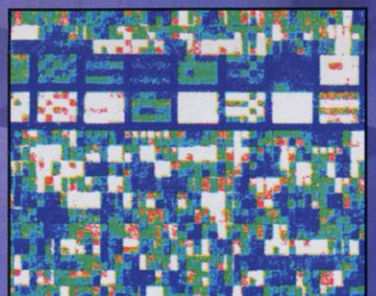
DNA within the framework of a landmark-based physical map. Ideally, sequencing would have been done on a clone-by-clone basis, with clones selected from the minimum BAC tiling path (i.e., a set of BACS that, with minimum overlap, stretched across the whole length of the genome). The working draft, although containing some gaps and ambiguities in order, will be extremely useful in such efforts as identifying disease-associated genes. The idealized strategy of Celera was to avoid the up-front mapping phase by subcloning random fragments of the human genome directly. Sequencing of both ends of fragments in libraries of different sizes facilitated ordering. While saving time and effort at the beginning, the Celera approach would make the assembly process much more dependent on algorithms and computer time.

In their efforts to reach their goals, the idealized strategies evolved into hybrids in which the HGP selected more clones arbitrarily and Celera made use of BAC maps and sequence generated by the HGP.

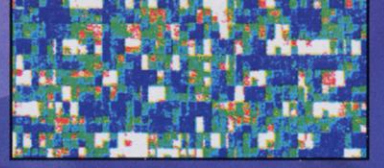


Medical Applications

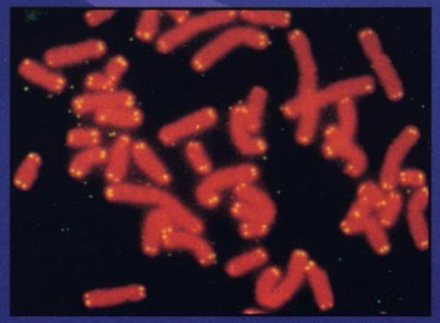
The last decade has seen great strides in the identification of the genetic contribution in diseases resulting from aberrations in single genes. The availability of a complete genome sequence will enormously facilitate the solution of the more difficult problem of identifying the genetic components of the more complex and more common disorders (such as many forms of diabetes, asthma, cancer, and mental illness) in which multiple genetic and environmental factors interact. Using techniques that can measure the expression of thousands of genes at a time



more common disorders (such as many forms of diabetes, asthma, cancer, and mental illness) in which multiple genetic and environmental factors interact. Using techniques that can measure the expression of thousands of genes at a time, scientists are now beginning to to look globally for differences in gene expression that are associated with, for example, the ability to respond to different drugs or pathological states such as cancer.



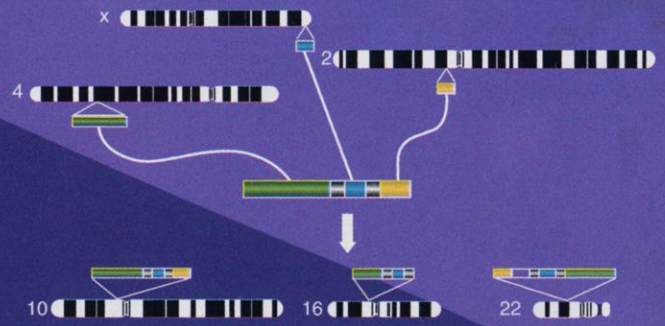
Gene expression in prostate cancer



Analysis of telomeres is leading to insights into chromosome structure and dynamics.

Chromosomal Landscape

The human genome is a complicated composite of many sequence features, such as regions of high-GC and of low-GC content, coding sequences, control elements and other kinds of noncoding functional elements, gene families, repeated sequences of many different types, repeat families, etc. The diversity and distribution of these sequences can shed light on genome evolution. Initial analyses of the human sequence indicate a spectacular range in the density of these features, and their organization provides new clues to the mechanisms that generated the current organization of genomic material. It may even be that we will ultimately discover that there is little or no "junk" in the genome, and that elements in these sequences may have highly evolved functions.



A two-step model for the origin and dispersal of recently duplicated segments in the human genome. Genomic segments of various lengths from different regions of the genome were duplicated to an ancestral pericentromeric region followed by the dispersal of a mosaic genomic segment to multiple pericentromeric regions. Green, 85 kb from 4q24; blue, 9.7 kb from Xq28; yellow, 10 kb from 2p12.

Human Diversity

Within a species, such as the human, there are relatively few differences; for example, the DNA sequences of any two humans differ by only 0.1%. Superficial differences that have had profound social implications, such as race, are not meaningful from a genetic viewpoint as a way of characterizing humanity. Studies of sequence polymorphisms can provide insight into such diverse areas as human migrations and the genetic basis for disease resistance. Genomic analysis has increased the numbers of sequence-based variants available for study, particularly single-nucleotide polymorphisms (SNPs), by orders of magnitude. There are anticipated to be several million common SNPs in the human population, and a significant fraction of those have already been discovered.

Comparative Genomi

With the ability to det
level of resolution of
bacteria has provided
and nonhuman primat
to what distinguishes
Initial examinations o
complement of the hu
vertebrates have not e
addition of new protei
novel ways of putting
make proteins. It is m
rather than the buildi
distinguishes us from

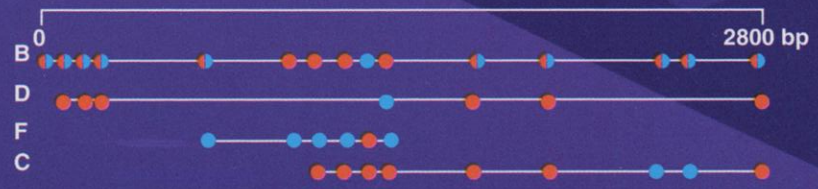
The evolutionary history of t
immunoglobulin (TIG) doma
transcription regulators suc
portions of receptors such a

Genes

Having the complete sequence is only the

Human Diversity

Within a species, such as the human, there are relatively few differences; for example, the DNA sequences of any two humans differ by only 0.1%. Superficial differences that have had profound social implications, such as race, are not meaningful from a genetic viewpoint as a way of characterizing humanity. Studies of sequence polymorphisms can provide insight into such diverse areas as human migrations and the genetic basis for disease resistance. Genomic analysis has increased the numbers of sequence-based variants available for study, particularly single-nucleotide polymorphisms (SNPs), by orders of magnitude. There are anticipated to be several million common SNPs in the human population, and a significant fraction of those have already been discovered.



Sequence variations of four donors, including one Caucasian, one Hispanic, one Chinese, and one African, in a 2800-bp region with 15 SNPs. The blue and orange circles are used to represent biallelic variations. Heterozygous sites in donor B were labeled as half-blue, half-orange circles.

Genes

Having the complete sequence is only the beginning of efforts to identify genes and determine their function. Examination of the sequence suggests that there are far fewer genes in the human genome than the long-expected 100,000—now the estimates indicate numbers closer to 30,000 to 40,000. Gene prediction is currently in a state of flux, with considerable ongoing research aimed at deriving the best algorithms to recognize a gene from the nucleotide sequence. Although various motifs can give clues, laboratory work will also be required to establish function. Of the open reading frames identified by sequence analysis, many have no predicted function at this time.

addition of new protei
novel ways of putting
make proteins. It is m
rather than the buildi
distinguishes us from

The evolutionary history of t
immunoglobulin (TIG) doma
transcription regulators such
portions of receptors such a

Web Resources

Genome Central
http://www.ncbi.nlm.n
http://www.ensembl.or

Celera Genomics
http://www.celera.com

National Human Genome R
http://www.nhgri.nih.g

ELSI (Ethical, Legal, and S
http://www.nhgri.nih.g