HC70A
Winter 2003
Professor Bob Goldberg

Learning Unit #6

How is the Human Genome organized & what changes can occur?

2.5 hr 2 lectures

2/24 / 1.5 hr lecture

2/25 Thur Lecture + 1.5 hr lecture 2/27/03

1. Movie - Genetic Prophesy (25')
2. Genome in Human Cells - Nuclear & Mitochondrial
3. Mitochondrial Genome & Maternal Inheritance
4. Sequencing the Human Genome - Top down vs. Bottom up - public vs. Private Effort
5. Features of Human Genome Sequence / Disease Genes
6. Human vs. Mouse Genome - what makes a "man a man" & "mouse a mouse"?

Stop @ 2:26

Discussion - Summary Last Lecture (60')

7. Organization of the Human Genome
8. Repeats / VNTRs
9. Packaging Human Genome into Chromosomes

Stop

10. Gross Changes in Human Genome & Diseases
11. Mutational Changes in Human Genome / SNPs

Movie - Perfect Baby (30")

HUMAN GENES ARE PRESENT IN TWO COMPARTMENTS -- THE NUCLEUS & THE MITOCHONDRIA

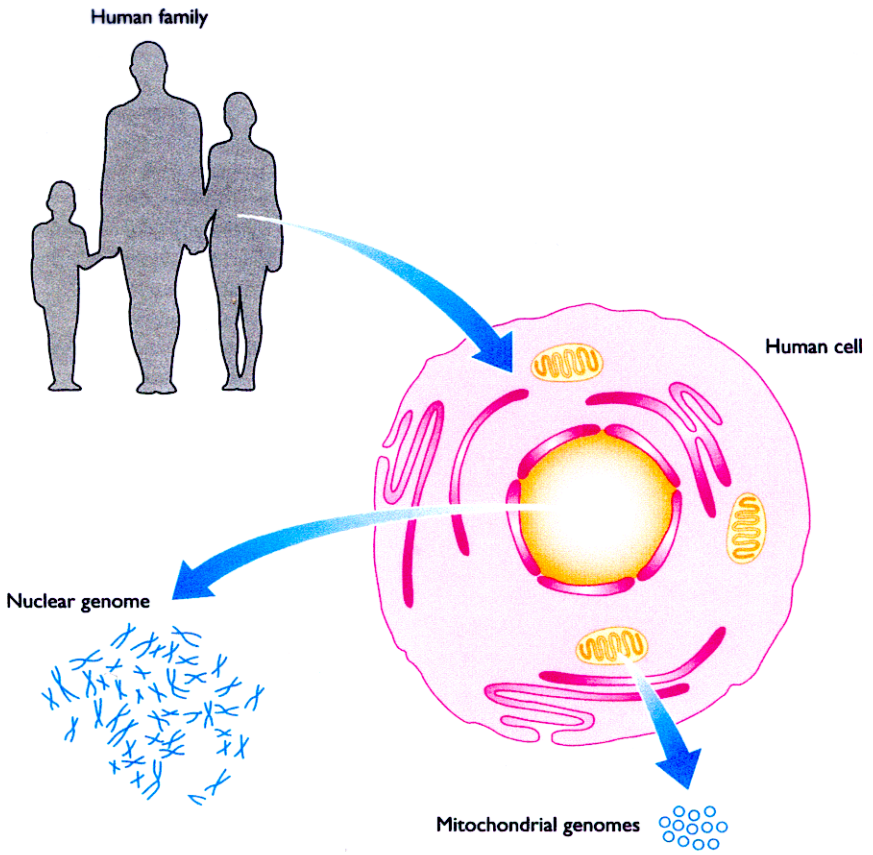
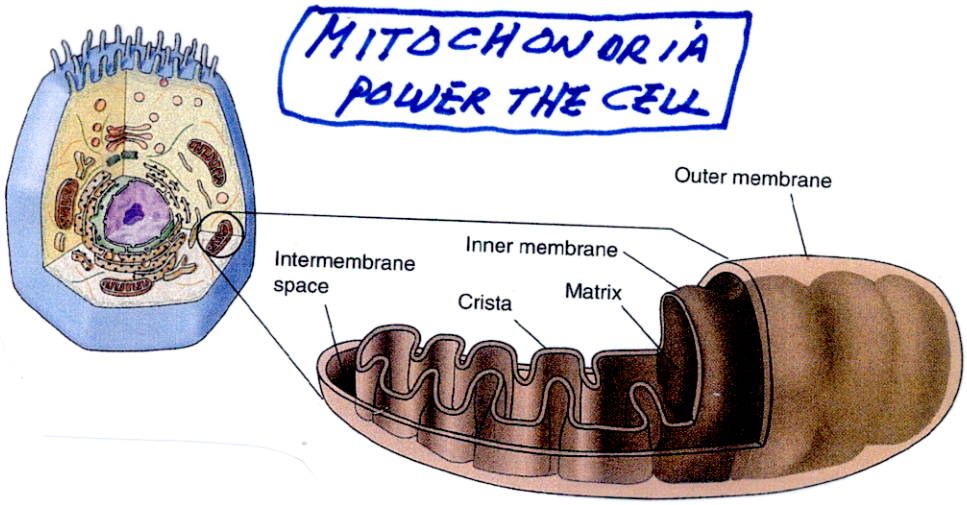


Figure 1.3 The nuclear and mitochondrial components of the human genome.

For more details on the anatomy of the human genome, see Section 6.1.

Genes in BOTH compartments are critical for human development.

MITOCHONDRIA POWER THE CELL



(D)

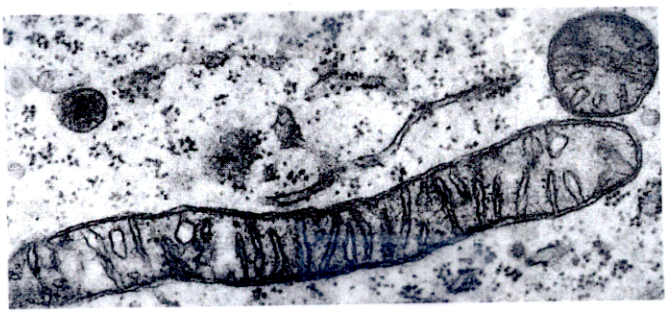


FIGURE 5.21
Mitochondria. (*a*) The inner membrane of a mitochondrion is shaped into folds called cristae, which greatly increase the surface area for oxidative metabolism. (*b*) Mitochondria in cross-section and cut lengthwise (70,000×).

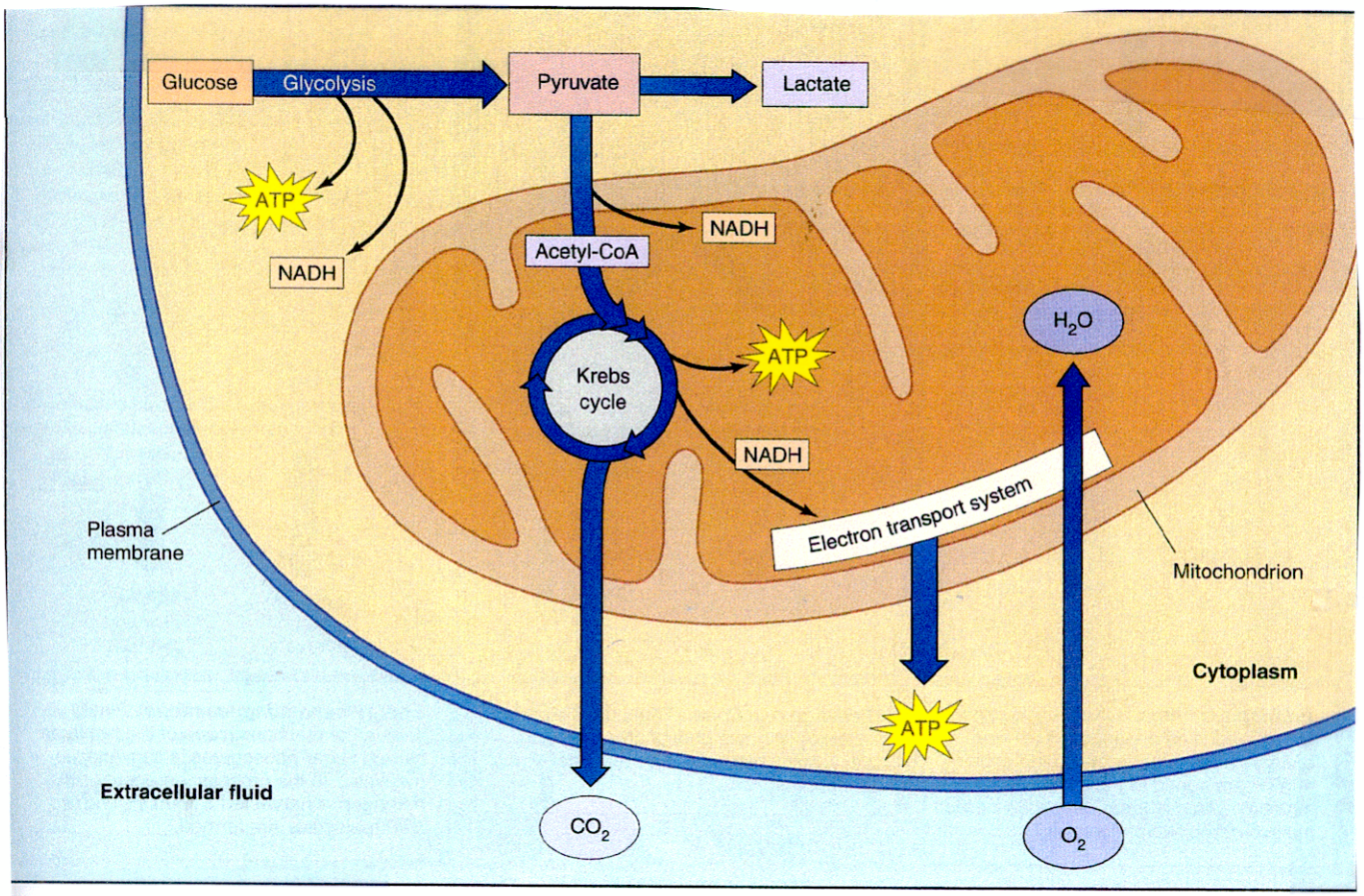


FIGURE 9.6
An overview of aerobic respiration.

What ARE THE CHARACTERISTICS of The HUMAN Nuclear & Mitochondrial Genomes?

Table 7.1: The human nuclear and mitochondrial genomes

	Nuclear genome	Mitochondrial genome
Size	3300 Mb	16.6 kb
No. of different DNA molecules	23 (in XX) or 24 (in XY) cells, all linear	One circular DNA molecule
Total no. of DNA molecules per cell	23 in haploid cells; 46 in diploid cells	Several thousand
Associated protein	Several classes of histone and nonhistone protein	Largely free of protein
Number of genes	~65 000–80 000	37
Gene density	~1/40 kb	1/0.45 kb
Repetitive DNA	Large fraction, see *Figure 7.1*.	Very little
Transcription	The great bulk of genes are transcribed individually	Continuous transcription of multiple genes
Introns	Found in most genes	Absent
% of coding DNA	~3%	~93%
Codon usage	See *Figure 1.22*	See *Figure 1.22*
Recombination	At least once for each pair of homologs at meiosis	Not evident
Inheritance	Mendelian for sequences on X and autosomes; paternal for sequences on Y	Exclusively maternal

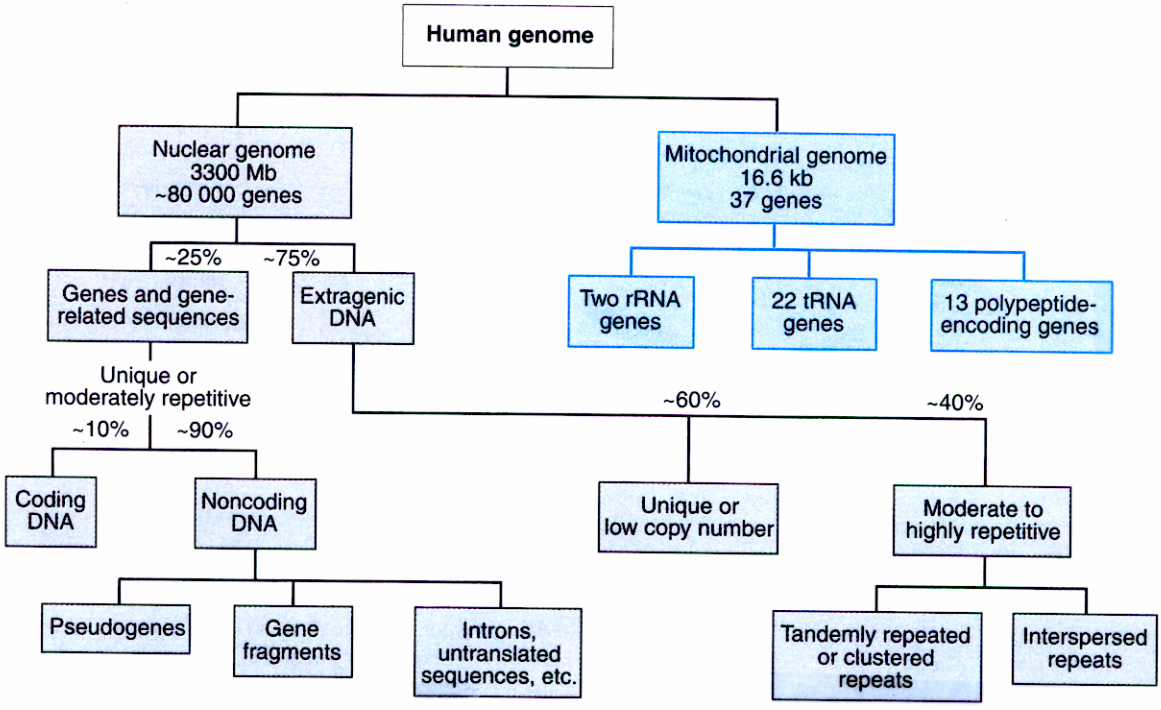


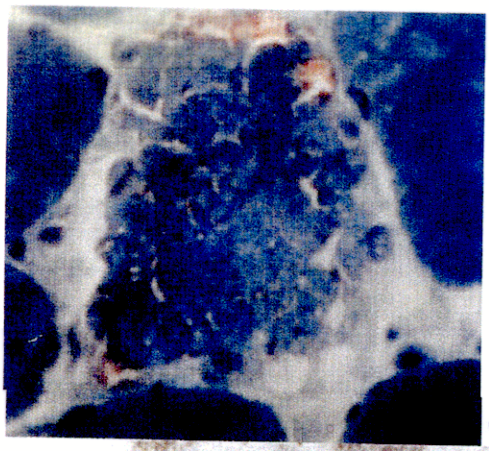
Figure 7.1: Organization of the human genome.

Several Mitochondrial Diseases Occur in Humans

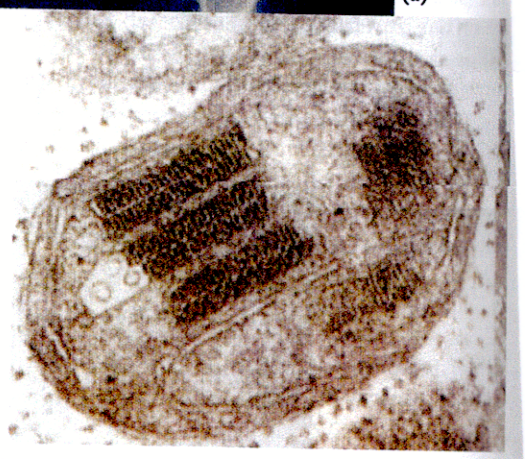
In order for a human disorder to be attributable to genetically altered mitochondria, several criteria must be met.

1. Inheritance must exhibit a maternal rather than a Mendelian pattern.
2. The disorder must reflect a deficiency in the bioenergetic function of the organelle.
3. There must be a specific genetic mutation in one of the mitochondrial genes.

Thus far, several cases are known to demonstrate these characteristics. For example, **myoclonic epilepsy and ragged red fiber disease (MERRF)** demonstrates a pattern of inheritance consistent with maternal inheritance. Only offspring of affected mothers inherit the disorder; the offspring of affected fathers are all normal. Individuals with this rare disorder express deafness, dementia, and seizures. Both muscle fibers and mitochondria are affected; the aberrant mitochondria characterize what are described as ragged red fibers (**RRFs**) of skeletal muscle (Figure 8.5). Analysis of mtDNA has revealed a mutation in one of the mitochondrial genes encoding a transfer RNA. This genetic alteration apparently interferes with translation within the organelle, which in turn leads to the various manifestations of the disorder.



(a)



(b)

FIGURE 8.5 Mitochondrial myopathy in skeletal muscle cells of a patient with MERFF. Part (a) shows a ragged red fiber with abnormal mitochondria. Part (b) shows an abnormal mitochondrion revealing paracrystalline arrays within it.

A second disorder, **Leber's hereditary optic neuropathy (LHON)**, also exhibits maternal inheritance as well as mtDNA lesions. The disorder is characterized by sudden bilateral blindness. The average age of vision loss is 27, but onset is quite variable. Four mutations have been identified, all of which disrupt normal oxidative phosphorylation. Over 50 percent of cases are due to a mutation at a specific position in the mitochondrial gene encoding a subunit of NADH dehydrogenase so that the amino acid arginine is converted to histidine. This mutation is transmitted to all maternal offspring. It is interesting to note that in many instances of LHON, there is no family history; a significant number of cases appear to result from "new" mutations.

Individuals severely affected by a third disorder, **Kearns-Sayre syndrome (KSS)**, lose their vision, undergo hearing loss, and display heart conditions. The genetic basis of KSS involves deletions at various positions within mtDNA. Many KSS patients are symptom-free as children but display progressive symptoms as adults. The proportion of mtDNAs that reveal deletions increases as the severity of symptoms increases.

The study of hereditary mitochondrial-based disorders provides insights into the importance and genetic basis of this organelle during normal development, as well as the relationship between mitochondrial function and neuromuscular disorders. Such study has also suggested a hypothesis for aging based on the progressive accumulation of mtDNA mutations and the accompanying loss of mitochondrial function.

The Mitochondrial Genome is A SMALL CIRCLE containing only 37 Genes

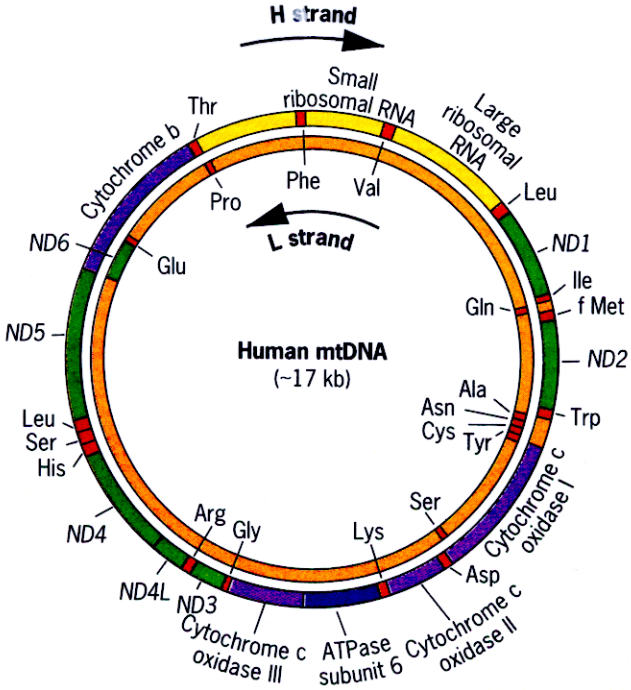
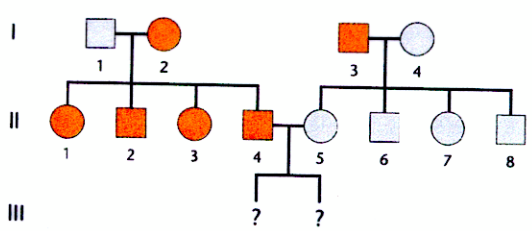


Figure 19.14 Map of human mtDNA showing the pattern of transcription. Genes on the inner circle are transcribed from the L strand of the DNA, whereas genes on the outer circle are transcribed from the H strand of the DNA. Arrows show the direction of transcription. *ND1–6* are genes encoding subunits of the enzyme NADH reductase; the tRNA genes in the mtDNA are indicated by abbreviations for the amino acids.

Mitochondrial Genes Are Inherited MATERNALLY



← PASSED DIRECTLY FROM Mother to Children

MITOCHONDRIAL DNA CAN BE USED FOR FINGERPRINT ANALYSIS

HUMAN GENETICS SIDELIGHT

Using Mitochondrial DNA to Study Human Evolution

In biology few subjects are more fascinating than that of human evolution. Who are we? Where did we come from? Where are we going? Before the advent of molecular biology, the study of human evolution depended on the analysis of rare fossils—fragments of bone, a few teeth, an occasional weapon or tool. Today, human evolution can be studied by comparing DNA sequences. Each DNA sequence is descended from a sequence that was present in an ancestral organism. Thus, the DNA sequences that we find today are, in effect, living fossils—records of ancient DNA that has been transmitted through many generations to organisms currently alive. Because mutations may have occurred during this time, a modern DNA sequence is not likely to be an exact replica of its ancestor. However, by comparing modern DNA sequences, we can sometimes reconstruct features of the evolutionary process that produced them.

Some of the most insightful studies of human evolution have involved the analysis of mitochondrial DNA. There are two reasons why mtDNA is so useful: (1) it evolves faster than nuclear DNA, and (2) it is transmitted exclusively through the female. The rapidity of mtDNA evolution allows a scientist to detect significant genetic changes over a relatively short period of time (in evolutionary terms), and the strict maternal transmission of mtDNA allows a researcher to trace modern DNA sequences back to a common female ancestor.

Pioneering studies of human mtDNA were carried out in the 1980s by Allan Wilson, Rebecca Cann, Mark Stoneking, and their colleagues. These studies established that there is relatively little variation in the mtDNA from different human populations and that the greatest variation is found in the mtDNA from populations in Africa. Given the rate at which mtDNA is known to evolve, these discoveries suggested that modern human beings originated rather recently, probably within the last 200,000 years, and probably in Africa. Although these conclusions were initially controversial, later work has reinforced them.[1] Wilson's laboratory collected mtDNA samples from more than 200 individuals representing many different racial and ethnic groups. The mtDNA sequences in this collection were determined biochemically and then analyzed by a computer program that arranges the sequences in a phylogenetic, or evolutionary, tree. Wilson's conclusion was startling. The mtDNA in all modern groups of humans is descended from an mtDNA molecule that existed in a single woman who lived in Africa about 200,000 years ago. Applying a biblical metaphor, the popular press nicknamed this woman "Mitochondrial Eve."

By focusing on the evolution of mtDNA, Wilson's laboratory traced human ancestry back to a point where the maternal lineages of all modern mtDNA sequences coalesce in a single common ancestor—the mitochondrial mother of us all. However, these researchers never meant to imply that a single woman alone gave rise to all modern human beings. The mass of human nuclear DNA, which is inherited equally from males and females, and which varies among the members of a breeding population, cannot be traced to a single individual.

The work of Wilson and his colleagues strongly argues that all modern humans evolved from individuals who lived in Africa less than 200,000 years ago, and possibly as recently as 120,000 years ago. Migrants from this original African population presumably founded the archaic human populations of Europe and Asia, which, in turn, founded the early human populations of Australia, Oceana, and the Americas. This evolutionary scenario has been called the "Out of Africa" hypothesis. Another hypothesis proposes that humans evolved simultaneously in many regions of the world, from groups that were long established in those regions, perhaps for many hundreds of thousands of years, and that these groups probably interbred with other archaic populations such as the Neanderthals of Europe and western Asia.

The Neanderthals have always been an enigmatic group for students of human evolution. Fossil remains indicate that they were quite different from modern humans; thicker bones, greater musculature, and different body proportions clearly set them apart. Were the Neanderthals ancestral to modern humans? Did they interbreed with the populations that ultimately produced modern humans, or were they a separate and distinct species altogether?

In 1997 Matthias Krings, Anne Stone, Ralf Schmitz, Heike Krainitzki, Mark Stoneking, and Svante Pääbo published the sequence of 379 base pairs of mtDNA extracted from a fossilized Neanderthal arm bone.[2] This particular fossil, discovered in 1856 near Dusseldorf, Germany, has been the subject of many intensive studies. After lengthy negotiations, the fossil's custodians granted Krings and co-workers permission to remove a 3.5-g piece of bone from the right humerus. Small fragments from this piece were pulverized, and the DNA remnants within them were carefully extracted. Because of the fossil's age (between 30,000 and 100,000 years), most of the DNA was expected to be degraded. However, because mtDNA is much more abundant than any particular sequence of nuclear DNA, Krings and co-workers hoped that some of it had survived. Their first step was to use a technique called the polymerase chain reaction (PCR, see Chapter 20) to amplify small segments of surviving mtDNA molecules. PCR allows a researcher to generate millions of identical DNA molecules from just a few molecules by *in vitro* replication with a bacterial DNA polymerase. The sequence of the amplified DNA can then be determined biochemically.

In carefully controlled experiments, Krings and coworkers succeeded in amplifying mtDNA remnants extracted from the fossil. Biochemical analysis of this ampli-

Neanderthal DNA Sequences Obtained By PCR Show That Humans Did Not descend FROM Neanderthals

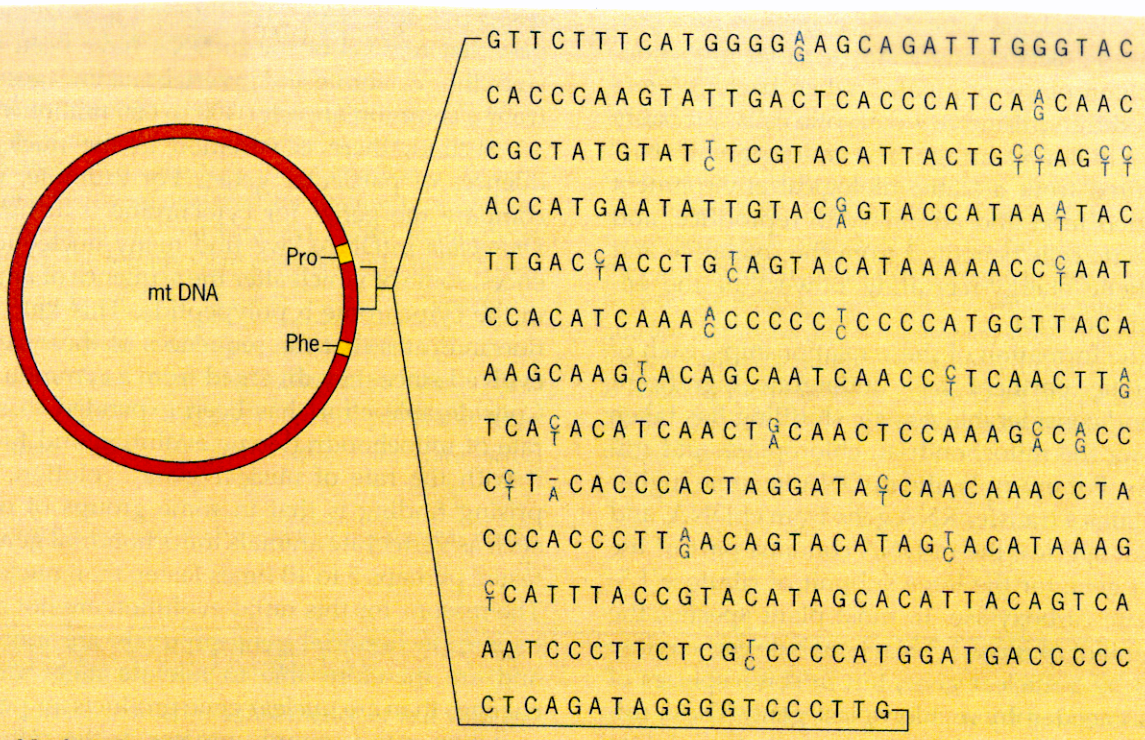


Figure 1. Nucleotide differences within a 379-bp noncoding region of the mtDNA of a Neanderthal fossil and that of a modern human being. The sequenced region lies between the genes for the phenylalanine (Phe) and proline (Pro) tRNAs. For each nucleotide difference (highlighted), the upper nucleotide is found in modern human mtDNA and the lower one is found in the Neanderthal mtDNA.

fied material showed that Neanderthal mtDNA differs from modern human mtDNA in 28 of the 379 nucleotides that were analyzed (Figure 1). The mtDNA isolated from different modern humans typically shows only 8 nucleotide substitutions in this region. Thus, Neanderthal mtDNA is significantly unlike that of modern humans. Computer analysis of the DNA sequences suggested that the human and Neanderthal mtDNA lineages began to evolve separately between 550,000 and 690,000 years ago, and that modern human mtDNAs originated between 120,000 and 150,000 years ago, apparently in Africa. Thus, Neanderthals were almost certainly not ancestral to modern humans. Rather, they evolved separately and, in the end, became extinct.

In the discussion section of their paper, Krings and coauthors concluded that "The Neanderthal mtDNA sequence thus supports a scenario in which modern humans arose recently in Africa as a distinct species and replaced Neanderthals with little or no interbreeding." They also added a caveat: "It must be emphasized that the above conclusions are based on a single individual sequence; the retrieval and analysis of mtDNA sequences from additional Neanderthal specimens is obviously desirable."[3] Of course, obtaining mtDNA sequences from other Neanderthals will entail the destruction of rare fossil material. Thus, the decision to collect such data should not be taken lightly. The benefit of collecting data from several individuals may not outweigh the cost of sacrificing so many valuable fossils. However, obtaining the sequence from at least one more Neanderthal does seem worthwhile, since this sequence could reinforce or invalidate the inferences that have to be made from the single sequence now available. We will have to wait and see if another Neanderthal fossil suitable for DNA analysis can be found. If it can, then the issue will be whether or not to allow part of that fossil to be destroyed to obtain a few molecules of mtDNA.

[1]Wilson, A. C., and R. L. Cann. 1992. The recent African genesis of humans. *Sci. Amer.*, 266(4):68–73.

[2]Krings, M., A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking, and S. Pääbo. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19–30.

[3]ibid., p. 27.

AMPLIFYING ANCIENT DNA USING PCR

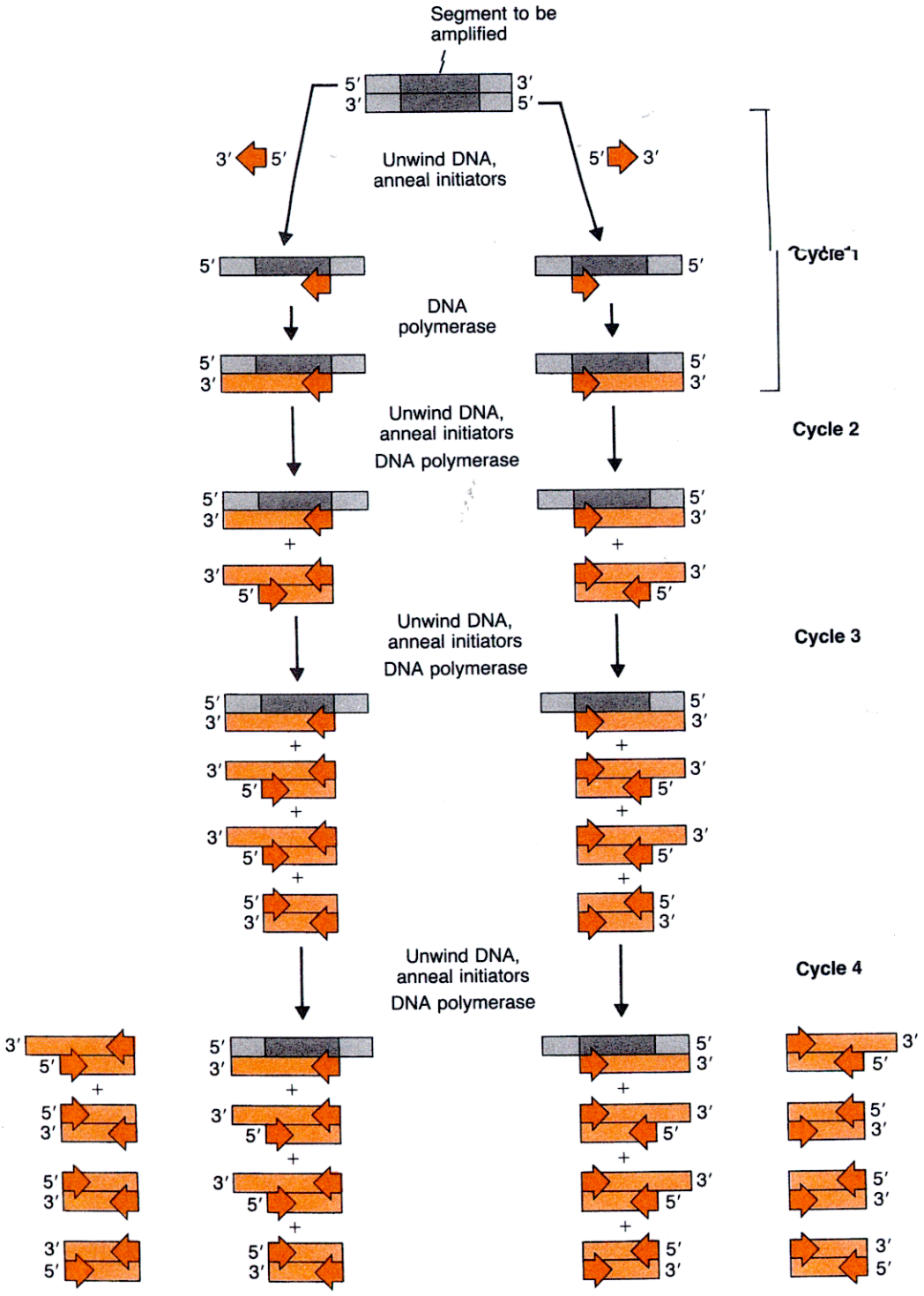


Figure 7.10 Amplifying defined segments of DNA using the polymerase chain reaction (PCR).

THE HUMAN GENOME SEQUENCE

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

* *A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.*

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

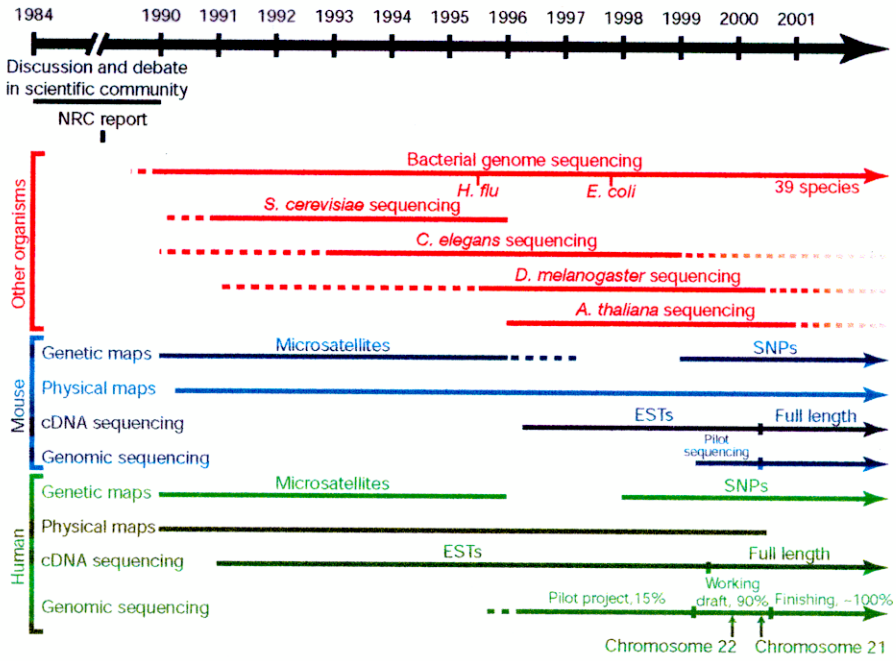


Figure 1 Timeline of large-scale genomic analyses. Shown are selected components of work on several non-vertebrate model organisms (red), the mouse (blue) and the human (green) from 1990; earlier projects are described in the text. SNPs, single nucleotide polymorphisms; ESTs, expressed sequence tags.

9

WITHOUT AUTOMATION THE HUMAN GENOME COULD NOT HAVE BEEN SEQUENCED

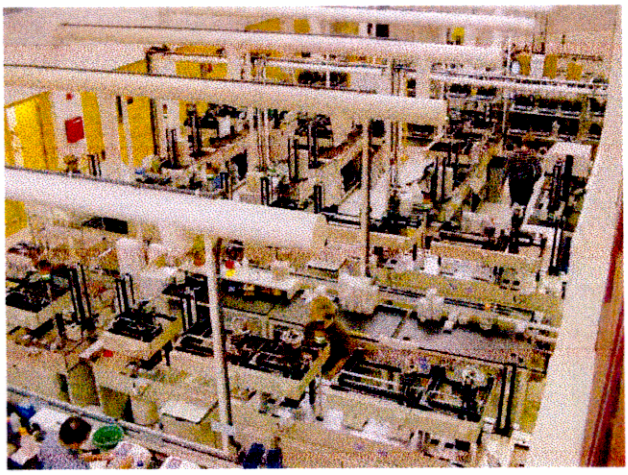


Figure 3 The automated production line for sample preparation at the Whitehead Institute, Center for Genome Research. The system consists of custom-designed factory-style conveyor belt robots that perform all functions from purifying DNA from bacterial cultures through setting up and purifying sequencing reactions.

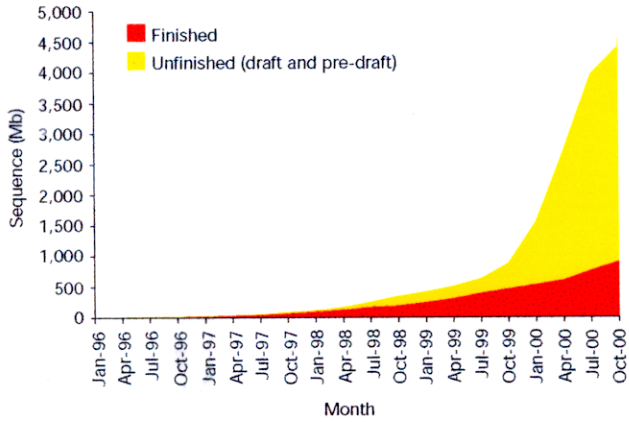


Figure 4 Total amount of human sequence in the High Throughput Genome Sequence (HTGS) division of GenBank. The total is the sum of finished sequence (red) and unfinished (draft plus predraft) sequence (yellow).

THE HUMAN GENOME SEQUENCE IS THE RESULT OF AN INTERNATIONAL COLLABORATION

Genome Sequencing Centres (Listed in order of total genomic sequence contributed, with a partial list of personnel. A full list of contributors at each centre is available as Supplementary Information.)

Whitehead Institute for Biomedical Research, Center for Genome Research: Eric S. Lander[1]*, Lauren M. Linton[1], Bruce Birren[1]*, Chad Nusbaum[1]*, Michael C. Zody[1]*, Jennifer Baldwin[1], Keri Devon[1], Ken Dewar[1], Michael Doyle[1], William FitzHugh[1]*, Roel Funke[1], Diane Gage[1], Katrina Harris[1], Andrew Heaford[1], John Howland[1], Lisa Kann[1], Jessica Lehoczky[1], Rosie LeVine[1], Paul McEwan[1], Kevin McKernan[1], James Meldrim[1], Jill P. Mesirov[1]*, Cher Miranda[1], William Morris[1], Jerome Naylor[1], Christina Raymond[1], Mark Rosetti[1], Ralph Santos[1], Andrew Sheridan[1], Carrie Sougnez[1], Nicole Stange-Thomann[1], Nikola Stojanovic[1], Aravind Subramanian[1] & Dudley Wyman[1]

The Sanger Centre: Jane Rogers[2], John Sulston[2]*, Rachael Ainscough[2], Stephan Beck[2], David Bentley[2], John Burton[2], Christopher Clee[2], Nigel Carter[2], Alan Coulson[2], Rebecca Deadman[2], Panos Deloukas[2], Andrew Dunham[2], Ian Dunham[2], Richard Durbin[2]*, Lisa French[2], Darren Grafham[2], Simon Gregory[2], Tim Hubbard[2]*, Sean Humphray[2], Adrienne Hunt[2], Matthew Jones[2], Christine Lloyd[2], Amanda McMurray[2], Lucy Matthews[2], Simon Mercer[2], Sarah Milne[2], James C. Mullikin[2]*, Andrew Mungall[2], Robert Plumb[2], Mark Ross[2], Ratna Shownkeen[2] & Sarah Sims[2]

Washington University Genome Sequencing Center: Robert H. Waterston[3]*, Richard K. Wilson[3], LaDeana W. Hillier[3]*, John D. McPherson[3], Marco A. Marra[3], Elaine R. Mardis[3], Lucinda A. Fulton[3], Asif T. Chinwalla[3]*, Kymberlie H. Pepin[3], Warren R. Gish[3], Stephanie L. Chissoe[3], Michael C. Wendl[3], Kim D. Delehaunty[3], Tracie L. Miner[3], Andrew Delehaunty[3], Jason B. Kramer[3], Lisa L. Cook[3], Robert S. Fulton[3], Douglas L. Johnson[3], Patrick J. Minx[3] & Sandra W. Clifton[3]

US DOE Joint Genome Institute: Trevor Hawkins[4], Elbert Branscomb[4], Paul Predki[4], Paul Richardson[4], Sarah Wenning[4], Tom Slezak[4], Norman Doggett[4], Jan-Fang Cheng[4], Anne Olsen[4], Susan Lucas[4], Christopher Elkin[4], Edward Uberbacher[4] & Marvin Frazier[4]

Baylor College of Medicine Human Genome Sequencing Center: Richard A. Gibbs[5]*, Donna M. Muzny[5], Steven E. Scherer[5], John B. Bouck[5]*, Erica J. Sodergren[5], Kim C. Worley[5]*, Catherine M. Rives[5], James H. Gorrell[5], Michael L. Metzker[5], Susan L. Naylor[6], Raju S. Kucherlapati[7], David L. Nelson, & George M. Weinstock[8]

RIKEN Genomic Sciences Center: Yoshiyuki Sakaki[9], Asao Fujiyama[9], Masahira Hattori[9], Tetsushi Yada[9], Atsushi Toyoda[9], Takehiko Itoh[9], Chiharu Kawagoe[9], Hidemi Watanabe[9], Yasushi Totoki[9] & Todd Taylor[9]

Genoscope and CNRS UMR-8030: Jean Weissenbach[10], Roland Heilig[10], William Saurin[10], Francois Artiguenave[10], Philippe Brottier[10], Thomas Bruls[10], Eric Pelletier[10], Catherine Robert[10] & Patrick Wincker[10]

GTC Sequencing Center: Douglas R. Smith[11], Lynn Doucette-Stamm[11], Marc Rubenfield[11], Keith Weinstock[11], Hong Mei Lee[11] & JoAnn Dubois[11]

Department of Genome Analysis, Institute of Molecular Biotechnology: André Rosenthal[12], Matthias Platzer[12], Gerald Nyakatura[12], Stefan Taudien[12] & Andreas Rump[12]

Beijing Genomics Institute/Human Genome Center: Huanming Yang[13], Jun Yu[13], Jian Wang[13], Guyang Huang[14] & Jun Gu[15]

Multimegabase Sequencing Center, The Institute for Systems Biology: Leroy Hood[16], Lee Rowen[16], Anup Madan[16] & Shizen Qin[16]

Stanford Genome Technology Center: Ronald W. Davis[17], Nancy A. Federspiel[17], A. Pia Abola[17] & Michael J. Proctor[17]

Stanford Human Genome Center: Richard M. Myers[18], Jeremy Schmutz[18], Mark Dickson[18], Jane Grimwood[18] & David R. Cox[18]

University of Washington Genome Center: Maynard V. Olson[19], Rajinder Kaul[19] & Christopher Raymond[19]

Department of Molecular Biology, Keio University School of Medicine: Nobuyoshi Shimizu[20], Kazuhiko Kawasaki[20] & Shinsei Minoshima[20]

University of Texas Southwestern Medical Center at Dallas: Glen A. Evans[21]†, Maria Athanasiou[21] & Roger Schultz[21]

University of Oklahoma's Advanced Center for Genome Technology: Bruce A. Roe[22], Feng Chen[22] & Huaqin Pan[22]

Max Planck Institute for Molecular Genetics: Juliane Ramser[23], Hans Lehrach[23] & Richard Reinhardt[23]

Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center: W. Richard McCombie[24], Melissa de la Bastide[24] & Neilay Dedhia[24]

GBF—German Research Centre for Biotechnology: Helmut Blöcker[25], Klaus Hornischer[25] & Gabriele Nordsiek[25]

* **Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):** Richa Agarwala[26], L. Aravind[26], Jeffrey A. Bailey[27], Alex Bateman[2], Serafim Batzoglou[1], Ewan Birney[28], Peer Bork[29,30], Daniel G. Brown[1], Christopher B. Burge[31], Lorenzo Cerutti[28], Hsiu-Chuan Chen[26], Deanna Church[26], Michele Clamp[2], Richard R. Copley[30], Tobias Doerks[29,30], Sean R. Eddy[32], Evan E. Eichler[27], Terrence S. Furey[33], James Galagan[1], James G. R. Gilbert[2], Cyrus Harmon[34], Yoshihide Hayashizaki[35], David Haussler[36], Henning Hermjakob[28], Karsten Hokamp[37], Wonhee Jang[26], L. Steven Johnson[32], Thomas A. Jones[32], Simon Kasif[38], Arek Kaspryzk[28], Scot Kennedy[39], W. James Kent[40], Paul Kitts[26], Eugene V. Koonin[26], Ian Korf[3], David Kulp[34], Doron Lancet[41], Todd M. Lowe[42], Aoife McLysaght[37], Tarjei Mikkelsen[38], John V. Moran[43], Nicola Mulder[28], Victor J. Pollara[1], Chris P. Ponting[44], Greg Schuler[26], Jörg Schultz[30], Guy Slater[28], Arian F. A. Smit[45], Elia Stupka[28], Joseph Szustakowki[38], Danielle Thierry-Mieg[26], Jean Thierry-Mieg[26], Lukas Wagner[26], John Wallis[3], Raymond Wheeler[34], Alan Williams[34], Yuri I. Wolf[26], Kenneth H. Wolfe[37], Shiaw-Pyng Yang[3] & Ru-Fang Yeh[31]

Scientific management: National Human Genome Research Institute, US National Institutes of Health: Francis Collins[46]*, Mark S. Guyer[46], Jane Peterson[46], Adam Felsenfeld[46]* & Kris A. Wetterstrand[46]; **Office of Science, US Department of Energy:** Aristides Patrinos[47]; **The Wellcome Trust:** Michael J. Morgan[48]

BUT IT WAS ALSO DONE INDEPENDENTLY BY A COMPANY – CELERA®

The Sequence of the Human Genome

J. Craig Venter,[1*] Mark D. Adams,[1] Eugene W. Myers,[1] Peter W. Li,[1] Richard J. Mural,[1] Granger G. Sutton,[1] Hamilton O. Smith,[1] Mark Yandell,[1] Cheryl A. Evans,[1] Robert A. Holt,[1] Jeannine D. Gocayne,[1] Peter Amanatides,[1] Richard M. Ballew,[1] Daniel H. Huson,[1] Jennifer Russo Wortman,[1] Qing Zhang,[1] Chinnappa D. Kodira,[1] Xiangqun H. Zheng,[1] Lin Chen,[1] Marian Skupski,[1] Gangadharan Subramanian,[1] Paul D. Thomas,[1] Jinghui Zhang,[1] George L. Gabor Miklos,[2] Catherine Nelson,[3] Samuel Broder,[1] Andrew G. Clark,[4] Joe Nadeau,[5] Victor A. McKusick,[6] Norton Zinder,[7] Arnold J. Levine,[7] Richard J. Roberts,[8] Mel Simon,[9] Carolyn Slayman,[10] Michael Hunkapiller,[11] Randall Bolanos,[1] Arthur Delcher,[1] Ian Dew,[1] Daniel Fasulo,[1] Michael Flanigan,[1] Liliana Florea,[1] Aaron Halpern,[1] Sridhar Hannenhalli,[1] Saul Kravitz,[1] Samuel Levy,[1] Clark Mobarry,[1] Knut Reinert,[1] Karin Remington,[1] Jane Abu-Threideh,[1] Ellen Beasley,[1] Kendra Biddick,[1] Vivien Bonazzi,[1] Rhonda Brandon,[1] Michele Cargill,[1] Ishwar Chandramouliswaran,[1] Rosane Charlab,[1] Kabir Chaturvedi,[1] Zuoming Deng,[1] Valentina Di Francesco,[1] Patrick Dunn,[1] Karen Eilbeck,[1] Carlos Evangelista,[1] Andrei E. Gabrielian,[1] Weiniu Gan,[1] Wangmao Ge,[1] Fangcheng Gong,[1] Zhiping Gu,[1] Ping Guan,[1] Thomas J. Heiman,[1] Maureen E. Higgins,[1] Rui-Ru Ji,[1] Zhaoxi Ke,[1] Karen A. Ketchum,[1] Zhongwu Lai,[1] Yiding Lei,[1] Zhenya Li,[1] Jiayin Li,[1] Yong Liang,[1] Xiaoying Lin,[1] Fu Lu,[1] Gennady V. Merkulov,[1] Natalia Milshina,[1] Helen M. Moore,[1] Ashwinikumar K Naik,[1] Vaibhav A. Narayan,[1] Beena Neelam,[1] Deborah Nusskern,[1] Douglas B. Rusch,[1] Steven Salzberg,[12] Wei Shao,[1] Bixiong Shue,[1] Jingtao Sun,[1] Zhen Yuan Wang,[1] Aihui Wang,[1] Xin Wang,[1] Jian Wang,[1] Ming-Hui Wei,[1] Ron Wides,[13] Chunlin Xiao,[1] Chunhua Yan,[1] Alison Yao,[1] Jane Ye,[1] Ming Zhan,[1] Weiqing Zhang,[1] Hongyu Zhang,[1] Qi Zhao,[1] Liansheng Zheng,[1] Fei Zhong,[1] Wenyan Zhong,[1] Shiaoping C. Zhu,[1] Shaying Zhao,[12] Dennis Gilbert,[1] Suzanna Baumhueter,[1] Gene Spier,[1] Christine Carter,[1] Anibal Cravchik,[1] Trevor Woodage,[1] Feroze Ali,[1] Huijin An,[1] Aderonke Awe,[1] Danita Baldwin,[1] Holly Baden,[1] Mary Barnstead,[1] Ian Barrow,[1] Karen Beeson,[1] Dana Busam,[1] Amy Carver,[1] Angela Center,[1] Ming Lai Cheng,[1] Liz Curry,[1] Steve Danaher,[1] Lionel Davenport,[1] Raymond Desilets,[1] Susanne Dietz,[1] Kristina Dodson,[1] Lisa Doup,[1] Steven Ferriera,[1] Neha Garg,[1] Andres Gluecksmann,[1] Brit Hart,[1] Jason Haynes,[1] Charles Haynes,[1] Cheryl Heiner,[1] Suzanne Hladun,[1] Damon Hostin,[1] Jarrett Houck,[1] Timothy Howland,[1] Chinyere Ibegwam,[1] Jeffery Johnson,[1] Francis Kalush,[1] Lesley Kline,[1] Shashi Koduru,[1] Amy Love,[1] Felecia Mann,[1] David May,[1] Steven McCawley,[1] Tina McIntosh,[1] Ivy McMullen,[1] Mee Moy,[1] Linda Moy,[1] Brian Murphy,[1] Keith Nelson,[1] Cynthia Pfannkoch,[1] Eric Pratts,[1] Vinita Puri,[1] Hina Qureshi,[1] Matthew Reardon,[1] Robert Rodriguez,[1] Yu-Hui Rogers,[1] Deanna Romblad,[1] Bob Ruhfel,[1] Richard Scott,[1] Cynthia Sitter,[1] Michelle Smallwood,[1] Erin Stewart,[1] Renee Strong,[1] Ellen Suh,[1] Reginald Thomas,[1] Ni Ni Tint,[1] Sukyee Tse,[1] Claire Vech,[1] Gary Wang,[1] Jeremy Wetter,[1] Sherita Williams,[1] Monica Williams,[1] Sandra Windsor,[1] Emily Winn-Deen,[1] Keriellen Wolfe,[1] Jayshree Zaveri,[1] Karena Zaveri,[1] Josep F. Abril,[14] Roderic Guigó,[14] Michael J. Campbell,[1] Kimmen V. Sjolander,[1] Brian Karlak,[1] Anish Kejariwal,[1] Huaiyu Mi,[1] Betty Lazareva,[1] Thomas Hatton,[1] Apurva Narechania,[1] Karen Diemer,[1] Anushya Muruganujan,[1] Nan Guo,[1] Shinji Sato,[1] Vineet Bafna,[1] Sorin Istrail,[1] Ross Lippert,[1] Russell Schwartz,[1] Brian Walenz,[1] Shibu Yooseph,[1] David Allen,[1] Anand Basu,[1] James Baxendale,[1] Louis Blick,[1] Marcelo Caminha,[1] John Carnes-Stine,[1] Parris Caulk,[1] Yen-Hui Chiang,[1] My Coyne,[1] Carl Dahlke,[1] Anne Deslattes Mays,[1] Maria Dombroski,[1] Michael Donnelly,[1] Dale Ely,[1] Shiva Esparham,[1] Carl Fosler,[1] Harold Gire,[1] Stephen Glanowski,[1] Kenneth Glasser,[1] Anna Glodek,[1] Mark Gorokhov,[1] Ken Graham,[1] Barry Gropman,[1] Michael Harris,[1] Jeremy Heil,[1] Scott Henderson,[1] Jeffrey Hoover,[1] Donald Jennings,[1] Catherine Jordan,[1] James Jordan,[1] John Kasha,[1] Leonid Kagan,[1] Cheryl Kraft,[1] Alexander Levitsky,[1] Mark Lewis,[1] Xiangjun Liu,[1] John Lopez,[1] Daniel Ma,[1] William Majoros,[1] Joe McDaniel,[1] Sean Murphy,[1] Matthew Newman,[1] Trung Nguyen,[1] Ngoc Nguyen,[1] Marc Nodell,[1] Sue Pan,[1] Jim Peck,[1] Marshall Peterson,[1] William Rowe,[1] Robert Sanders,[1] John Scott,[1] Michael Simpson,[1] Thomas Smith,[1] Arlan Sprague,[1] Timothy Stockwell,[1] Russell Turner,[1] Eli Venter,[1] Mei Wang,[1] Meiyuan Wen,[1] David Wu,[1] Mitchell Wu,[1] Ashley Xia,[1] Ali Zandieh,[1] Xiaohong Zhu[1]

AND COMPLETED IN ONLY Nine Months!

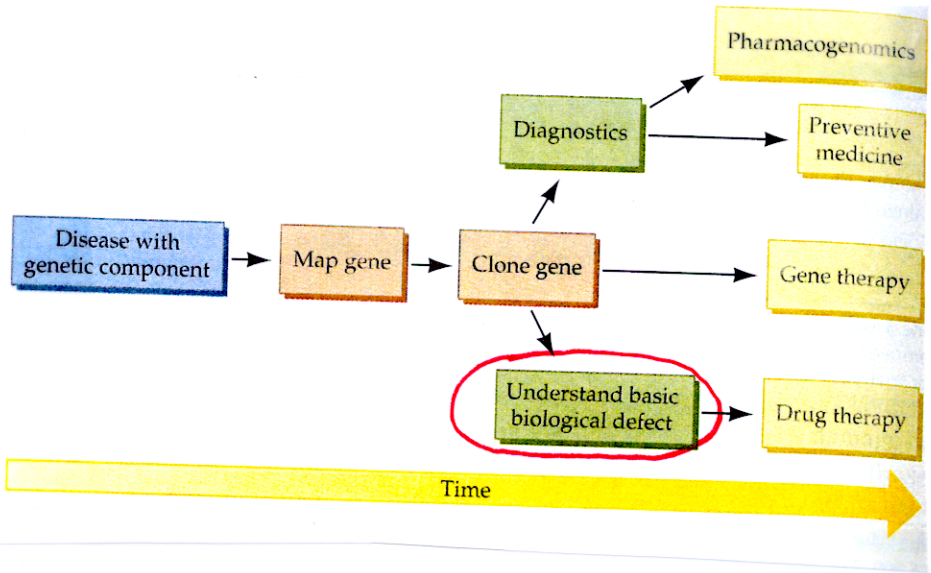
IT WAS A RACE!

A 2.91-billion base pair (bp) consensus sequence of the euchromatic portion of the human genome was generated by the whole-genome shotgun sequencing method. The 14.8-billion bp DNA sequence was generated over 9 months from 27,271,853 high-quality sequence reads (5.11-fold coverage of the genome) from both ends of plasmid clones made from the DNA of five individuals. Two assembly strategies—a whole-genome assembly and a regional chromosome assembly—were used, each combining sequence data from Celera and the publicly funded genome effort. The public data were shredded into 550-bp segments to create a 2.9-fold coverage of those genome regions that had been sequenced, without including biases inherent in the cloning and assembly procedure used by the publicly funded group. This brought the effective coverage in the assemblies to eightfold, reducing the number and size of gaps in the final assembly over what would be obtained with 5.11-fold coverage. The two assembly strategies yielded very similar results that largely agree with independent mapping data. The assemblies effectively cover the euchromatic regions of the human chromosomes. More than 90% of the genome is in scaffold assemblies of 100,000 bp or more, and 25% of the genome is in scaffolds of 10 million bp or larger. Analysis of the genome sequence revealed 26,588 protein-encoding transcripts for which there was strong corroborating evidence and an additional ~12,000 computationally derived genes with mouse matches or other weak supporting evidence. Although gene-dense clusters are obvious, almost half the genes are dispersed in low G+C sequence separated by large tracts of apparently noncoding sequence. Only 1.1% of the genome is spanned by exons, whereas 24% is in introns, with 75% of the genome being intergenic DNA. Duplications of segmental blocks, ranging in size up to chromosomal lengths, are abundant throughout the genome and reveal a complex evolutionary history. Comparative genomic analysis indicates vertebrate expansions of genes associated with neuronal function, with tissue-specific developmental regulation, and with the hemostasis and immune systems. DNA sequence comparisons between the consensus sequence and publicly funded genome data provided locations of 2.1 million single-nucleotide polymorphisms (SNPs). A random pair of human haploid genomes differed at a rate of 1 bp per 1250 on average, but there was marked heterogeneity in the level of polymorphism across the genome. Less than 1% of all SNPs resulted in variation in proteins, but the task of determining which SNPs have functional consequences remains an open challenge.

BUT The International Public Sequence is More complete than the Private one That contains Gaps

KNOWLEDGE OF THE HUMAN GENOME WILL REVOLUTIONIZE MEDICINE

BASIC KNOWLEDGE DRIVES APPLICATIONS!



DRUG USE

PREVENTION

GEN THERAPY

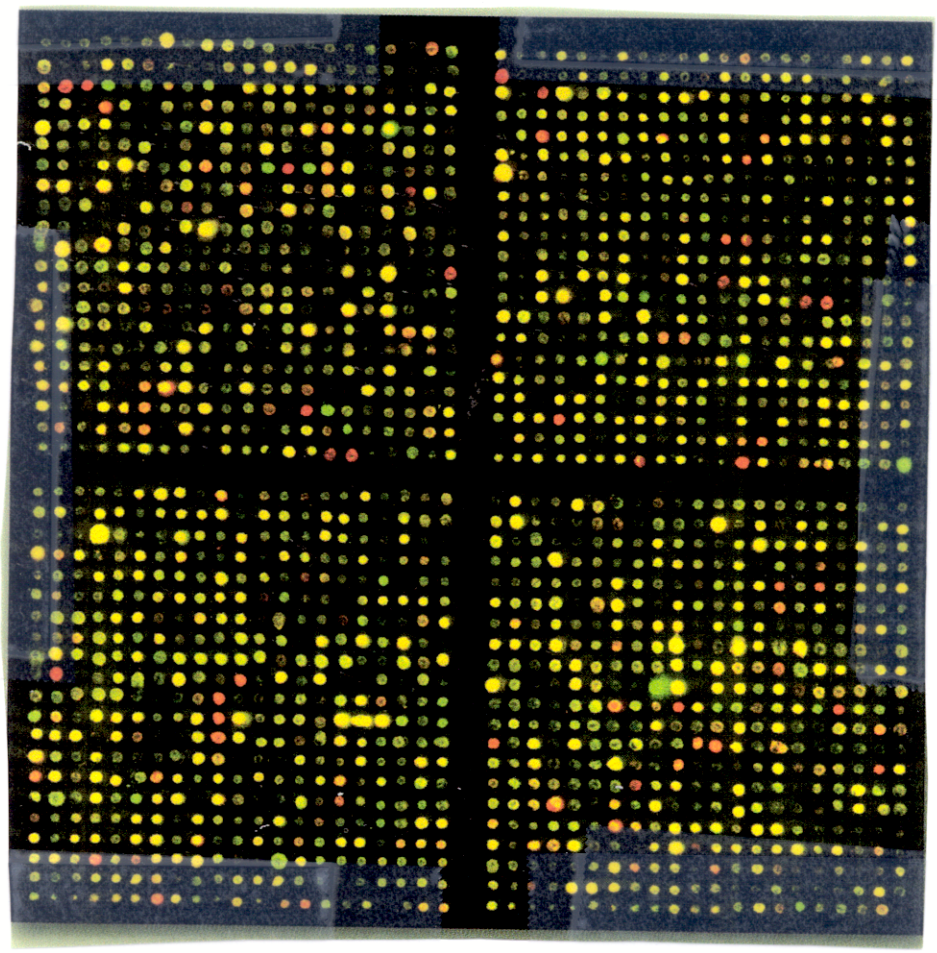
DRUG (DNA, RNA, protein) THERAPY

18.22 Is This the Future of Medicine?
The elucidation of the human genome sequence may result in an approach to medicine that is oriented to the genetic and functional individuality of each patient.

PERSONAL, PROACTIVE, PREVENTIVE

ASSAY FOR PERSON-SPECIFIC GENES!

INDIVIDUAL GENE PROFILES WILL BECOME POSSIBLE



Heart Disease? Predisposition!

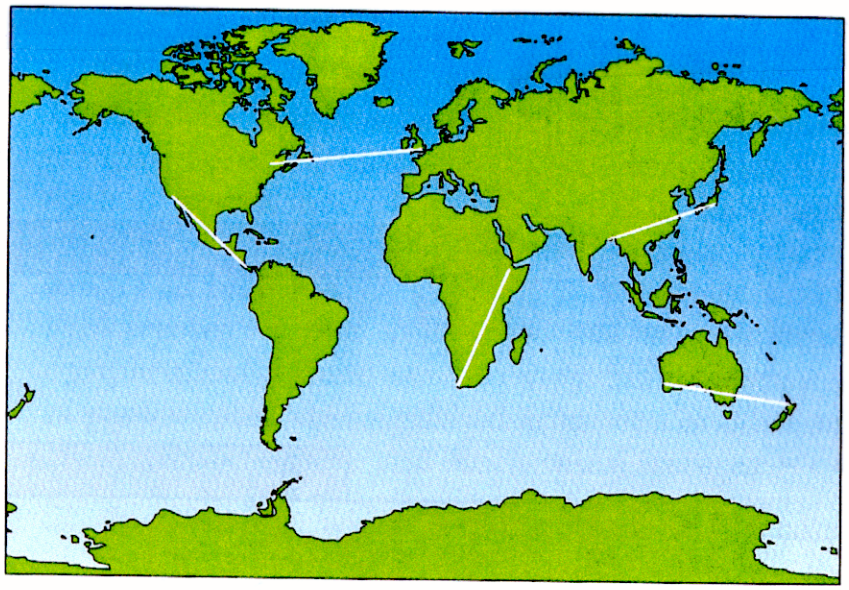
Drug utility? Metabolism!

CANCER? Early detection / Therapy!

etc

The HUMAN GENOME IS LARGE — BUT NOT THE LARGEST GENOME!

Scale = 60 nts / 10CM



What is the HUMAN GENOME LENGTH? 3×10^9 bp in meters?

1 meter per haploid genome!

Figure 1.4 The immense length of the human genome.

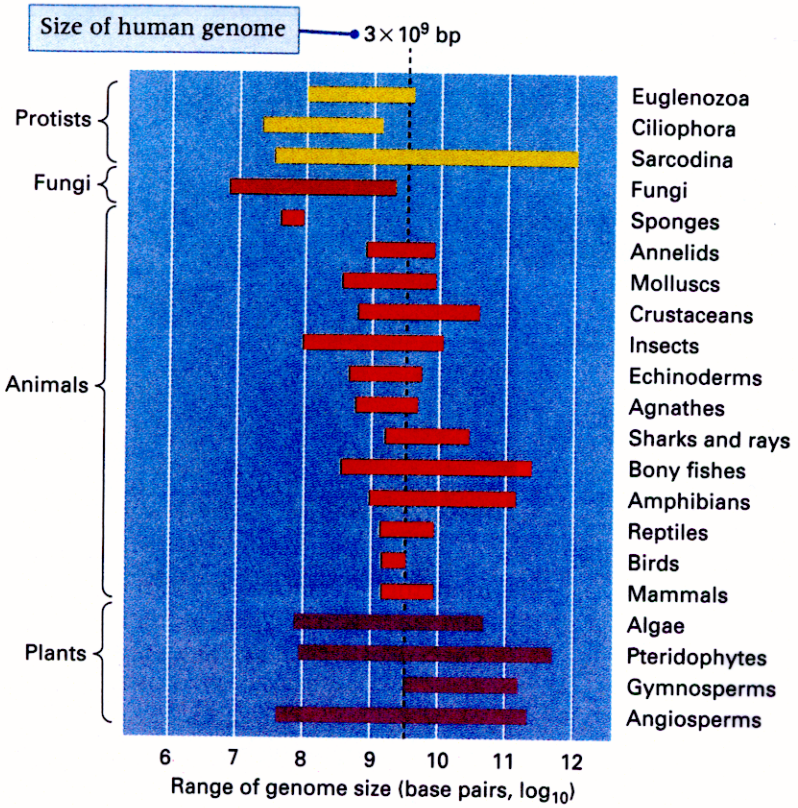


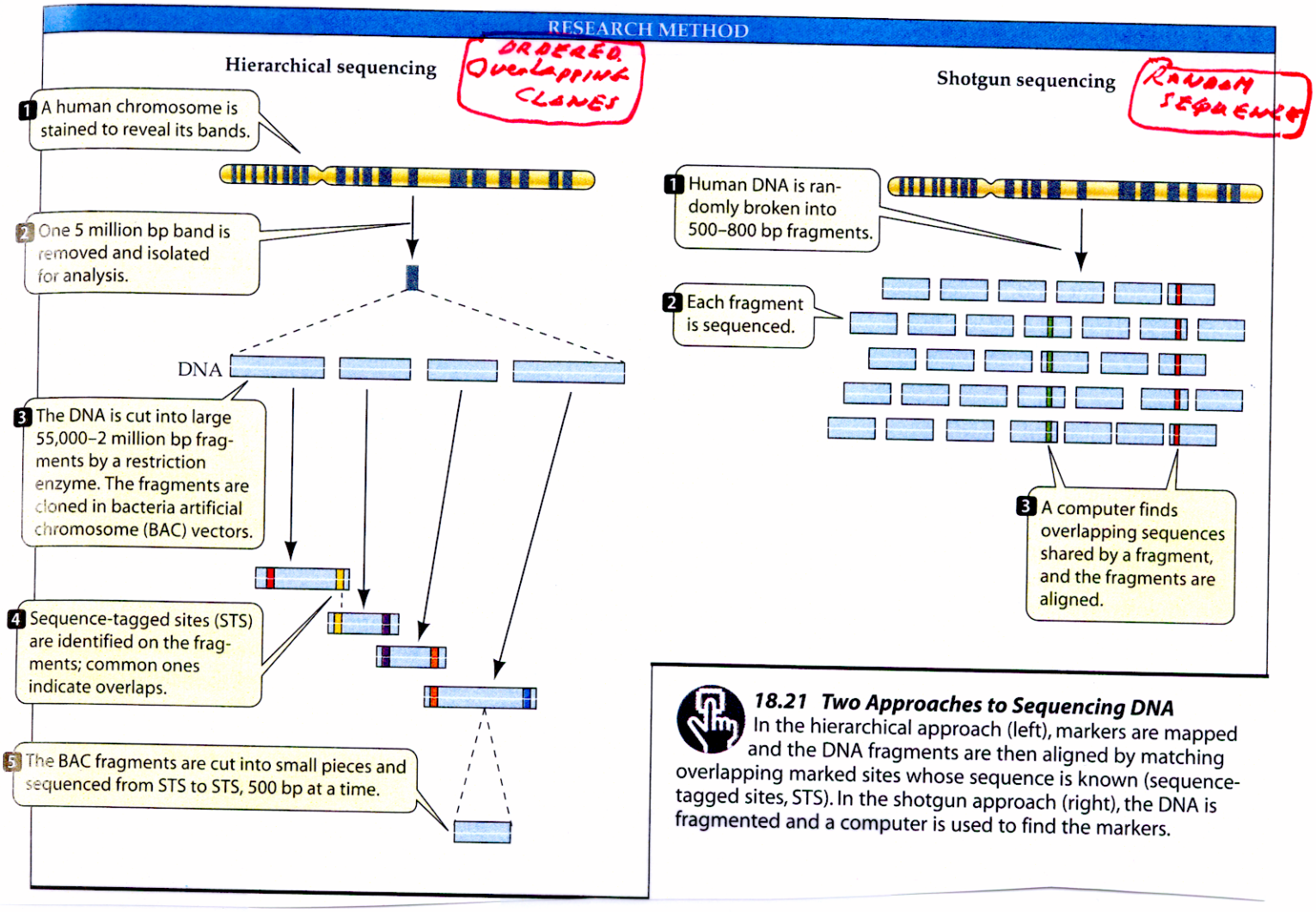
Figure 8.1 Genome size ranges over several orders of magnitude in some groups of organisms, and genome size is not correlated with developmental, metabolic, or behavioral complexity.

1. *kilobase (kb)*	10^3 nucleotide pairs (double-stranded) or 10^3 nucleotides (single-stranded)
2. *megabase (Mb)*	10^6 nucleotide pairs (double-stranded) or 10^6 nucleotides (single-stranded)
3. *gigabase (Gb)*	10^9 nucleotide pairs (double-stranded) or 10^9 nucleotides (single-stranded)

HOW WAS THE HUMAN GENOME SEQUENCED?

TOP DOWN

BOTTOM UP



18.21 Two Approaches to Sequencing DNA In the hierarchical approach (left), markers are mapped and the DNA fragments are then aligned by matching overlapping marked sites whose sequence is known (sequence-tagged sites, STS). In the shotgun approach (right), the DNA is fragmented and a computer is used to find the markers.

PUBLIC EFFORT

CHROMOSOME WALKS → SEQUENCE

BEST
MOST COMPLETE
ENTIRE
SLOW

PRIVATE EFFORT

SHOTGUN

FAST
MANY GAPS
SKELETON

NEEDS PUBLIC DATA TO ASSEMBLE → CHROMOSOMES

The PUBLIC TOP DOWN APPROACH

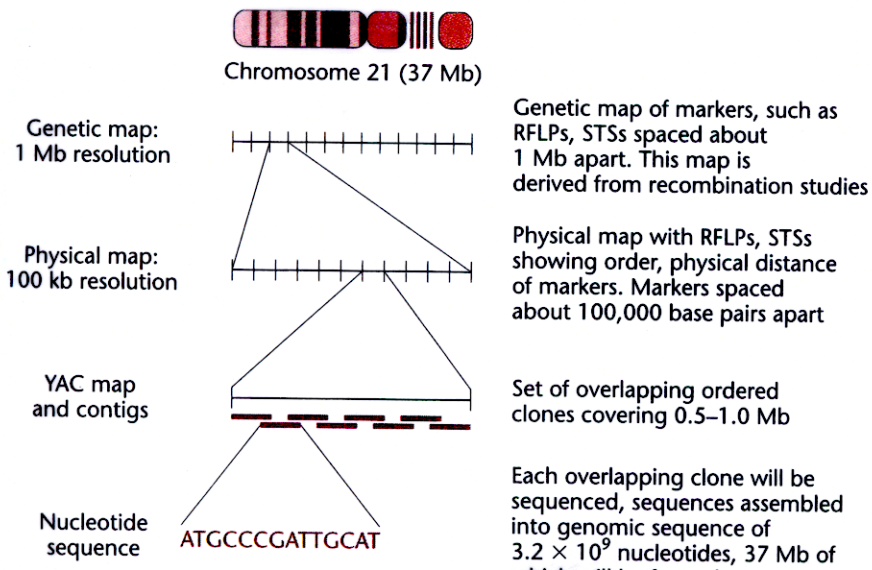


FIGURE 21.17 An overview of the strategy used in the Human Genome Project. The first goal, achieved in 1995, was to have a genetic map of each chromosome, with markers spaced at distances of about 1 Mb (1 million base pairs of DNA). This work was accomplished by finding markers such as RFLPs and STSs and assigning them to chromosomes. Once assigned to chromosomes, the markers' inheritance was observed in heterozygous families to establish the order and distance between them (a genetic map). In the second stage, the goal was to prepare a physical map of each chromosome (our example uses chromosome 21, the smallest chromosome) containing the location of markers spaced about 100,000 base pairs apart. This goal has now been achieved. The third stage involves the construction of a set of overlapping clones, in yeast artificial chromosomes (YACs) or other vectors that cover the length of the chromosome. The last stage will be the sequencing of the entire genome. Sequencing on selected parts of the genome has started.

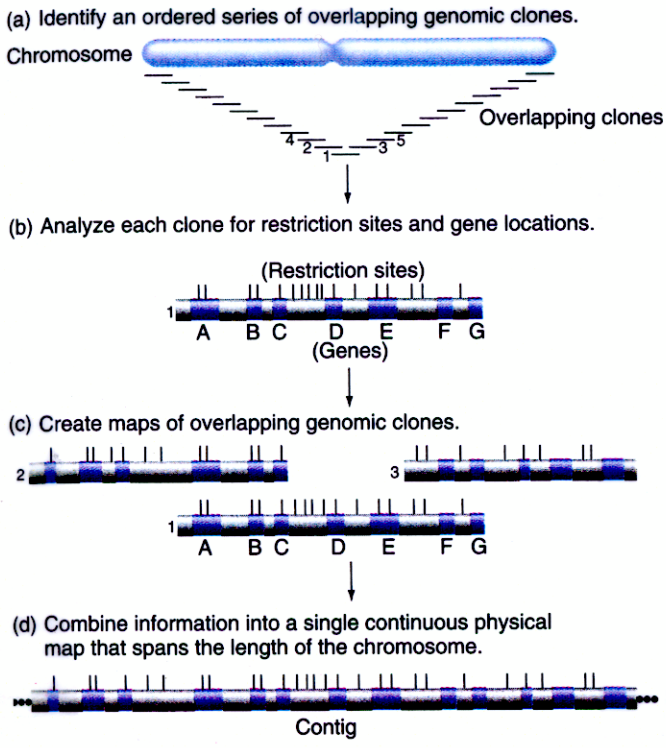


Figure 10.5 Building a whole-chromosome physical map. (a) To produce a whole-chromosome physical map, you first order a set of overlapping genomic clones that extend from one end of the chromosome to the other. Subsequent figures describe various methods of obtaining this ordered set of clones. (b) You next map the restriction sites of each clone in the set through restriction analysis, and analyze individual restriction fragments in other ways, such as Northern blot analysis, to identify transcription units. (c) Computers overlay the different types of maps for each clone onto the overlapping clones to obtain a continuous map. (d) The result is a single continuous map extending the length of the chromosome.

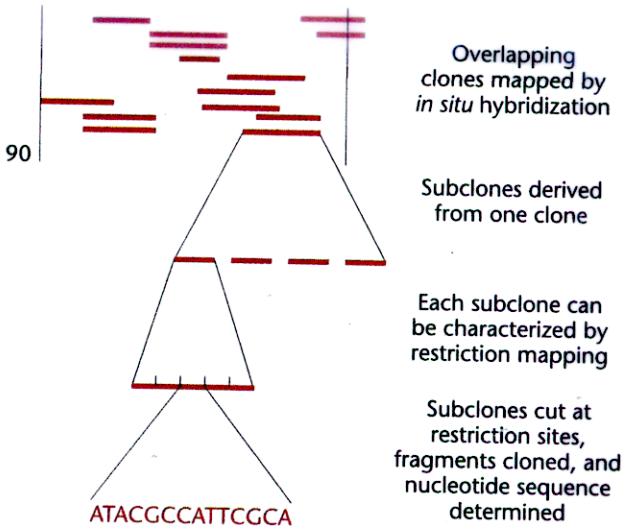
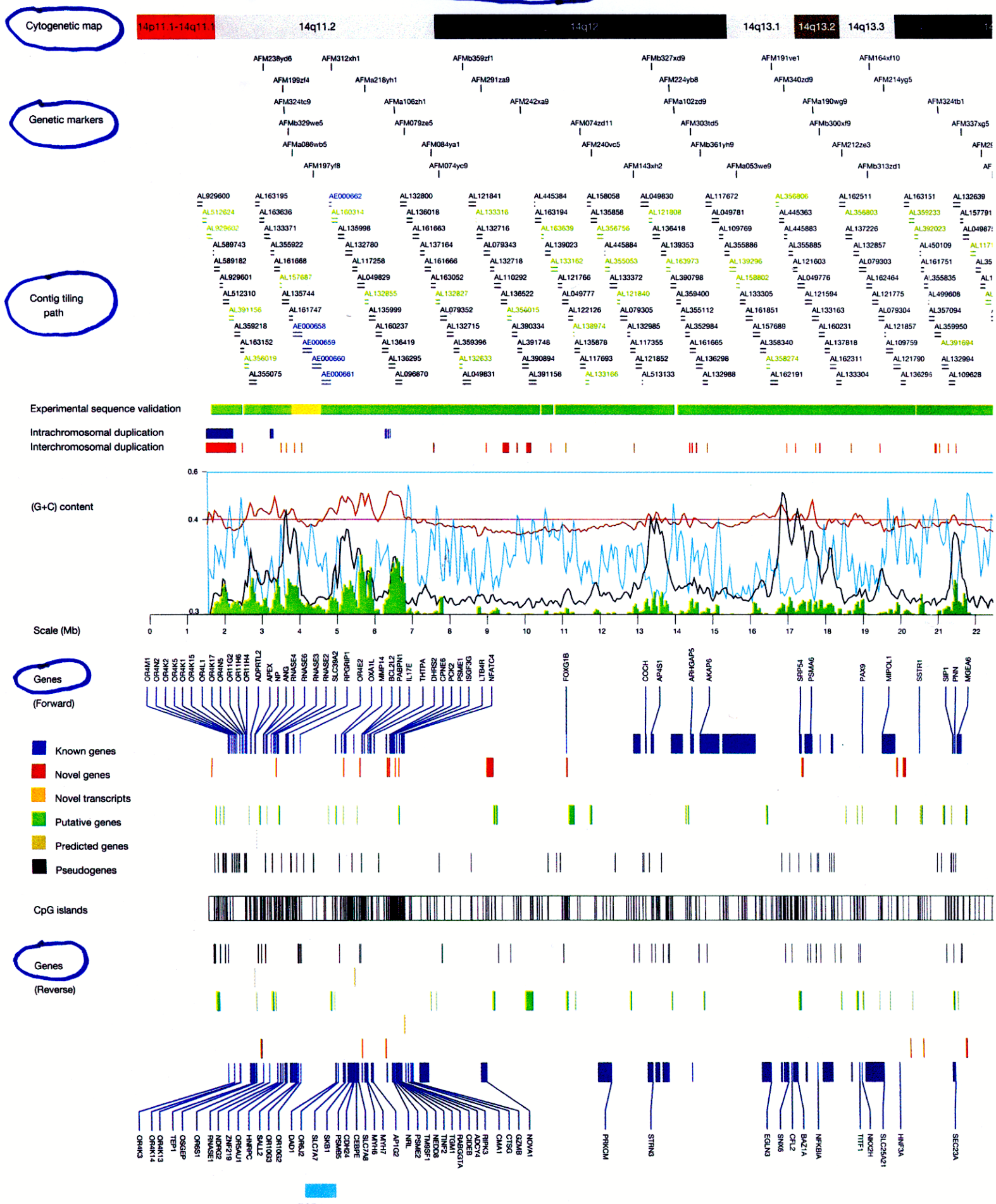


FIGURE 20.2 The top-down approach for the *Drosophila* genome project. A genome library is constructed with very large fragments (~200 kb) in a special vector. The physical location of each is mapped to the polytene chromosomes. Each clone is then broken down into subclones, which are characterized by restriction mapping for DNA sequence analysis.

SEQUENCING CHROMOSOME 14 TOP DOWN



THE "PRIVATE" BOTTOM UP APPROACH

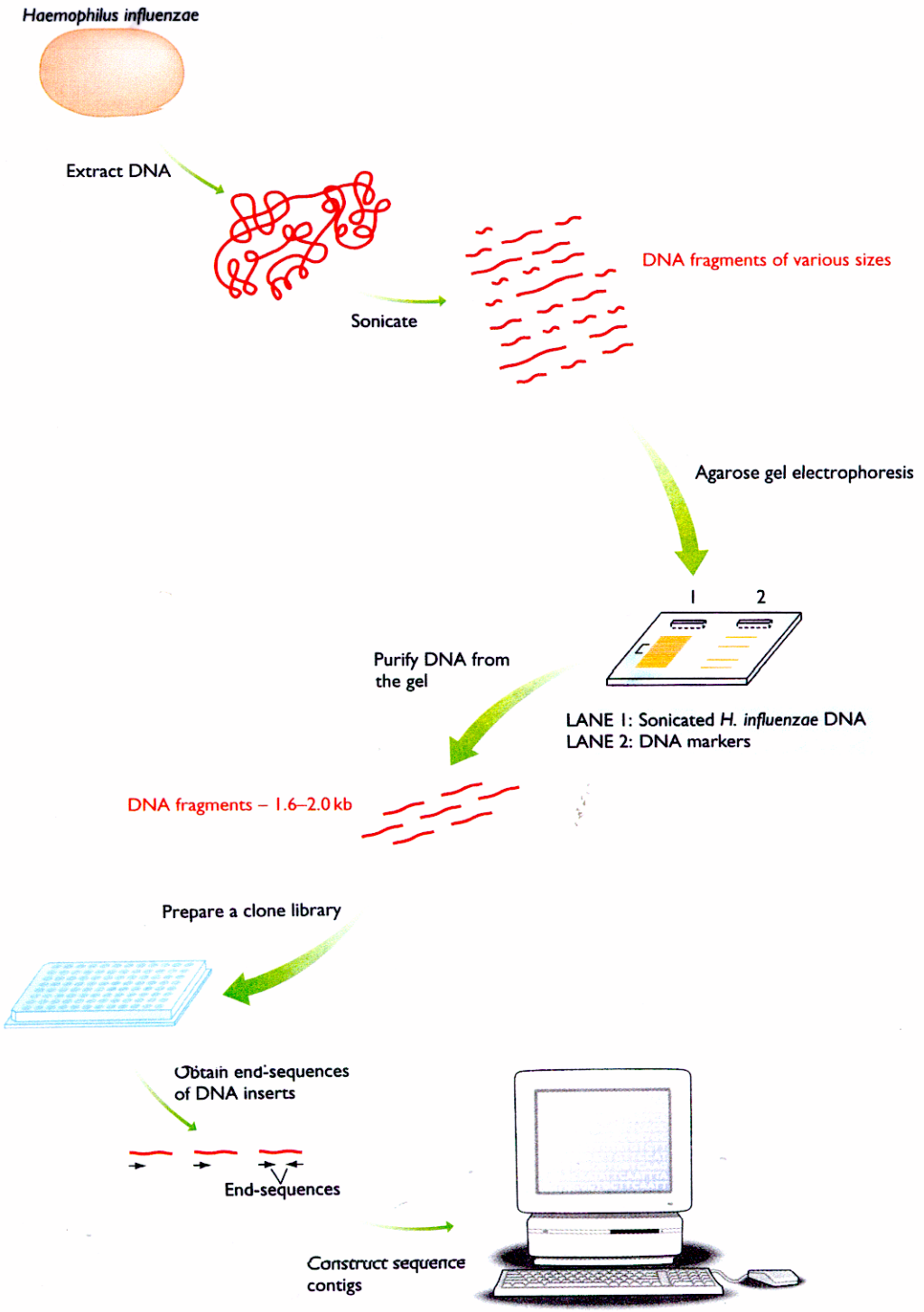


Figure 4.10 The way in which the shotgun approach was used to obtain the DNA sequence of the *Haemophilus influenzae* genome.

H. influenzae DNA was sonicated and fragments with sizes between 1.6 and 2.0 kb purified from an agarose gel and ligated into a plasmid vector to produce a clone library. End-sequences were obtained from clones taken from this library, and a computer used to identify overlaps between sequences. This resulted in 140 sequence contigs, which were assembled into the complete genome sequence as shown in *Figure 4.11*. For further details, see Fleischmann *et al.*, 1995.

APPROXIMATELY 30,000 GENES HAVE BEEN IDENTIFIED IN THE HUMAN GENOME

Table 11. Genome overview.

Size of the genome (including gaps)	2.91 Gbp
Size of the genome (excluding gaps)	2.66 Gbp
Longest contig	1.99 Mbp
Longest scaffold	14.4 Mbp
Percent of A+T in the genome	54
Percent of G+C in the genome	38
Percent of undetermined bases in the genome	9
Most GC-rich 50 kb	Chr. 2 (66%)
Least GC-rich 50 kb	Chr. X (25%)
Percent of genome classified as repeats	35
Number of annotated genes	26,383
Percent of annotated genes with unknown function	42
Number of genes (hypothetical and annotated)	39,114
Percent of hypothetical and annotated genes with unknown function	59
Gene with the most exons	Titin (234 exons)
Average gene size	27 kbp
Most gene-rich chromosome	Chr. 19 (23 genes/Mb)
Least gene-rich chromosomes	Chr. 13 (5 genes/Mb), Chr. Y (5 genes/Mb)
Total size of gene deserts (>500 kb with no annotated genes)	605 Mbp
Percent of base pairs spanned by genes	25.5 to 37.8*
Percent of base pairs spanned by exons	1.1 to 1.4*
Percent of base pairs spanned by introns	24.4 to 36.4*
Percent of base pairs in intergenic DNA	74.5 to 63.6*
Chromosome with highest proportion of DNA in annotated exons	Chr. 19 (9.33)
Chromosome with lowest proportion of DNA in annotated exons	Chr. Y (0.36)
Longest intergenic region (between annotated + hypothetical genes)	Chr. 13 (3,038,416 bp)
Rate of SNP variation	1/1250 bp

*In these ranges, the percentages correspond to the annotated gene set (26, 383 genes) and the hypothetical + annotated gene set (39,114 genes), respectively.

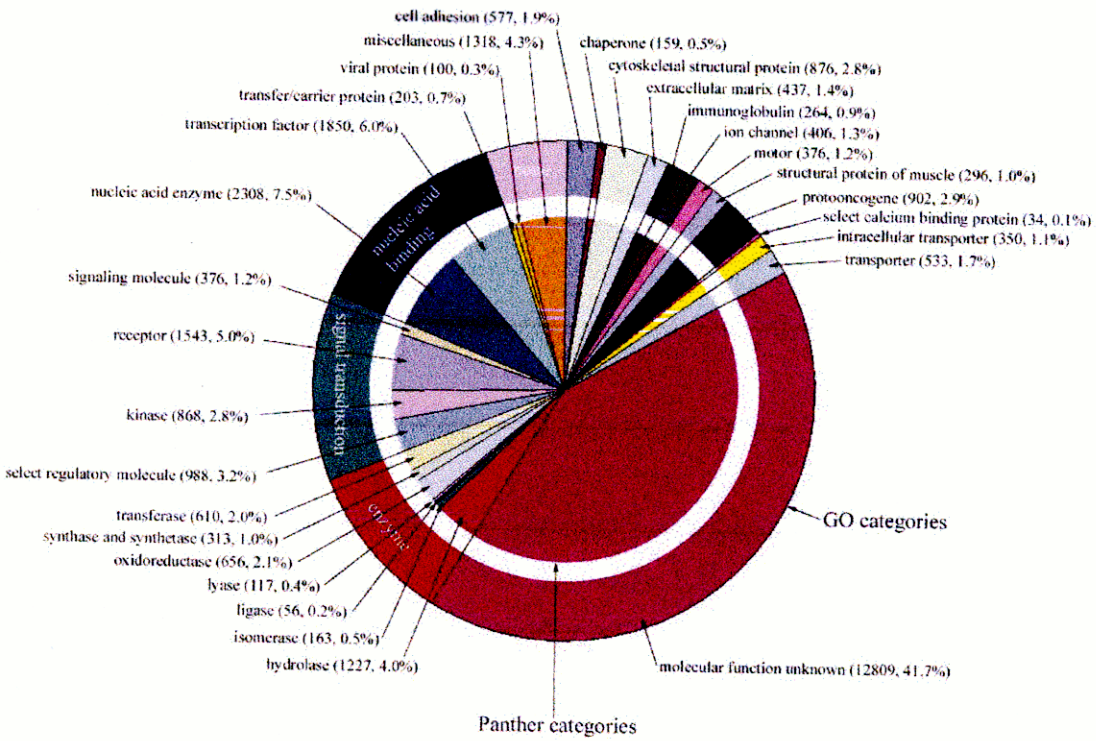


Fig. 15. Distribution of the molecular functions of 26,383 human genes. Each slice lists the numbers and percentages (in parentheses) of human gene functions assigned to a given category of molecular function. The outer circle shows the assignment to molecular function categories in the Gene Ontology (GO) (*179*), and the inner circle shows the assignment to Celera's Panther molecular function categories (*116*).

The HUMAN GENOME CONTAINS The SAME NUMBER OF GENES AS A WEED!!

Table 23 Properties of genome and proteome in essentially completed eukaryotic proteomes

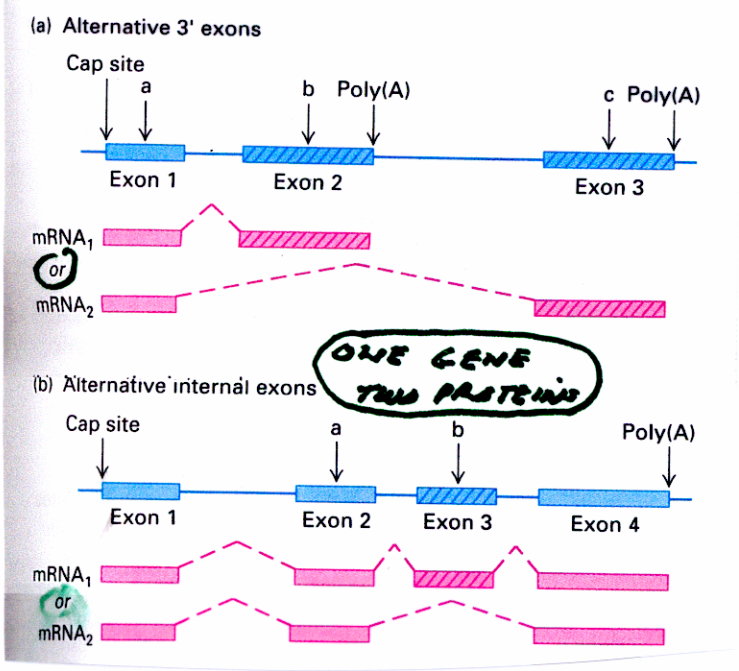
	Human	Fly	Worm	Yeast	Mustard weed
Number of identified genes	~32,000*	13,338	18,266	6,144	25,706
% with InterPro matches	51	56	50	50	52
Number of annotated domain families	1,262	1,035	1,014	851	1,010
Number of InterPro entries per gene	0.53	0.84	0.63	0.6	0.62
Number of distinct domain architectures	1,695	1,036	1,018	310	–
Percentage of 1-1-1-1	1.40	4.20	3.10	9.20	–
% Signal sequences	20	20	24	11	–
% Transmembrane proteins	20	25	28	15	–
% Repeat-containing	10	11	9	5	–
% Coiled-coil	11	13	10	9	–

The numbers of distinct architectures were calculated using SMART[339] and the percentages of repeat-containing proteins were estimated using Prospero[452] and a *P*-value threshold of 10^{-5}. The protein sets used in the analysis were taken from http://www.ebi.ac.uk/proteome/ for yeast, worm and fly. The proteins from mustard weed were taken from the TAIR website (http:// www.arabidopsis.org/) on 5 September 2000. The protein set was searched against the InterPro database (http://www.ebi.ac.uk/interpro/) using the InterProscan software. Comparison of protein sequences with the InterPro database allows prediction of protein families, domain and repeat families and sequence motifs. The searches used Pfam release 5.2[307], Prints release 26.1[326], Prosite release 16[327] and Prosite preliminary profiles. InterPro analysis results are available as Supplementary Information. The fraction of 1-1-1-1 is the percentage of the genome that falls into orthologous groups composed of only one member each in human, fly, worm and yeast.

* The gene number for the human is still uncertain (see text). Table is based on 31,778 known genes and gene predictions.

BUT REMEMBER – PROTEINS PRODUCE the Phenotype

∴ The potential to make MANY MORE Thousands of Proteins EXISTS ..... Alternate Splicing!



▲ **FIGURE 9-2 Two examples of complex eukaryotic transcription units and the effect of mutations on expression of the encoded proteins.** The RNA transcribed from a complex transcription unit (blue) can be processed in alternative ways to yield two or more functional monocistronic mRNAs. Dashed lines indicate spliced-out introns. (a) A complex transcription unit whose primary transcript has two poly(A) sites produces two mRNAs with alternative 3′ exons. (b) A complex transcription unit whose primary transcript undergoes exon skipping during processing produces alternative mRNAs with the same 5′ and 3′ exons. In this example, some cell types would express the mRNA including exon 3, whereas in other cell types, exon 2 is spliced to exon 4, producing an mRNA lacking exon 3 and the protein sequence it encodes. In (a) and (b), mutations (designated *a*) within exons shared by the alternative mRNAs (solid red) affect the proteins encoded by both alternatively processed mRNAs. In contrast, mutations (designated *b* and *c*) within exons unique to one of the alternatively processed mRNAs (red with diagonal lines) affect only the protein encoded by that mRNA.

HUMAN GENES CAN BE VERY LARGE!

WITH MANY INTRONS!

articles

Table 21 Characteristics of human genes

	Median	Mean	Sample (size)
Internal exon	122 bp	145 bp	RefSeq alignments to draft genome sequence, with confirmed intron boundaries (43,317 exons)
Exon number	7	8.8	RefSeq alignments to finished sequence (3,501 genes)
Introns	1,023 bp	3,365 bp	RefSeq alignments to finished sequence (27,238 introns)
3′ UTR	400 bp	770 bp	Confirmed by mRNA or EST on chromosome 22 (689)
5′ UTR	240 bp	300 bp	Confirmed by mRNA or EST on chromosome 22 (463)
Coding sequence (CDS)	1,100 bp 367 aa	1,340 bp 447 aa	Selected RefSeq entries (1,804)
Genomic extent	14 kb	27 kb	Selected RefSeq entries (1,804)

Median and mean values for a number of properties of human protein-coding genes. The 1,804 selected RefSeq entries were those that could be unambiguously aligned to finished sequence over their entire length.

Table 7.7: Average sizes of exons and introns in human genes

Gene product	Size of gene (kb)	Number of exons	Average size of exon (bp)	Average size of intron (bp)
$tRNA^{tyr}$	0.1	2	50	20
Insulin	1.4	3	155	480
β-Globin	1.6	3	150	490
Class I HLA	3.5	8	187	260
Serum albumin	18	14	137	1100
Type VII collagen	31	118	77	190
Complement C3	41	29	122	900
Phenylalanine hydroxylase	90	26	96	3500
Factor VIII	186	26	375	7100
CFTR (cystic fibrosis)	250	27	227	9100
Dystrophin	2400	79	180	30 000

NOTE – SMALL EXONS LARGE IN INTRONS!

MANY DISEASE GENES HAVE BEEN IDENTIFIED

AND what their Proteins Are!

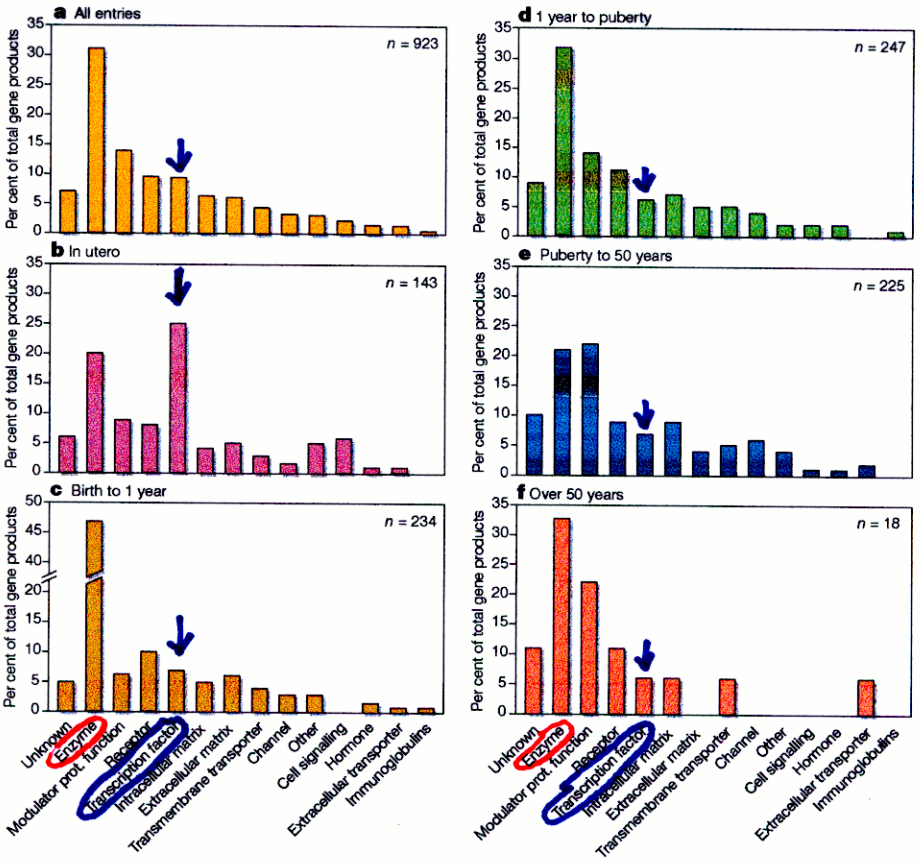


Figure 1 The functions of the protein products of disease genes. **a**, The entire disease gene set. **b–f**, Disease genes stratified according to the typical age of onset of the disease phenotype. The fraction of disease genes encoding transcription factors in the *in utero* onset disorders (25%) differs from the fraction encoding transcription factors for disorders with onset after birth (6%; $\chi^2 = 49.4$, $P < 0.001$). Similarly, the fraction of disease genes encoding enzymes causing a disorder with onset in the first year of life (47%) is different from the fraction encoding enzymes causing disorders with other ages of onset (25.8%; $\chi^2 = 35.8$, $P < 0.001$).

AND HOW THEY ARE INHERITED, WHEN DISEASE BEGINS, & HOW Life Expectancy Affected!!

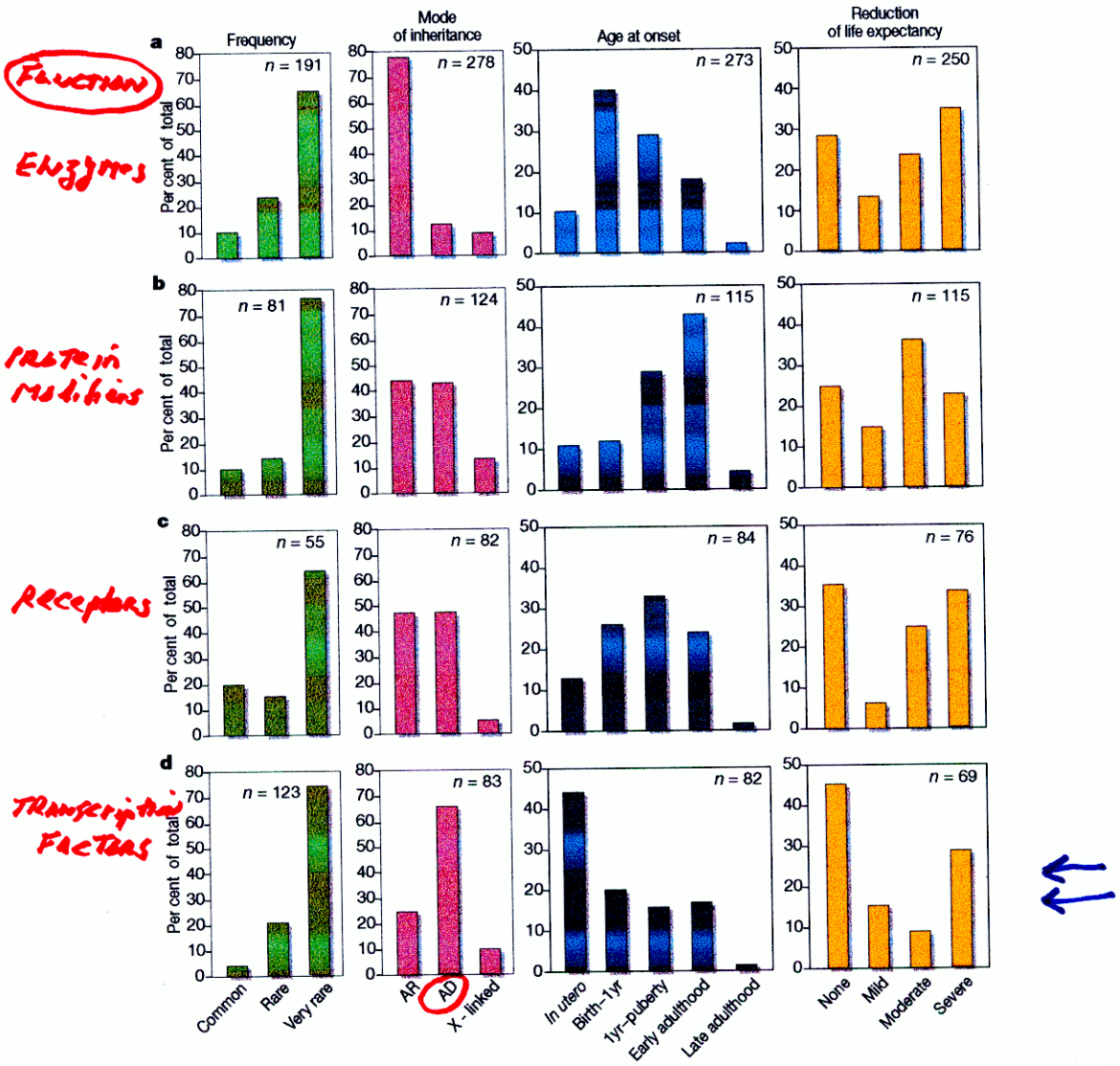


Figure 2 Characteristics of disease arranged by function of the protein encoded by the disease gene. **a**, Disease genes encoding enzymes; **b**, disease genes encoding modifiers of protein function; **c**, disease genes encoding receptors; **d**, disease genes encoding transcription factors. The columns of disease features are labelled at the top. AR, autosomal recessive; AD, autosomal dominant; early adulthood, puberty to <50 years; late adulthood, >50 years.

Clearly – These Genes Can BE ASSAYED using Probes/Markers!

What Makes A MOUSE A MOUSE AND A PERSON A PERSON?!

75,000,000 years apart

TABLE 22.1 Comparison of Mice and Humans

Trait	Mice	Humans
Average weight	30 g	77,000 g (170 lb)
Average length	10 cm (without tail)	175 cm
Genome size	~3,000,000,000 bp	~3,000,000,000 bp
Haploid gene number	~100,000	~100,000
Number of chromosomes	19 autosomes + X and Y	22 autosomes + X and Y
Gestation period	3 weeks	38 weeks (8.9 months)
Age at puberty	5–6 weeks	624–728 weeks (12–14 years)
Estrus cycle	4 days	28 days
Life span	2 years	78 years

LARGE BLOCKS OF MOUSE GENES ARE FOUND IN HUMAN CHROMOSOMES!

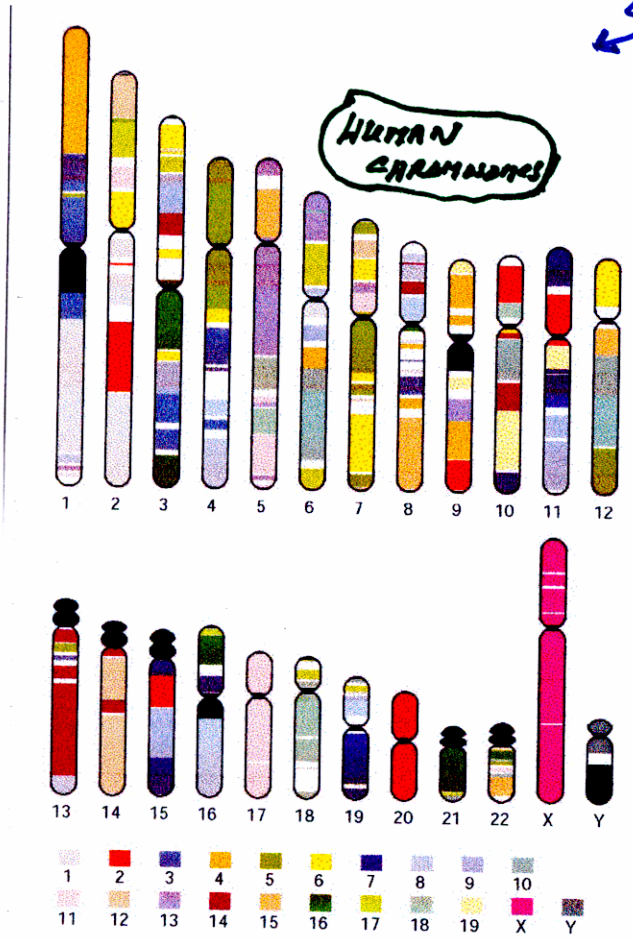


Figure 46 Conserved segments in the human and mouse genome. Human chromosomes, with segments containing at least two genes whose order is conserved in the mouse genome as colour blocks. Each colour corresponds to a particular mouse chromosome. Centromeres, subcentromeric heterochromatin of chromosomes 1, 9 and 16, and the repetitive short arms of 13, 14, 15, 21 and 22 are in black.

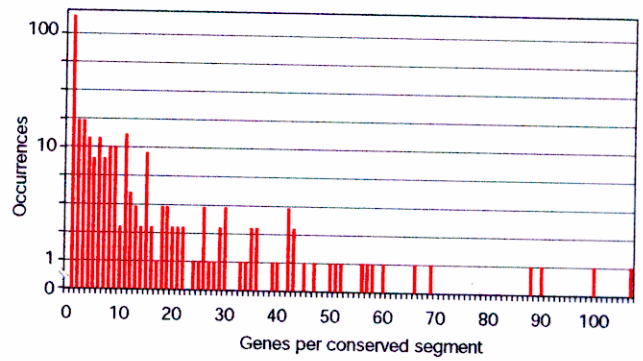


Figure 47 Distribution of number of genes per conserved segment between human and mouse genomes.

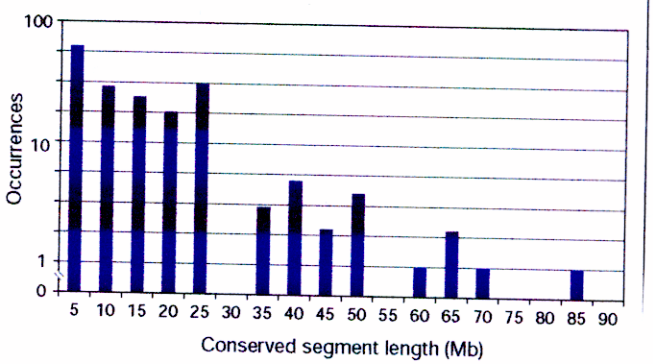


Figure 48 Distribution of lengths (in 5-Mb bins) of conserved segments between human and mouse genomes, omitting singletons.

99% of ALL HUMAN GENES ARE FOUND in the MOUSE GENOME!

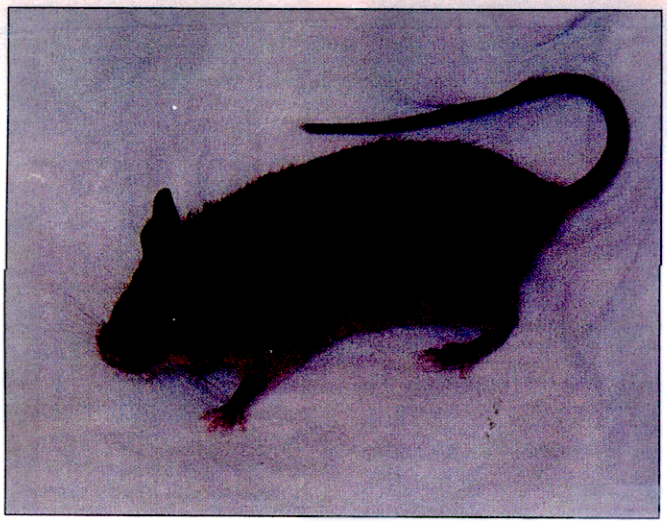
Initial sequencing and comparative analysis of the mouse genome

Mouse Genome Sequencing Consortium*

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

NATURE | VOL 420 | 5 DECEMBER 2002 | www.nature.com/nature

WHAT MAKES a HUMAN?



A member of the 129 strain of inbred mice commonly used in targeted mutagenesis studies.

MOUSE CHROMOSOMES

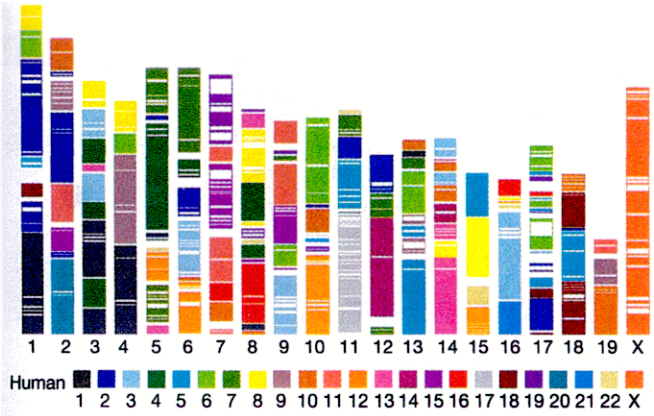


Figure 3 Segments and blocks >300 kb in size with conserved synteny in human are superimposed on the mouse genome. Each colour corresponds to a particular human chromosome. The 342 segments are separated from each other by thin, white lines within the 217 blocks of consistent colour.

& VICE VERSA!

Table 10 **Gene count in human and mouse genomes**

Genome feature	Human		Mouse	
	Initial (Feb. 2001)	Current (Sept. 2002)	Initial* (this paper)	Extended† (this paper)
Predicted transcripts	44,860	27,048	28,097	29,201
Predicted genes	31,778	22,808	22,444	22,011
Known cDNAs	14,882	17,152	13,591	12,226
New predictions	16,896	5,656	8,853	9,785
Mean exons/transcript‡	4.2 (3)	8.7 (6)	8.2 (6)	8.4 (6)
Total predicted exons	170,211	198,889	191,290	213,562

MICE ARE POWERFUL "TOOLS" FOR STUDYING HUMAN DISEASES

WHAT IS THE OVERALL ORGANIZATION OF THE HUMAN GENOME?

TABLE 9-1 Classification of Eukaryotic DNA

Protein-coding genes
Solitary genes
Duplicated and diverged genes (functional gene families and nonfunctional pseudogenes)
Tandemly repeated genes encoding rRNA, 5S rRNA, tRNA, and histones
Repetitious DNA
Simple-sequence DNA
Moderately repeated DNA (mobile DNA elements)
Transposons
Viral retrotransposons
Long interspersed elements (LINES; nonviral retrotransposons)
Short interspersed elements (SINES; nonviral retrotransposons)
Unclassified spacer DNA

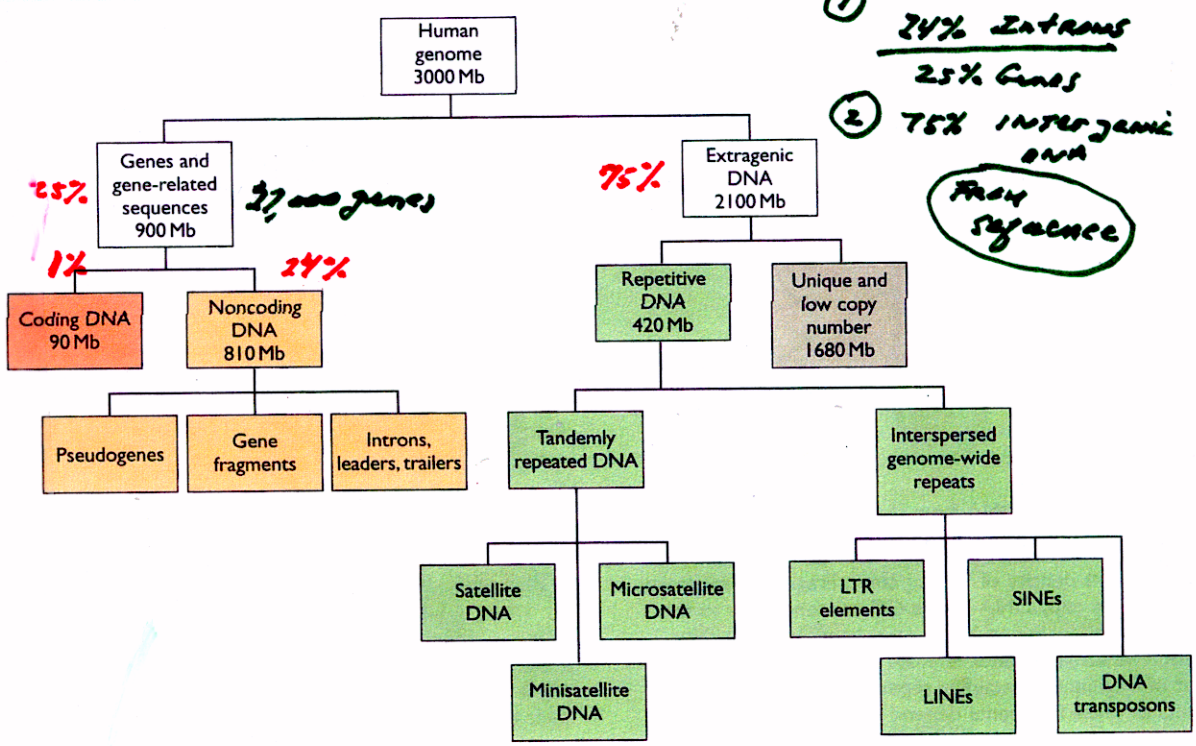
UNIQUE SEQUENCES 65%

REPEATED SEQUENCES 35%

→ Genes!

→ NO KNOWN FUNCTION!

Box 6.4: The organization of the human genome



Based on Strachan and Read (1996).

① 1% Exons
24% Introns
25% Genes
② 75% INTERGENIC DNA
FROM SEQUENCE

THE HUMAN GENOME CONTAINS DIFFERENT CLASSES OF REPEATED SEQUENCES

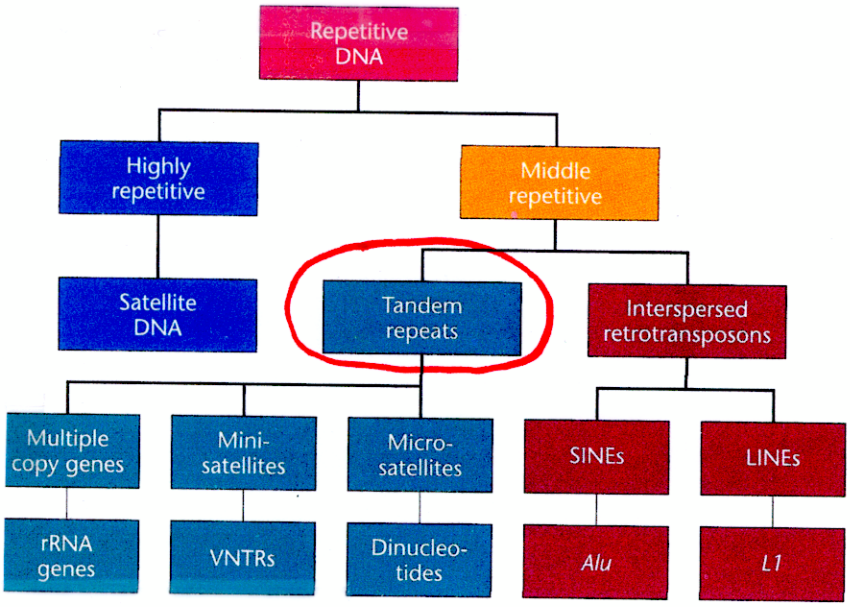


FIGURE 19.16 An overview of the various categories of repetitive DNA.

VNTRs Useful for Individual Identity

Table 7.11: Major classes of tandemly repeated human DNA

USEFUL AS VNTRS

Class	Size of repeat	Major chromosomal location(s)
'Megasatellite' DNA (blocks of hundreds of kb in some cases)	several kb	Various locations on selected chromosomes
RS447	4.7 kb	~50–70 copies on 4p15 plus several copies on distal 8p
untitled	2.5 kb	~400 copies on 4q31 and 19q13
untitled	3.0 kb	~50 copies on the X chromosome
Satellite DNA (blocks often from 100 kb to several Mb in length)	5–171 bp	Especially at centromeres
α (alphoid DNA)	171 bp	Centromeric heterochromatin of all chromosomes
β (*Sau3* A family)	68 bp	Centromeric heterochromatin of 1, 9, 13, 14, 15, 21, 22 and Y
Satellite 1 (AT-rich)	25–48 bp	Centromeric heterochromatin of most chromosomes and other heterochromatic regions
Satellites 2 and 3	5 bp	Most, possibly all, chromosomes
Minisatellite DNA (blocks often within the 0.1–20 kb range)	6–64 bp	At or close to telomeres of all chromosomes
telomeric family	6 bp	All telomeres
hypervariable family	9–64 bp	All chromosomes, often near telomeres
Microsatellite DNA (blocks often less than 150 bp)	1–4 bp	Dispersed throughout all chromosomes

HUMAN DNA SEQUENCE ORGANIZATION

REPETITIVE SEQUENCES Scattered Thru GENOME!

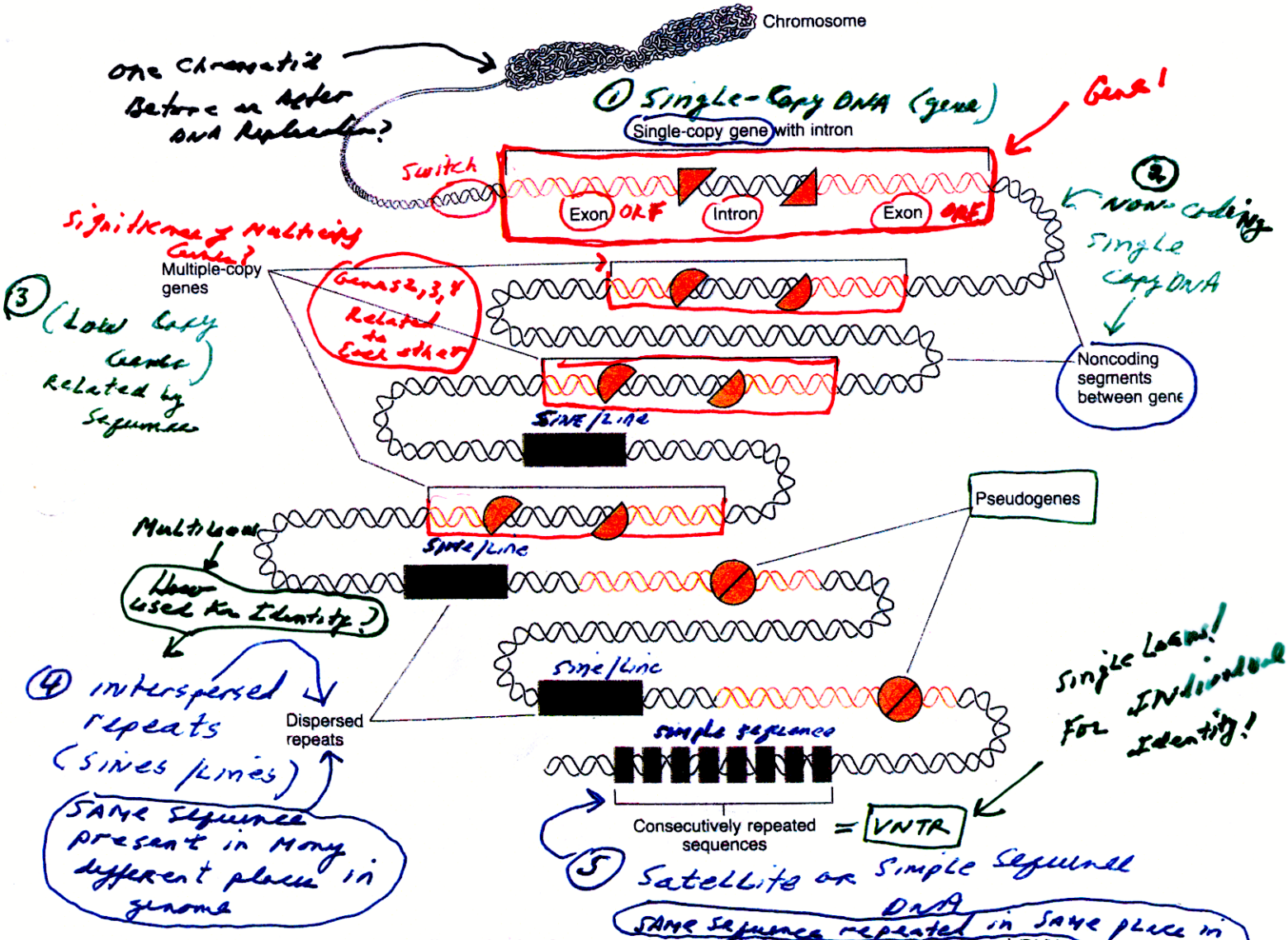


Figure 7.1 Occurrence of different kinds of unique and repeated DNA segments on chromosomal DNA.

① Single-Copy Gene Sequence & ② Single-Copy Spacer Sequence
③ Low-Copy Gene Sequence (~ 10 copies)
④ Dispersed Repeated Sequences
⑤ TANDEM Repeated Sequences
} DNA Fingerprinting "Tools"

Note:
Gene order Reflects DNA Sequence!!

VNTRs Are Tandem Repeats & Give Rise to Allelic Variability

Cleavage sites conserved

Variable # Tandem Repeats

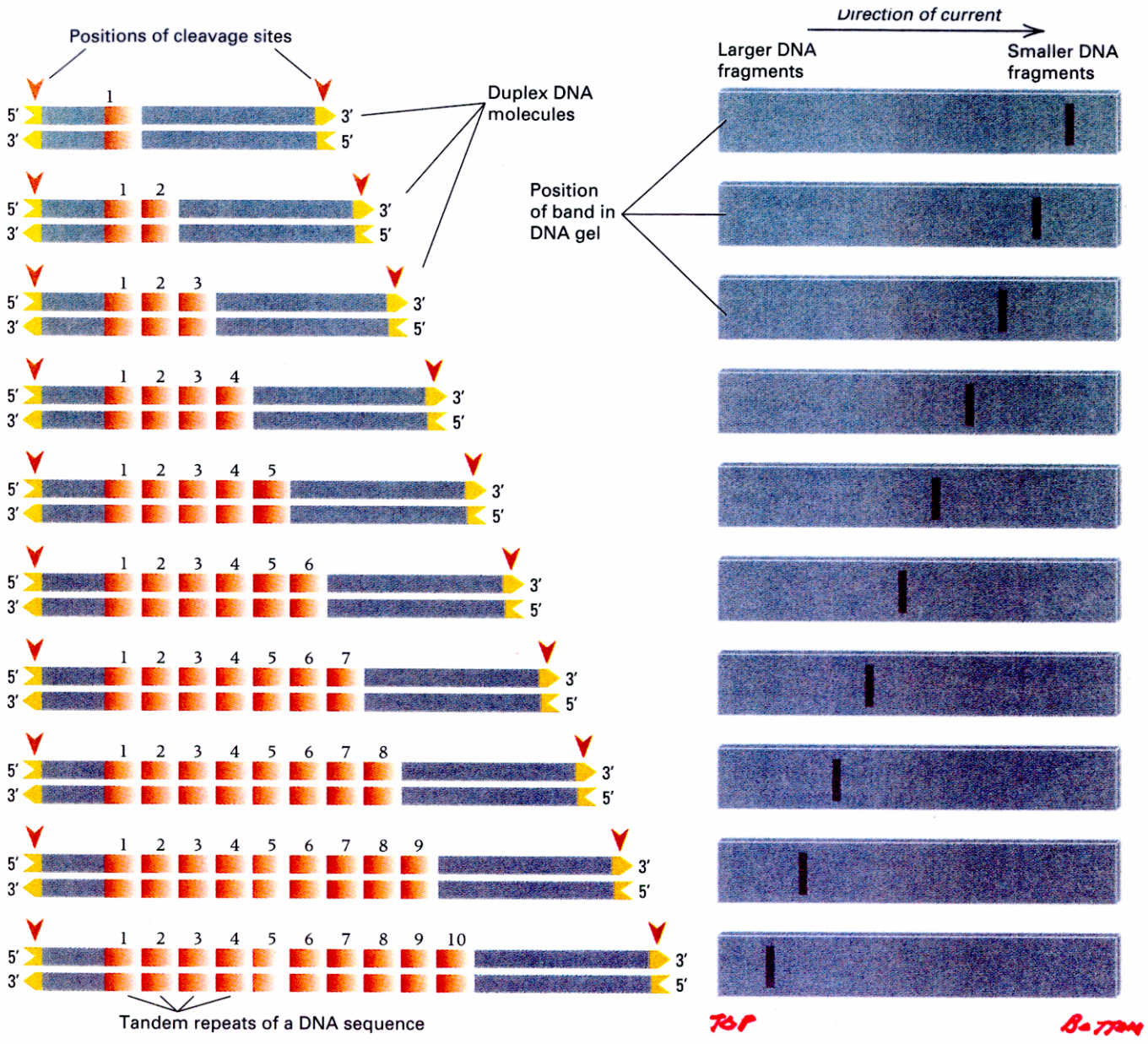


Figure 2.28 In a simple tandem repeat polymorphism (STRP), the alleles in a population differ in the number of copies of a short sequence (typically 2–60 bp) that is repeated in tandem along the DNA molecule. This example shows alleles in which the repeat number varies from 1 to 10. Cleavage at restriction sites flanking the STRP yields a unique fragment length for each allele. The alleles can also be distinguished by the size of the fragment amplified by PCR using primers that flank the STRP.

Size Varies Between Conserved Regions

Like an Accordian — AT SAME LOCUS on Chromosome Location

VNTRs Are Sequence-Specific Tandem Repeats Present Throughout the Genome

Repeat = (GGAGG)(GGAGG)$_n$

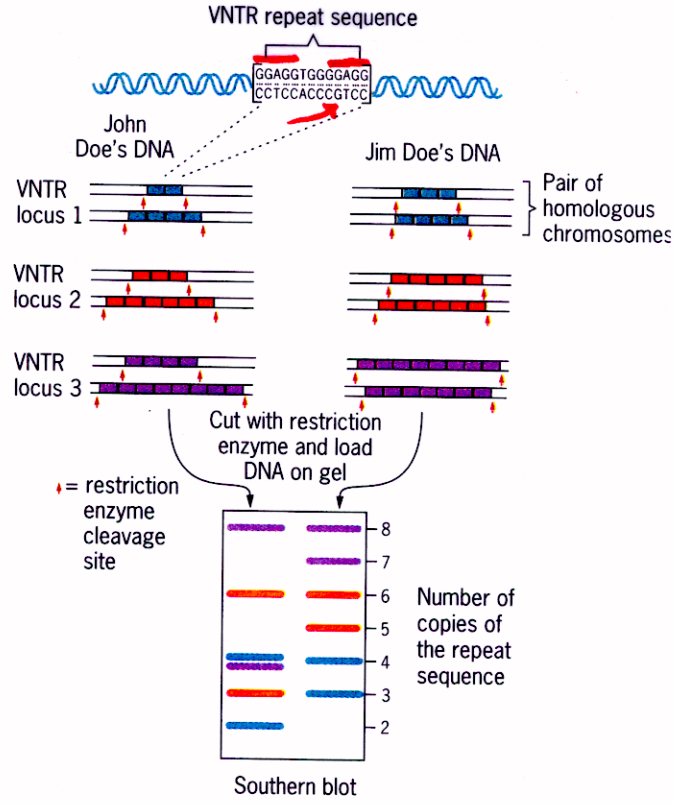


Figure 22.8 Simplified diagram of the use of variable number tandem repeats in preparing DNA fingerprints.

How are these Experiments done?

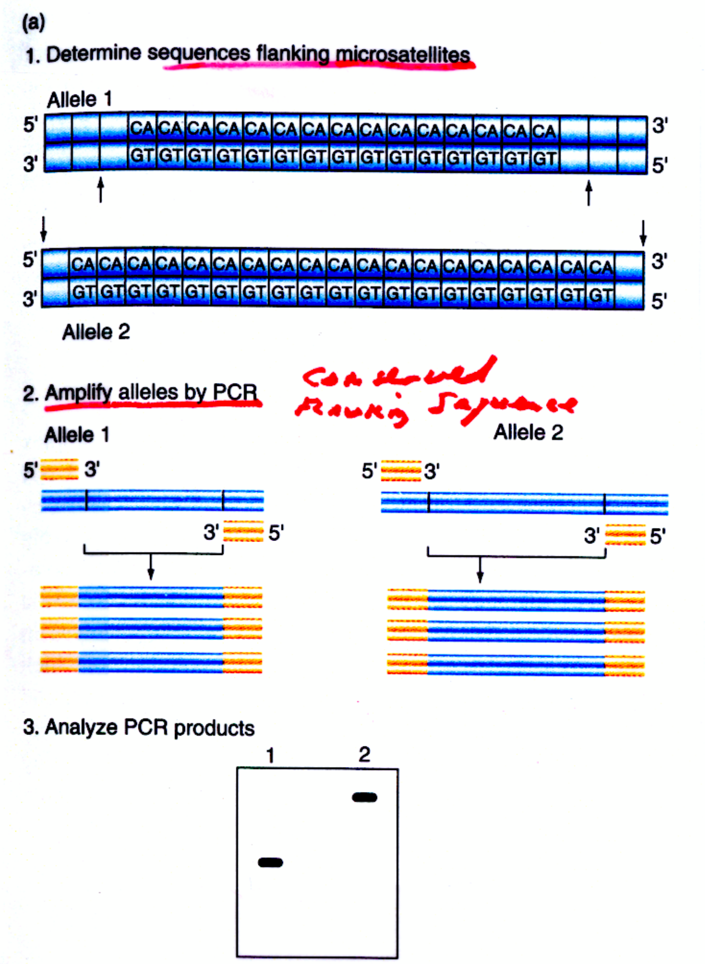
MANY different types! Differ in Sequence & Location!

VARY in Repeat Length (2bp & 4bp!)

VNTRs Generally Have Many Different Alleles at a Given Locus

Repeat = (CA)n

Population of Alleles!



Conserved Flanking Sequence

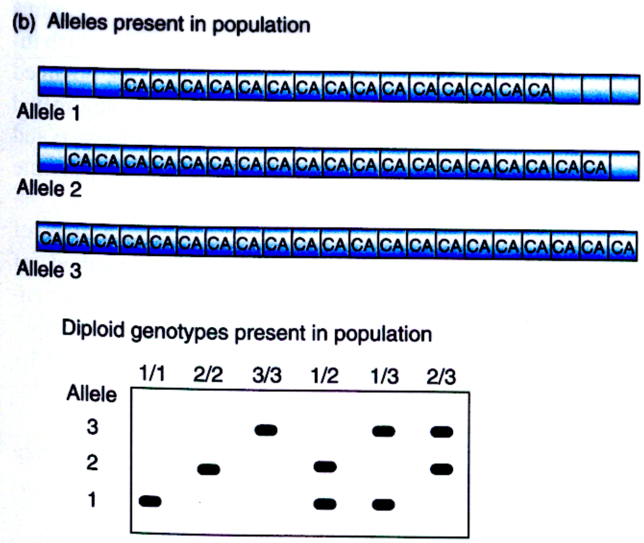
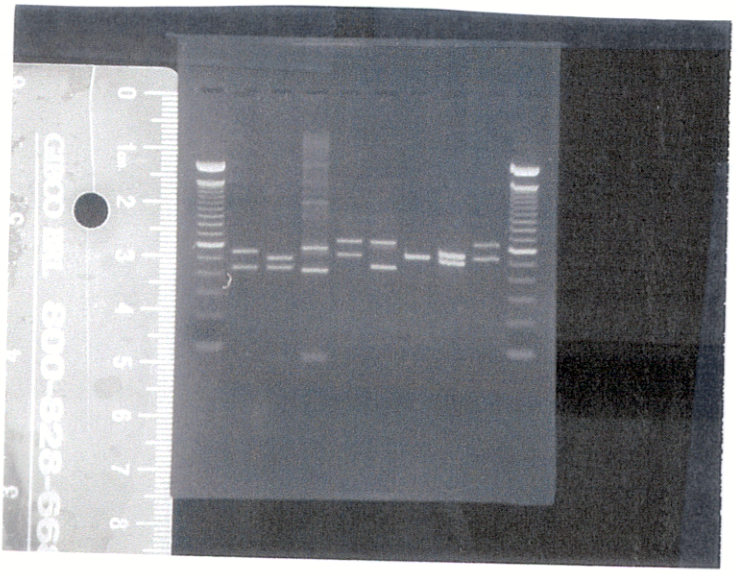


Figure 9.12 Detection of microsatellite polymorphisms by PCR and gel electrophoresis. (a.1) Microsatellite alleles differ from one another in length. (2) Sequence determination from both sides of a microsatellite enables the construction of primers that can be used to amplify the microsatellite by PCR. (3) Gel electrophoresis and ethidium bromide staining distinguish the alleles from each other. (b) Microsatellites are often highly polymorphic with many different alleles present in a population. With just three alleles, there are six possible genotypes. With N (any number of) alleles, there will be $\frac{N}{2}(N + 1)$ genotypes.

Useful for Comparing Individuals + Populations (e.g., HGP?)

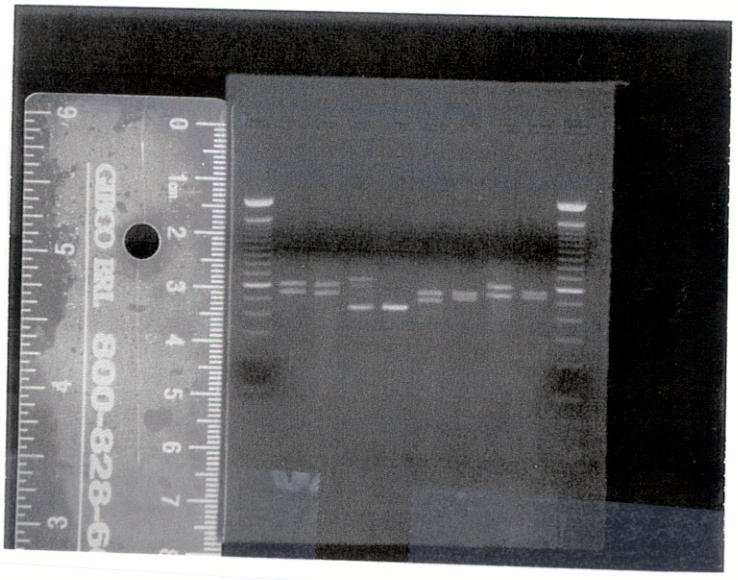
are there races?

Alleles at one VNTR Locus in HG70A Class

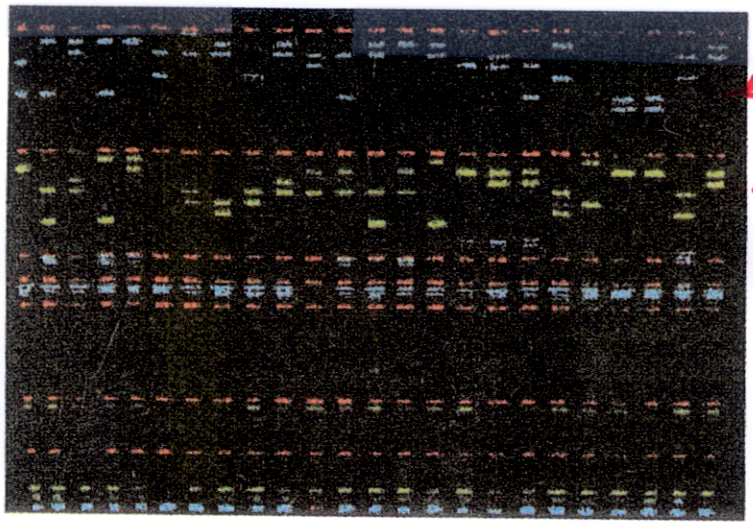


Most individuals are Heterozygous

VNTR DIS80



USING "MICROSATELLITE" VNTRs IN GENETIC PROFILING



How many VNTRs scored for in this gel?

Figure 6.17 The use of microsatellite analysis in genetic profiling.

In this example, microsatellites located on the short arm of chromosome 6 have been amplified by PCR. The PCR products are labeled with a blue or green fluorescent marker and run in a polyacrylamide gel, each lane showing the genetic profile of a different individual. No two individuals have the same genetic profile because each person has a different set of microsatellite alleles, the alleles giving rise to bands of different sizes after PCR. The red bands are DNA size markers. Image supplied courtesy of PE Biosystems, Warrington, UK, and reproduced with permission.

Multiple Single-Locus VNTRs used in a Criminal Case

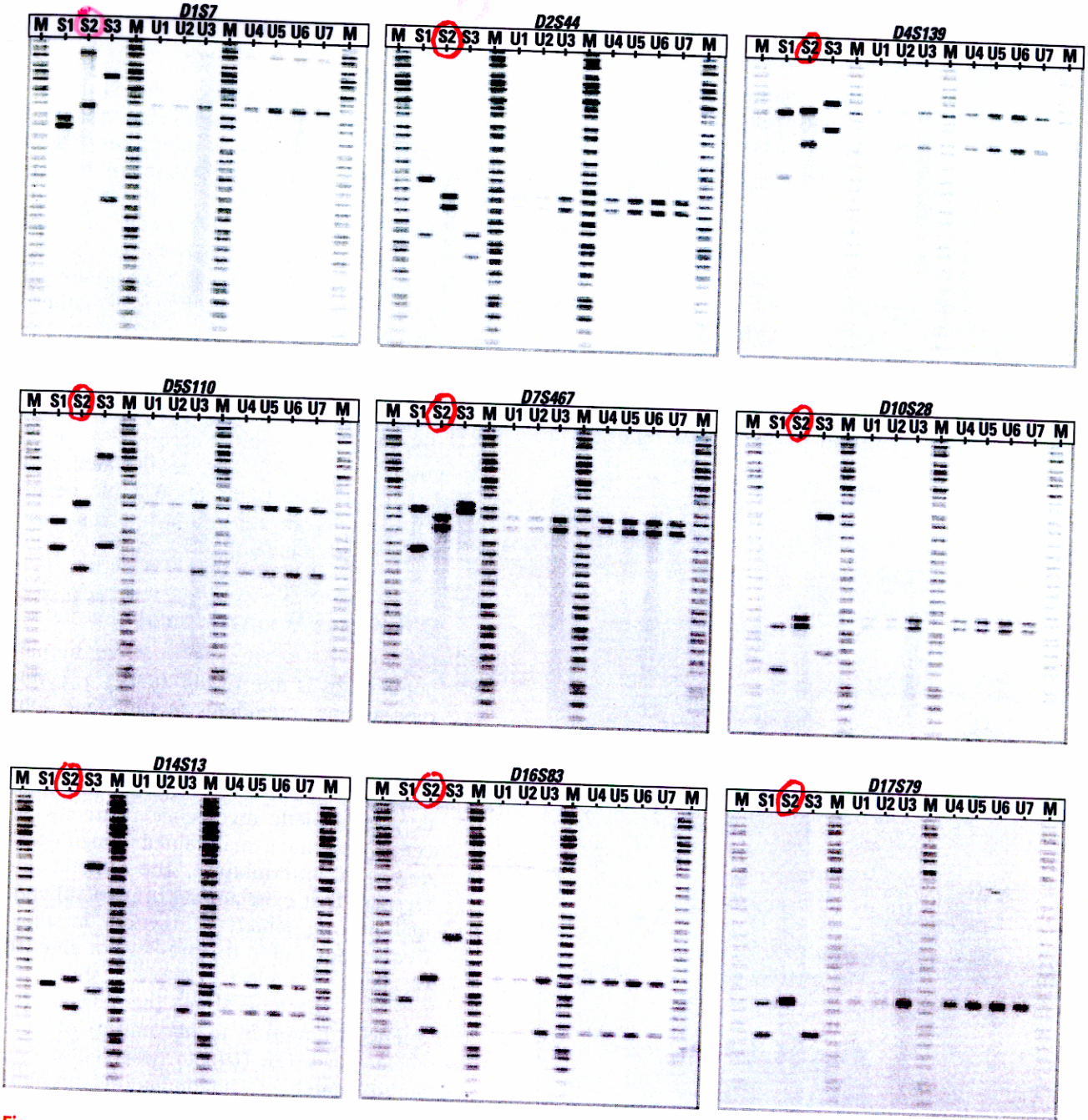


Figure 17.13 An example of DNA typing in a criminal case. Each panel is the result of DNA typing for a different VNTR. The lanes marked S1, S2, and S3 contain DNA from blood samples of three male suspects; those in columns U1 through U7 contain DNA from semen samples collected from seven female victims of rape. The lanes marked M contain molecular-weight markers. In each case, the DNA from suspect S2 matches the samples obtained from the victims. [Courtesy of Steven J. Redding, Office of the Hennepin County District Attorney, Minneapolis, and Lowell C. Van Berkom and Carla J. Finis, Minnesota Bureau of Criminal Apprehension.]

Who done it!! But also who is innocent?

ORIGINS OF VNTR VARIABILITY

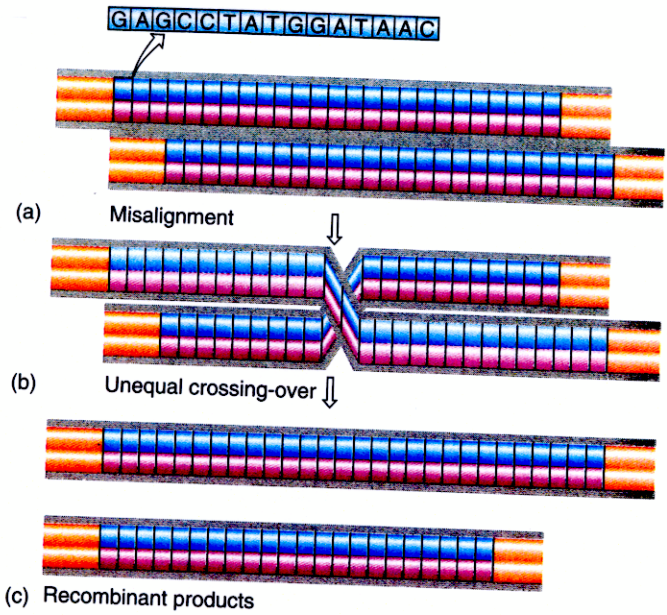
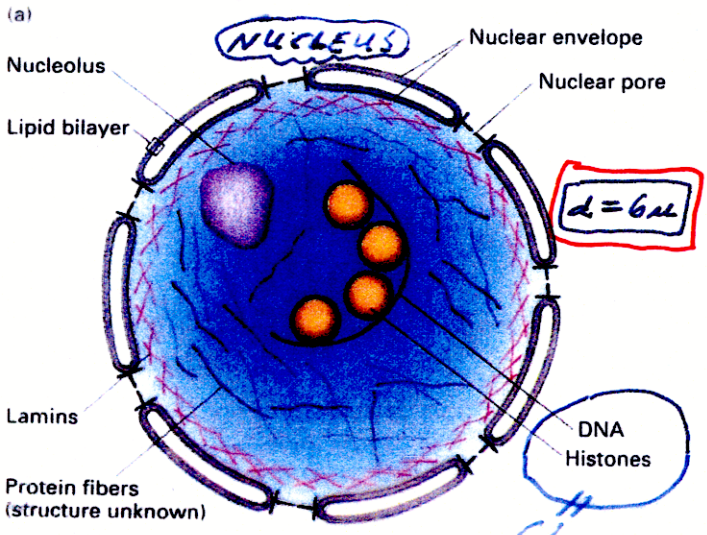


Figure 9.4 Minisatellites are highly polymorphic because of their potential for misalignment and unequal crossing-over. Minisatellites are composed of relatively long tandem repeating units of identical sequence. (a) Misalignment and (b) unequal crossing-over produce (c) recombinant products that contain different numbers of repeating units than either parental locus; each new recombinant product is a new allele.

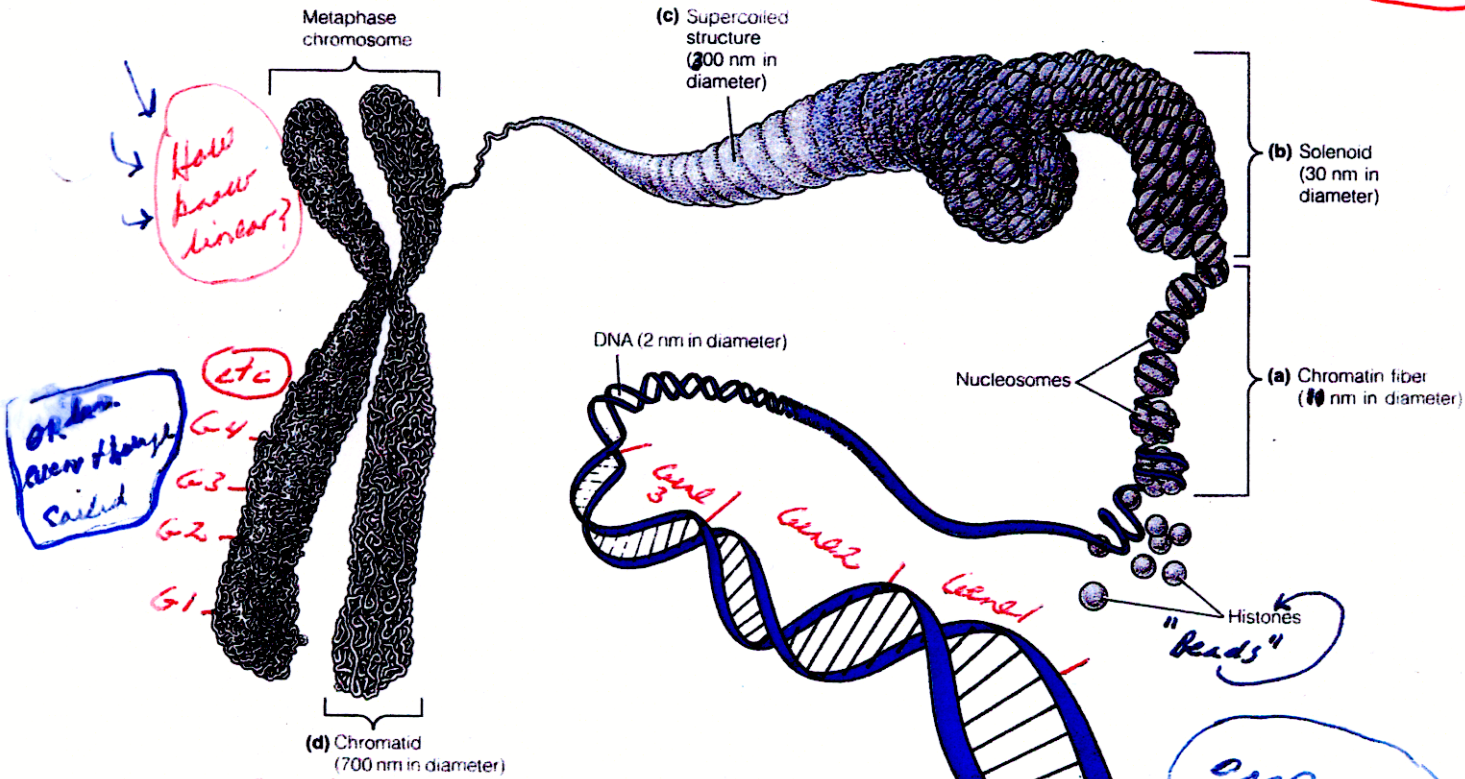
During Crossing over in Meiosis

The HUMAN GENOME IS PACKAGED INTO CHROMOSOMES



◀ **Figure 8-60** (a) Generalized diagram of eukaryotic nucleus showing identifiable structural elements. (b) Electron micrograph of nuclear pore–lamin complexes isolated from rat nuclei. Nuclear pores *(arrows)* are embedded in fibrous lamin proteins (la). (c) A transmission electron micrograph of a whole mount of a HeLa cell, showing a skeletal network within the nucleus. The cell was prepared by removing lipids and soluble factors with a mild detergent. The remaining skeletal struture was then treated to remove most of the DNA. The sample was fixed with glutaraldehyde, but no heavy-metal shadowing was done. [See S. Penman et al., 1982, *Cold Spring Harbor Symp. Quant. Biol.* 46:1013.] *Photograph (b) courtesy of N. Dwyer. Reproduced from the* Journal of Cell Biology, *1976, by copyright permission of Rockefeller University Press. Photograph (c) courtesy of S. Penman.*

Nucleus = 6μ diameter
DNA = 6×10^6 m in length!!



each chromatid is a duplicate of the other

Hundreds of Genes per chromosome

DNA tightly coiled why?

Dynamic / Coiling Chromosomes

Figure 28.10
Levels of chromatin structure. The beaded string structure is a 10-nm fiber, which folds into a "solenoidal" 30-nm fiber with about six nucleosomes per turn. This can further fold to form thick 200-nm fibers that can be observed in electron micrographs of chromosomes or nuclei.

HISTONE PROTEINS INTERACT WITH DNA TO MAKE A CHROMOSOME

DNA division occured?

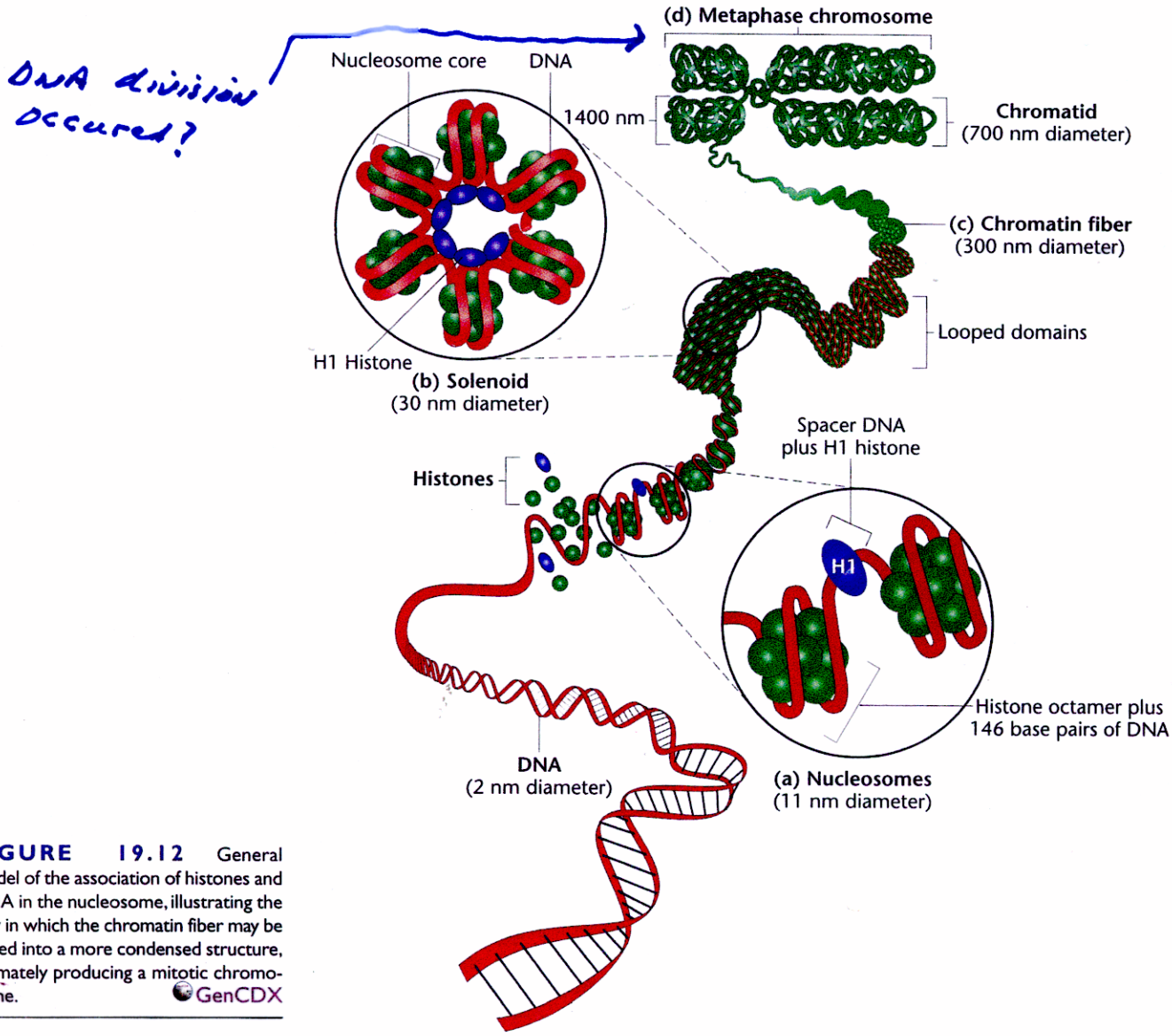


FIGURE 19.12 General model of the association of histones and DNA in the nucleosome, illustrating the way in which the chromatin fiber may be coiled into a more condensed structure, ultimately producing a mitotic chromosome. GenCDX

Significance of coiling?

CHROMOSOMES CAN BE CHARACTERIZED USING A MICROSCOPE AND CONSTRUCTING A Karyotype

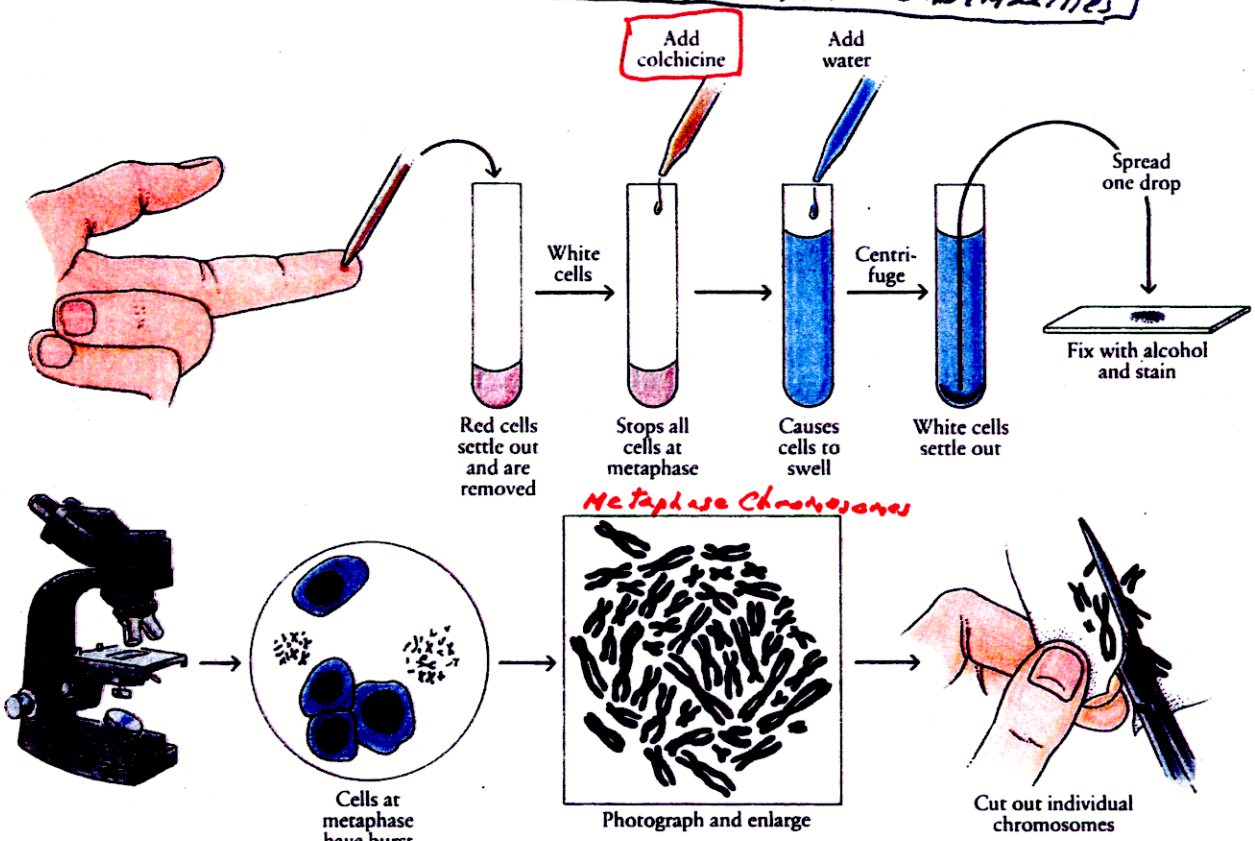
Preparation of a Karyotype

AT METAPHASE

why "short" then?

Chromosome typing for the identification of gross chromosomal abnormalities is being carried out at an increasing number of genetic counseling centers throughout the United States. The result of the procedure is a graphic display of the chromosome complement, known as a karyotype. The chromosomes shown in a karyotype are mitotic metaphase chromosomes, each consisting of two sister chromatids held together at their centromeres. To prepare a karyotype, cells in the process of dividing are interrupted at metaphase by the addition of colchicine, a drug that prevents the subsequent steps of mitosis from taking place by interfering with the spindle microtubules. After treating and staining, the chromosomes are photographed, enlarged, cut out, and arranged according to size. Chromosomes of the same size are paired according to centromere position, which results in different "arm" lengths. From the karyotype, certain abnormalities, such as an extra chromosome or piece of a chromosome, can be detected.

Karyotypes are Needed to detect Chromosomal Abnormalities



C- Metaphase Chromosomes

colchicine Metaphase!

Paste in order of diminishing size with centromere on pencil line

before or after DNA replication?

2 Chromatids after DNA replication = 2 DNA molecules

CHROMOSOMES HAVE STRUCTURES That ARE visible in Light and Electron Microscopes

Light Microscope

ight micrograph of
uman chromosomes
enlarged 600 times)

Scanning Electron Microscope

Electron micrographic 3-D image
(enlarged 30,000 times)

TRANSMISSION Electron Microscope

Electron micrograph
of fixed chromosome
(enlarged 30,000 times)

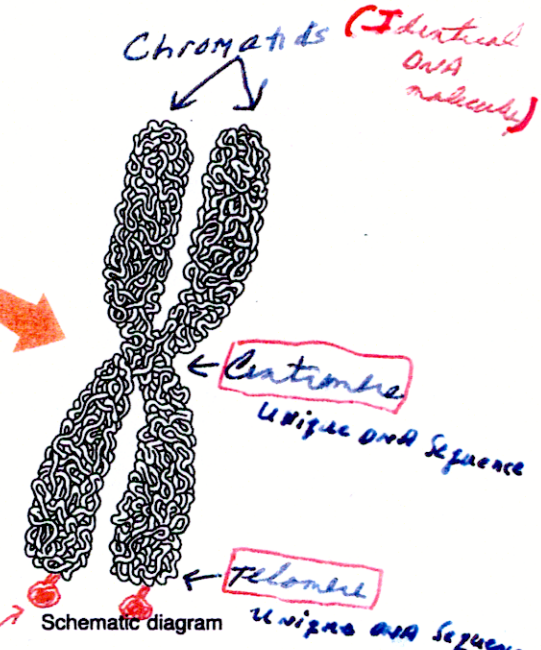


Figure 1.3 Human chromosomes.

Chromatids
centromere
telomere

A chromosome during division

EACH CHROMOSOME HAS A UNIQUE MORPHOLOGY + BANDING PATTERN

22 PAIRS OF AUTOSOMES (♂) X (♀) Y

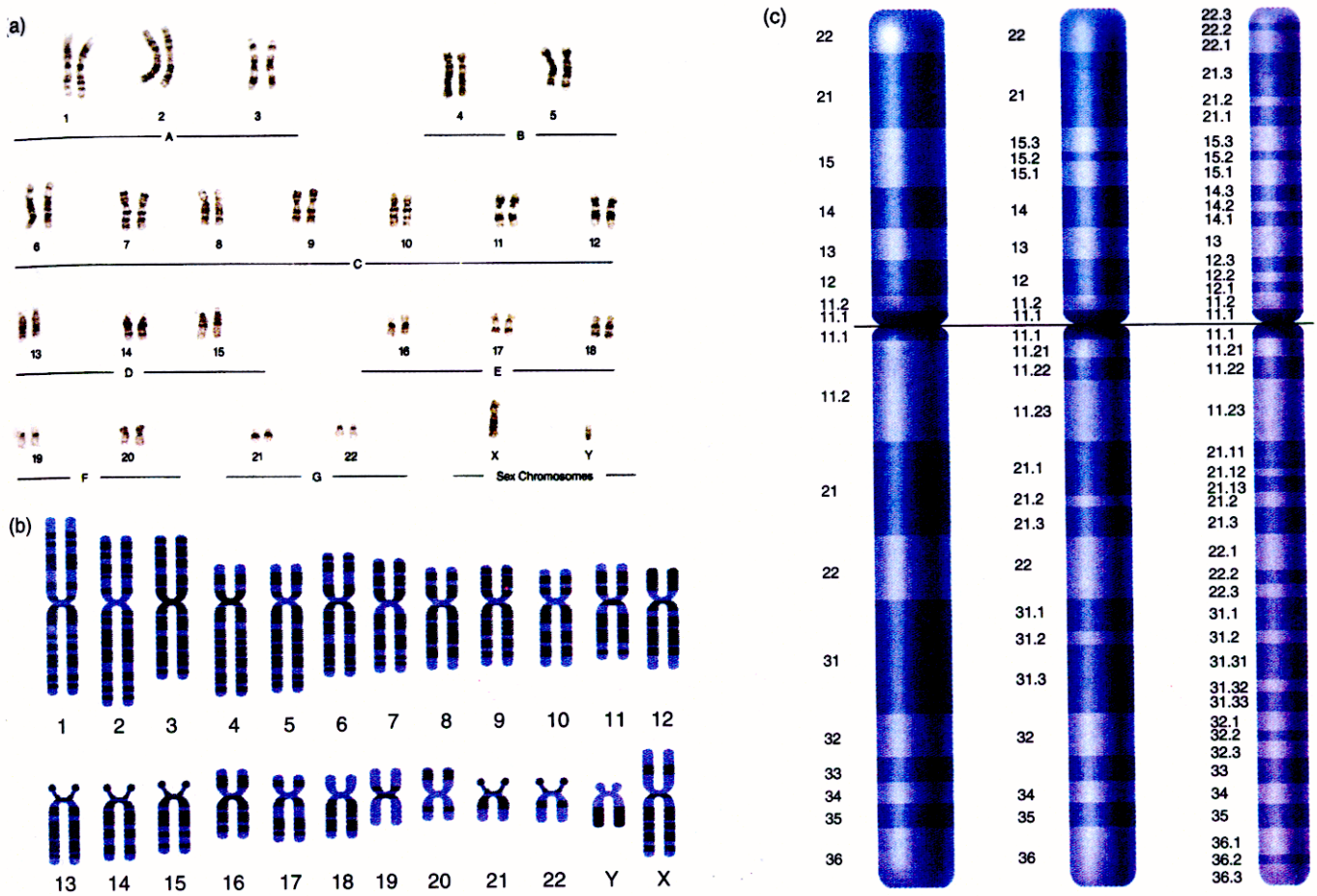


Figure 10.3 The human karyotype: Banding distinguishes the chromosomes. (a) Photograph of a complete set of human chromosomes at metaphase. Staining with Giemsa dye accentuates the bands and interbands. (b) Idiograms for the complete set of human chromosomes. An idiogram is an idealized diagram of the banding pattern associated with a stained chromosome. (c) Chromosome 7 at three different levels of banding resolution. As staining techniques improve, it becomes possible to resolve what previously appeared as a single band into a series of bands and interbands, producing more and more bands along each chromosome. Thus, at one resolution, 7q31 appears as one band. At a slightly higher resolution, 7q31 becomes two bands (7q31.1 and 7q31.3) flanking an interband (7q31.2); and at an even higher resolution, 7q31.3 itself appears as two bands (7q31.31 and 7q31.33) and an interband (7q31.32).

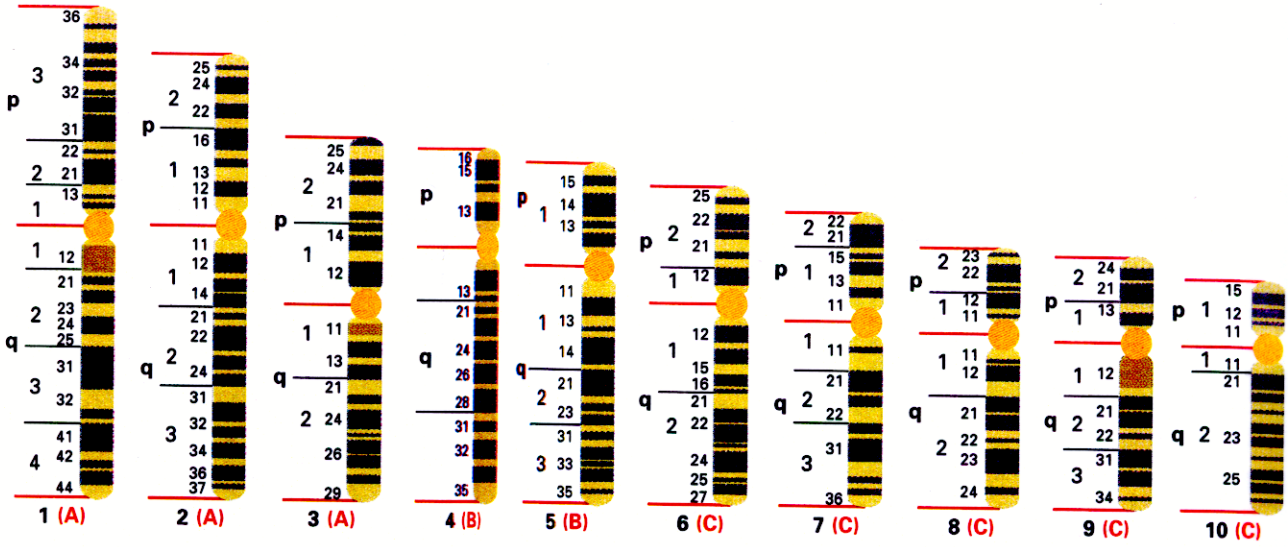
What causes banding patterns of chromosomes to be unique?

Size of bands?

CHROMOSOME NOMENCLATURE

Table 9.1 Conventional karyotype symbols used in human genetics

A–G	Chromosome groups
1–22	Autosome designations
X, Y	Sex-chromosome designations
p	Short arm of chromosome
q	Long arm of chromosome
ter	Terminal portion: pter refers to terminal portion of short arm, qter to terminal portion of long arm
+	Preceding a chromosome designation, indicates that the chromosome or arm is extra; following a designation, indicates that the chromosome or arm is larger than normal
−	Preceding a chromosome designation, indicates that the chromosome or arm is missing; following a designation, indicates that the chromosome or arm is smaller than normal
mos	Mosaic
/	Separates karyotypes of clones in mosaics—e.g., 47, XXX/45,X
dup	Duplication
dir dup	Direct duplication
inv dup	Inverted duplication
del	Deletion
inv	Inversion
t	translocation
rcp	Reciprocal translocation
rob	Robertsonian translocation
r	Ring chromosome
i	Isochromosome (two identical arms attached to a single centromere, like an attached-X chromosome in *Drosophila*)



BANDING PATTERNS CAN BE USED TO DISTINGUISH CHROMOSOMES & LOCATE GENES

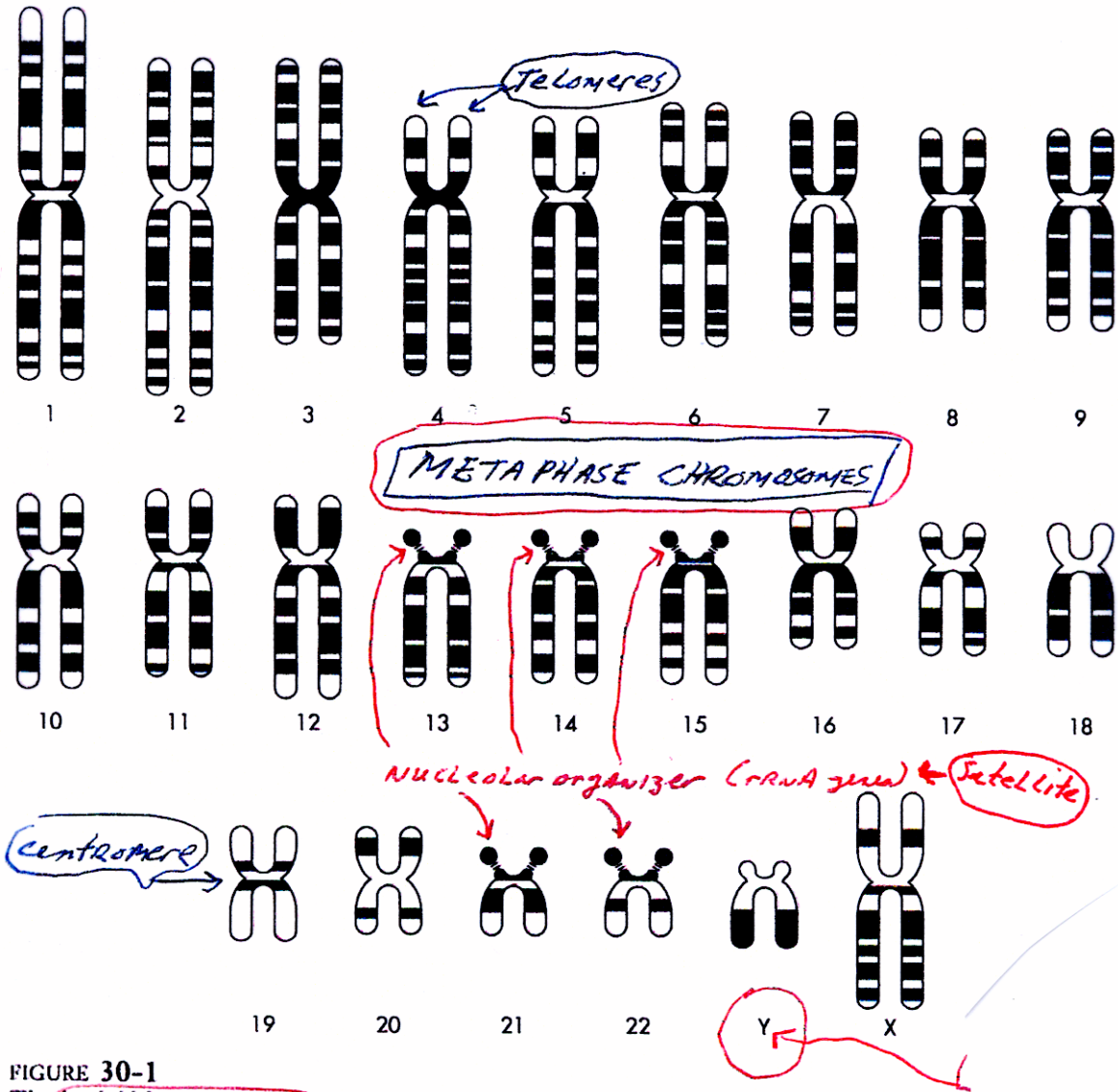
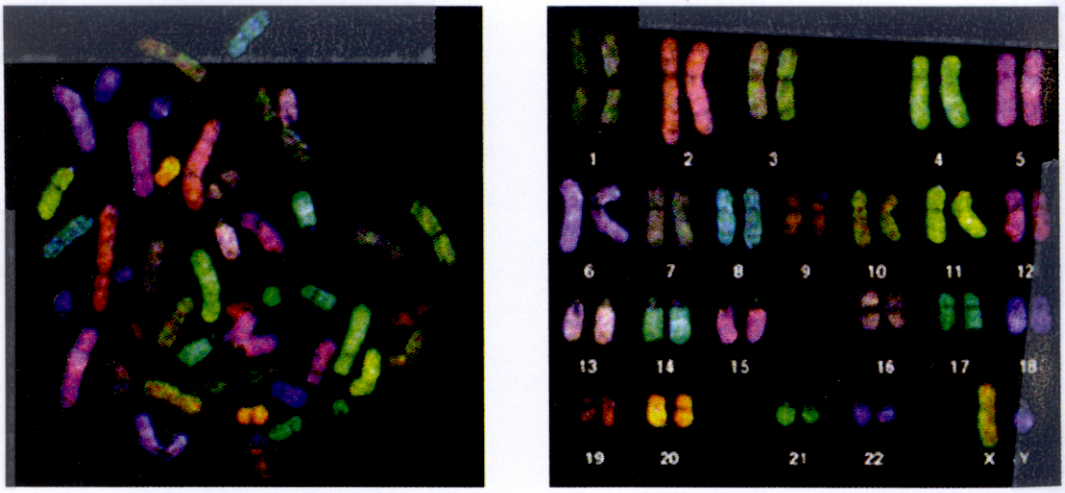


FIGURE **30-1**
The haploid human genome. This is a schematic drawing of 1 of each of the 23 human chromosomes, showing the pattern of staining seen with the Giemsa banding method. Chromosomes are first treated with trypsin and then stained with Giemsa. The patterns of light and dark bands are characteristic for each chromosome; and translocations, deletions, and other structural abnormalities can be identified. Typically 400 bands can be seen per haploid genome, and each band represents on average 7.5×10^6 bp, or twice as many base pairs as in the entire *E. coli* genome! Chromosome 1 constitutes 8.4 percent, and the Y chromosome about 2.0 percent, of the human genome. Taking the *E. coli* genome as a unit of genome size, a cytogenetic band is 2 genome units, and the Y chromosome is 15 genome units.

Ave band size = 7.5 Mb or 7.5×10^6 bp
larger than size of E. coli Genome!

HUMAN CHROMOSOMES CAN ALSO BE DISTINGUISHED BY THEIR SEQUENCES

How are these chromosomes "painted"?



(A) (B)

Figure 9.1 Human chromosome painting, in which each pair of chromosomes is labeled by hybridization with a different fluorescent probe. (A) Metaphase spread showing the chromosomes in a random arrangement as they were squashed onto the slide. (B) A karyotype, in which the chomosomes have been grouped in pairs and arranged in conventional order. Chromosomes 1–20 are arranged in order of decreasing size, but for historical reasons, chromosome 21 precedes chromosome 22, even though chromosome 21 is smaller. [Courtesy of Johannes Wienberg and Thomas Ried.]

Table 7.2: DNA content of human chromosomes[a]

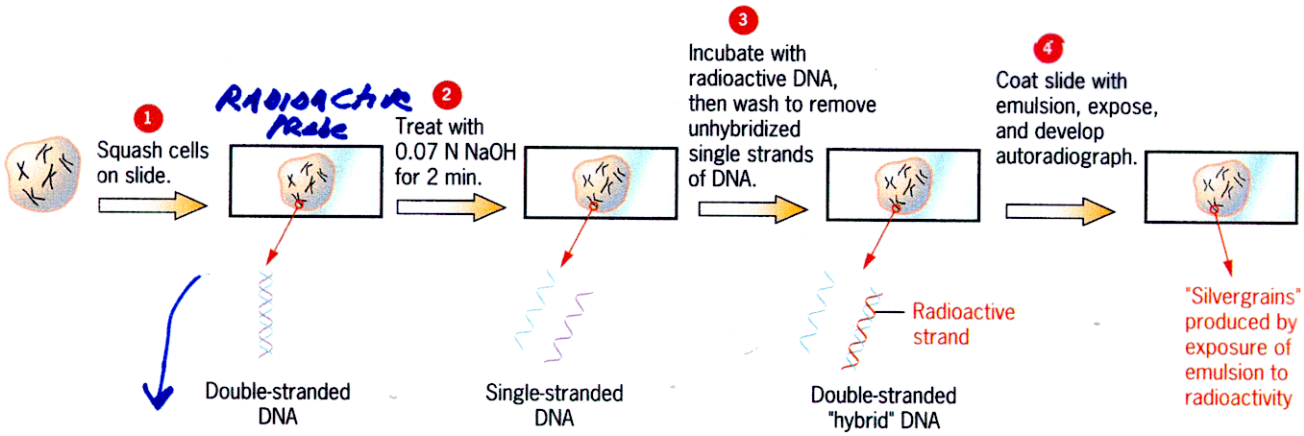
Chromosome	Amount of DNA (Mb)	Chromosome	Amount of DNA (Mb)
1	263	13	114
2	255	14	109
3	214	15	106
4	203	16	98
5	194	17	92
6	183	18	85
7	171	19	67
8	155	20	72
9	145	21	50
10	144	22	56
11	144	X	164
12	143	Y	59

[a] The DNA content is given for chromosomes prior to entering the S (DNA replication) phase of cell division (see *Figure 2.2*). Data abstracted from electronic reference 1.

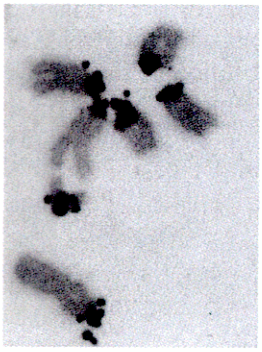
AND AMOUNT OF DNA!

FROM SEQUENCE DATA!

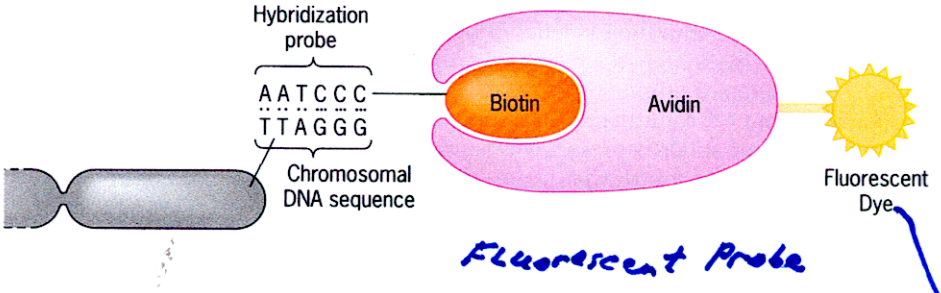
IN SITU HYBRIDIZATION WITH FLUORESCENT PROBES CAN IDENTIFY GENES & CHROMOSOMES



(a) Steps in performing *in situ* hybridization.



(b) Autoradiograph showing chromosomal locations of mouse satellite DNA sequences.

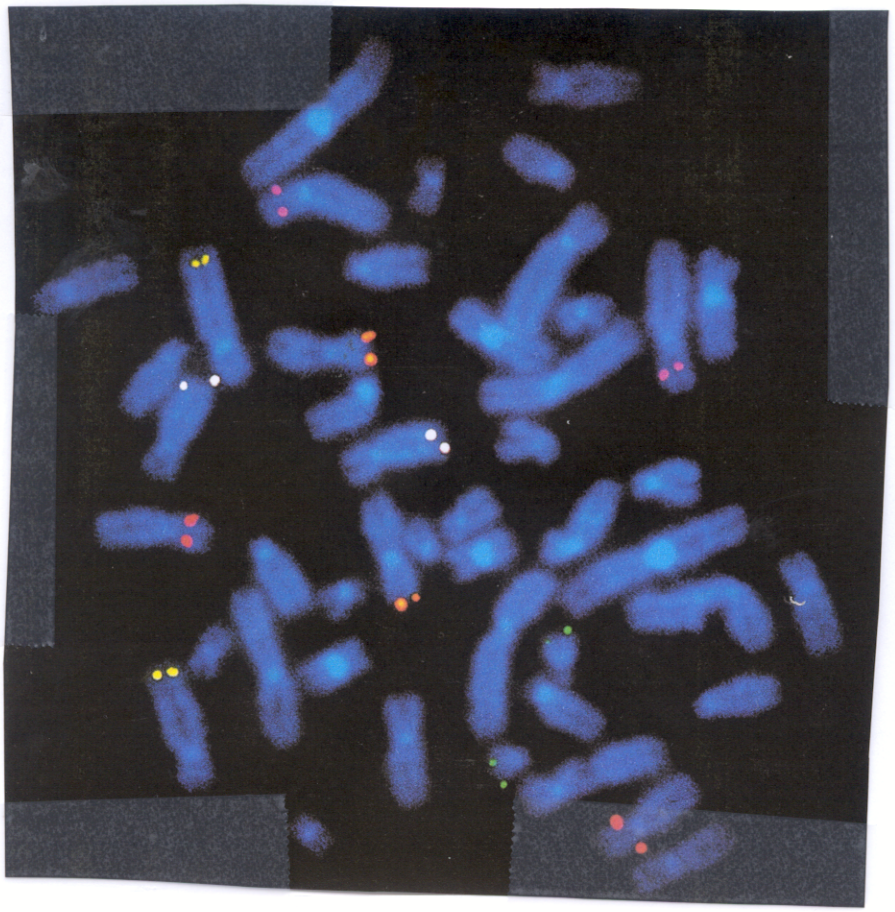


(c) Visualization of human telomeres by using fluorescent dyes and *in situ* hybridization.

Figure 10.15 Localization of repeated DNA sequences in chromosomes by *in situ* hybridization performed with radioactive probes (*a* and *b*) or fluorescent probes (*c* and *d*). The *in situ* hybridization procedure developed by Pardue and Gall is shown in (*a*), and one of their autoradiographs demonstrating the presence of the mouse satellite DNA sequence in centromeric heterochromatin is shown in (*b*). Use of fluorescent dyes to localize the TTAGGG repeat sequence to the telomeres of human chromosomes is illustrated in (*c*), and a photomicrograph demonstrating its telomeric location is shown in (*d*).

Visible Color in Microscope @ Specific wavelength

MAPPING GENES TO CHROMOSOMES AND SPECIFIC REGIONS



How correlate gene to chromosome position + band?

Pre-Sequence approach

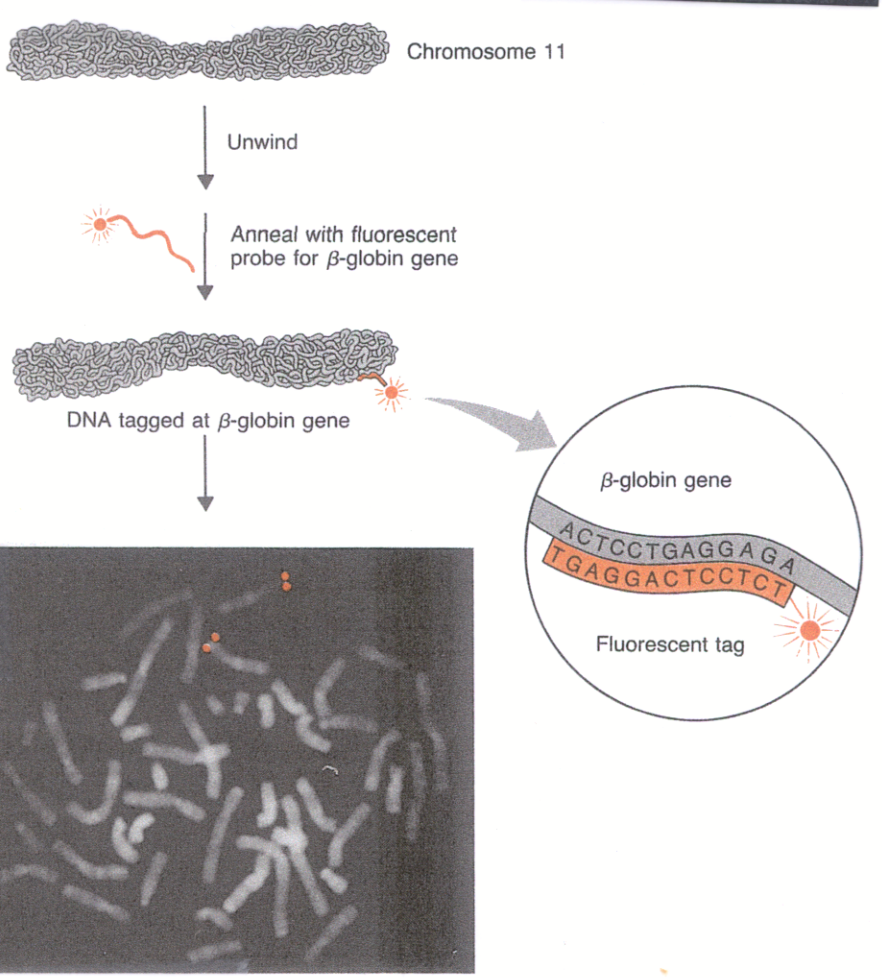


Figure 7.5 Locating the position of the β-globin gene on human chromosome 11.

GENES CAN BE MAPPED TO SPECIFIC BANDS OF EACH CHROMOSOME

How locate these Genes if no Probe or Sequence?

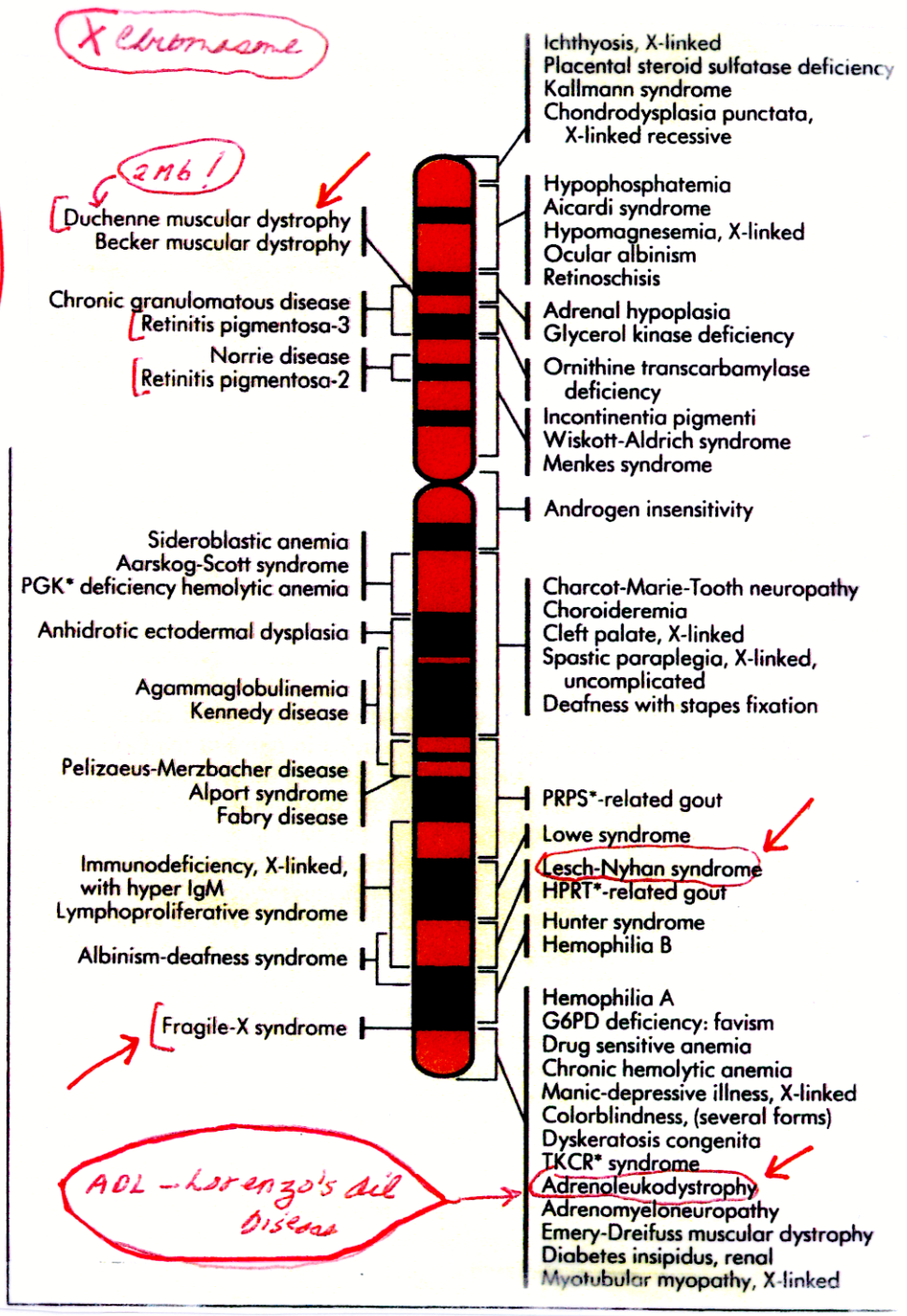
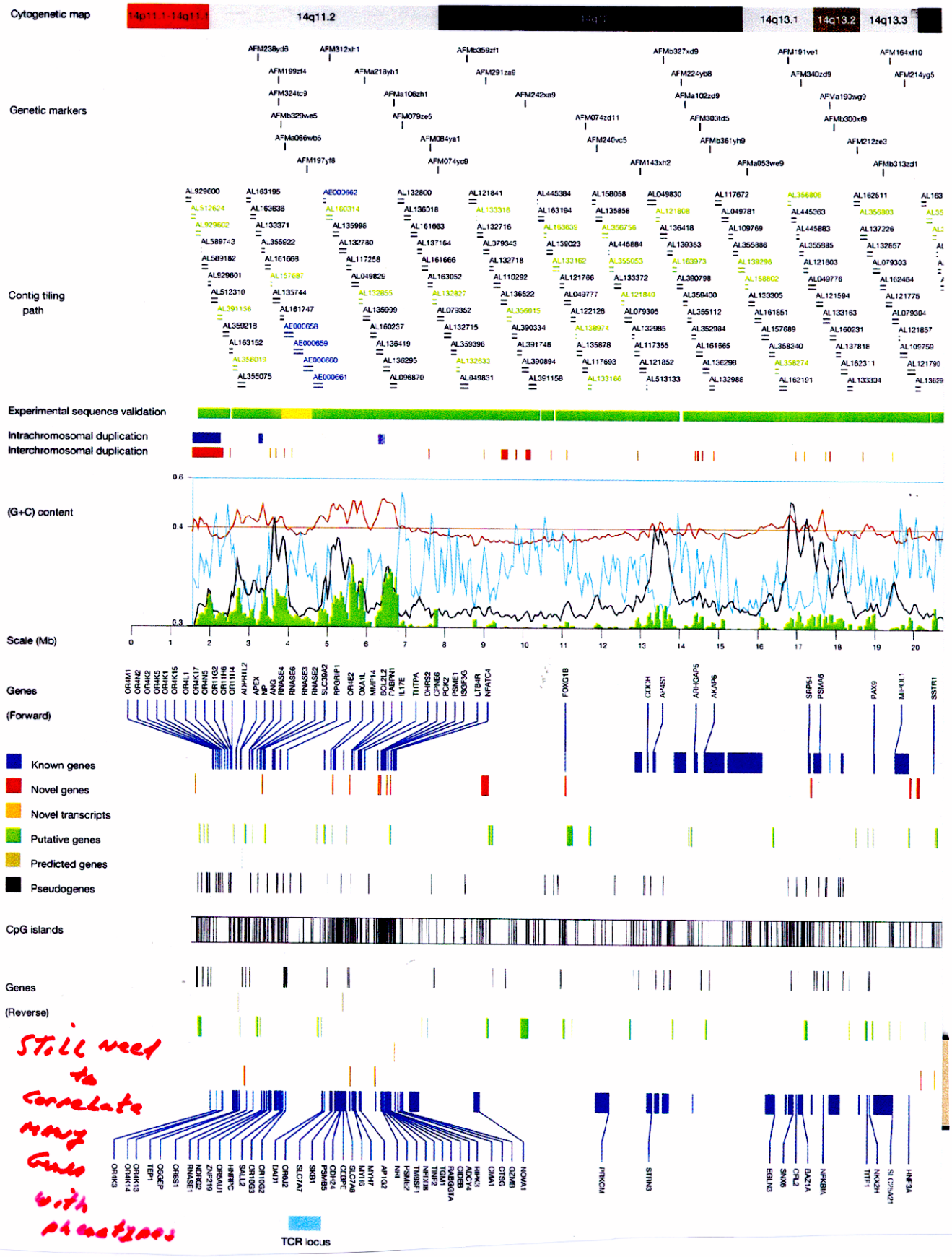


FIGURE 12-22
The human X-chromosome gene map. Over 59 diseases have now been traced to specific segments of the X-chromosome. Many of these disorders are also influenced by genes on other chromosomes. *KEY: *PGK*, phosphoglycerate kinase; *PRPS*, phosphoribosyl pyrophosphate synthetase; *HPRT*, hypoxanthine phosphorbibosyl transferase; *TKCR*, torticollis, keloids, cryptorchidism, and renal dysplasia.

This TASK IS NOW COMPLETE WITH THE COMPLETION OF THE HUMAN GENOME SEQUENCE



Still need to correlate many Genes with phenotypes

DISEASE GENES CAN BE LOCALIZED TO SPECIFIC CHROMOSOMES

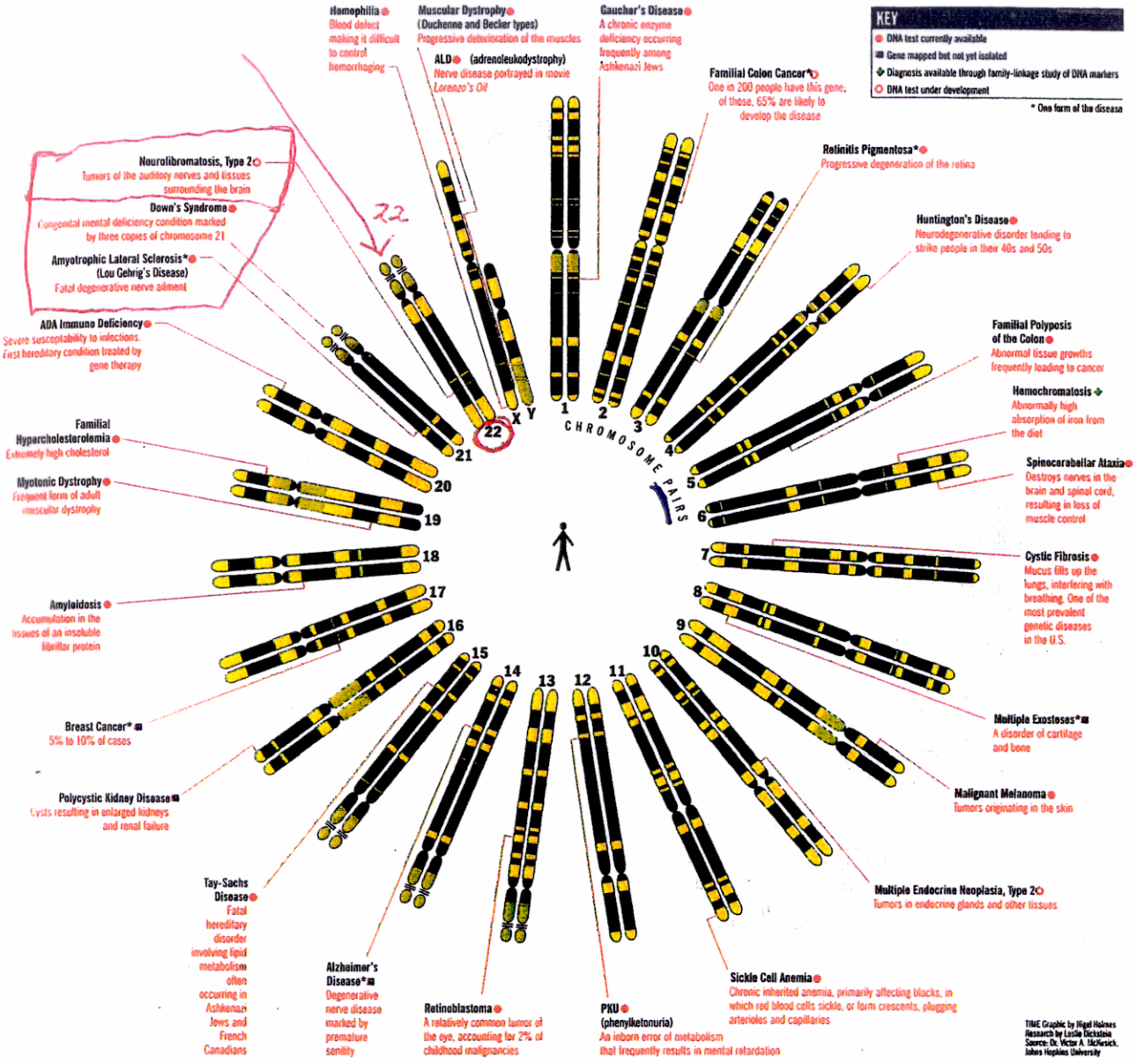


Figure 1-6 The 23 chromosomes of a human being, showing the positions of genes whose abnormal forms cause some of the better-known hereditary diseases. (*Time*)

BUT MANY MORE NEED to be correlated with disease! why difficult to do?

HOW CAN CHANGES OCCUR IN THE HUMAN GENOME?

LARGE GROSS CHANGES

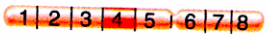
TABLE 12.1 Chromosomal Rearrangements and Changes in Chromosome Number (or Ploidy).

How detect?

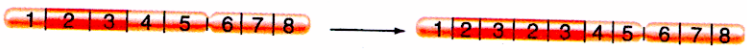
Chromosomal Rearrangements

Before → After

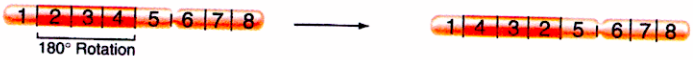
Deletion: Removal of a segment of DNA



Duplication: Increase in the number of copies of a chromosomal region

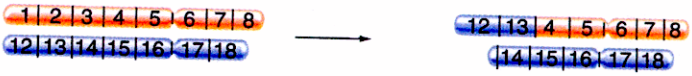


Inversion: Half-circle rotation of a chromosomal region



Translocations:

Nonreciprocal: Unequal exchanges between nonhomologous chromosomes



Reciprocal: Parts of two nonhomologous chromosomes trade places

Transposition: Movement of short DNA segments from one position in the genome to another

How detect?

How + when would these occur?

Changes in Chromosome Number or Ploidy

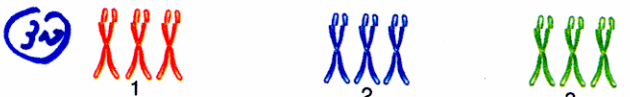
Euploidy: Cells that contain only complete sets of chromosomes

Diploidy (2*x*): Two copies of each homolog

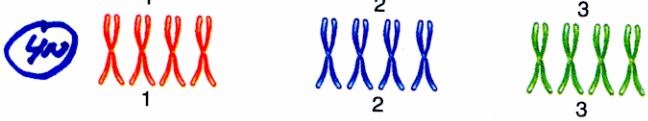
Monoploidy (*x*): One copy of each homolog

Polyploidy: More than the normal diploid number of chromosome sets

Triploidy (3*x*): Three copies of each homolog



Tetraploidy (4*x*): Four copies of each homolog



Aneuploidy: Loss or gain of one or more chromosomes producing a chromosome number that is not an exact multiple of the haploid number

Monosomy (2*n* − 1)

MONOSOMIC

− Chromosome 3

Trisomy (2*n* + 1)

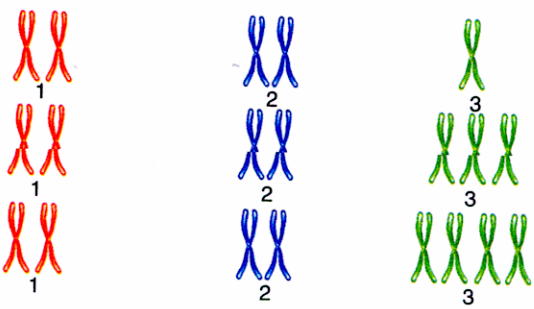
TRISOMIC

+ Chromosome 3

Tetrasomy (2*n* + 2)

TETRASOMIC

+ 2 Chromosome 3



Note that it is more accurate to denote monoploids, triploids, and tetraploids as multiples of *x*, which represents the number of different chromosomes in a complete set, rather than as multiples of *n*, the number of chromosomes in the gametes. In this table, as throughout the chapter, nonhomologous chromosomes are drawn in different colors. Different shades of the same color highlight different regions of the same chromosome.

ORIGINS OF LETHAL POLYPLOID Zygotes / Embryos

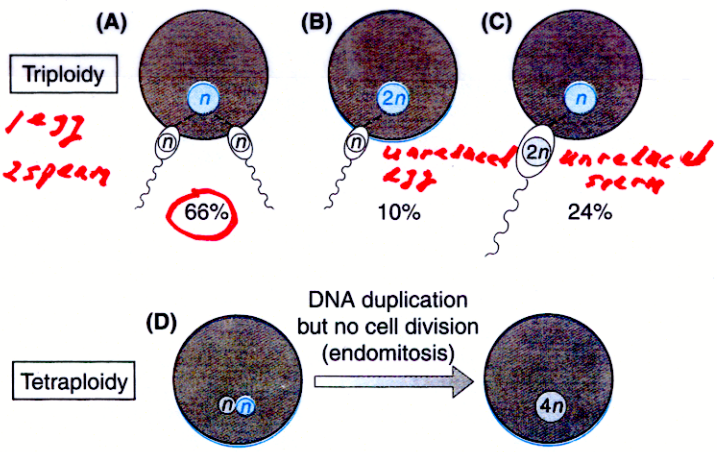


Figure 2.19: Origins of triploidy and tetraploidy.

About two-thirds of human triploids arise by fertilization of a single egg by two sperm (**A**). Other causes are a diploid egg (**B**) or sperm (**C**). Most human triploids abort spontaneously; very rarely they survive to term, but not beyond. Tetraploidy (**D**) results from failure of the first mitotic division after fertilization, and is incompatible with development.

What causes lethality with xtra genes/chromosomes?

How are these changes detected?

HUMAN GENETICS SIDELIGHT

Amniocentesis and Chorionic Biopsy: Procedures to Detect Aneuploidy in Human Fetuses

The Andersons, a couple living in Minneapolis, were expecting their first baby. Neither Donald nor Laura Anderson knew of any genetic abnormalities in their families, but because of Laura's age—38—they decided to have the fetus checked for aneuploidy.

Laura's physician performed a procedure called **amniocentesis.** A small amount of fluid was removed from the cavity surrounding the developing fetus by inserting a needle into Laura's abdomen (Figure 1). This cavity, called the amnionic sac, is enclosed by a membrane. To prevent discomfort during the procedure, Laura was given a local anesthetic. The needle was guided into position by following an ultrasound scan, and some of the amnionic fluid was drawn out. Because this fluid contains nucleated cells sloughed off from the fetus, it is possible to determine the fetus's karyotype (Figure 2). Usually the fetal cells are purified from the amnionic fluid by centrifugation, and then the cells are cultured for several days to a few weeks. Cytological analysis of these cells will reveal if the fetus is aneuploid. Additional tests may be performed on the fluid recovered from the amnionic sac to detect other sorts of abnormalities, including neural tube defects and some kinds of mutations. The results of all these tests may take up to three weeks. In Laura's case, no abnormalities of any sort were detected, and 20 weeks after the amniocentesis, she gave birth to a healthy baby girl.

Chorionic biopsy provides another way of detecting chromosomal abnormalities in the fetus. The chorion is a fetal membrane that interdigitates with the uterine wall, eventually forming the placenta. The minute chorionic projections into the uterine tissue are called *villi* (singular, villus). At 10–11 weeks of gestation, before the placenta has developed, a sample of chorionic villi can be obtained by passing a hollow plastic tube into the uterus through the cervix. This tube can be guided by an ultrasound scan, and when it is in place, a tiny bit of material can be drawn up into the tube by aspiration. The recovered material usually consists of a mixture of maternal and fetal tissue. After these tissues are separated by dissection, the fetal cells can be analyzed for chromosome abnormalities.

Chorionic biopsy can be performed earlier than amniocentesis (10–11 weeks gestation versus 14–16 weeks), but it is not as reliable. In addition, it seems to be associated with a slightly greater chance of miscarriage than amniocentesis, perhaps 2 to 3 percent. For these reasons, it tends to be used only in pregnancies where there is a strong reason to expect a genetic abnormality. In routine pregnancies, such as Laura Anderson's, amniocentesis is the preferred procedure.

Figure 1 A physician taking a sample of fluid from the amniotic sac of a pregnant woman for prenatal diagnosis of a chromosomal or biochemical abnormality.

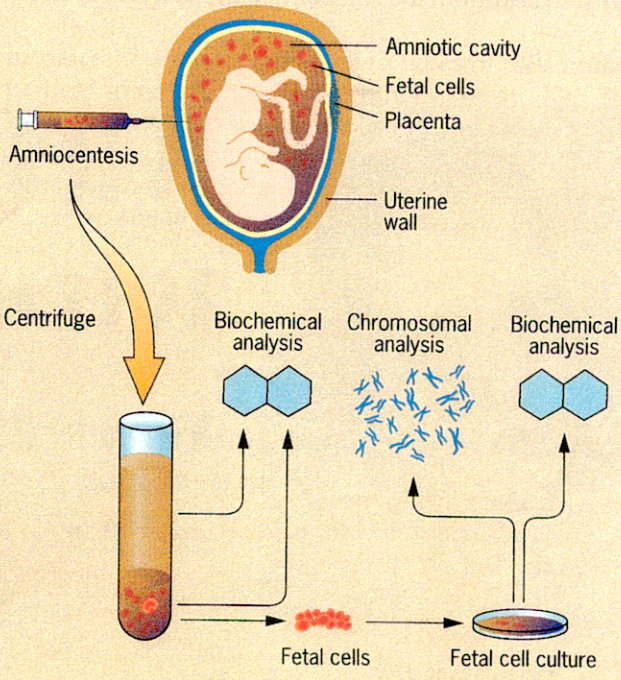


Figure 2 Amniocentesis and procedures for prenatal diagnosis of chromosomal and biochemical abnormalities.

PRENATAL Detection of CHROMOSOMAL ABNORMALITIES

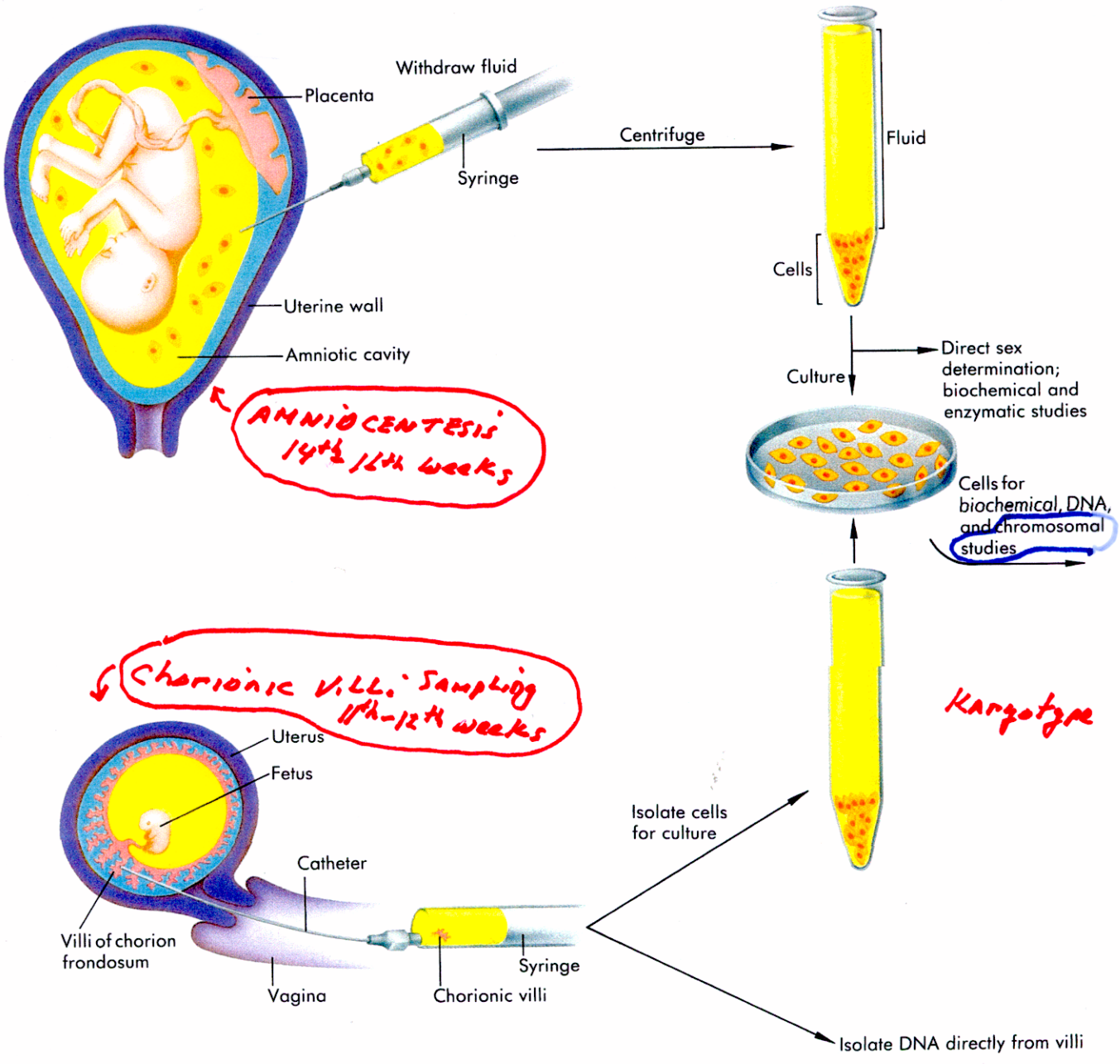
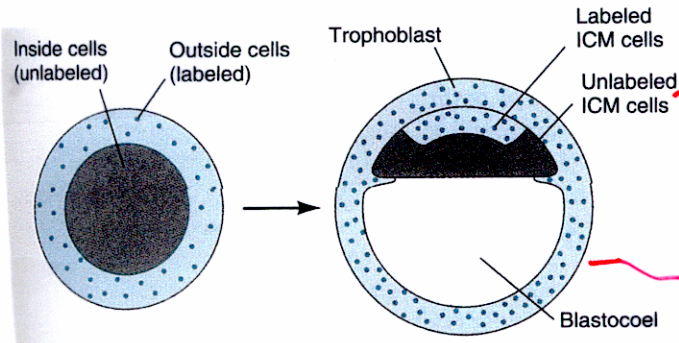


FIGURE 27-1
Amniocentesis and chorionic villus sampling. (a) A sample of amniotic fluid (mostly fetal urine and other secretions) is taken by inserting a needle into the amniotic cavity during or around the sixteenth week of gestation. The fetal cells are separated from the fluid by centrifugation. The cells can be used immediately, or more usually they are cultured so that a number of biochemical, enzymatic, and chromosomal analyses can be made. The cultured cells can also be a source of DNA. (b) Chorionic villus sampling is performed between the eighth and twelfth weeks of gestation. A catheter is introduced through the vagina or transabdominally, and a small sample of chorionic villi is drawn into the syringe. DNA can be isolated directly from the tissue, or cell cultures can be established. Note that the various elements of this figure are not drawn to scale.

ALSO CAN BE USED FOR DNA Testing

HUMAN EMBRYO FORMATION

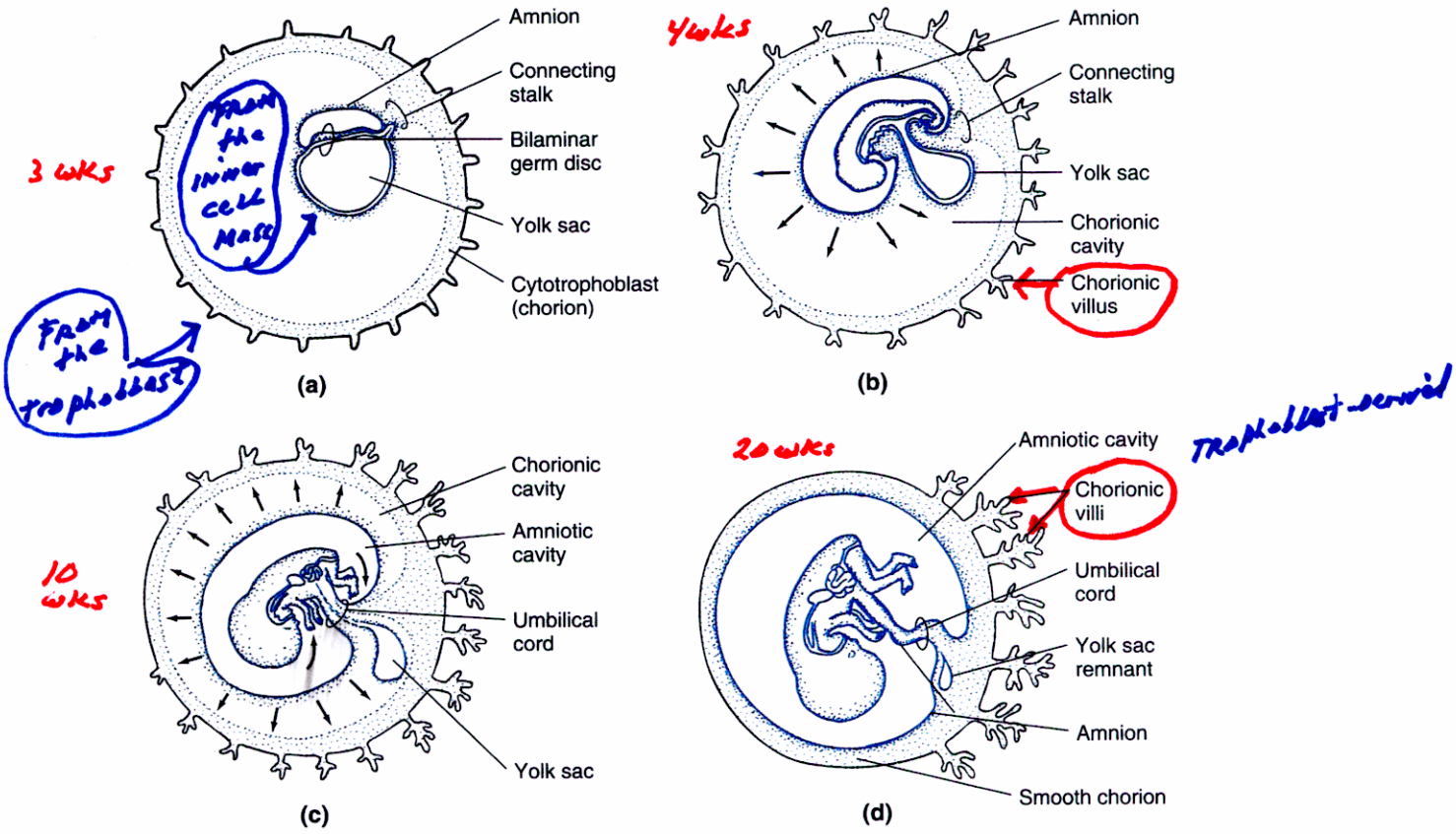


Embryo/Fetus

Embryo/Yolk, Chorion

Number of inside cells of 16-cell stage	Proportion of labeled cells in ICM
2	> 65%
5 (mean)	~ 25%
7	< 5%

Figure 5.13 Stepwise formation of the inner cell mass (ICM) in mammalian embryos. Most of the ICM cells are derived from those cells that are in an inside position at the morula stage. Thus, after selectively labeling cells on the outside of a morula, most ICM cells of the developing blastocyst are unlabeled. However, in embryos that have few inside morula cells, additional ICM cells are generated by differential cleavage of outside morula cells.



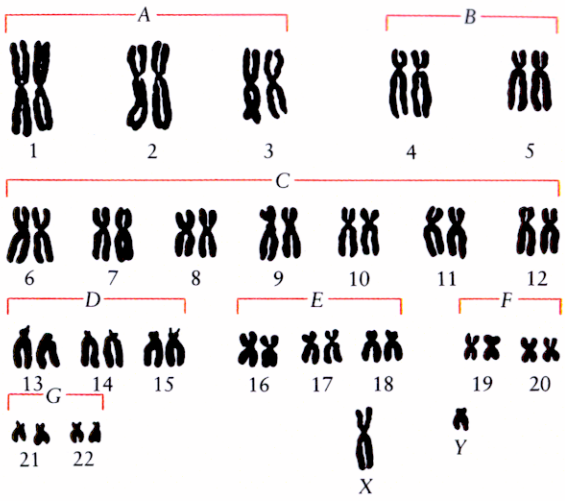
Trophoblast-derived

Figure 14.38 Extraembryonic membranes in human development: **(a)** at 3 weeks; **(b)** at 4 weeks; **(c)** at 10 weeks; **(d)** at 20 weeks. The connecting stalk develops into the umbilical cord. The amniotic cavity expands (arrows) until it completely fills the chorionic cavity and envelops the umbilical cord plus the remnant of the yolk sac. The chorionic villi near the umbilical cord branch and form the embryonic portion of the placenta. The other villi disappear.

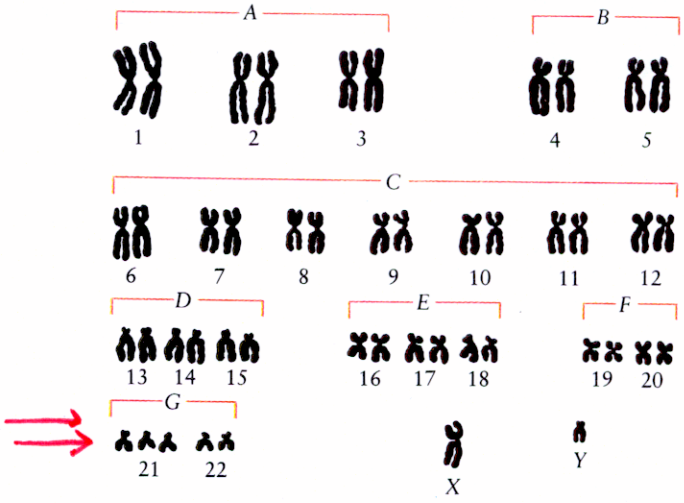
Karyotypes Reveal Chromosomal Abnormalities

19–2 *The normal diploid chromosome number of a human being is 46, 22 pairs of autosomes and two sex chromosomes. The autosomes are grouped by size (A, B, C, etc.), and then the probable homologues are paired. A normal woman has two X chromosomes and a normal man, shown here, an X and a Y.*

How know which chromosome is which?



Down's Syndrome



(b)

(a)

19–4 **(a)** *Although children with Down's syndrome share certain physical characteristics, there is a wide range of mental capacity among these individuals.* **(b)** *The karyotype of a male with Down's syndrome caused by nondisjunction. Note that there are three chromosomes 21.*

Mechanism causing abnormality?

CHROMOSOMAL ABNORMALITIES CAUSED BY ERRORS in EGG & SPERM FORMATION (MEIOSIS)

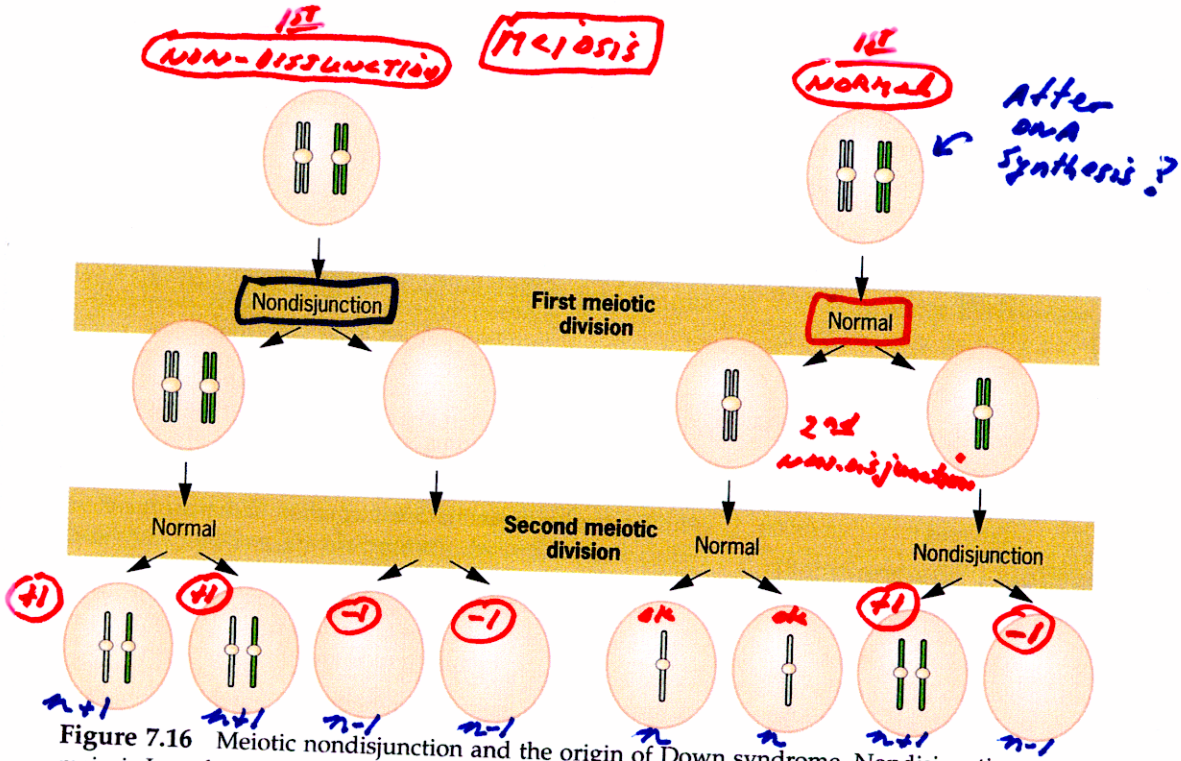


Figure 7.16 Meiotic nondisjunction and the origin of Down syndrome. Nondisjunction at meiosis I produces no normal gametes. Nondisjunction at meiosis II produces a gamete with two identical sister chromosomes, a gamete lacking chromosome 21, and two normal gametes.

ONLY PROBES or BANDING PATTERN identifies CHROMOSOME

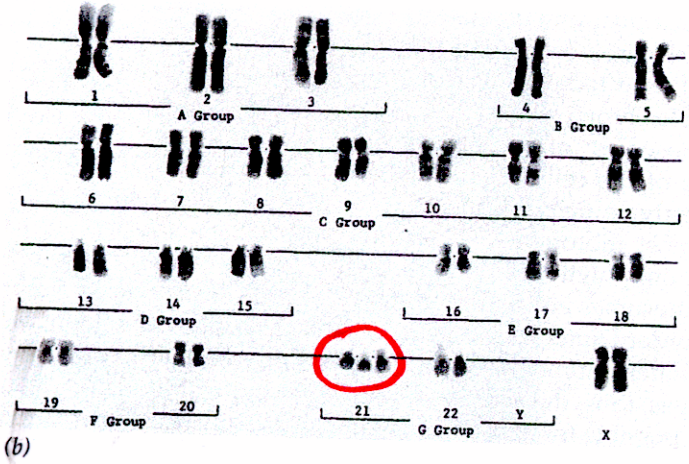


Figure 7.15 Down syndrome. (*a*) Facial features of a child with Down syndrome. (*b*) Karyotype of a child with Down syndrome, showing trisomy for chromosome 21 (47,XX, +21).

Detection of Extra Chromosome By In Situ Hybridization

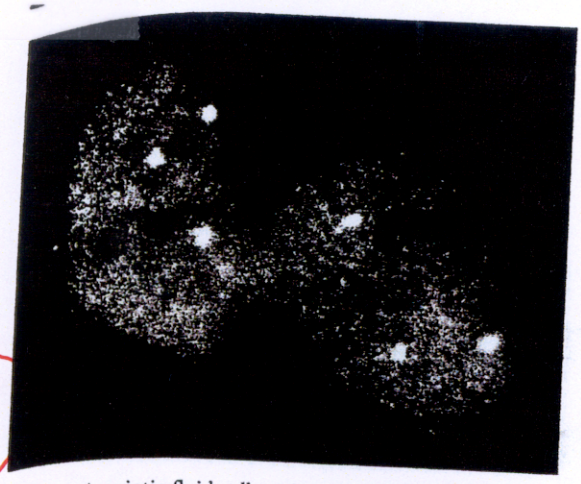


Fig. 3.13. Amniotic fluid cell nuclei of a fetus with trisomy 18 after CISS hybridization with the biotinylated Alu-PCR amplified YAC clone HTY 3045 (mapped to 18q23) detected with avidin-FITC. Nuclei were counterstained with propidium iodide

Note

3 chromosome 18's

Amniotic cells

Use of a Chromosome 18 Specific DNA Sequence

Major chromosomal Defects in Humans

TABLE 7.1
Aneuploidy Resulting from Nondisjunction in Human Beings

Karyotype	*Chromosome Formula*	*Clinical Syndrome*	*Estimated Frequency at Birth*	*Phenotype*
47,+21	$2n+1$	Down	1/700	Short, broad hands with palmar crease, short stature, hyperflexibility of joints, mental retardation, broad head with round face, open mouth with large tongue, epicanthal fold.
47,+13	$2n+1$	Patau	1/20,000	Mental deficiency and deafness, minor muscle seizures, cleft lip and/or palate, cardiac anomalies, posterior heel prominence.
47,+18	$2n+1$	Edward	1/8000	Congenital malformation of many organs, low-set, malformed ears, receding mandible, small mouth and nose with general elfin appearance, mental deficiency, horseshoe or double kidney, short sternum, 90 percent die within first six months after birth.
SEX CHROMOSOMES				
45,X	$2n-1$	Turner	1/2500 female births	Female with retarded sexual development, usually sterile, short stature, webbing of skin in neck region, cardiovascular abnormalities, hearing impairment.
47 XYY	2n+1	normal	1/10,000 births	
47,XXY	$2n+1$	Klinefelter	1/500 male births	Male, subfertile with small testes, developed breasts, feminine-pitched voice, knock knees, long limbs.
48,XXXY	$2n+2$			
48,XXYY	$2n+2$			
49,XXXXY	$2n+3$			
50,XXXXXY	$2n+4$			
47,XXX	$2n+1$	Triplo-X	1/700	Female with usually normal genitalia and limited fertility, slight mental retardation.

1. The XYY Story – Science goes wrong!
2. How obtain an XX male?

Most Changes in Chromosome Number Are Lethal

Table 9.2 Chromosome abnormalities per 100,000 recognized human pregnancies

	15,000 spontaneous abortions	85,000 live births
Trisomy		
A: 1	0	0
A: 2	159	0
A: 3	53	0
B: 4	95	0
B: 5	0	0
C: 6–12	561	0
D: 13	128	17
D: 14	275	0
D: 15	318	0
E: 16	1229	0
E: 17	10	0
E: 18	223	13
F: 19–20	52	0
G: 21	350	113
G: 22	424	0
Sex chromosomes		
XYY	4	46
XXY	4	44
XO	1350	8
XXX	21	44
Translocations		
Balanced	14	164
Unbalanced	225	52
Polyploid		
Triploid	1275	0
Tetraploid	450	0
Other (mosaics, etc.)	280	49
Total	7500	550

Explain why 1 Xtra chromosome CAN lead to Abortion?

Lead to Spontaneous Abortion

(1) 15% of conceptions Lead to Spontaneous Abortions

(2) Half of these Are Due to chromosome Abnormalities

(3) ~ 0.65% of Live Birth Mutations Due to chromosomal Abnormalities

(4) ~ 1.2% of Live Birth Mutations Due to DNA Changes / Point Mutations

~ 2% of Live Births Have Genetic Defects That are Visible

How obtain polyploid Fetus?

* The other 7500 are caused by Mutations!

FREQUENCY OF Gene and Chromosomal Mutations in Live Births

Live Births

Table 9-1. Relative Incidence of Human Ill Health due to Gene Mutation and to Chromosome Mutation

Type of mutation	Percentage of live births
Gene mutation (Gene)	
Autosomal dominant	0.90
Autosomal recessive	0.25
X-linked	0.05
Total gene mutation (1.2%)	1.20
Chromosome mutation (Chromosomal)	
Autosomal trisomies (mainly Down syndrome)	0.14
Other unbalanced autosomal aberrations	0.06
Balanced autosomal aberrations	0.19
Sex chromosomes	
XYY, XXY, and other ♂♂	0.17
XO, XXX, and ♀♀	0.05
Total chromosome mutation (0.61%)	0.61

Live Births

1.2% of Live Births = Mutations

0.61% of Live Births = Chromosome Defects

~2% of all Live Births have Genetic Defects

Note: 15% of conceptions lead to Spontaneous Abortions. only 85% give rise to live births

A Large # of Spontaneous Abortions Caused by Chromosome & DNA Changes. Mutations also Affect Large #'s of Children who are Born

Large Alterations of the Human Genome Lead to Death in Most Cases

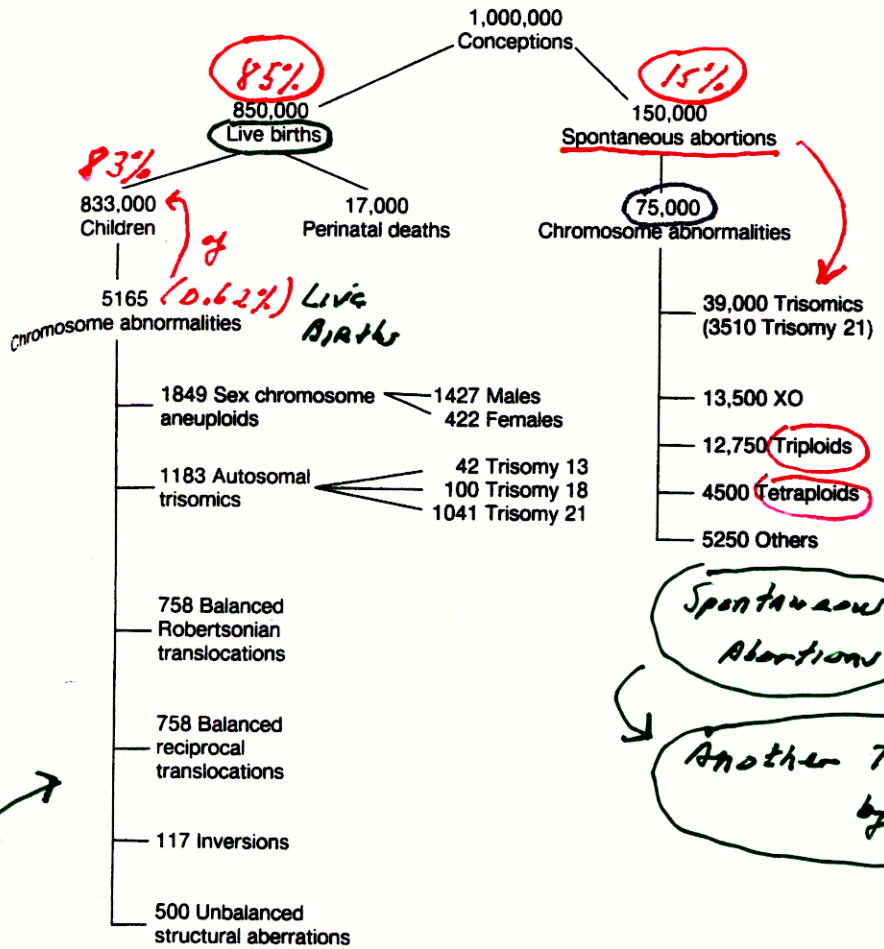


Figure 9-23 The fate of a million implanted human zygotes. (Robertsonian translocations involve fusion or dissociation of centromeres.) (From K. Sankaranarayanan, *Mutation Research* 61, 1979.)

75,000 chromosome abnormalities
75,000 point mutations

50% due to chromosome changes

50% due to mutations (lethal)

Spontaneous Abortions

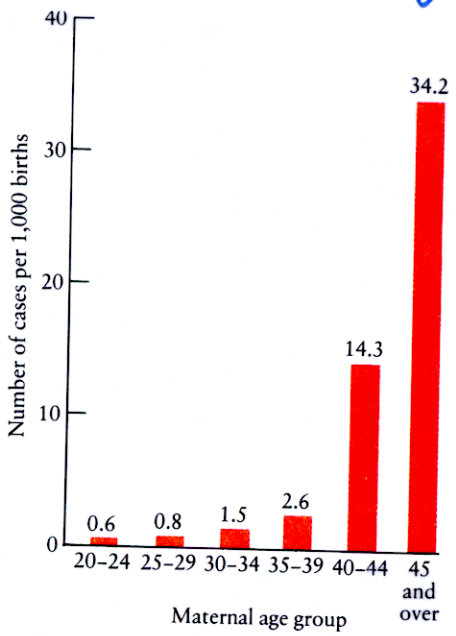
Another 75,000 caused by Mutations

another 1.2% of live births have visible mutations

≈ 2% of live births ($\frac{1}{50}$) with genetic defects

Note:
Many more mutations accumulate with no visible effects or correct combination (e.g. recessive allele) does not occur ∴ Large # of defective genes in carriers!

FREQUENCIES OF CHILDREN BORN WITH CHROMOSONAL DEFECTS INCREASES WITH AGE OF MOTHER



19–6 *The frequencies of births of infants with Down's syndrome in relation to the ages of the mothers. The number of cases shown for each age group represents the occurrence of Down's syndrome in every 1,000 live births by mothers in that group. As you can see, the risk of having a child with Down's syndrome increases rapidly after the mother's age exceeds 40. An increased risk is also thought to occur after the father's age exceeds 55.*

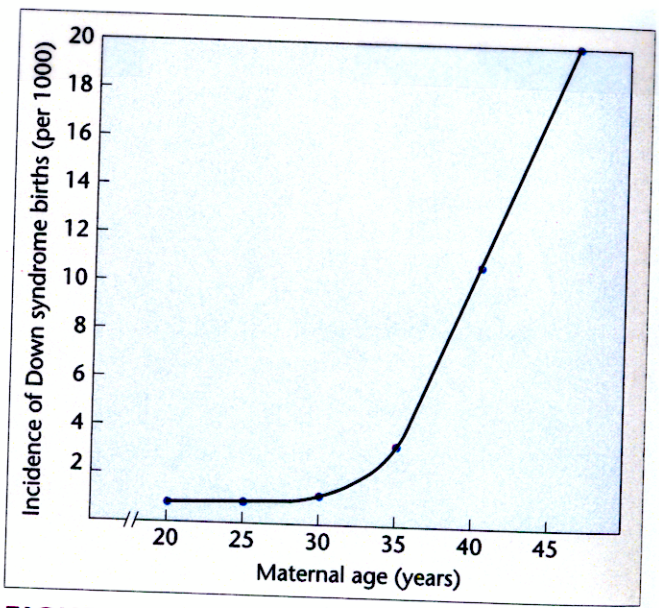
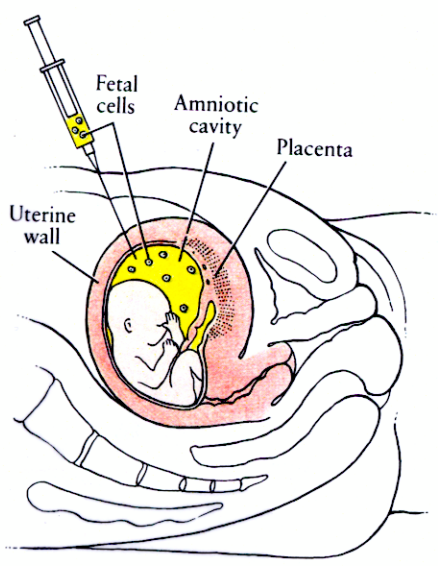


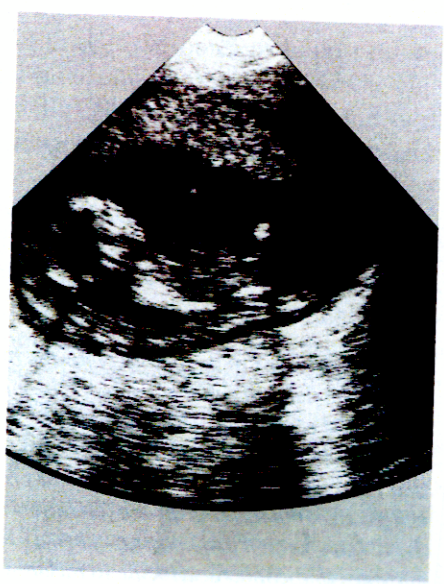
FIGURE 10.6 Incidence of Down syndrome births contrasted with maternal age.

nificantly older and arrested longer than those they ovulated 10 or 20 years previously. However, it is not yet known whether ovum age is the cause of the increased incidence of nondisjunction leading to Down syndrome.

These statistics are the basis of a serious issue facing parents when pregnancy occurs late in a woman's reproductive years. Genetic counseling early in such pregnan-



(a)



(b)

Sperm & Egg Formation

♂ ♀

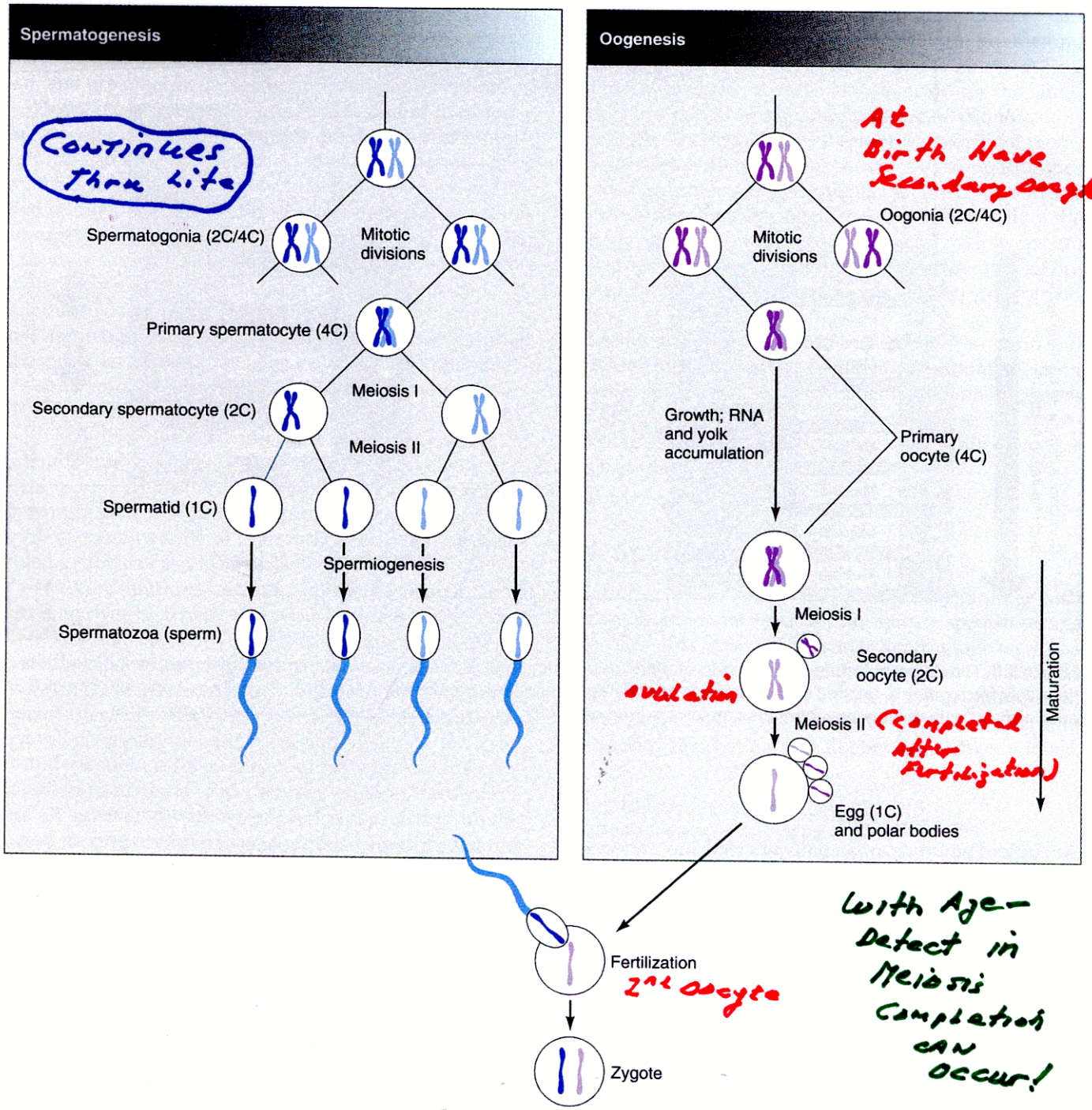
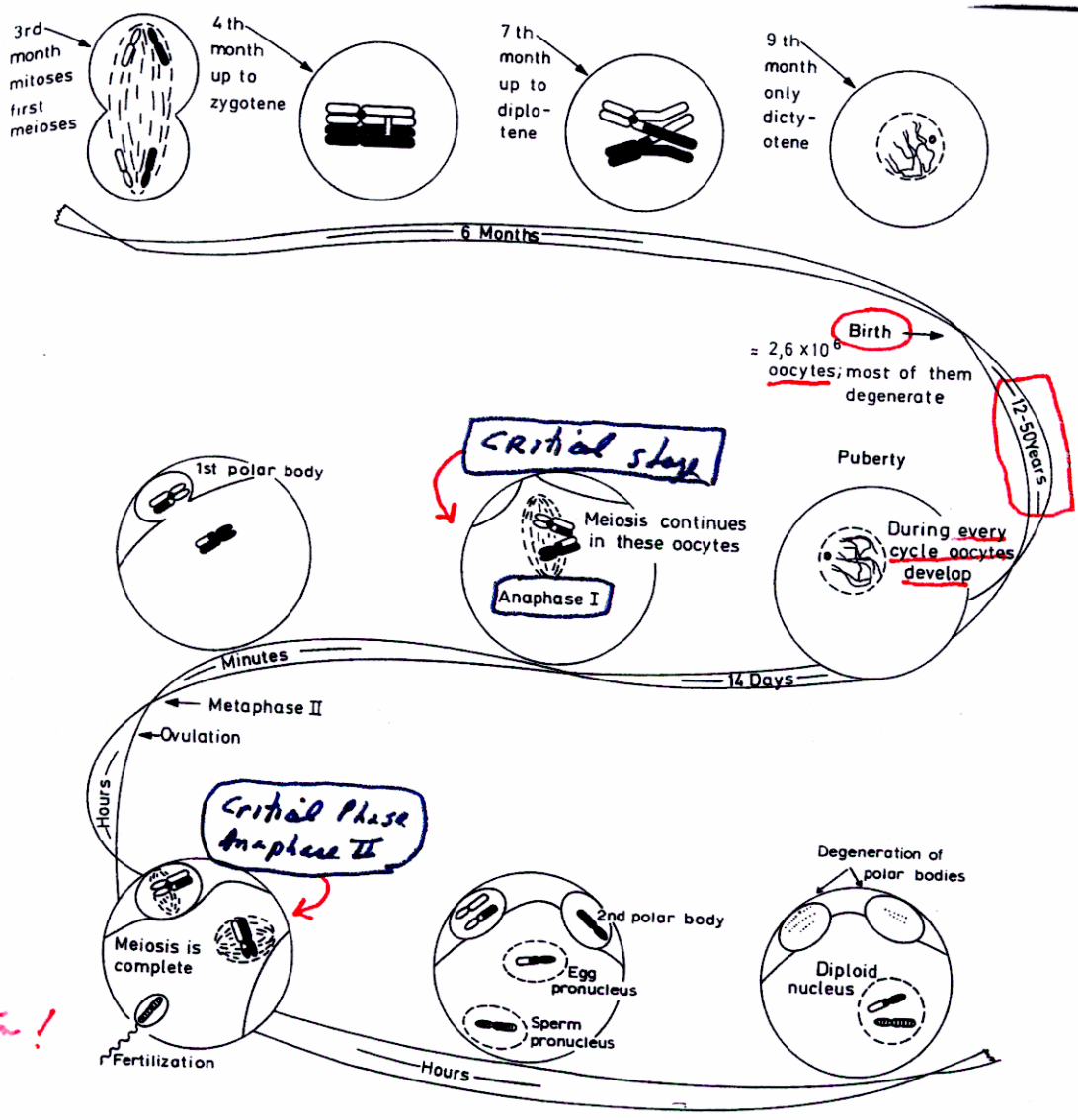


Figure 3.5 Comparison of spermatogenesis and oogenesis. Primordial germ cells divide mitotically, producing spermatogonia in males and oogonia in females. These cells are diploid, containing two or four genomic complements (2C or 4C), depending on their stage in the mitotic cycle. Before the gonia enter meiosis, their DNA replicates. They are then called primary spermatocytes or oocytes. After the first meiotic division, they contain two genomes (2C) and are called secondary spermatocytes or oocytes. After the second meiotic division, they are haploid (1C) spermatids or eggs. Note that the two rounds of meiosis produce four haploid spermatids, each of which develops into a spermatozoon, but only one egg. The egg's three small sister cells, known as polar bodies, have no known function and degenerate. Often the first polar body does not divide, so that only a total of two polar bodies is formed. Depending upon the species, eggs are fertilized at various stages of meiosis (see Fig. 3.18).

Meiosis in Egg Formation is Completed After Fertilization



Egg forms after fertilization!

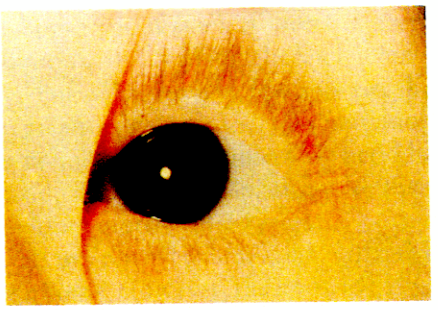
Fig. 2.21. Meiosis in the human female. Meiosis starts after 3 months of development. During childhood the cytoplasm of oocytes increases in volume, but the nucleus remains unchanged. About 90% of all oocytes degenerate at the onset of puberty. During the first half of every month the luteinizing hormone *(LH)* of the pituitary stimulates meiosis which is now almost completed (end of the prophase that began during embryonic age; metaphase I, anaphase I, telophase I and – within a few minutes – prophase II and metaphase II). Then meiosis stops again. A few hours after metaphase I is reached ovulation is induced by *LH*. Fertilization occurs in the fallopian tube. Then the second meiotic division is completed. Nuclear membranes are formed around the maternal and paternal chromosomes. After some hours the two "pronuclei" fuse, and the first cleavage division begins. (From Bresch and Hausmann 1972)

And all "Eggs" are present at Birth

∴ Meiotic Errors increase with Age!

Large deletions also cause genetic abnormalities

(a)



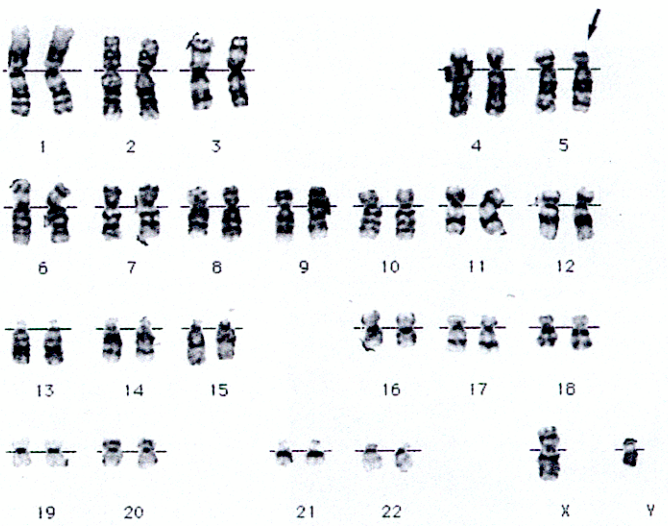
(b)

19–7 (a) *A chromosomal abnormality associated with cancer. The chromosomes shown here have been stained to reveal banding patterns. The chromosome on the left is normal. The one on the right has a deletion, shown by the smaller size of the bracket. Such deletions have been found in children with Wilms' tumor.* (b) *The left eye of a 15-year-old boy who has this chromosomal deletion and who developed Wilms' tumor in infancy. Note the absence of an iris. An older half-brother and a maternal aunt also had aniridia and developed Wilms' tumor at an early age. Another brother and the boy's mother are phenotypically normal. Analysis of the mother's chromosomes revealed that although she carries the deletion in chromosome 11, the missing segment is present in her cells in chromosome 2. Almost all other chromosomal abnormalities associated with cancer have occurred only in somatic cells and are not inherited.*

were used to correlate gene with chromosome region

How correlate gene with locus?

(a)



(b)

Figure 7.18 *Cri-du-chat* syndrome. (*a*) Patient with *cri-du-chat* syndrome. (*b*) Karyotype of infant with *cri-du-chat* syndrome, 46, XY(5p-). There is a deletion in the short arm of chromosome 5 (arrow).

Rearranged Chromosomes ALSO lead to Genetic Abnormalities

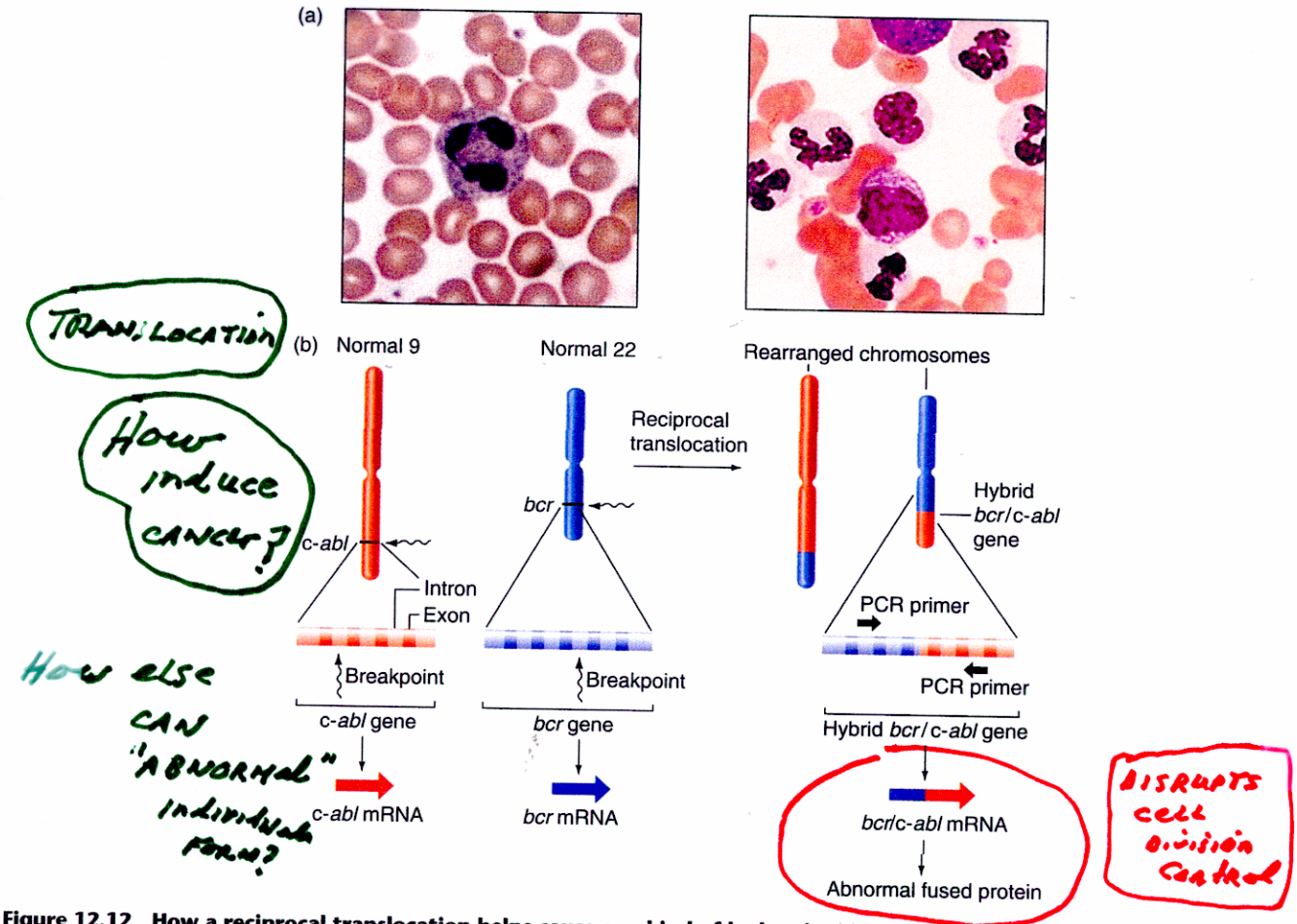
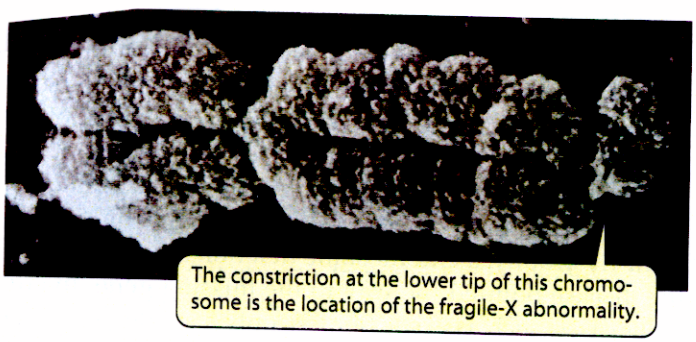


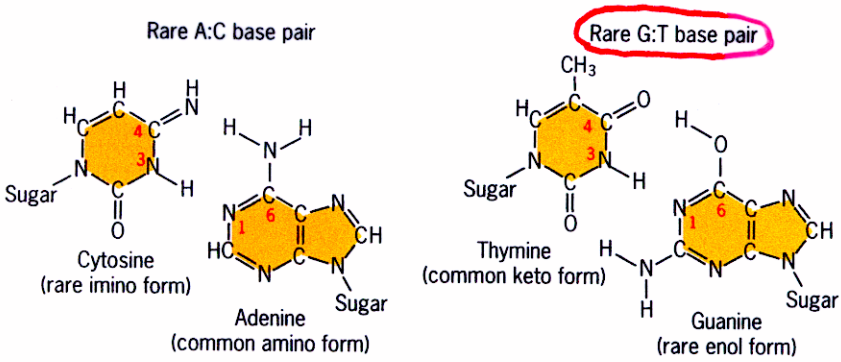
Figure 12.12 How a reciprocal translocation helps cause one kind of leukemia. (a) Uncontrolled divisions of large, dark-staining white blood cells in the blood of a leukemia patient (right) produce a higher than normal ratio of white to red blood cells than that of a normal individual (left). (b) A reciprocal translocation between chromosomes 9 and 22 contributes to chronic myelogenous leukemia. This rearrangement makes an abnormal hybrid gene composed of part of the c-*abl* gene on chromosome 9 and part of the *bcr* gene on chromosome 22. The hybrid gene produces a mRNA with sequences from both c-*abl* and *bcr*, and this hybrid mRNA is translated into an abnormal fused protein that disrupts controls on cell division. Black arrows indicate PCR primers that will generate a PCR product only in DNA containing the hybrid gene.



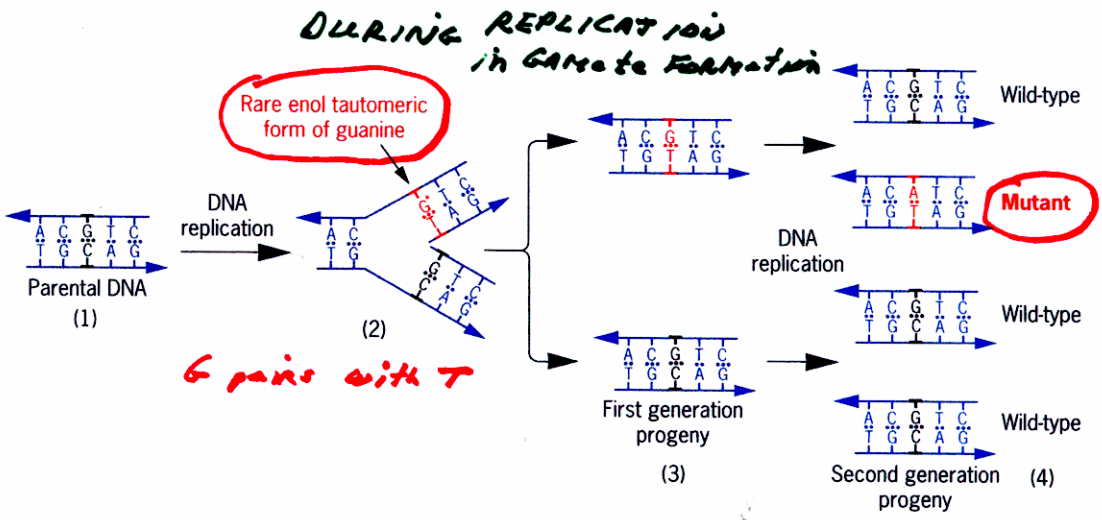
Chromosome Breakage/Deletion

18.5 A Fragile-X Chromosome at Metaphase
The chromosomal abnormality that causes the mental retardation symptomatic of fragile-X syndrome shows up physically as a constriction.

MANY CHANGES in DNA Sequence ALSO Occur



(a) Hydrogen-bonded A:C and G:T base pairs that form when cytosine and guanine are in their rare imino and enol tautomeric forms.



(b) Mechanism by which tautomeric shifts in the bases in DNA cause mutations.

Figure 14.14 The effects of tautomeric shifts in the nucleotides in DNA on (*a*) base-pairing and (*b*) mutation. Rare A:C and G:T base pairs like those shown in (*a*) also form when thymine and adenine are in their rare enol and imino forms, respectively. (*b*) A guanine (1) undergoes a tautomeric shift to its rare enol form (G*) at the time of replication (2). In its enol form, guanine pairs with thymine (2). During the subsequent replication (3 to 4), the guanine shifts back to its more stable keto form. The thymine incorporated opposite the enol form of guanine (2) directs the incorporation of adenine during the next replication (3 to 4). The net result is a G:C to A:T base-pair substitution.

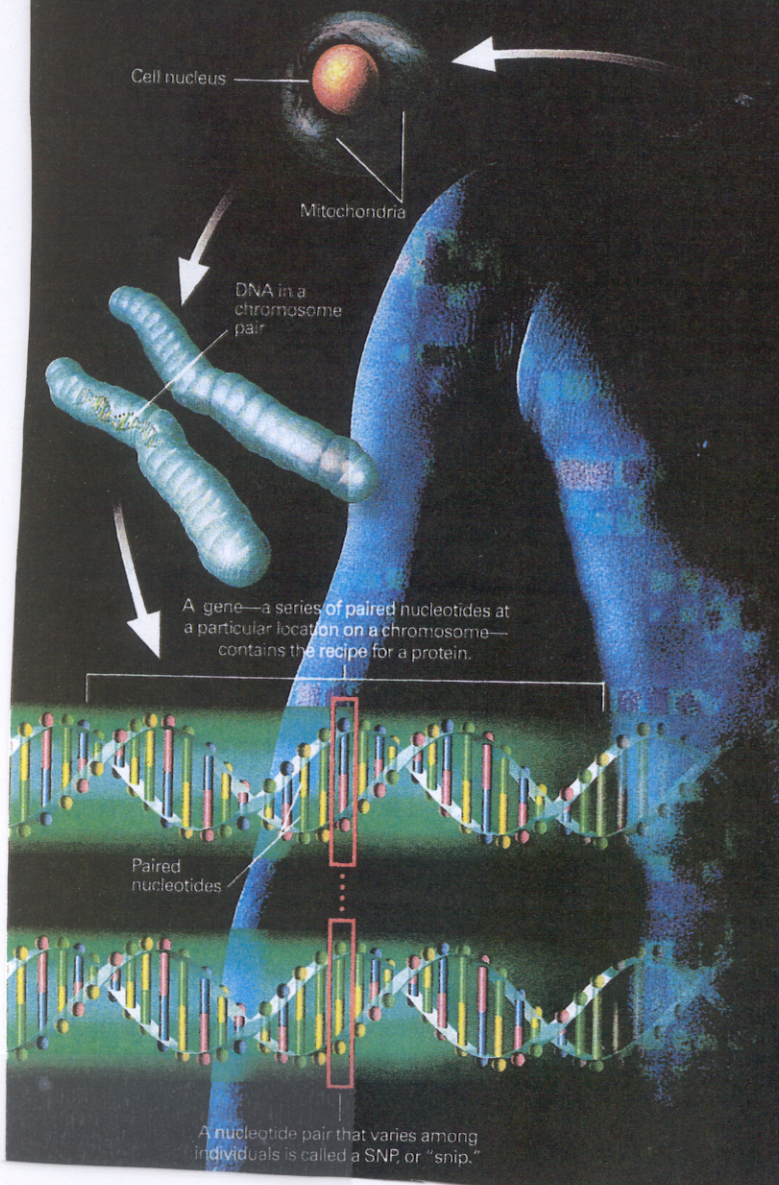
1.2% of Live Births Affected by these Mutations

ONLY MOLECULAR APPROACHES CAN detect these Mutations

Lead to RFLP NOT VNTR

The hidden language of cells

One-quarter of our DNA's three-billion-unit code has been spelled out by teams of scientists working on the Human Genome Project—an international effort spearheaded by the U.S. government—and by the project's corporate competitors. Researchers are on track to finish the rest by 2003. Once DNA has been sequenced, the 80,000 to 100,000 genes that make the proteins vital to human life will be easier to pinpoint. New treatments for disease and possibly cures won't be far behind.



RFLP

CELLS

Most cells contain complete instructions, in the form of DNA, for building a human being. Most DNA is in the nucleus; additional DNA is in cells' mitochondria.

CHROMOSOMES

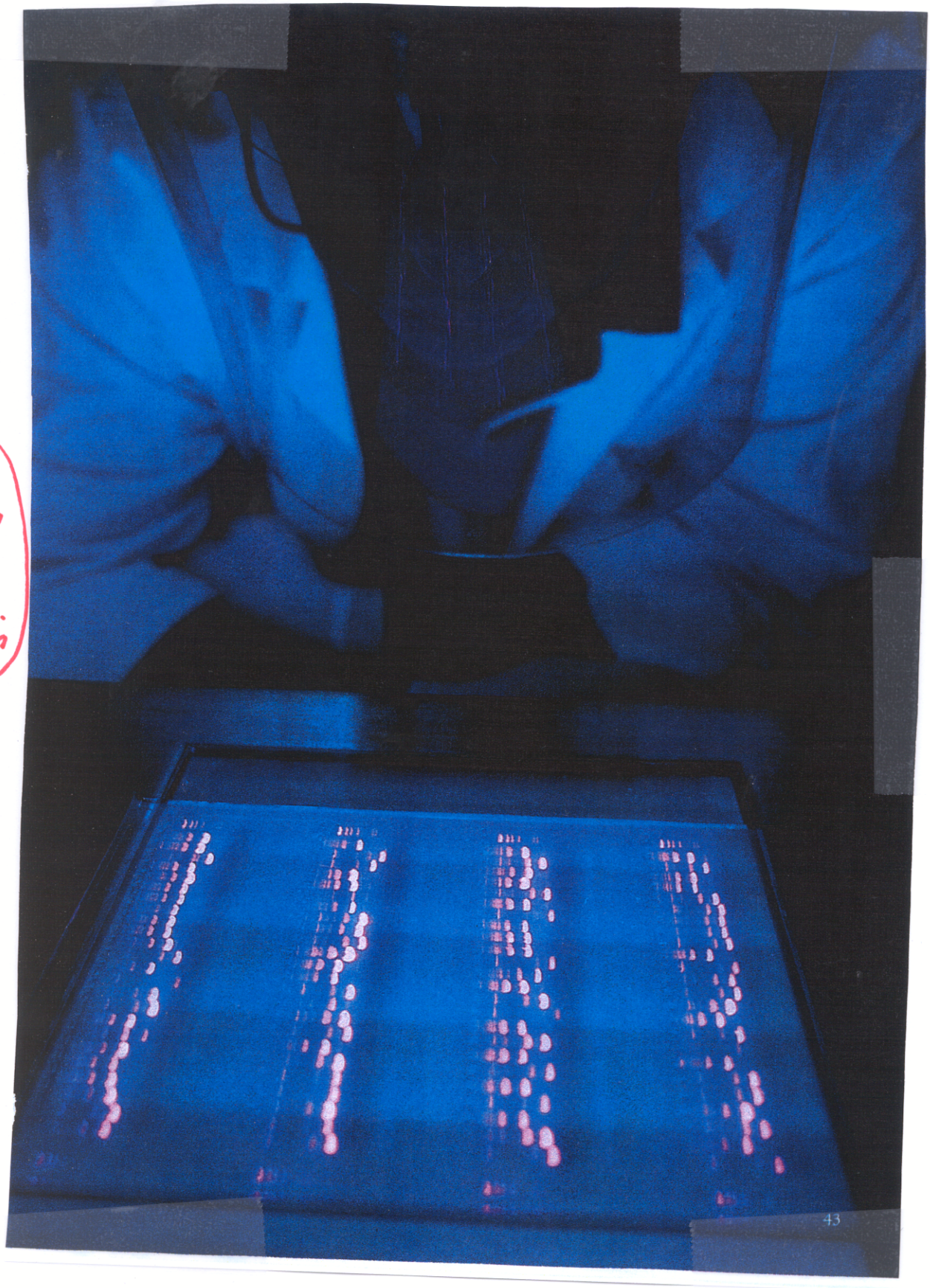
Within the nucleus, DNA is arranged in 23 pairs of rod-like packages called chromosomes—one set from the mother and one set from the father. Each chromosome contains a tightly packed strand of DNA. If unwound, the DNA in one cell's chromosomes would be more than six feet long.

DNA

DNA's structure is simple: Four chemical subunits called nucleotides pair up to form a twisted ladder. Mutations are common and mostly harmless, but even a one-nucleotide error can cause problems if it disrupts a critical gene. Less than 10 percent of our DNA makes up genes; some of the rest serves to regulate genes or plays important roles in chromosome function.

(D)

Visualization of Polymorphisms Due to Mutations



RFLP due to point Mutation

NIL

VNTR

OR By Sequencing.....

A 2.91-billion base pair (bp) consensus sequence of the euchromatic portion of the human genome was generated by the whole-genome shotgun sequencing method. The 14.8-billion bp DNA sequence was generated over 9 months from 27,271,853 high-quality sequence reads (5.11-fold coverage of the genome) from both ends of plasmid clones made from the DNA of five individuals. Two assembly strategies—a whole-genome assembly and a regional chromosome assembly—were used, each combining sequence data from Celera and the publicly funded genome effort. The public data were shredded into 550-bp segments to create a 2.9-fold coverage of those genome regions that had been sequenced, without including biases inherent in the cloning and assembly procedure used by the publicly funded group. This brought the effective coverage in the assemblies to eightfold, reducing the number and size of gaps in the final assembly over what would be obtained with 5.11-fold coverage. The two assembly strategies yielded very similar results that largely agree with independent mapping data. The assemblies effectively cover the euchromatic regions of the human chromosomes. More than 90% of the genome is in scaffold assemblies of 100,000 bp or more, and 25% of the genome is in scaffolds of 10 million bp or larger. Analysis of the genome sequence revealed 26,588 protein-encoding transcripts for which there was strong corroborating evidence and an additional ~12,000 computationally derived genes with mouse matches or other weak supporting evidence. Although gene-dense clusters are obvious, almost half the genes are dispersed in low G+C sequence separated by large tracts of apparently noncoding sequence. Only 1.1% of the genome is spanned by exons, whereas 24% is in introns, with 75% of the genome being intergenic DNA. Duplications of segmental blocks, ranging in size up to chromosomal lengths, are abundant throughout the genome and reveal a complex evolutionary history. Comparative genomic analysis indicates vertebrate expansions of genes associated with neuronal function, with tissue-specific developmental regulation, and with the hemostasis and immune systems. DNA sequence comparisons between the consensus sequence and publicly funded genome data provided locations of 2.1 million single-nucleotide polymorphisms (SNPs). A random pair of human haploid genomes differed at a rate of 1 bp per 1250 on average, but there was marked heterogeneity in the level of polymorphism across the genome. Less than 1% of all SNPs resulted in variation in proteins, but the task of determining which SNPs have functional consequences remains an open challenge.

1 bp change per 1250 bp on Average between 2 Genomes

21% cause protein change - why?

SNPs

Single Nucleotide Polymorphisms

SNPs Are MARKERS

- use SNPs to determine Linkage with disease Genes - Markers
- Associate with adverse drug reactions
- Associate with predisposition to heart disease, etc.

Nature 409, 928-933 (2001)

A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms

The International SNP Map Working Group*

** A full list of authors appears at the end of this paper.*

We describe a map of 1.42 million single nucleotide polymorphisms (SNPs) distributed throughout the human genome, providing an average density on available sequence of one SNP every 1.9 kilobases. These SNPs were primarily discovered by two projects: The SNP Consortium and the analysis of clone overlaps by the International Human Genome Sequencing Consortium. The map integrates all publicly available SNPs with described genes and other genomic features. We estimate that 60,000 SNPs fall within exon (coding and untranslated regions), and 85% of exons are within 5 kb of the nearest SNP. Nucleotide diversity varies greatly across the genome, in a manner broadly consistent with a standard population genetic model of human history. This high-density SNP map provides a public resource for defining haplotype variation across the genome, and should help to identify biomedically important genes for diagnosis and therapy.

Inherited differences in DNA sequence contribute to phenotypic variation, influencing an individual's anthropometric characteristics, risk of disease and response to the environment. A central goal of genetics is to pinpoint the DNA variants that contribute most significantly to population variation in each trait. Genome-wide linkage analysis and positional cloning have identified hundreds of genes for human diseases[1] (http://ncbi.nlm. nih.gov/OMIM), but nearly all are rare conditions in which mutation of a single gene is necessary and sufficient to cause disease. For common diseases, genome-wide linkage studies have had limited success, consistent with a more complex genetic architecture. If each locus contributes modestly to disease aetiology, more powerful methods will be required.

One promising approach is systematically to explore the limited set of common gene variants for association with disease[2–4]. In the human population most variant sites are rare, but the small number of common polymorphisms explain the bulk of heterozygosity[3] (see also refs 5–11). Moreover, human genetic diversity appears to be limited not only at the level of individual polymorphisms, but also in the specific combinations of alleles (haplotypes) observed at closely linked sites[8,11–14]. As these common variants are responsible for most heterozygosity in the population, it will be important to assess their potential impact on phenotypic trait variation.

If limited haplotype diversity is general, it should be practical to define common haplotypes using a dense set of polymorphic markers, and to evaluate each haplotype for association with disease. Such haplotype-based association studies offer a significant advantage: genomic regions can be tested for association without requiring the discovery of the functional variants. The required density of markers will depend on the complexity of the local haplotype structure, and the distance over which these haplotypes extend, neither of which is yet well defined.

Current estimates (refs 13–17) indicate that a very dense marker map (30,000–1,000,000 variants) would be required to perform haplotype-based association studies. Most human sequence variation is attributable to SNPs, with the rest attributable to insertions or deletions of one or more bases, repeat length polymorphisms and rearrangements. SNPs occur (on average) every 1,000–2,000 bases when two human chromosomes are compared[5,6,9,18–20], and are thus present at sufficient density for comprehensive haplotype analysis. SNPs are binary, and thus well suited to automated, high-throughput genotyping. Finally, in contrast to more mutable markers, such as microsatellites[21], SNPs have a low rate of recurrent mutation, making them stable indicators of human history. We have constructed a SNP map of the human genome with sufficient density to study human haplotype structure, enabling future study of human medical and population genetics.

Identification and characteristics of SNPs

The map contains all SNPs that were publicly available in November 2000. Over 95% were discovered by The SNP Consortium (TSC) and the public Human Genome Project (HGP). TSC contributed 1,023,950 candidate SNPs (http:// snp.cshl.org) identified by shotgun sequencing of genomic fragments drawn from a complete (45% of data) or reduced (55% of data) representation of the human genome[18,22]. Individual contributions were: Whitehead Institute, 589,209 SNPs from 2.57 million (M) passing reads; Sanger Centre, 262,279 SNPs from 1.16M passing reads; Washington University, 172,462 SNPs from 1.69M passing reads. TSC SNPs were discovered using a publicly available panel of 24 ethnically diverse individuals[23]. Reads were aligned to one another and to the available genome sequence, followed by detection of single base differences using one of two validated algorithms: Polybayes[24] and the neighbourhood quality standard (NQS[18,22]).

An additional 971,077 candidate SNPs were identified as sequence differences in regions of overlap between large-insert clones (bacterial artificial chromosomes (BACs) or P1-derived artificial chromosomes (PACs)) sequenced by the HGP. Two groups (NCBI/Washington University (556,694 SNPs): G.B., P.Y.K. and S.S.; and The Sanger Centre (630,147SNPs): J.C.M. and D.R.B.) independently analysed these overlaps using the two detection algorithms. This approach contributes dense clusters of SNPs throughout the genome. The remaining 5% of SNPs were discovered in gene-based studies, either by automated detection of single base differences in clusters of overlapping expressed sequence tags[24–28] or by targeted resequencing efforts (see ftp://ncbi.nlm.nih.gov/snp/human/submit_format/*/*publicat.rep. gz).

It is critical that candidate SNPs have a high likelihood of representing true polymorphisms when examined in population studies. Although many methods and contributors are represented on the map (see above), most SNPs (>95%) were contributed by two large-scale efforts that uniformly applied automated methods.