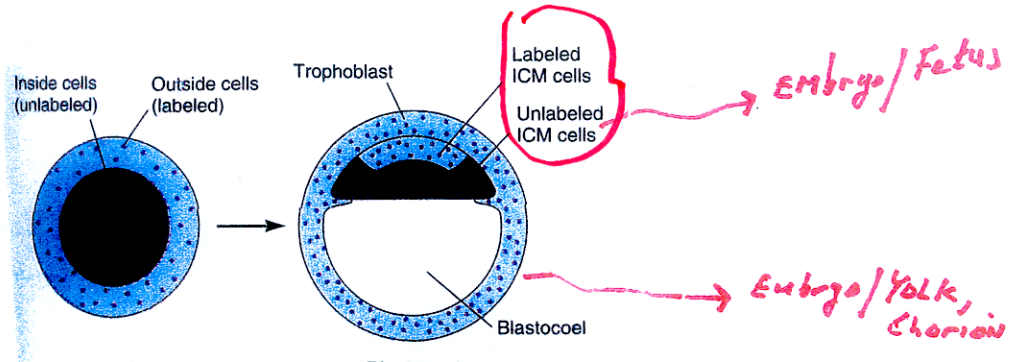
HUMAN EMBRYO FORMATION



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.

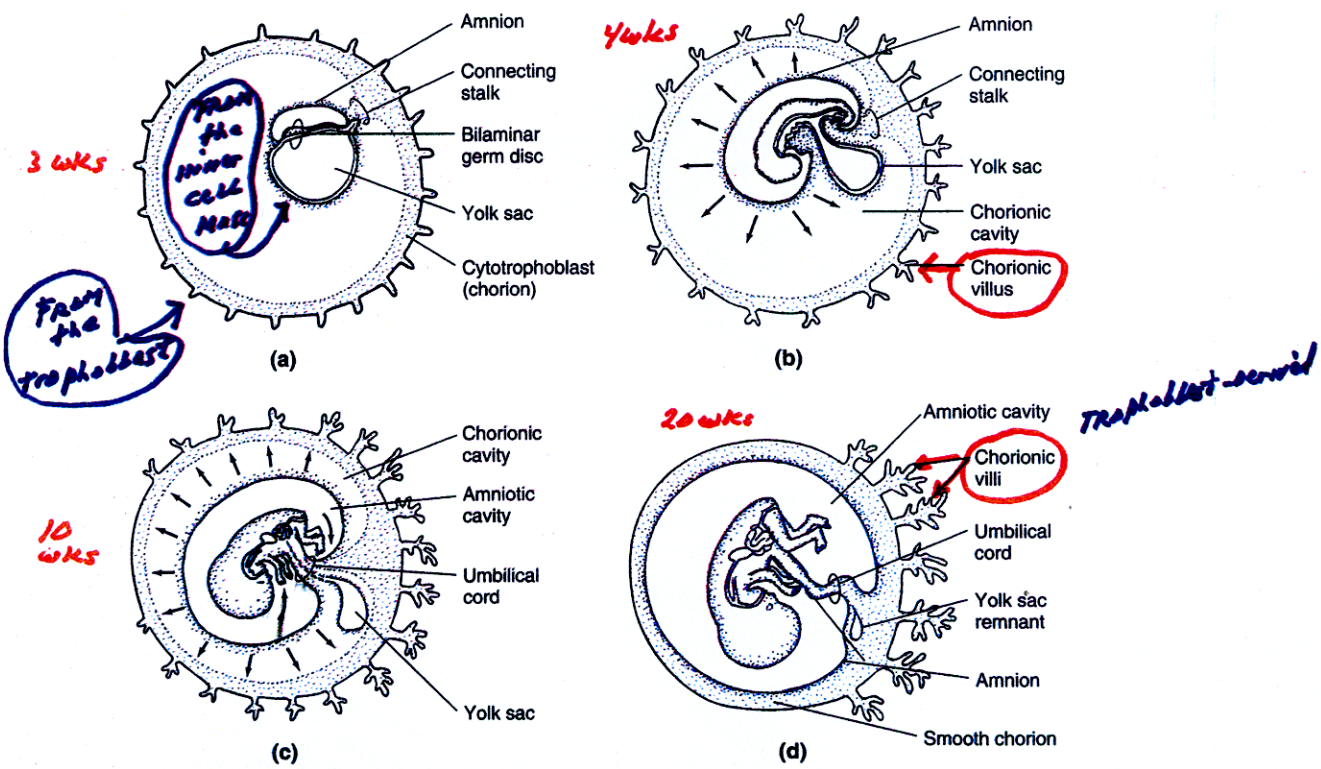


Figure 14.38 Extraembryonic membranes in human development: **(a)** at 3 weeks; **(b)** at 4 weeks; **(c)** at 10 weeks; **(d)** at 20 weeks. The connecting stalk develops into the umbilical cord. The amniotic cavity expands (arrows) until it completely fills the chorionic cavity and envelops the umbilical 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.

PREIMPLANTATION GENE DIAGNOSIS USING PCR - DEFECTIVE GENES

USE TO ASSAY GENES NOT LARGE CHROMOSOME ALTERATIONS

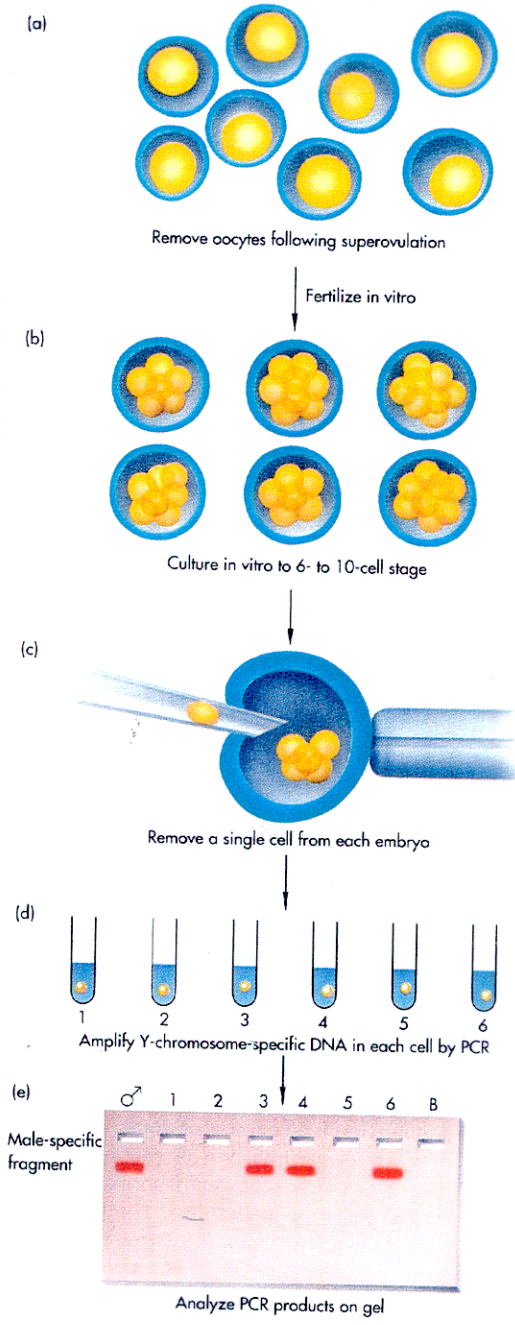
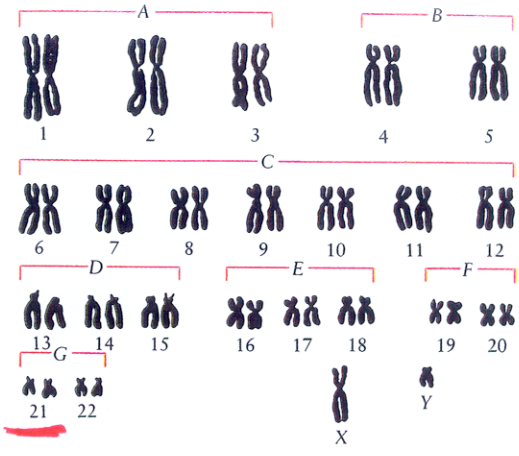


FIGURE 6-11
Determining sex of fetuses at risk for X-linked inherited disorders. (a) Oocytes are removed from the mother following superovulation and fertilized in vitro. (b) The oocytes that are fertilized successfully are cultured in vitro until there are 6 to 10 cells in each embryo. (c) A hole is made in the zona pellucida and a single cell removed from each embryo. (d) Amplification of the DYZ1 sequence is attempted. (e) Only in DNA from males is the male-specific DYZ1 sequence amplified by PCR, giving rise to a 149-bp, male-specific fragment. The lane marked with the male symbol is a positive control showing the expected fragment; the lane marked B (for "Blank") is from a PCR that included all the reagents but no DNA and is used to detect any contamination. Female embryos are negative (lanes 1, 2, and 5) and are implanted into the mothers.

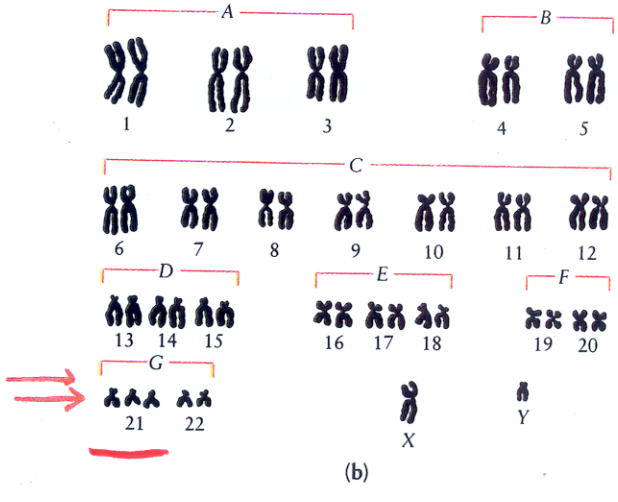
Karyotypes Reveal Chromosomal Abnormalities

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)

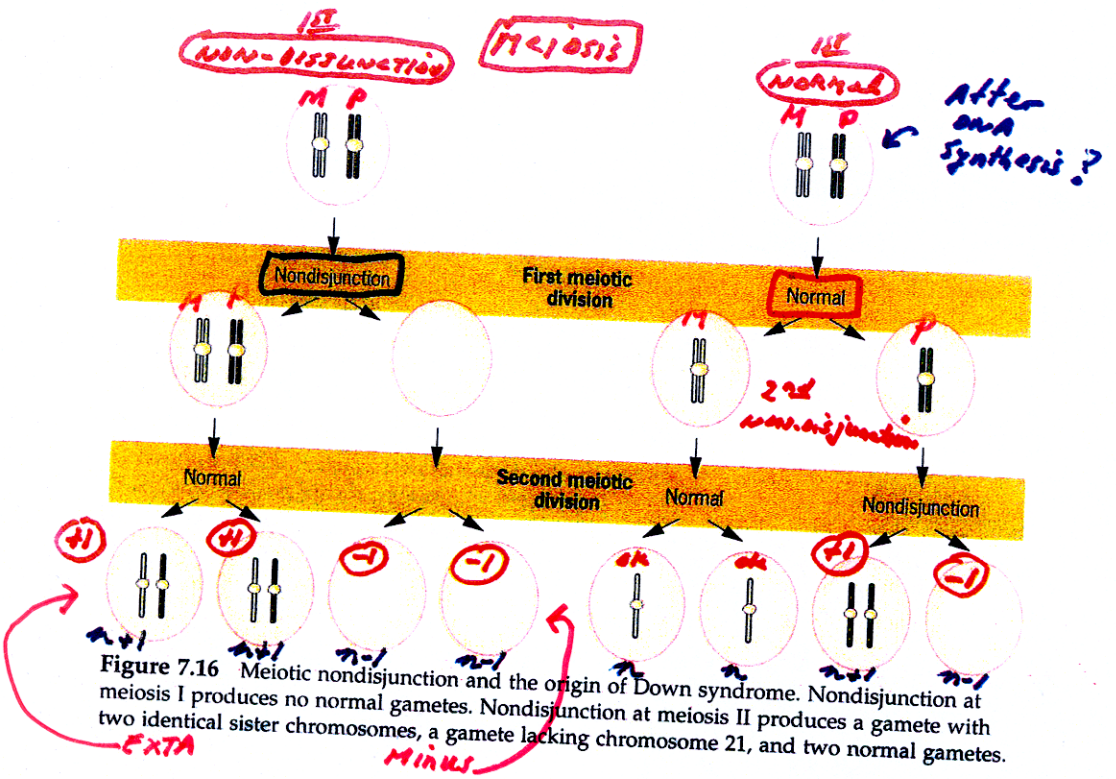


Figure 7.16 Meiotic nondisjunction and the origin of Down syndrome. Nondisjunction at meiosis I produces no normal gametes. Nondisjunction at meiosis II produces a gamete with two identical sister chromosomes, a gamete lacking chromosome 21, and two normal gametes.

ONLY PROBES OR BANDING PATTERN IDENTIFIES CHROMOSOME

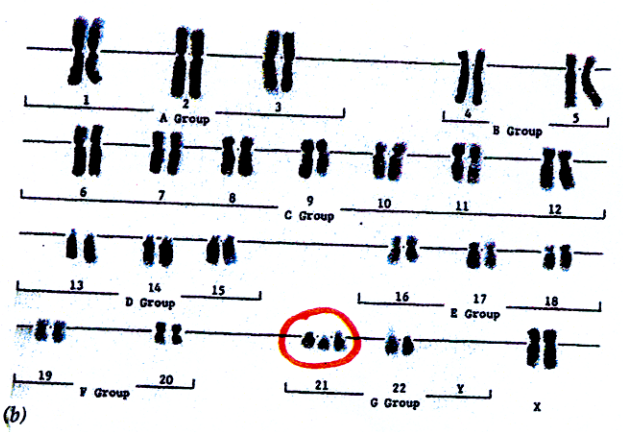


Figure 7.15 Down syndrome. (*a*) Facial features of a child with Down syndrome. (*b*) Karyotype of a child with Down syndrome, showing trisomy for chromosome 21 (47,XX, +21).

Detection of Extra Chromosome By In Situ Hybridization

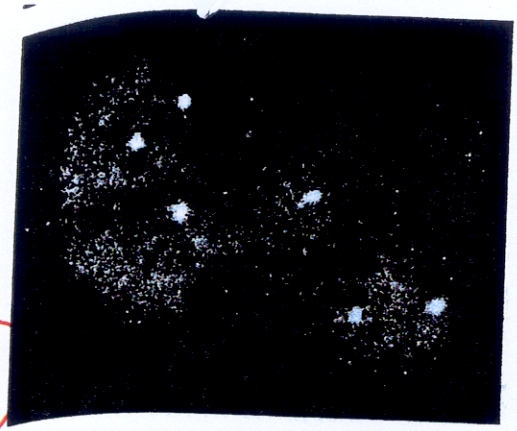


Fig. 3.13. Amniotic fluid cell nuclei of a fetus with trisomy 18 after CISS hybridization with the biotinylated Alu-PCR amplified YAC clone HTY 3045 (mapped to 18q23) detected with avidin-FITC. Nuclei were counterstained with propidium iodide

Note

3 chromosome 18's

Use of a Chromosome 18 Specific DNA Sequence

Amniotic cells

Recounting a genetic story

Roger H. Reeves

The DNA sequence of human chromosome 21, now published, provides indications that the total number of human genes has been overestimated, and is a valuable resource for research into Down syndrome.

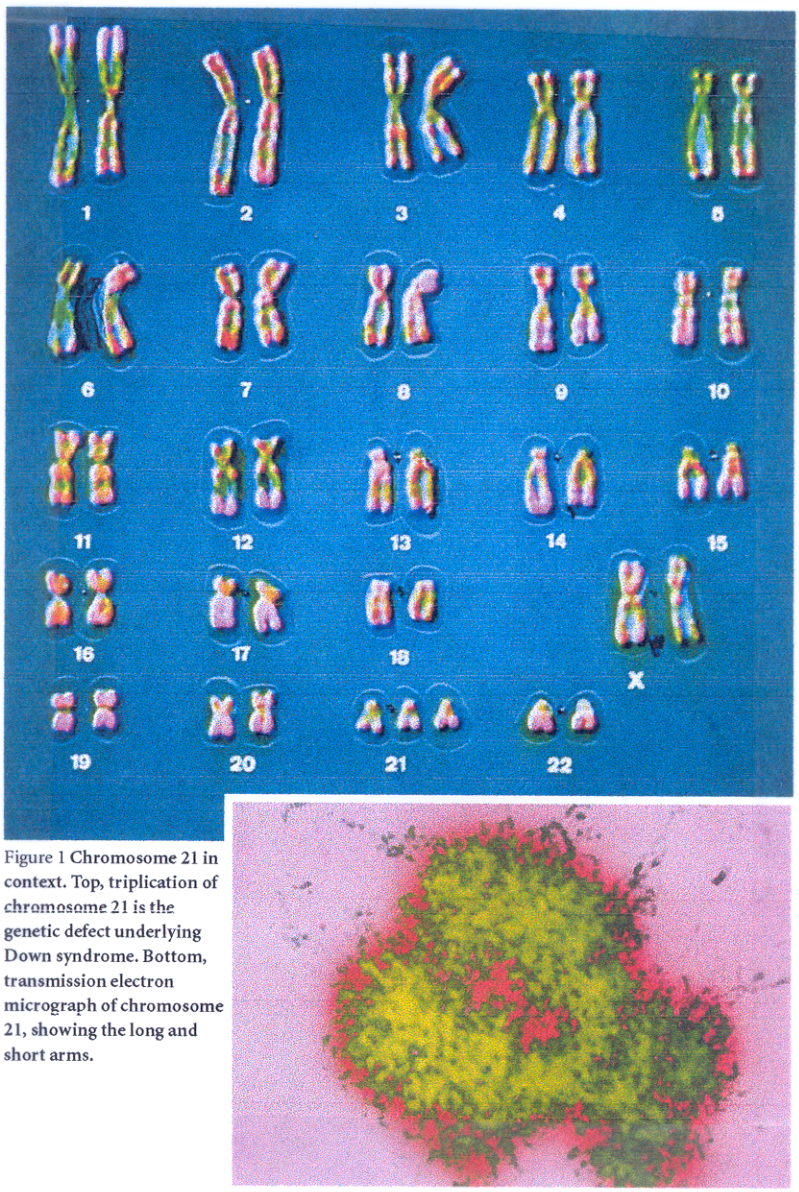


Figure 1 Chromosome 21 in context. Top, triplication of chromosome 21 is the genetic defect underlying Down syndrome. Bottom, transmission electron micrograph of chromosome 21, showing the long and short arms.

A gene expression map of human chromosome 21 orthologues in the mouse

The HSA21 expression map initiative
***Group 1:Yorick Gitton†, Nadia Dahmane‡, Sonya Baik†**
Ariel Ruiz i Altaba†
***Group 2:Lorenz Neidhardt§, Manuela Scholze§, Bernhard G. Herrmann§**
***Group 3:Pascal Kahlem||, Alia Benkahla||, Sabine Schrinner||,**
Reha Yildirimman||, Ralf Herwig||, Hans Lehrach|| & Marie-Laure Yaspo||

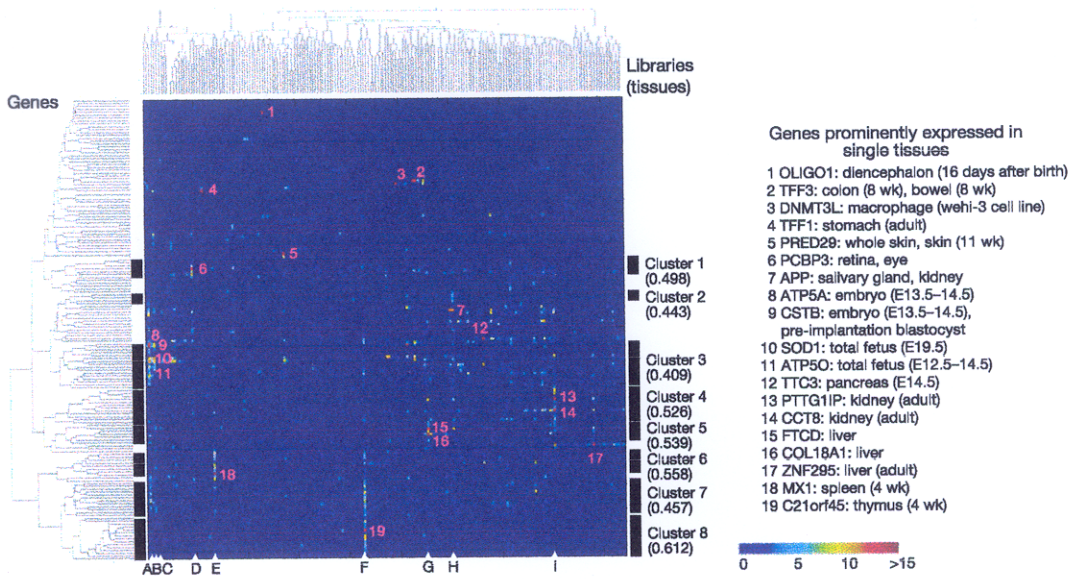


Figure 4 EST analysis. Matrix displaying the expression profiles of 159 mmu21 genes (rows) within 190 cDNA libraries (columns). Dendrograms used to reorder libraries (top) and genes (left) are shown, together with eight significant gene clusters (solid bars, left and right of the matrix together with correlation coefficients). The cluster composition is shown with corresponding libraries (A-I and indicated by white arrows at the bottom of the matrix). Coloured dots represent the number of ESTs found in a given library for each gene (see scale). Numbers in red refer to genes prominently expressed in single tissues, listed at the right. Interactive figure and details are given in the Supplementary Information. Gene symbols of human orthologues are used.

75c

The DNA sequence of human chromosome 21

The chromosome 21 mapping and sequencing consortium
M. Hattori*, A. Fujiyama*, T. D. Taylor*, H. Watanabe*, T. Yada*, H.-S. Park*, A. Toyoda*, K. Ishii*, Y. Totoki*, D.-K. Choi*, E. Soeda†, M. Ohki‡, T. Takagi§, Y. Sakaki*§; S. Taudien‖, K. Blechschmidt‖, A. Polley‖, U. Menzel‖, J. Delabar¶, K. Kumpf‖, R. Lehmann‖, D. Patterson#, K. Reichwald‖, A. Rump‖, M. Schillhabel‖, A. Schudy‖, W. Zimmermann‖, A. Rosenthal‖; J. Kudoh☆, K. Shibuya☆, K. Kawasaki☆, S. Asakawa☆, A. Shintani☆, T. Sasaki☆, K. Nagamine☆, S. Mitsuyama☆, S. E. Antonarakis, S. Minoshima☆, N. Shimizu☆, G. Nordsiek††, K. Hornischer††, P. Brandt††, M. Scharfe††, O. Schön††, A. Desario‡‡, J. Reichelt††, G. Kauer††, H. Blöcker††; J. Ramser§§, A. Beck§§, S. Klages§§, S. Hennig§§, L. Riesselmann§§, E. Dagand§§, T. Haaf§§, S. Wehrmeyer§§, K. Borzym§§, K. Gardiner#, D. Nizetic‖‖, F. Francis§§, H. Lehrach§§, R. Reinhardt§§ & M.-L. Yaspo§§**

Consortium Institutions:
** RIKEN, Genomic Sciences Center, Sagamihara 228-8555, Japan*
‖ Institut für Molekulare Biotechnologie, Genomanalyse, D-07745 Jena, Germany
☆ Department of Molecular Biology, Keio University School of Medicine, Tokyo 160-8582, Japan
†† GBF (German Research Centre for Biotechnology), Genome Analysis, D-38124 Braunschweig, Germany
§§ Max-Planck-Institut für Molekulare Genetik, D-14195 Berlin-Dahlem, Germany
Collaborating Institutions:
† RIKEN, Life Science Tsukuba Research Center, Tsukuba 305-0074, Japan
‡ Cancer Genomics Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan
§ Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan
¶ UMR 8602 CNRS, UFR Necker Enfants-Malades, Paris 75730, France
 # Eleanor Roosevelt Institute, Denver, Colorado 80206, USA
*** Medical Genetics Division, University of Geneva Medical School, Geneva 1211, Switzerland*
‡‡ CNRS UPR 1142, Institut de Biologie, Montpellier, 34060, France
‖‖ School of Pharmacy, University of London, London WC1N 1AX, UK

Chromosome 21 is the smallest human autosome. An extra copy of chromosome 21 causes Down syndrome, the most frequent genetic cause of significant mental retardation, which affects up to 1 in 700 live births. Several anonymous loci for monogenic disorders and predispositions for common complex disorders have also been mapped to this chromosome, and loss of heterozygosity has been observed in regions associated with solid tumours. Here we report the sequence and gene catalogue of the long arm of chromosome 21. We have sequenced 33,546,361 base pairs (bp) of DNA with very high accuracy, the largest contig being 25,491,867 bp. Only three small clone gaps and seven sequencing gaps remain, comprising about 100 kilobases. Thus, we achieved 99.7% coverage of 21q. We also sequenced 281,116 bp from the short arm. The structural features identified include duplications that are probably involved in chromosomal abnormalities and repeat structures in the telomeric and pericentromeric regions. Analysis of the chromosome revealed 127 known genes, 98 predicted genes and 59 pseudogenes.

Nature
May 2000
Vol 405

Chromosome 21

(1) ~ 200-300 genes

(2) Down's Syndrome Results From over-expression of these genes

(3) SOD / Superoxide Dismutase Gene

SOD converts oxygen radicals to H_2O_2
if SOD activity increase above enzyme activity to remove H_2O_2 (peroxidases) – peroxidase damage can occur to brain cells

(4) Amyloid Protein Gene

Amyloid plagues in brain → Down's individuals + Alzheimer individuals

(5) Ts65Dn mouse – TRIsomy 16 – Genes similar to those on Human 21 – : Model to understand Molecular Basis of Down's Syndrome.

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				
−1 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.
+1 47,XXY	2n+1	Klinefelter	1/500 male births	Male, subfertile with small testes, developed breasts, feminine-pitched voice, knock knees, long limbs.
+2 48,XXXY	2n+2			
+2 48,XXYY	2n+2			
+3 49,XXXXY	2n+3			
+4 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

1/7 or 15%!

Live Birth's

	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

Lead to Spontaneous Abortion

Explain why 1 Xtra chromosome can lead to Abortion?

Explain How Chromosomes 13, 18, & 21 can lead to live births with severe abnormalities & other chromosomes ↳ death

How obtain polyploid fetus?

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 chromosome abnormalities
4. ~1.2% of Live Birth Mutations due to DNA Changes / Point Mutations

550/85000 = 0.65%

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

1/50!

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

Live Births

Type of mutation	Percentage of live births
Gene mutation (Gene) GENE	
Autosomal dominant	0.90
Autosomal recessive	0.25
X-linked	0.05
Total gene mutation	1.20
Chromosome mutation (Chromosomal) 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

total 1.81%

GENE MUTATION →

CHROMOSOME ABNORMALITIES →

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

HOW RELATE TO DNA TESTING - ELIMINATE "lottery"?

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

SPERM + EGG FORMATION

♂ ♀

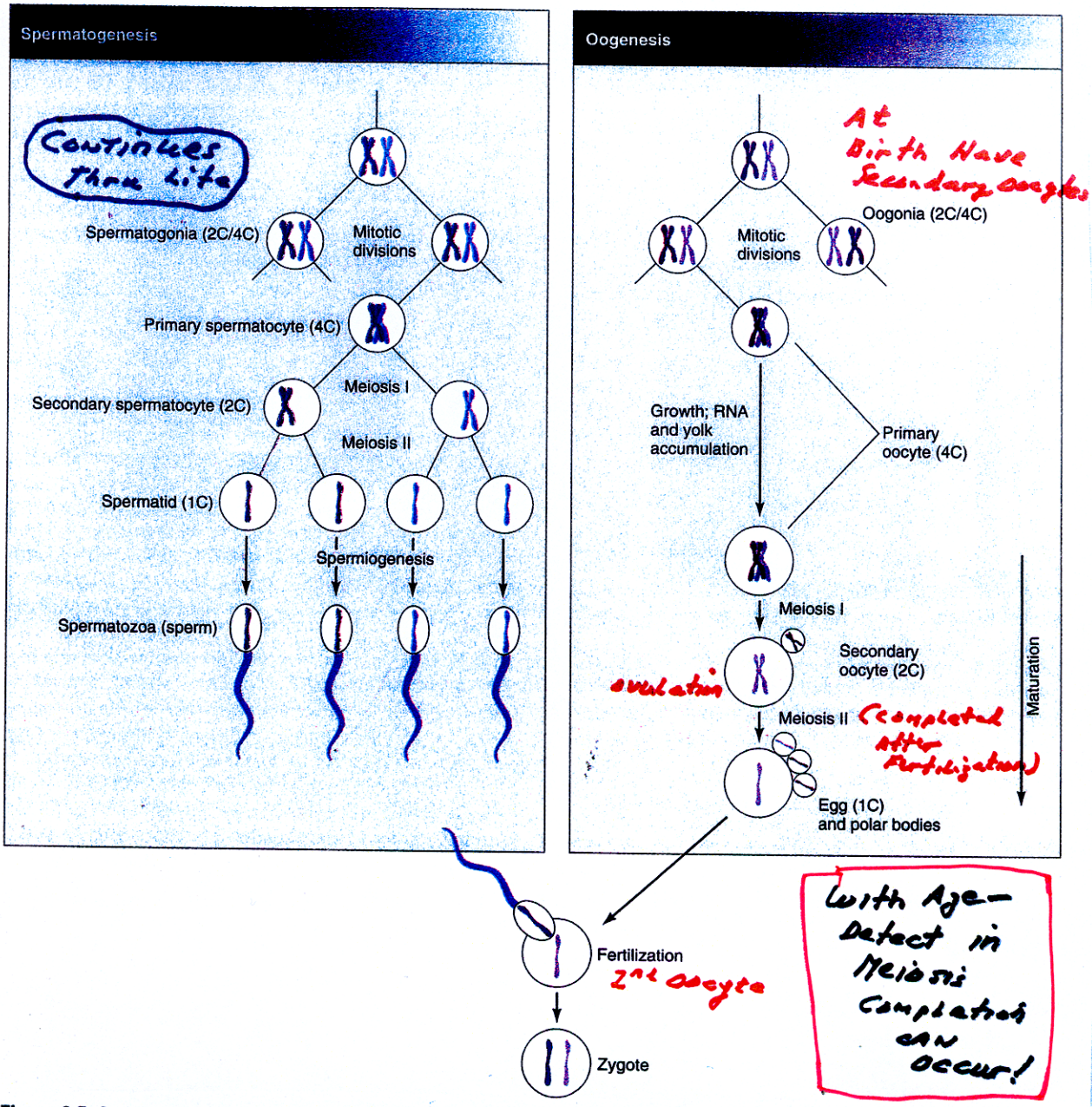
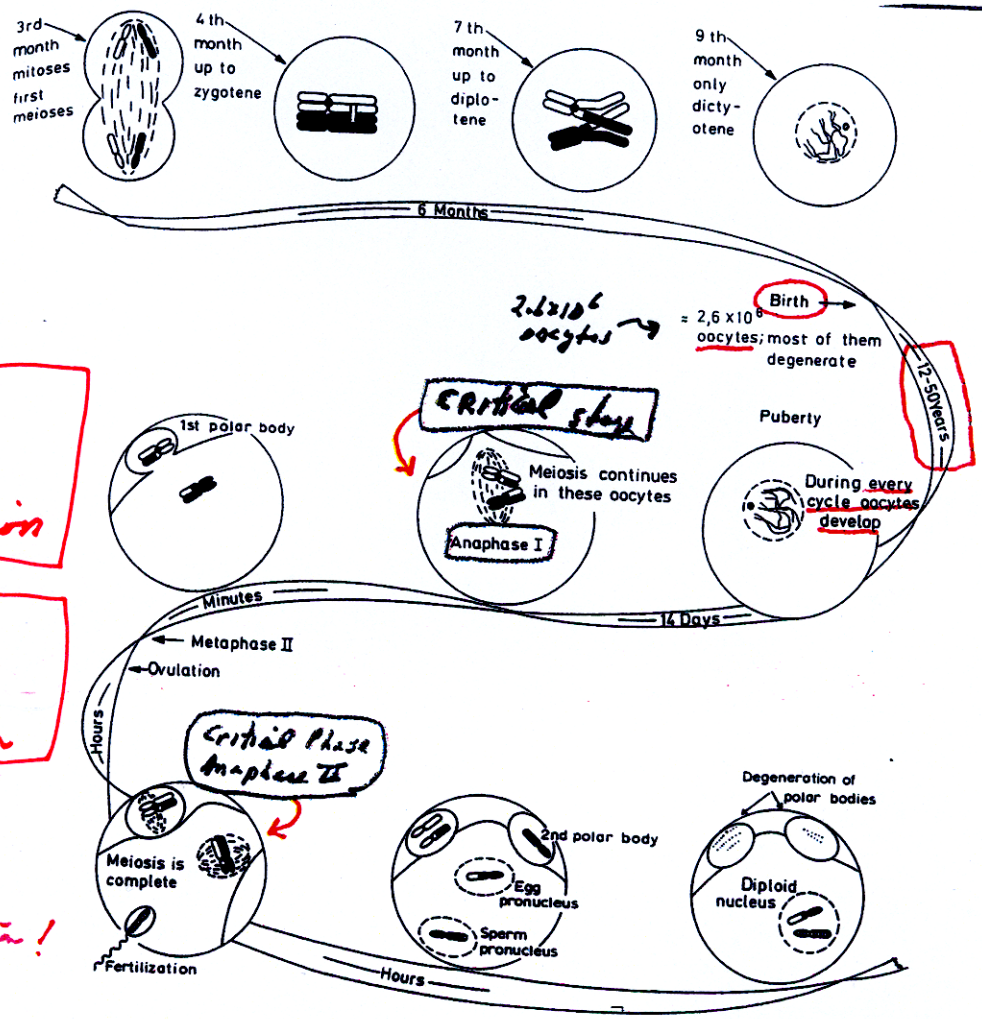


Figure 3.5 Comparison of spermatogenesis and oogenesis. Primordial germ cells divide mitotically, producing spermatogonia in males and oogonia in females. These cells are diploid, containing two or four genomic complements (2C or 4C), depending on their stage in the mitotic cycle. Before the gonia enter 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).

DEFECTS AGE RELATED — ♀ + AND ♂

Meiosis in Egg Formation is Completed After Fertilization



Meiosis I PRIOR to ovulation

MEIOSIS II After Fertilization

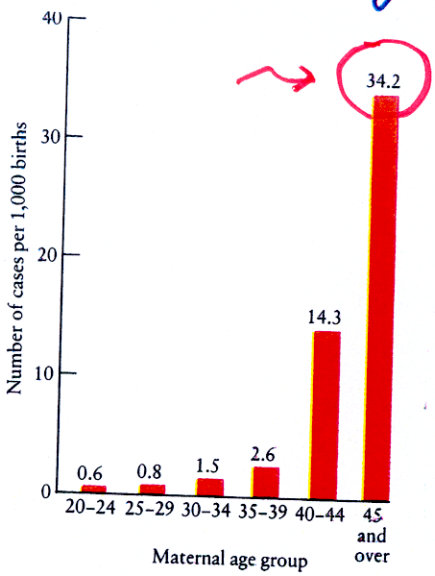
Egg forms after fertilization!

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!

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.*

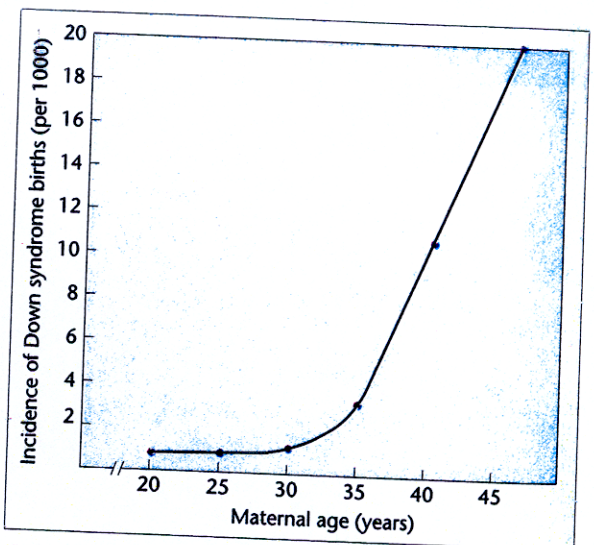
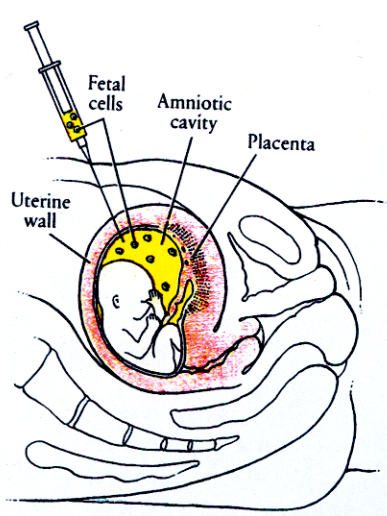


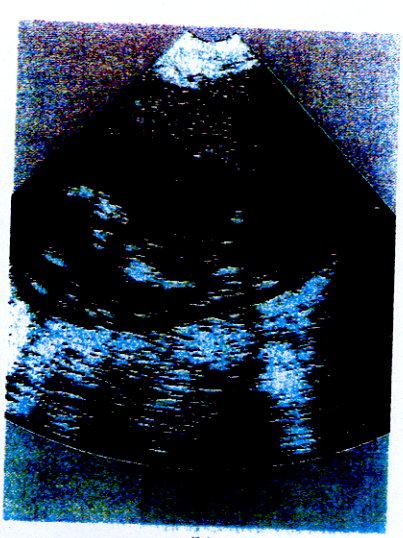
FIGURE 10.6 Incidence of Down syndrome births contrasted with maternal age.

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)

IMPORTANCE OF PRENATAL CHROMOSOME TESTING

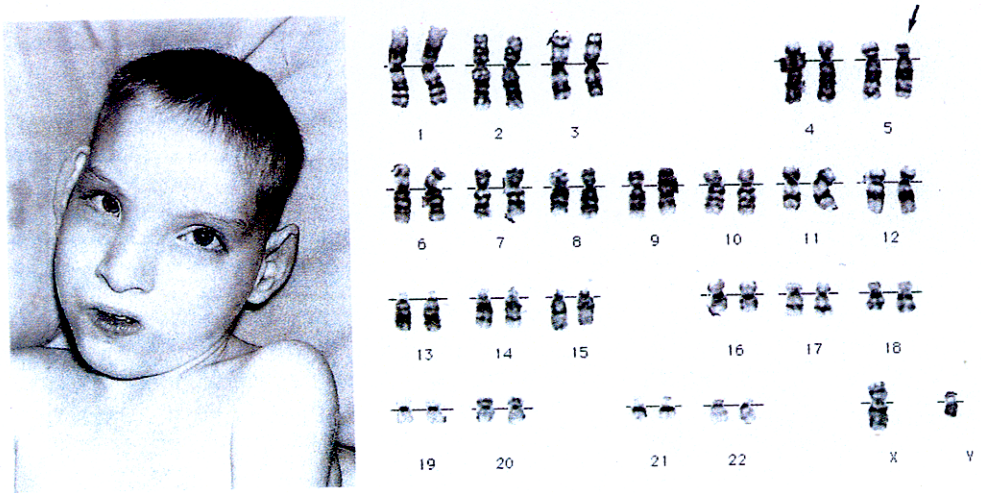
Large deletions also cause genetic abnormalities

(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

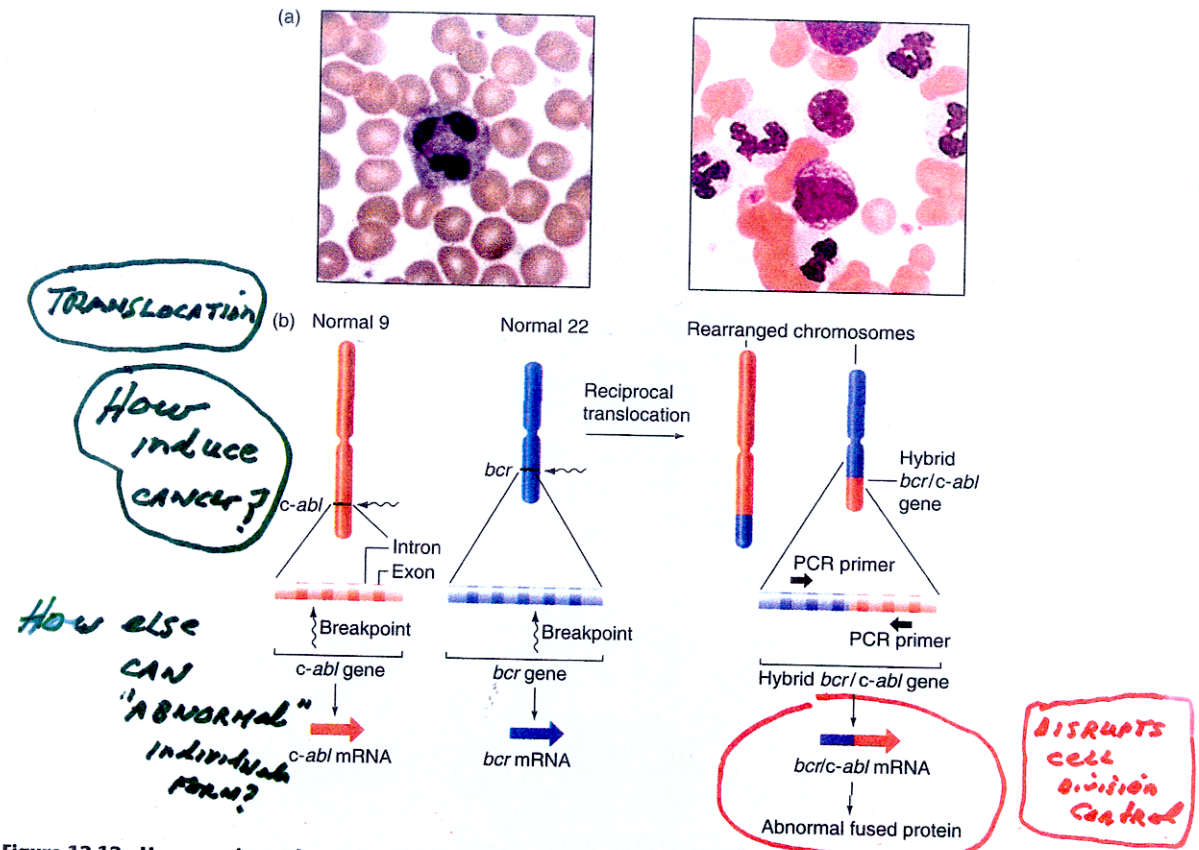
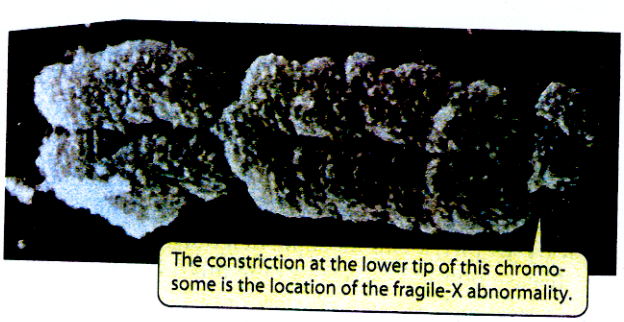


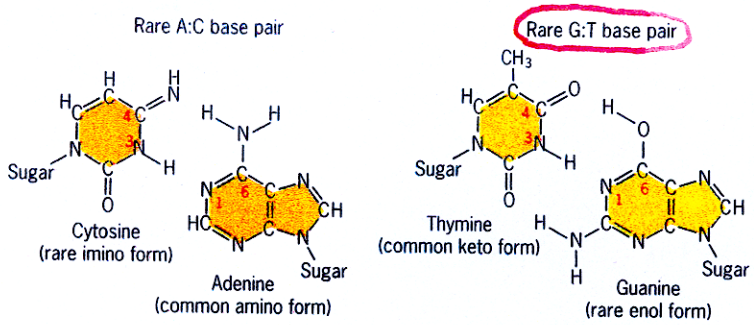
Figure 12.12 How a reciprocal translocation helps cause one kind of leukemia. (a) Uncontrolled divisions of large, dark-staining white blood cells in the blood of a leukemia patient (right) produce a higher than normal ratio of white to red blood cells than that of a normal individual (left). (b) A reciprocal 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.



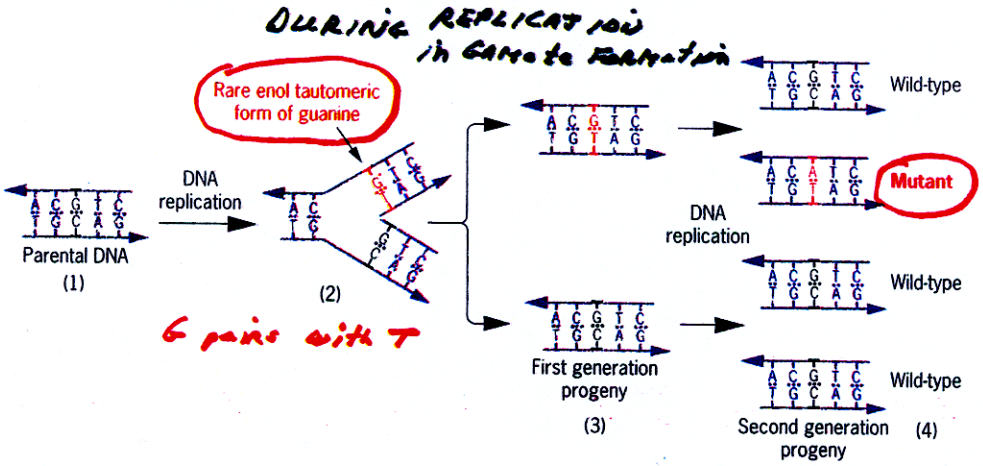
18.5 A Fragile-X Chromosome at Metaphase
The chromosomal abnormality that causes the mental retardation symptomatic of fragile-X syndrome shows up physically as a constriction.

Chromosome Breakage/Deletion

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

MOST CHANGES IN GENOME IN NON-CODING DNA & DO NOT LEAD TO MUTATIONS

RECALL - ONLY 1% of GENOME = EXONS!

SINGLE BASE PAIR CHANGES (SNPs) ARE FREQUENT IN GENOME

TABLE 9.1 Five Classes of DNA Polymorphism

Class	Cause	Rate of Mutation per Locus per Gamete	Frequency in Genome	Number per Human Genome (on average)
Single base	Mutagens or replication errors	10^{-8}–10^{-9}	1/700 bp	3 million
Microsatellite	Slippage during replication	10^{-3}	1/30,000 bp	100,000
Minisatellite	Unequal crossovers	10^{-3}	Unknown; discovered by chance	Fewer than 100 families known, yielding 1000 copies in all
Deletions	Mutagens; unequal crossovers	Extremely rare	Very low	0 – a few
Duplications	Mutagens; unequal crossovers	Extremely rare	Very low	0 – a few
Other insertions (excluding those resulting from micro- or minisatellite recombination)	Transposable elements	Extremely rare	Very low	0 – a few
Complex haplotype (any locus of 5 kb or more)	Any of the above	Combination of the above	Not applicable	Not applicable

DETECTED USING RESTRICTION ENZYMES (The OLD FASHIONED WAY) OR BY DIRECT DNA sequencing of 2 individuals' genes / genomes

ONLY MOLECULAR APPROACHES CAN DETECT SNPs

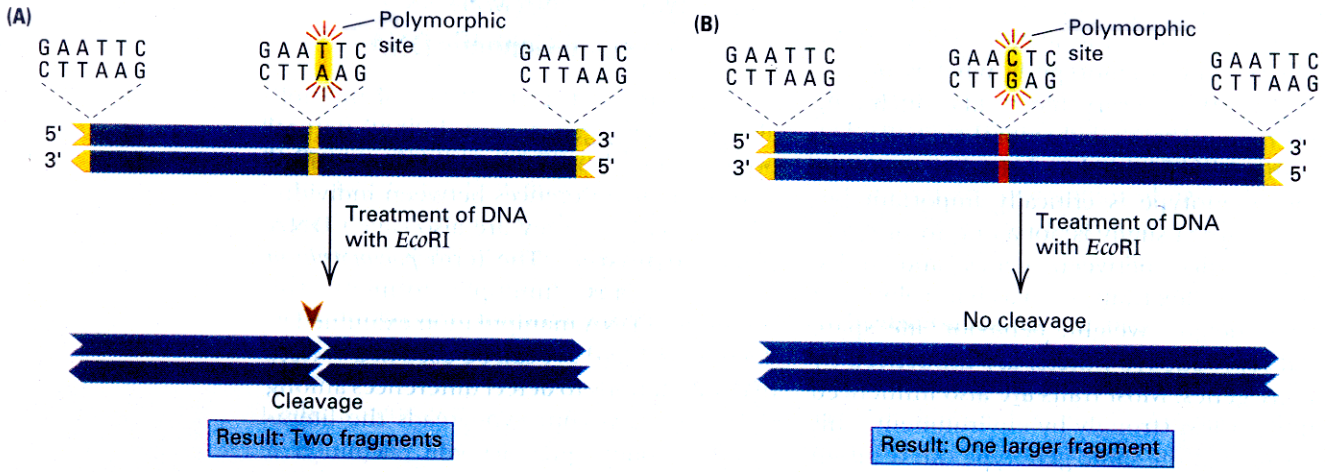
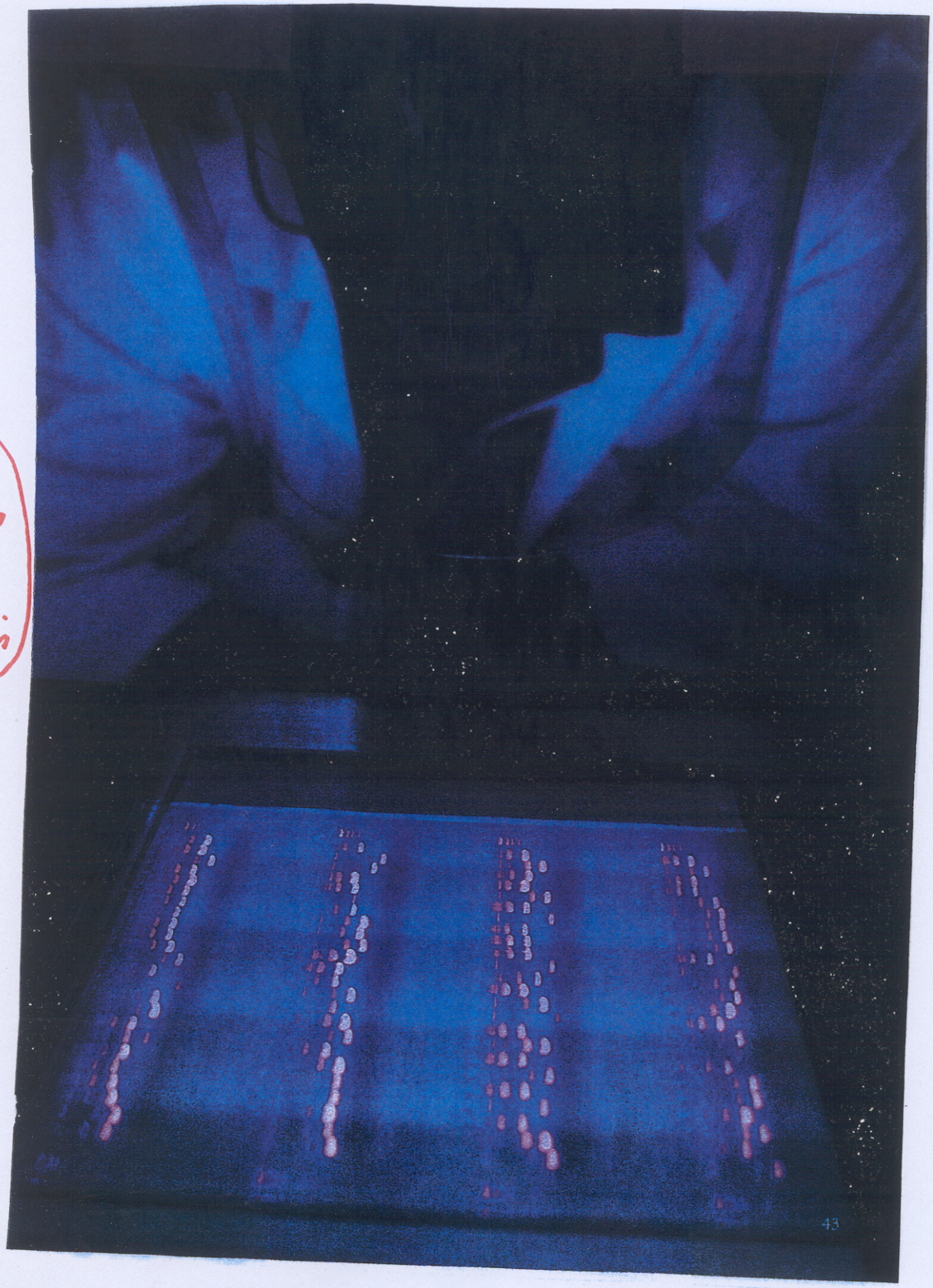


Figure 2.23 A minor difference in the DNA sequence of two molecules can be detected if the difference eliminates a restriction site. (A) This molecule contains three restriction sites for *Eco*RI, including one at each end. It is cleaved into two fragments by the enzyme. (B) This molecule has an altered *Eco*RI site in the middle, in which 5'-GAATTC-3' becomes 5'-GAACTC-3'. The altered site cannot be cleaved by *Eco*RI, so treatment of this molecule with *Eco*RI results in one larger fragment.

The old fashioned way to detect RFLPs —

DNA Blot
PCR
ASO

Visualization of Polymorphisms Due to Mutations



RFLP due to point Mutation

NOT VNTR

OR By Sequencing .....

THE CURRENT METHOD

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

MARKERS !!

SNPs Are MARKERS

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

PEOPLE, GROUPS, DISEASES, ORIGINS

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.

(1) Did not know who

(2) All gave consent & agreed by IRB

cannot use for group specific studies

(P9)

MOST SNPs ARE NOT IN CODING SEQUENCES & HAVE NO PHENOTYPIC EFFECTS!

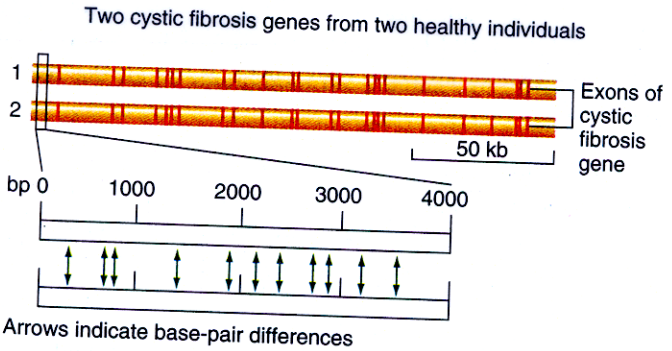


Figure 9.2 Base-pair differences between DNA cloned from the cystic fibrosis locus of two healthy individuals. These base-pair differences have no phenotypic effect; apparently they neither encode nor regulate expressed regions of the gene.

$$\frac{3\times10^{9}\text{ bp/genome}}{1250\text{ bp/SNP}} = 2.4\times10^{6}\text{ SNPs/genome!}$$

EACH OF US DIFFERS BY 2×10^{6}-3×10^{6} nts.!!

~0.07% of genome!

MOST OCCUR IN INTERGENIC/INTRON REGIONS & ARE USEFUL IN FORENSICS & DISEASE MARKING

USING SNPs as GENE MARKERS

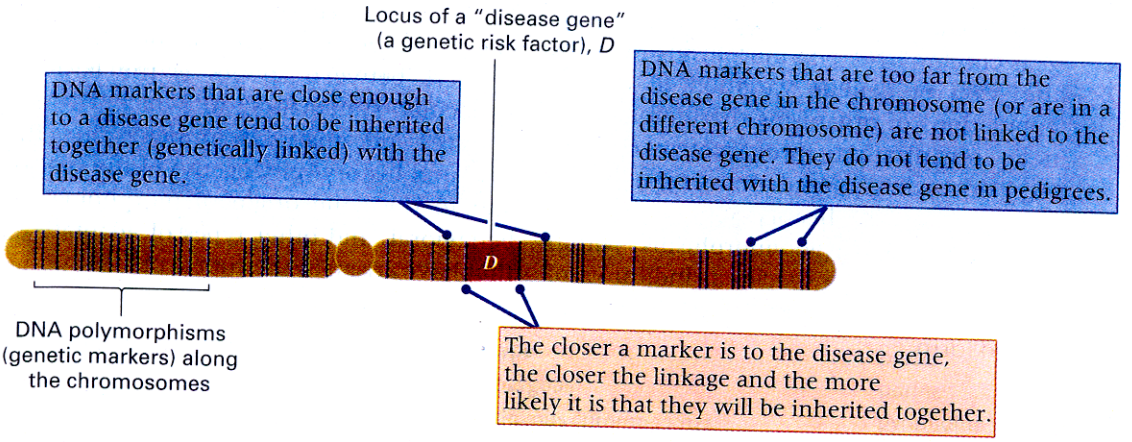


Figure 2.29 Concepts in genetic localization of genetic risk factors for disease. Polymorphic DNA markers (indicated by the vertical lines) that are close to a genetic risk factor (*D*) in the chromosome tend to be inherited together with the disease itself. The genomic location of the risk factor is determined by examining the known genomic locations of the DNA polymorphisms that are linked with it.

MAP GENES FOR DISEASE SUSCEPTIBILITY — GENES ENCODING COMPLEX (MULTI-GENIC) TRAITS (e.g., Heart Disease, Depression, DRUG Sensitivity, Obesity)

SNPs = INDIVIDUAL GENE PROFILE

↳ INDIVIDUAL Medicine!