

HC70A Winter 2005

Professor Bob Goldberg

Lecture # 9

Analysis of Sequence Variation in
the Human Genome: Are There
Races?

Themes/Concepts

- 1 Visions of the Human Genome Project
- 2 DNA Sequence Polymorphisms / SNPs / Review / How Arise?
- 3 Markers & SNPs & Multiple Alleles / Review
- 4 Detecting SNPs in the Human Genome - For Variants & Markers & Disease Genes - RFLP, ASO, Sequencing, Chips
- 5 Ethical Issues in Genetic Screening Programs
- 6 Are there "Group" Differences in the Human Genome?
- 7 Haplotypes & Groups & Disease
- 8 Variation in Human Genes & "Races"
- 9 Are there Races? What is the History & the Science?

GOALS OF THE HUMAN GENOME PROJECT

TABLE 17-5 Eight Goals of Human Genome Project (1998-2003)

1. Human DNA sequence	<ul style="list-style-type: none">• Finish the complete human genome sequence by the end of 2003. DAVE• Achieve coverage of at least 90% of the genome in a working draft based on mapped clones by the end of 2001.• Make the sequence totally and freely accessible.
2. Sequencing technology	<ul style="list-style-type: none">• Continue to increase the throughput and reduce the cost of current sequencing technology.• Support research on novel technologies that can lead to significant improvements in sequencing technology.• Develop effective methods for the development and introduction of new sequencing technologies.
3. Human genome sequence variation	<ul style="list-style-type: none">• Develop technologies for rapid, large-scale identification and/or scoring of single-nucleotide polymorphisms and other DNA sequence variants.• Identify common variants in the coding regions of the majority of identified genes during this five-year period.• Create a SNP map of at least 100,000 markers.• Create public resources of DNA samples and cell lines.
4. Functional genomics technology	<ul style="list-style-type: none">• Generate sets of full-length cDNA clones and sequences that represent human genes and model organisms.• Support research on methods for studying functions of nonprotein-coding sequences.• Develop technology for comprehensive analysis of gene expression.• Improve methods for genomewide mutagenesis.• Develop technology for large-scale protein analyses.
5. Comparative genomics	<ul style="list-style-type: none">• Complete the sequence of the roundworm <i>C. elegans</i> genome and the fruitfly <i>Drosophila</i> genome. DAVE• Develop an integrated physical and genetic map for the mouse, generate additional mouse cDNA resources, and complete the sequence of the mouse genome by 2008. DAVE
6. Ethical, legal, and social issues	<ul style="list-style-type: none">• Examine issues surrounding completion of the human DNA sequence and the study of <u>genetic variation</u>.• Examine issues raised by the integration of genetic technologies and information into health care and public health activities.• Examine issues raised by the integration of knowledge about <u>genomics</u> and <u>gene-environment</u> interactions in nonclinical settings.• Explore how new genetic knowledge may interact with a variety of philosophical, theological, and ethical perspectives.• Explore how racial, ethnic, and socioeconomic factors affect the use, understanding, and interpretation of genetic information, the use of genetic services, and the development of policy.
7. Bioinformatics and computational biology	<ul style="list-style-type: none">• Improve content and utility of databases.• Develop better tools for data generation, capture, and annotation.• Develop and improve tools and databases for comprehensive functional studies.• Develop and improve tools for representing and analyzing sequence similarity and variation.• Create mechanisms to support effective approaches for producing robust, exportable software that can be widely shared.
8. Training and manpower	<ul style="list-style-type: none">• Nurture the training of scientists skilled in genomics research.• Encourage the establishment of academic career paths for genomic scientists.• Increase the number of scholars who are knowledgeable in both <u>genomic</u> and <u>genetic sciences</u> and in ethics, law, or the social sciences.

Source: Adapted from <http://www.ornl.gov/TechResources/Human.Genome/hg5yp/goal.html>.

A vision for the future of genomics research

A blueprint for the genomic era.

Francis S. Collins, Eric D. Green, Alan E. Guttmacher and Mark S. Guyer on behalf of the US National Human Genome Research Institute*

I Genomics to biology

Elucidating the structure and function of genomes

① BIOLOGY

Grand Challenge I-1 Comprehensively identify the structural and functional components encoded in the human genome

Grand Challenge I-2 Elucidate the organization of genetic networks and protein pathways and establish how they

contribute to cellular and organismal phenotypes

Grand Challenge I-3 Develop a detailed understanding of the heritable variation in the human genome

Grand Challenge I-4 Understand evolutionary variation across species and the mechanisms underlying it

② HEALTH

II Genomics to health

Translating genome-based knowledge into health benefits

Grand Challenge I-5 Develop policy options that facilitate the widespread use of genome information in both research and clinical settings

Grand Challenge II-1 Develop robust strategies for identifying the genetic contributions to disease and drug response

Grand Challenge II-2 Develop strategies to identify gene variants that contribute to good health and resistance to disease

Grand Challenge II-3 Develop genome-based approaches to prediction of disease susceptibility and drug response, early detection of illness, and molecular taxonomy of disease states

Grand Challenge II-4 Use new understanding of genes and pathways to develop powerful new therapeutic approaches to disease

Grand Challenge II-5 Investigate how genetic risk information is conveyed in clinical settings, how that information influences health strategies and behaviours, and how these affect health outcomes and costs

Grand Challenge II-6 Develop genome-based tools that improve the health of all. Disparities in health status constitute a significant global issue, but can genome-based approaches to health and disease help to reduce this problem?

③ SOCIETY

III Genomics to society

Promoting the use of genomics to maximize benefits and minimize harms

Grand Challenge III-1 Develop policy options for the uses of genomics in medical and non-medical settings

Grand Challenge III-2 Understand the relationships between genomics, race and ethnicity, and the consequences of uncovering these relationships

Grand Challenge III-3 Understand the consequences of uncovering the genomic contributions to human traits and behaviours

Grand Challenge III-4 Assess how to define the ethical boundaries for uses of genomics

②

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 diseases¹ (<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 disease²⁻⁴. In the human population most variant sites are rare, but the small number of common polymorphisms explain the bulk of heterozygosity⁵ (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^{6,11-14}. 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^{15,18,20}, 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 microsatellites²¹, 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 individuals²³. 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: PolyBayes²⁴ 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 (630,147 SNPs): 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 tags²⁴⁻²⁸ or by targeted resequencing efforts (see http://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 have current + removed by IRO
③
cannot be for group. Specific studies

**SINGLE BASE PAIR CHANGES (SNPs)
ARE FREQUENT IN GENOME**

TABLE 9.1 Five Classes of DNA Polymorphism

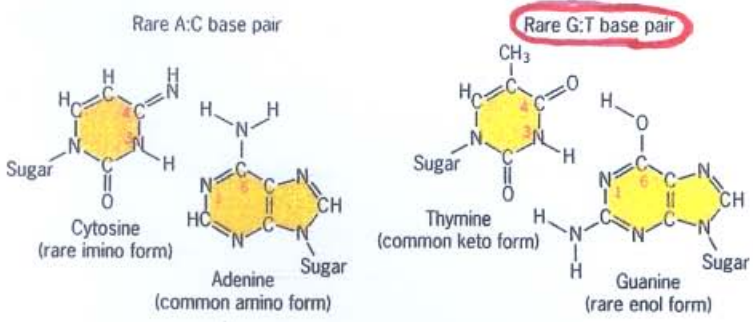
Class	Cause	Rate of Mutation per Locus per Gamete	Frequency in Genome	Number per Human Genome (on average)
<u>Single base</u>	Mutagens or replication errors	10^{-8} - 10^{-9}	1/700 bp	3 million
Microsatellite	Slippage during replication	10^{-3}	1/30,000 bp	100,000
Minisatellite	Unequal crossovers	10^{-3}	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

$\sim 6 \times 10^6$
in 2005!

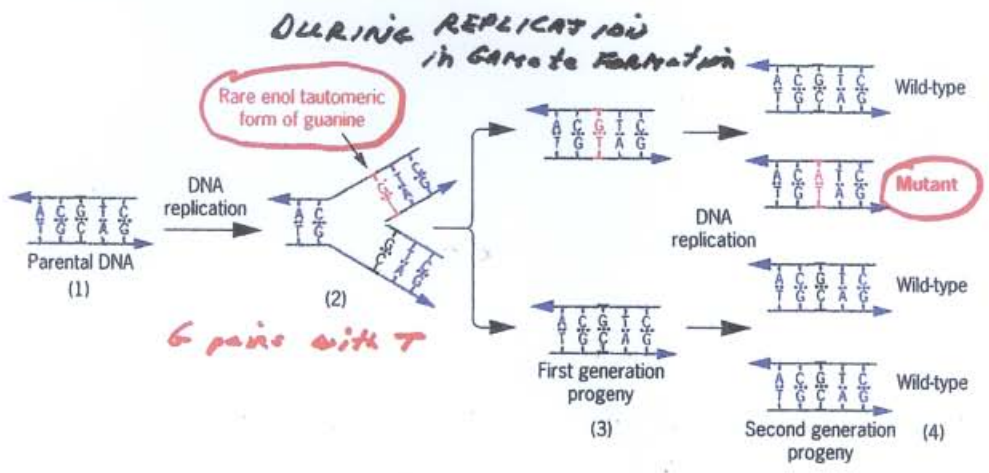
DETECTED USING RESTRICTION
ENZYMES (THE OLD FASHIONED WAY)
OR BY DIRECT DNA SEQUENCING OF
2 INDIVIDUALS' GENES/GENOMES

MANY CHANGES in DNA Sequence ALSO OCCUR

Random Base Pair Change during Replication



(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.

42% of live births affected by these mutations

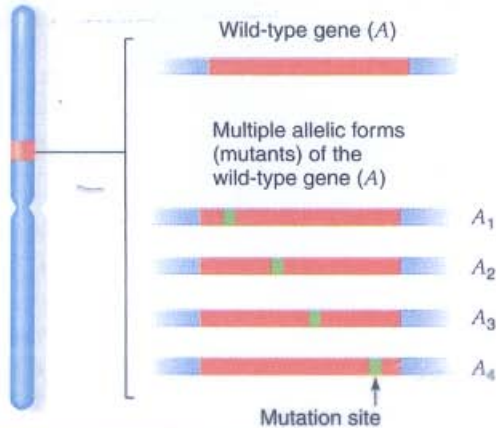
MOST CHANGES IN GENOME in NON-CODING DNA & DO NOT LEAD TO MUTATIONS

RECALL - ONLY 1% of GENOME = EXONS!

SNPs Generate Multiple Alleles in a Population

Figure 4.1

Allelic forms of a gene.



Recall - 2 alleles/individual
Many in POPULATIONS!

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

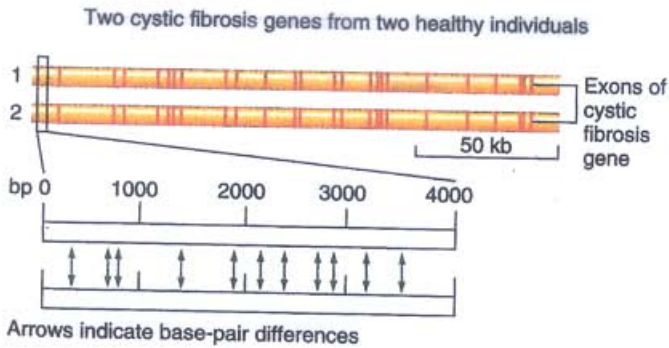


Figure 9.2 Base-pair differences between DNA cloned from the cystic fibrosis locus of two healthy individuals. These base-pair differences have no phenotypic effect; apparently they neither encode nor regulate expressed regions of the gene.

NOTE Differences
Between Two
Healthy
Cystic Fibrosis
Genes!

$\sim 6 \times 10^6$ SNPs (2005) !!

$$3.3 \times 10^9 / 6 \times 10^6 = 500-600 \text{ bp/SNP on average}$$

EACH OF US DIFFERS BY 6×10^6 bp, $\sim 0.2\%$ of genome!!

MOST in NON-CODING/INTRON & Intergenic Regions & Useful As Markers For Forensics, Disease Genes, Populations

SNPs or Single Nucleotide Polymorphisms occur randomly in the Human Genome

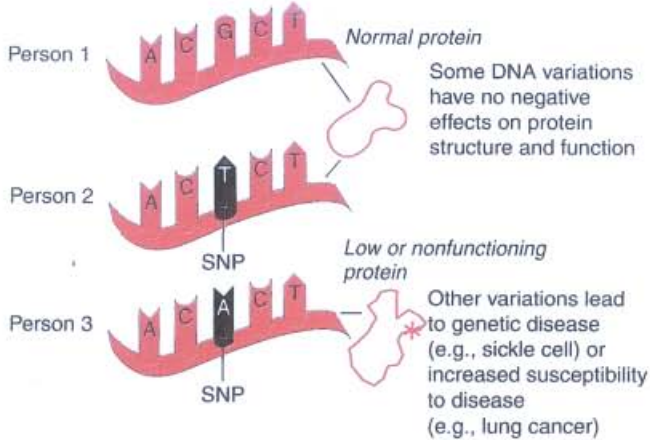


Figure 1.10 Single Nucleotide Polymorphisms A small piece of a gene sequence for three different individuals is represented. For simplicity, only one strand of a DNA molecule is shown. Notice how person 2 has a SNP in this gene which has no effect on protein structure and function. Person 3 however, has a different SNP in the same gene. This subtle genetic change may affect how this person responds to a medical drug or influence the likelihood that person 3 will develop a genetic disease.

Detected directly by sequencing

Used to individualize genomes
*
associate with disease genes (i.e., disease gene markers)

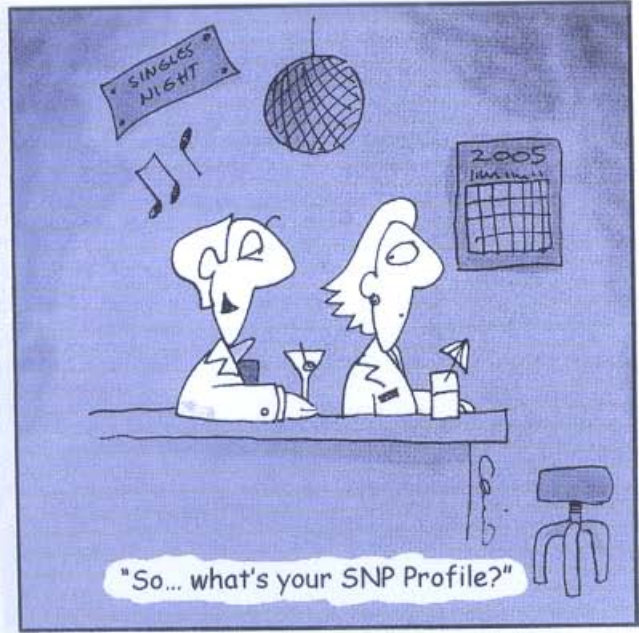


Figure 1.11 Secrets of the Human Genome In the future, we will have unprecedented knowledge of our genetic make-up including SNPs and other markers of genetic diseases. Can you think of possible ethical, legal, and social implications of such information?

LOCATION of SNPs Relative to Genes

TABLE 18-11 dbSNP Statistics (NCBI Genome Build 30, November 2002)

See http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi. SNP count is the number of distinct RefSNPs having the noted functional relationship to at least one mRNA in the current assembly. Gene count is the number of distinct locus.id(s) having at least one variation of the noted functional class. (Genes with multiple variations may be counted in multiple classes.)

Functional Classification	SNP Count	Gene Count
Locus region	291,459	26,482
Allele synonymous to contig nucleotide	12,322	7,147
Allele nonsynonymous to contig nucleotide	16,251	8,496
Untranslated region	131,987	13,208
Intron ←	904,573	22,113
Splice site	277	268
Allele is same as contig nucleotide	28,491	11,621
Coding: synonymy unknown	13,501	3,584

TABLE 18-10 SNP Resources

Resource	Comment	URL
dbSNP	At NCBI	▶ www.ncbi.nlm.nih.gov/SNP
Human SNP database	At the Whitehead Institute	▶ http://www-genome.wi.mit.edu/snp/human/
The SNP Consortium (TSC)	A collaboration of industrial and academic laboratories	▶ http://snp.cshl.org

USES OF SNPs

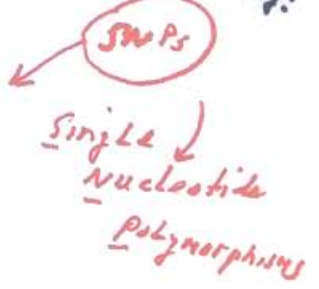
- ① Gene Identity / Allele Marker
- ② Disease Gene Identity / Pedigrees / Testing
- ③ Group Identity / Population History / Human origins
- ④ Individual Identity
 - ↳ Pharmacogenomics
 - ↳ Disease Prevention / Preventative Medicine
 - ↳ DNA Fingerprinting
- ⑤ Group Susceptibility to Disease / Drugs
 - ↳ Identify Genes Involved
 - ↳ "Resistance" Genes

SNPs Are Markers

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

2% cause protein change - why?



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

Using SNPs to Personalize Medicine Pharmacogenomics

Individuals respond differently to the anti-leukemia drug 6-mercaptopurine.

Most people metabolize the drug quickly. Doses need to be high enough to treat leukemia and prevent relapses.



Others metabolize the drug slowly and need lower doses to avoid toxic side effects of the drug.



A small portion of people metabolize the drug so poorly that its effects can be fatal.



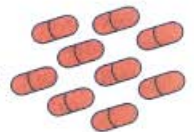
The diversity in responses is due to variations (mutations, ■ or ✱) in the gene for an enzyme called TPMT, or thiopurine methyltransferase.



After a simple blood test, individuals can be given doses of medication that are tailored to their genetic profile.



Normal dose

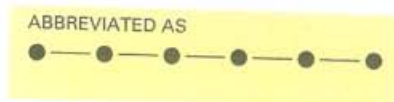
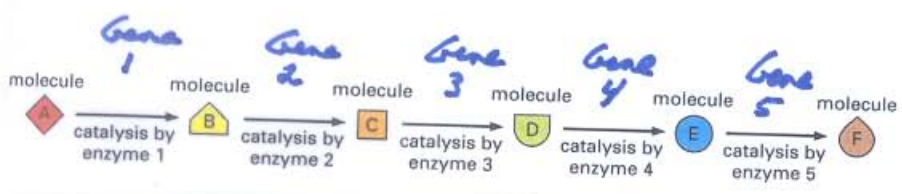


Dose for an extra slow metabolizer (TPMT deficient)



Figure 11.7 Pharmacogenomics Different individuals with the same disease often respond differently to a drug treatment because of subtle differences in gene expression. The dose that works for one person may be toxic for another—this is a basic problem of conventional medicine. Pharmacogenomics holds the promise of customizing medical treatment by determining the appropriate dosage for each individual based on the genes that person expresses.

Using SNPs in Pharmacogenetics



Different alleles can give rise to enzymes that differ slightly in activity.

Figure 2-34 How a set of enzyme-catalyzed reactions generates a metabolic pathway. Each enzyme catalyzes a particular chemical reaction, leaving the enzyme unchanged. In this example, a set of enzymes acting in series converts molecule A to molecule F, forming a metabolic pathway.

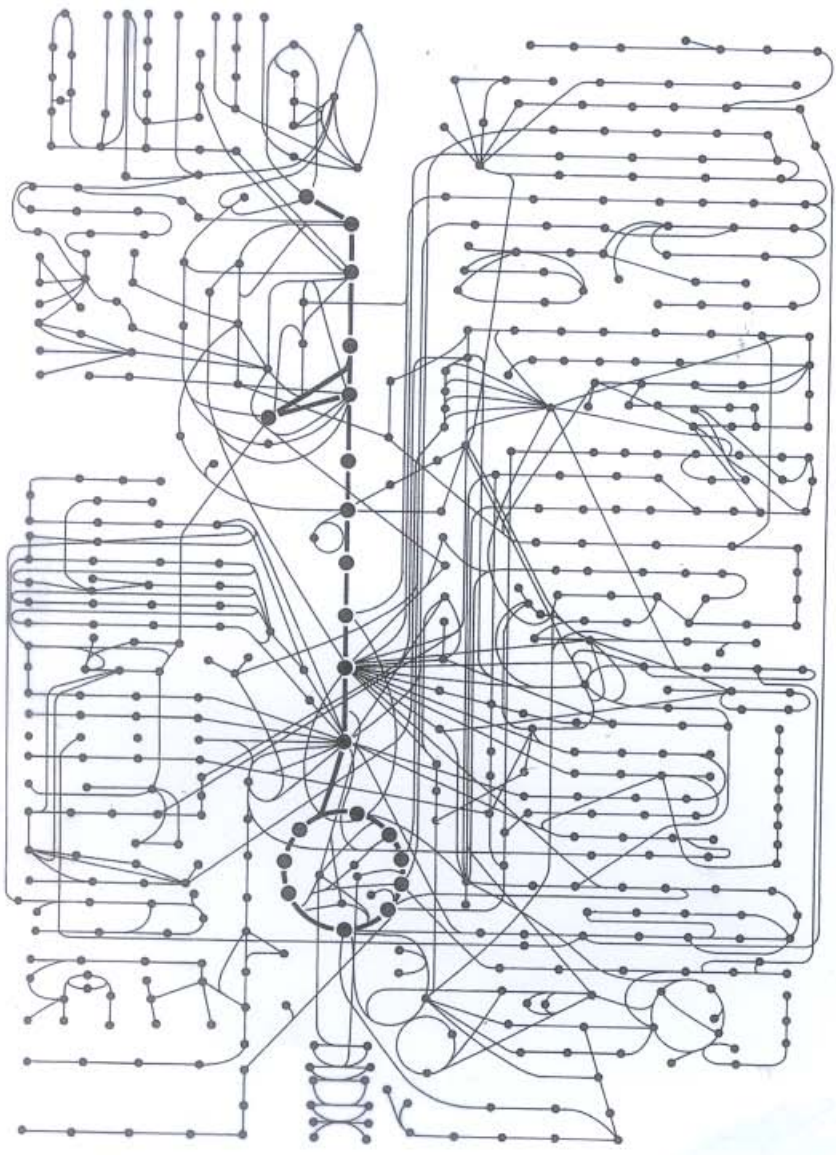


Figure 2-35 Some of the metabolic pathways and their interconnections in a typical cell. About 500 common metabolic reactions are shown diagrammatically, with each molecule in a metabolic pathway represented by a filled circle, as in the yellow box in Figure 2-34.

Goal: to find genes that encode enzymes that efficiently metabolize drugs
∴ no side effects

A Variety of Markers Not Just SNPs

Box 11.1

The development of human genetic markers

Type of marker	No. of loci	Features
Blood groups 1910-1960	~20	May need fresh blood, rare antisera Genotype cannot always be inferred from phenotype because of dominance No easy physical localization
Electrophoretic mobility variants of serum proteins 1960-1975	~30	May need fresh serum, specialized assays No easy physical localization Often limited polymorphism
HLA tissue types 1970-	1 (haplotype)	One linked set Highly informative Can only test for linkage to 6p21.3
DNA RFLPs 1975-	$>10^5$ (potentially)	Two allele markers, maximum heterozygosity 0.5 Initially required Southern blotting, now PCR Easy physical localization
DNA VNTRs (minisatellites) 1985-	$>10^4$ (potentially)	<i>Many alleles, highly informative</i> Type by Southern blotting Easy physical localization Tend to cluster near ends of chromosomes
DNA VNTRs (microsatellites) (di-, tri- and tetranucleotide repeats) 1989-	$>10^5$ (potentially)	Many alleles, highly informative Can type by automated multiplex PCR Easy physical localization Distributed throughout genome
DNA SNPs (single nucleotide polymorphisms) 1998-	$>10^6$ (potentially)	Less informative than microsatellites Can be typed on a very large scale by automated equipment without gel electrophoresis

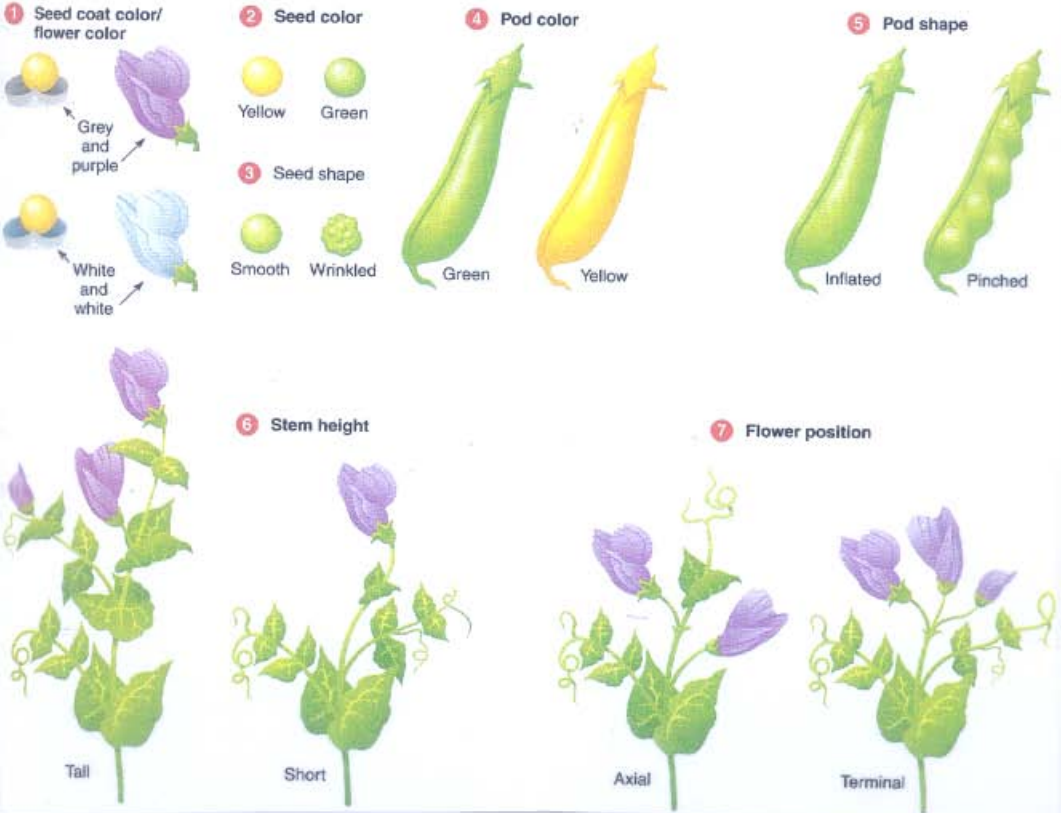
VNTR, variable number of tandem repeats

SNPs the Most Powerful
↳ Genome-wide

Morphological Markers

Figure 2.4

Seven character pairs in the garden pea that Mendel studied in his breeding experiments.

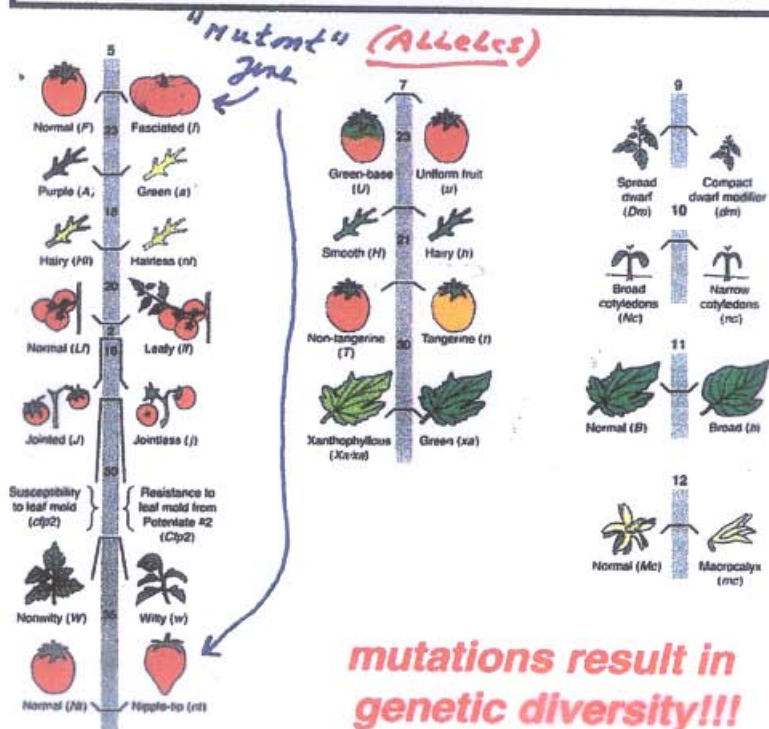


Represent only a very minor fraction of genes!

What is used for group classification?

VARIABILITY due to Morphological Markers

Alternative Forms of the Same Gene Lead to Genetic Diversity



HUMAN BLOOD GROUP MARKERS

Table 4.1 ABO Blood Groups in Humans, Determined by the Alleles I^A , I^B , and I

Phenotype (Blood Group)	Genotype
O	i/i
A	I^A/I^A or I^A/i
B	I^B/I^B or I^B/i
AB	I^A/I^B

Serum from blood type	Antibodies present in serum	Cells from blood type			
		O	A	B	AB
O	Anti-A Anti-B				
A	Anti-B				
B	Anti-A				
AB	—				

Minor Fraction of Genes in Genome

Blood Group Antigens are Surface Carbohydrates

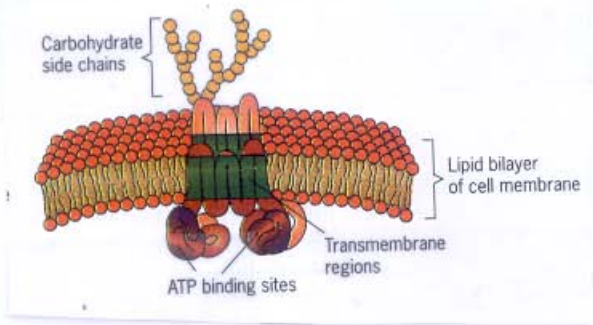
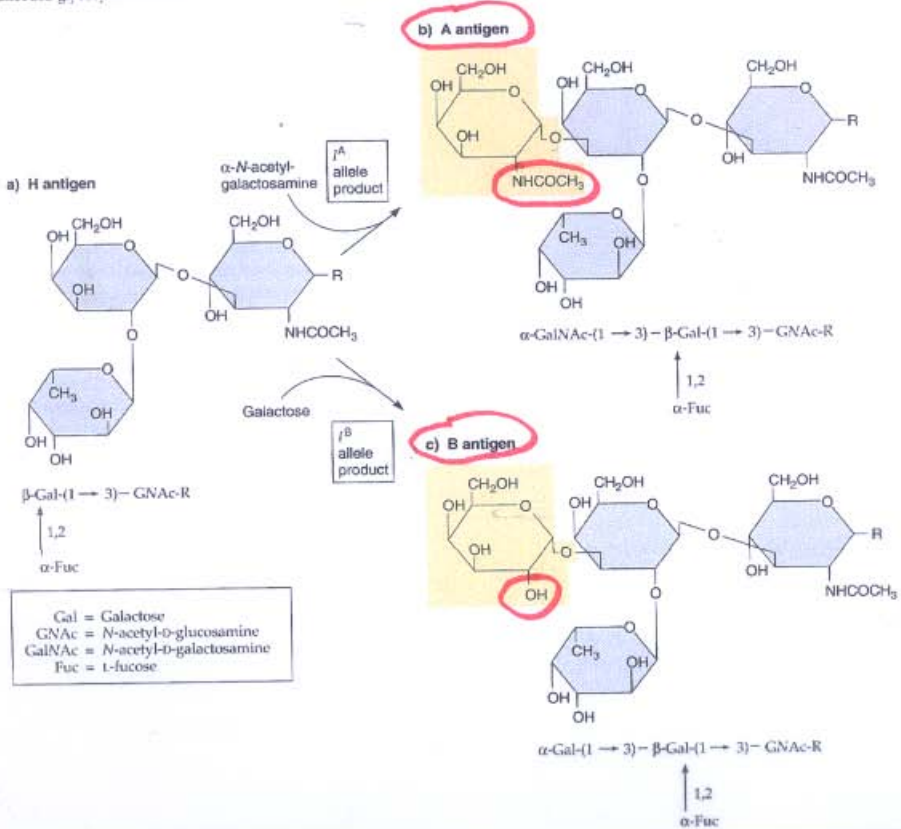


Figure 10.9
Human ABO blood type antigens. Conversion of (a) the H antigen to (b) the A antigen by the I^A allele-encoded glycosyltransferase and to (c) the B antigen by the I^B allele-encoded glycosyltransferase.



ALTERNATIVE FORMS OF SAME ANTIGEN!
 ENCODED BY SAME GENE!

Markers at the Biochemical level

Figure 10.5

Colorized scanning electron micrographs of (left) normal and (right) sickled red blood cells.

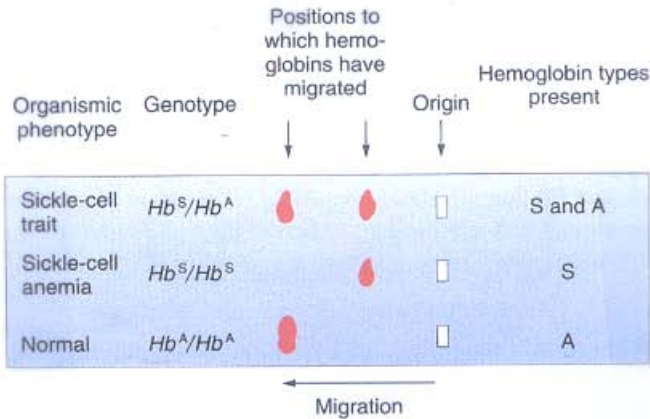
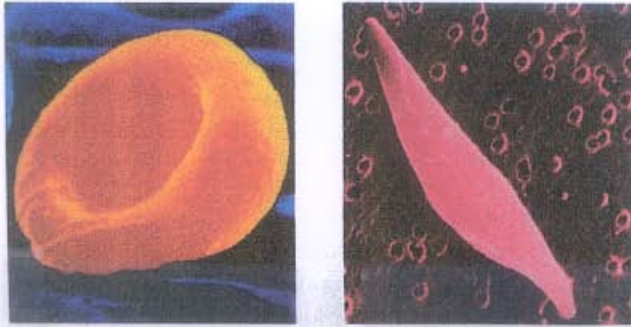


Figure 4-6 Electrophoresis of hemoglobin from a person with sickle-cell anemia, a heterozygote (called sickle-cell trait), and a normal person. The smudges show the positions to which the hemoglobins migrate on the starch gel.

Protein Forms

Two Alleles of SAME Gene!

MOLECULAR SNP DETECTION

- ① RFLPs → BLOTS
→ PCR
- ② ASOs → PCR + Hybridization / "Blot"
- ③ Comparative Sequencing / Genes / Genome
- ④ Chips / Whole Genome / Individuals
+ Groups

MOLECULAR DETECTION OF MARKERS

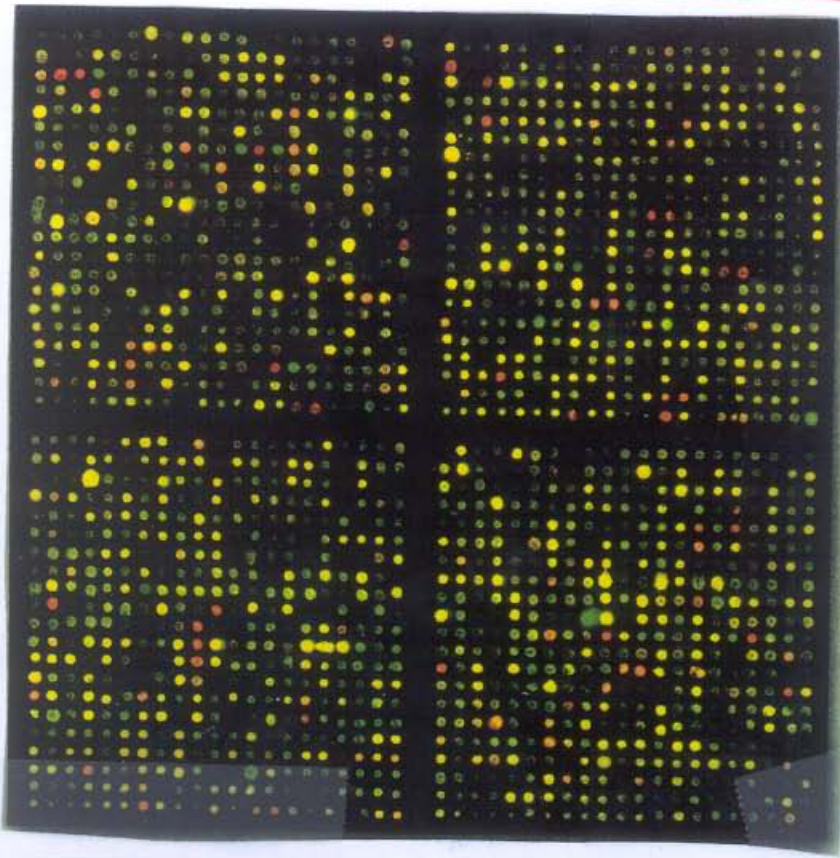
TABLE 9.2 Protocols for Detecting Single-Base Variations

Protocol	Advantages	Disadvantages	Main Uses
1 RFLP analysis By Southern blots and labeled probes	No requirement for sequence information	Requires a large amount of DNA	Genotyping in the <u>absence of sequence information</u>
By restriction enzyme digestion and electrophoresis of PCR products	Rapid; no blotting and hybridization	Substitution must affect restriction site; requires sequence information from locus	Rapid analysis of restriction site polymorphisms among individuals within a population
2 PCR analysis using ASOs	No need to run a gel; can detect all single-base variations	Must know sequence of alternative alleles	Standard genotyping

4 Chips

3 Direct Sequencing - but not practical for routine work!

GENOME CHIPS WILL BE USED FOR Genotyping



They work like ASOs but use MANY Probes at a time

Genotyping of The Future

USE OF RFLPs For
Detecting SNPs & Genes

"The Old Fashioned Way"

SNP that Generates An RFLP

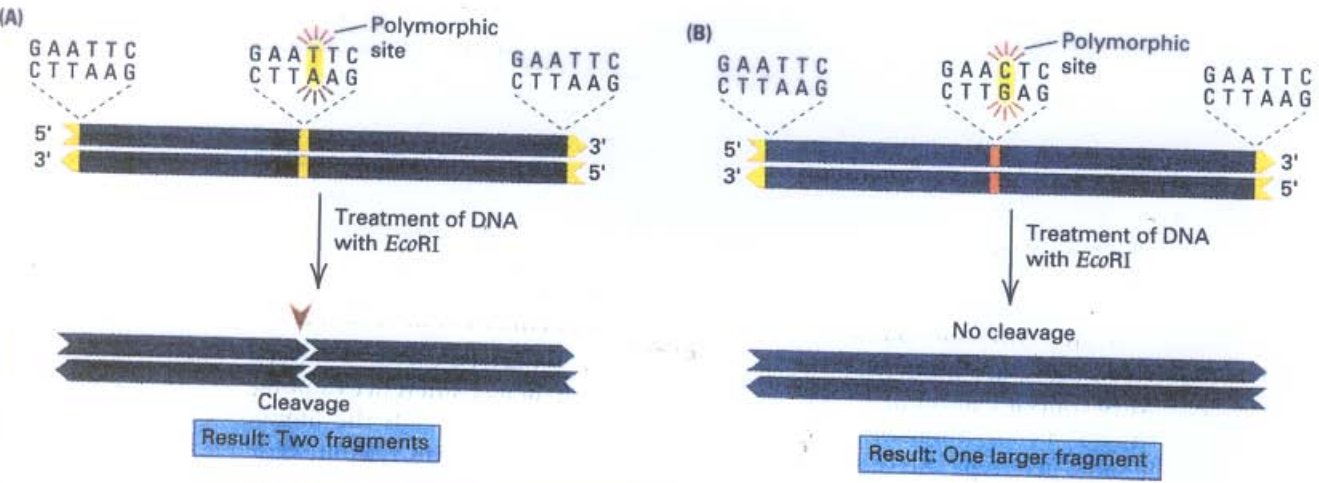
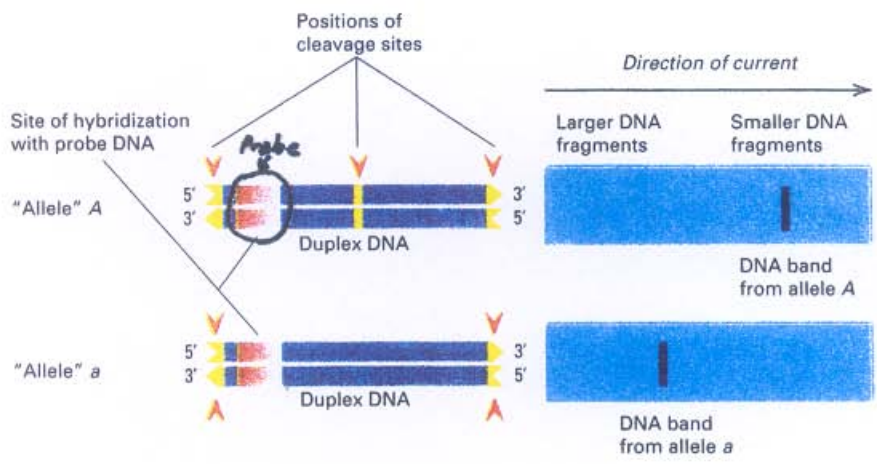


Figure 2.23 A minor difference in the DNA sequence of two molecules can be detected if the difference eliminates a restriction site. (A) This molecule contains three restriction sites for *Eco*RI, including one at each end. It is cleaved into two fragments by the enzyme. (B) This molecule has an altered *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.

A SNP can generate an RFLP!
 But too labor-intensive to
 detect on a whole genome level!

need blots or PCR!

RFLPs/SNPs CAN BE USE TO detect ALLELIC VARIABILITY at a LOcus



Probe Specific for left half of DNA only
∴ 1 band !!!

Pattern After Blotting

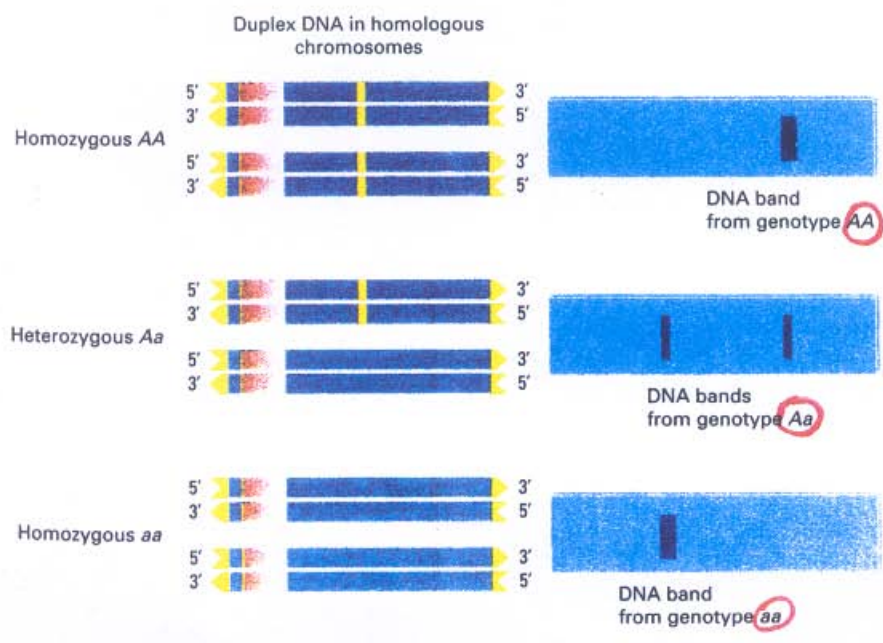
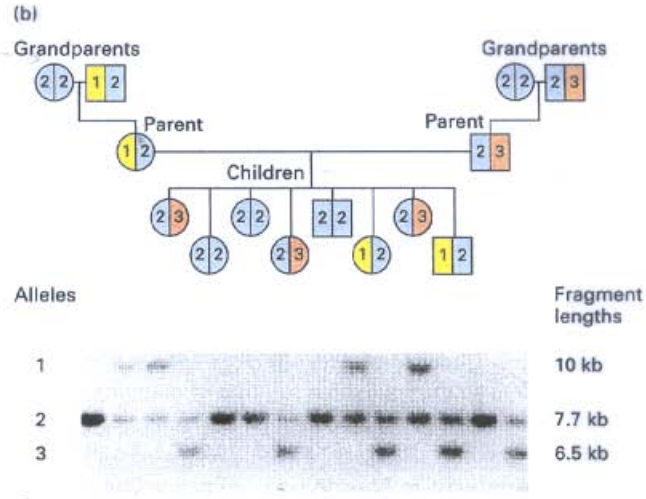
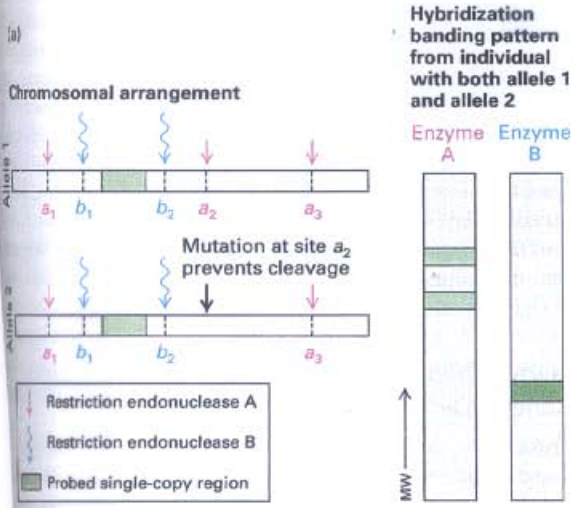


Figure 2.24 In a restriction fragment length polymorphism (RFLP), alleles may differ in the presence or absence of a cleavage site in the DNA. In this example, the a allele lacks a restriction site that is present in the DNA of the A allele. The difference in fragment length can be detected by Southern blotting. RFLP alleles are codominant, which means (as shown at the bottom) that DNA from the heterozygous Aa genotype yields each of the single bands observed in DNA from homozygous AA and aa genotypes.

As well as Genotypes

RFLPs CAN BE USED TO IDENTIFY INDIVIDUALS



EXPERIMENTAL FIGURE 9-46 Restriction fragment length polymorphisms (RFLPs) can be followed like genetic markers. (a) In the example shown, DNA from an individual is treated with two different restriction enzymes (A and B), which cut DNA at different sequences (*a* and *b*). The resulting fragments are subjected to Southern blot analysis (see Figure 9-38) with a radioactive probe that binds to the indicated DNA region (green) to detect the fragments. Since no differences between the two homologous chromosomes occur in the sequences recognized by the B enzyme, only one fragment is recognized by the probe, as indicated by a single hybridization band. However, treatment with enzyme A produces fragments of

two different lengths (two bands are seen), indicating that a mutation has caused the loss of one of the *a* sites in one of the two chromosomes. (b) Pedigree based on RFLP analysis of the DNA from a region known to be present on chromosome 5. The DNA samples were cut with the restriction enzyme *TaqI* and analyzed by Southern blotting. In this family, this region of the genome exists in three allelic forms characterized by *TaqI* sites spaced 10, 7.7, or 6.5 kb apart. Each individual has two alleles; some contain allele 2 (7.7 kb) on both chromosomes, and others are heterozygous at this site. Circles indicate females; squares indicate males. The gel lanes are aligned below the corresponding subjects. [After H. Donis-Keller et al., 1987, *Cell* 51:319.]

RFLPs CAN BE USED AS MARKERS FOR disease genes if linked to the mutant allele

RFLPs are caused by SNPs!

Molecular Diagnosis of Human Diseases

Sickle Cell Anemia

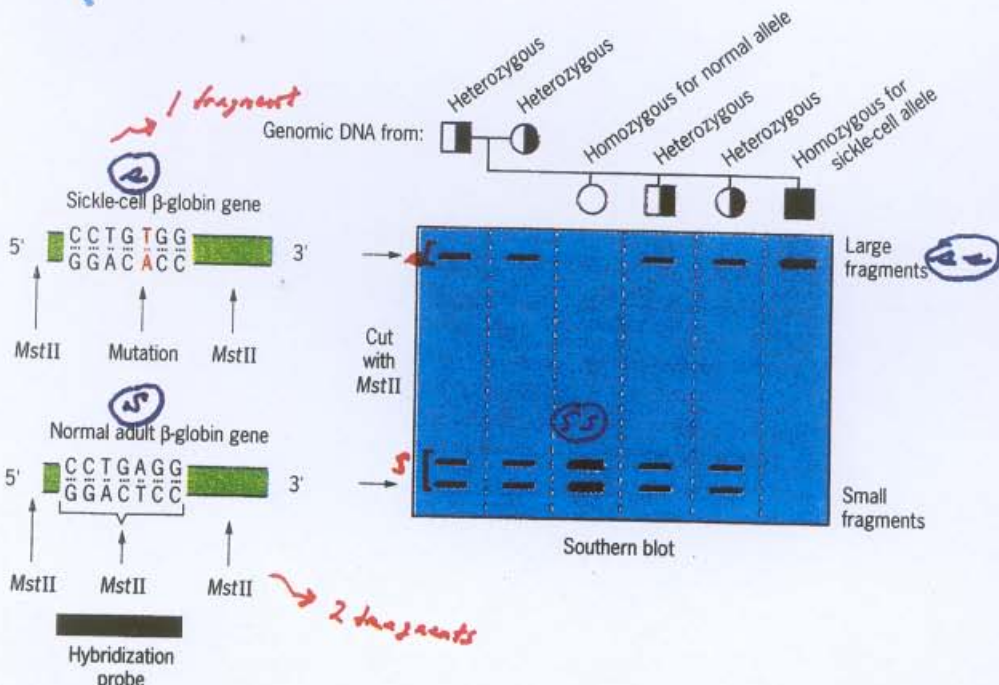


Figure 22.7 Detection of the sickle-cell hemoglobin mutation by Southern blot analysis of genomic DNAs cut with restriction enzyme MstII.

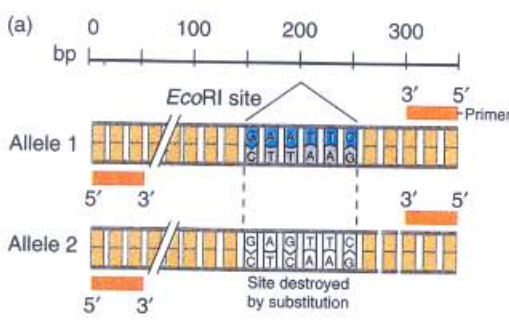
Marker is (N) Gene! \therefore always linked with phenotype!
 \therefore 1 fragment at this locus is a marker for the normal S allele.

RFLPs CAN BE DETECTED USING PCR

SNPs cause RFLPs!

NO NEED FOR BLOT!
ONLY NEED TINY AMOUNT OF DNA

- ① AMPLIFY Gene Region containing the Polymorphism
- ② Digest with Relevant Restriction Enzyme
- ③ Visualize fragments directly on Gel!



why different polarities? for primers?

→ 2 fragments (A1)

→ 1 fragment (A2)

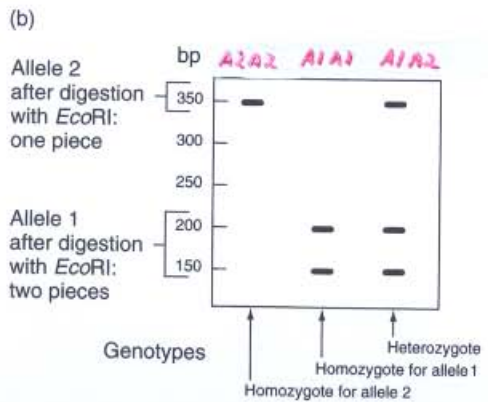
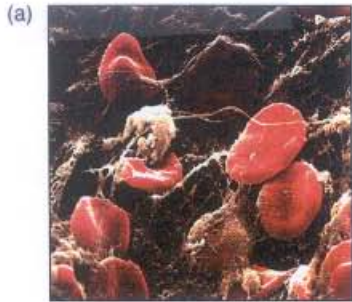


Figure 9.7 Restriction site polymorphisms can be detected most efficiently with PCR-based protocols. (a) PCR amplification of two alleles of a DNA locus with a restriction site polymorphism. Allele 1 has an EcoRI site that is eliminated in allele 2. The PCR products amplified from both alleles are identical in size. (b) Exposure of these PCR products to EcoRI causes cleavage of the allele 1 product but not the allele 2 product. Gel electrophoresis and ethidium bromide staining distinguish the three genotypes possible with the two alleles at this locus.

USING RFLPs TO Detect a Mutant Factor VIII Gene

A PCR Approach



(b) Factor VIII gene

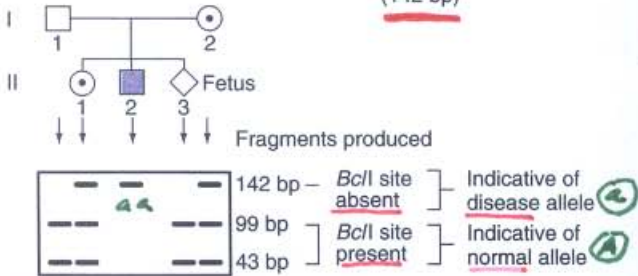
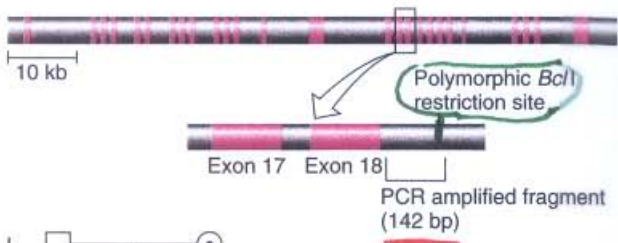


Figure 9.18 Diagnosis of hemophilia through the indirect detection of genotype at the factor VIII locus. The factor VIII protein participates in a cascade of reactions that result in formation of a blood clot. (a) A polymorphic *BclI* restriction site within intron 18 of the factor VIII gene has no effect on gene function but can provide a marker to follow the segregation of the gene from parents to children. (b) The family described by the pedigree has two healthy parents, but the mother is an obligate carrier of the disease mutation because she has passed this X-linked disease on to her son; her carrier status is signified by a circle with a dot in the middle. By comparing the RFLP pattern obtained from the mother's DNA with the pattern from her son's DNA, you can see that the disease allele is associated with the 142-bp *BclI* restriction fragment, and the wild-type allele in the mother's genome contains a *BclI* restriction site that causes this fragment to be cut into two pieces, one 43 bp and the other 99 bp in length. Using this information, you can determine that the firstborn sister is a carrier like her mother, while the male fetus will be disease free.

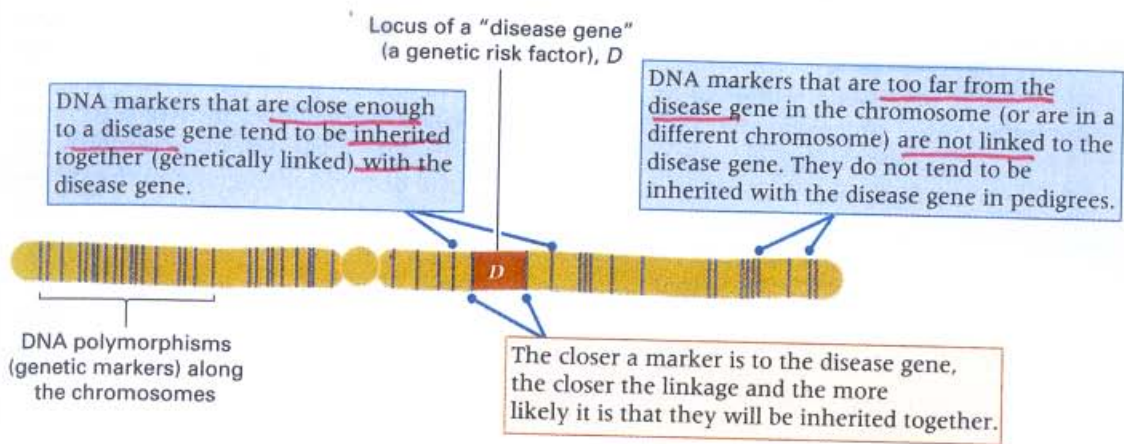
AAa Aa

Disease Gene - *BclI* site Absent (A)

NORMAL Gene - *BclI* site Present (a)

Markers tightly linked to disease gene can be used to identify what the disease gene is

Positional cloning in absence of known protein/gene product



How find Gene?

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.

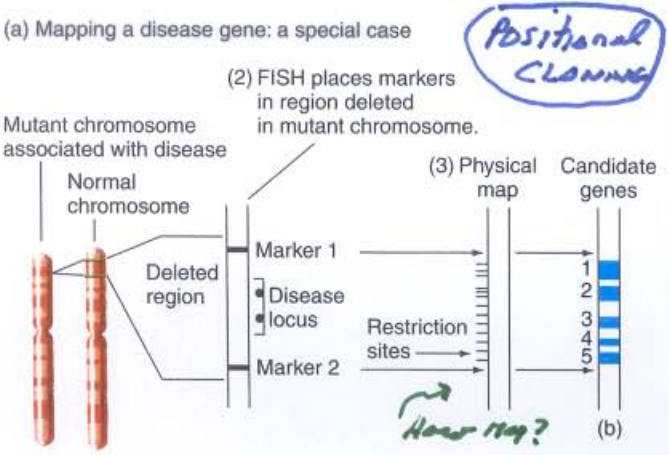


Figure 10.11 Positional cloning: From phenotype to gene.

(a) Correlating the expression of a phenotype with one small segment of the genome. (1) Some diseases, such as Duchenne muscular dystrophy, are caused by a deletion. It is sometimes possible to observe directly the absence or shortening of a band in a chromosome from an affected individual as compared to the same chromosome from a healthy individual. Even when it is not possible to observe the deletion directly, the FISH protocol can detect it. (2) A marker in the deleted region will hybridize to the chromosome from the healthy individual, but not to the same chromosome from the diseased individual. Markers associated with the disease can be used in linkage analyses of families carrying mutant disease alleles that are not deletions. (3) When linkage analysis shows that a specific chromosomal region contains the disease locus, researchers can subject the marked region to physical analysis. (b) Investigators next analyze the region between recombination sites that define the smallest area within which the disease locus can lie for the presence of candidate genes (as described later in this chapter). (c) They then compare the structure and expression of each candidate gene in many diseased and nondiseased individuals. A correlation between a mutant structure or expression for a particular candidate gene and the disease phenotype can provide evidence that a particular gene is responsible for the disease phenotype. Proof of the association, however, requires further functional studies, which we describe later in this chapter.

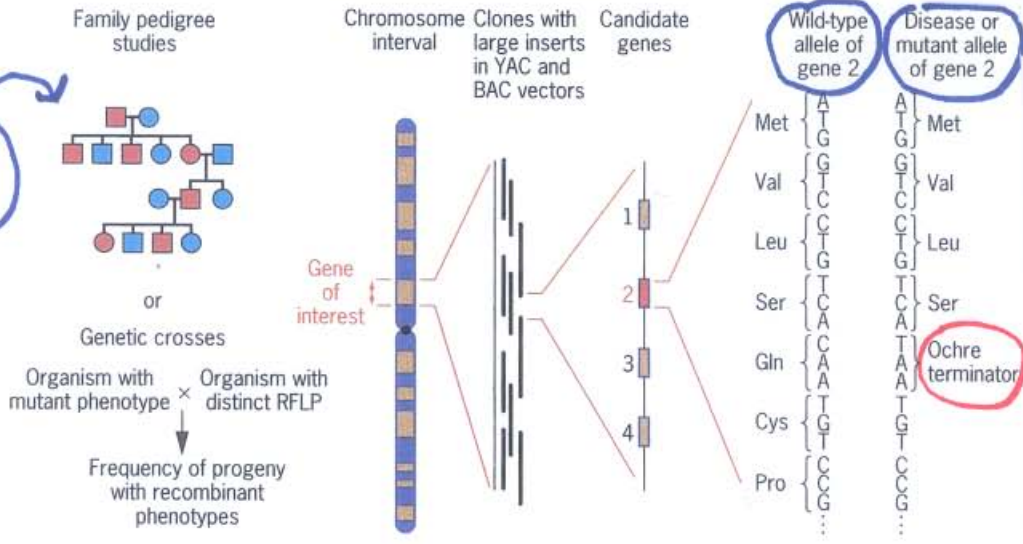
(c) Examine expression of genes 1-5 in normal and diseased individuals.

Gene	Expression	
	Normal	Diseased
1	+	+
2	+	+
3	+	-
4	+	+
5	+	+

Conclusion: The tested diseased individual shows a correlation between the disease phenotype and the absence of expression from the number 3 gene.

USING LINKED MARKERS TO IDENTIFY A DISEASE GENE

Dominant or Recessive?



- 1 Genetic mapping
- 2 Physical mapping
- 3 Transcript mapping
- 4a Gene sequencing

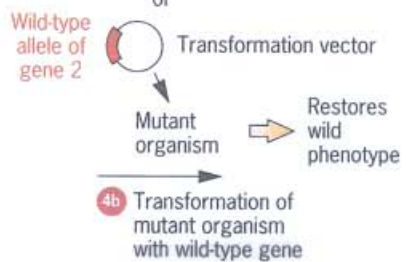
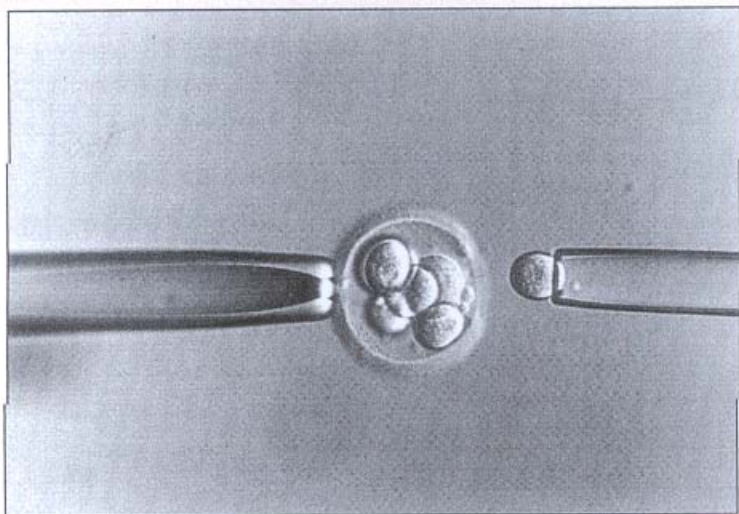


Figure 21.8 Steps involved in the positional cloning of genes. In humans, genetic mapping must be done by pedigree analysis, and candidate genes must be screened by sequencing mutant and wild-type alleles (step 4a). In other species, the gene of interest is mapped by appropriate genetic crosses, and the candidate genes are screened by transforming the wild-type alleles into mutant organisms and determining whether or not they restore the wild-type phenotype (step 4b).

How Find the Gene Responsible for the Disease?

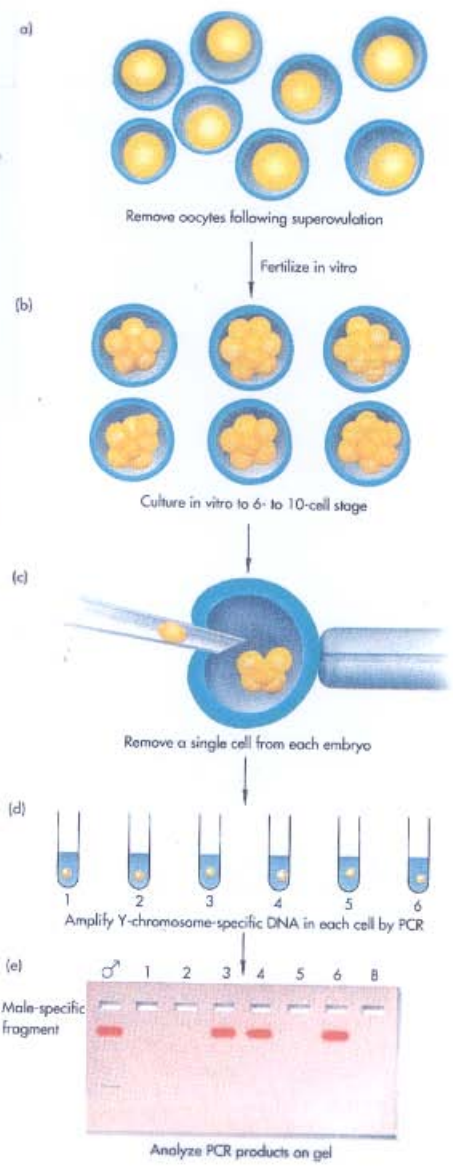
THE DIRECT DETECTION OF GENOTYPE



Plucking one cell from an 8-cell stage human embryo for direct determination of genotype.

PCR CAN BE USED TO DETECT DISEASE GENES PRIOR TO IMPLANTATION

PRE-IMPLANTATION
Genetic
DIAGNOSIS
PGD



..... IN ADDITION TO AMNIOCENTESIS
 * Chorionic Villi Sampling

Figure 10.12

Chorionic villus sampling, a procedure used for early prenatal diagnosis of genetic defects.

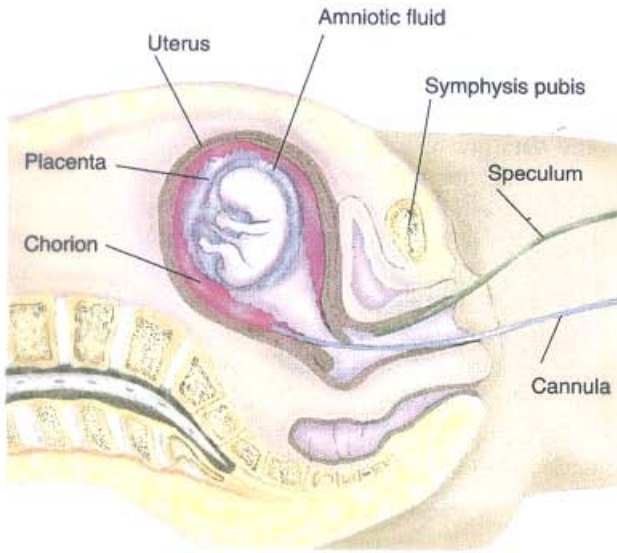
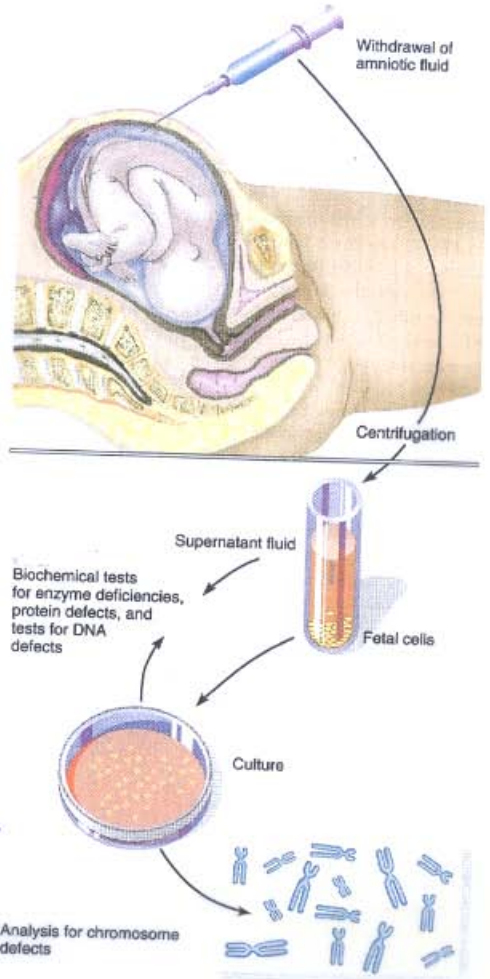


Figure 10.11

Amniocentesis, a procedure used for prenatal diagnosis of genetic defects.



PRE-IMPLANTATION GENETIC TESTING

FIGURE 59.19

The meiotic events of oogenesis in humans. A primary oocyte is diploid. At the completion of the first meiotic division, one division product is eliminated as a polar body, while the other, the secondary oocyte, is released during ovulation. The secondary oocyte does not complete the second meiotic division until after fertilization; that division yields a second polar body and a single haploid egg, or ovum. Fusion of the haploid egg with a haploid sperm during fertilization produces a diploid zygote.

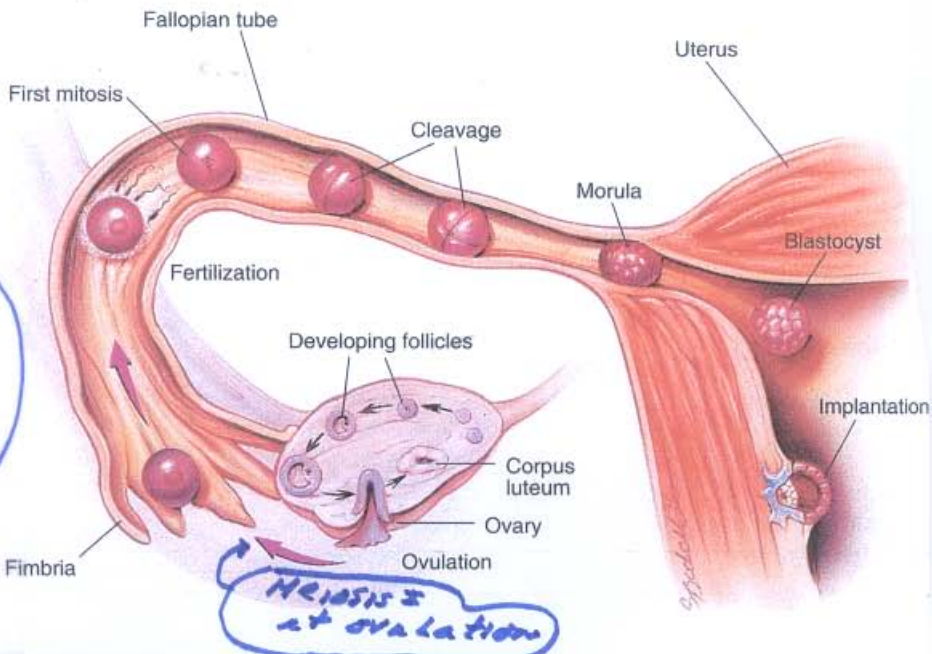
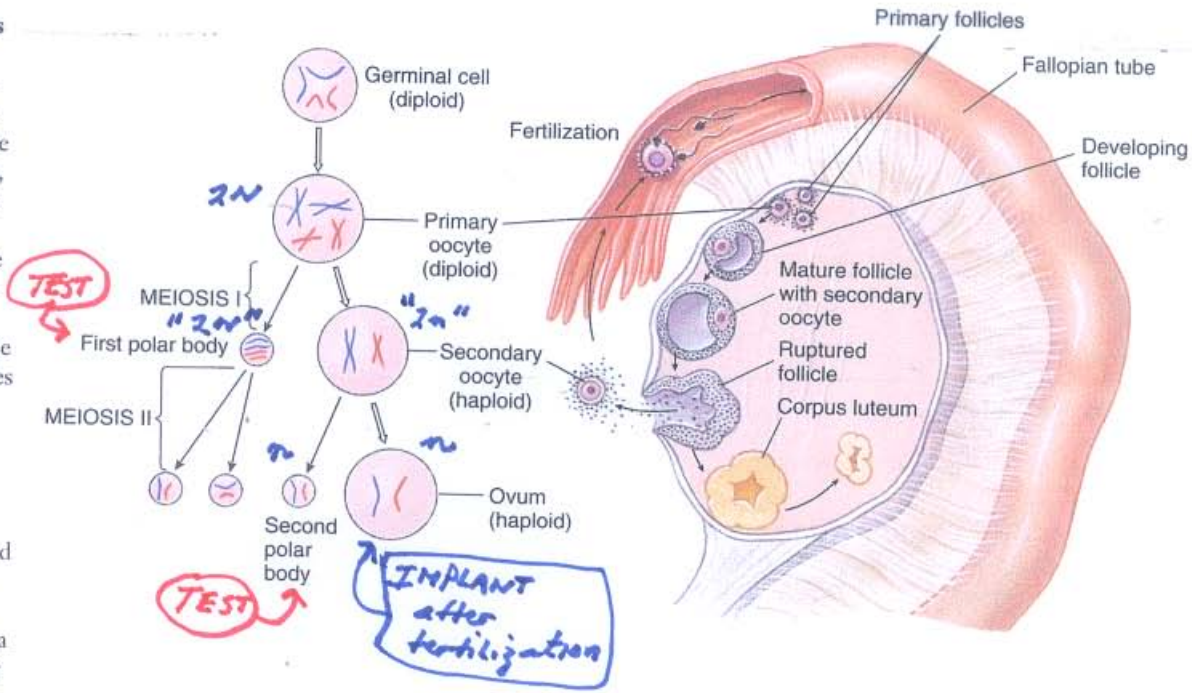


FIGURE 59.20

The journey of an egg. Produced within a follicle and released at ovulation, an egg is swept into a fallopian tube and carried along by waves of ciliary motion in the tube walls. Sperm journeying upward from the vagina fertilize the egg within the fallopian tube. The resulting zygote undergoes several mitotic divisions while still in the tube, so that by the time it enters the uterus, it is a hollow sphere of cells called a blastocyst. The blastocyst implants within the wall of the uterus, where it continues its development. (The egg and its subsequent stages have been enlarged for clarification.)

RFLPs can also be used to detect disease genes in embryo cells counted with PCR!

The New York Times ON THE WEB

February 27, 2002

Baby Spared Mother's Fate by Genetic Tests as Embryo

By DENISE GRADY

A 30-year-old woman who is very likely to develop a rare form of Alzheimer's disease before she turns 40 has had a baby girl who will be spared that fate because she was genetically screened as an embryo before being implanted in her mother's womb, doctors are reporting.

The case is a medical milestone, the first use of genetic testing to prevent an early onset form of Alzheimer's disease. But some find it ethically disturbing because within a few years the mother will probably become unable to take care of her daughter, who will witness her deterioration and death.

The 30-year-old woman, who carried a rare gene making it almost certain she will develop Alzheimer's, wanted to have a child but hoped to avoid passing on the bad gene. So she sought preimplantation diagnosis. In that procedure, embryos are created in the laboratory from the mother's eggs and the father's sperm and are tested genetically. Only healthy embryos are implanted in the mother's uterus.

The case is being described today in The Journal of the American Medical Association by Dr. Yury Verlinsky and his associates from two private clinics in Chicago.

In a commentary article accompanying Dr. Verlinsky's report, Dr. Roberta Springer Loewy and Dr. Dena Towner wrote that the mother "most likely will not be able to care for or even recognize her child in a few years." Dr. Towner is director of the prenatal diagnosis center at the University of California at Davis, and Dr. Loewy is a bioethicist there.

In an interview, Dr. Loewy said that it was laudable for the mother to try to protect her child from illness but that the child's physical health was not her only responsibility.

"I'm not trying to pounce on this poor woman," Dr. Loewy said, but she added that she thought it would be traumatic for the child to watch the mother's slow decline and that if this same woman "wanted to adopt, our society essentially holds the position that, gee, no, we wouldn't let her do this, wouldn't let her subject a child to this. But because it's coming from her own loins, this is something we shouldn't have a say about?"

William Thies, vice president of medical and scientific affairs for the Alzheimer's Association, said: "It's a discussion being held in absence of any data. Does this end up being harmful to the child? Nobody knows."

Dr. Loewy also said that society as a whole should consider whether this kind of treatment was the best use of limited medical resources.

In a case like that of Dr. Verlinsky's patient, each effort to begin a pregnancy costs more than \$12,000. Insurance coverage varies.

Dr. Verlinsky said he had no qualms about helping the woman and her husband. "It's totally up to the patient," he said.

As for the child's losing her mother, he said that many children were brought up by single parents, and that this family would be no different. Dr. Verlinsky said, too, that the couple had not made their decision in "a moment of emotion," but had had weeks and months to mull it over, because the medical procedures they went through in order to conceive take time.

Asked whether the couple would have gone ahead and had children anyway if preimplantation diagnosis was not available, Dr. Verlinsky said he did not know.

The couple were not available for interviews, he said.

The new report does not apply to most families with Alzheimer's disease, because the form the woman has accounts for less than 1 percent of all cases. The type of testing the woman had is not done for the more common types of Alzheimer's disease, which develop much later in life, and in which the role of particular genes is far less clear.

Dr. Thies said, "It's important for people to recognize that this does not represent a strategy that will have much of a public health impact."

The woman described by Dr. Verlinsky and his colleagues carries a rare genetic mutation that makes it almost certain she will develop Alzheimer's disease in her 30's. Her sister, who also has the mutation, developed Alzheimer's symptoms at age 38. The gene is involved in the formation of the tangled protein deposits, known as amyloid plaques, found in the brain in people with Alzheimer's disease. People who carry certain mutations in the gene nearly always develop dementia.

The sister with Alzheimer's declined so much mentally that she had to be moved to an assisted-living center. A brother also had the mutation and began suffering memory problems at 35. Their father died at 42, with psychological and memory problems.

People who do not want the expense and trouble of preimplantation diagnosis can become pregnant naturally and then have the fetus tested via amniocentesis or another procedure, chorionic villus sampling. But if the fetus has a genetic disease, the parents must decide whether to end the pregnancy. Preimplantation testing is particularly appealing to people who want to avoid abortion.

The technique has been used to prevent many diseases, including hemophilia, sickle cell anemia, muscular dystrophy, Tay-Sachs disease, cystic fibrosis and Huntington's disease. It should not be used just to pick embryos of a particular sex, according to an opinion issued this month by a professional group, the American Society for Reproductive Medicine.

http://www.nytimes.com/2002/02/27/health/27BABY.html?todayshoelines=&pagewanted=print

Advertisement for TechnoScout with various tech-related headlines like 'If the world was flat, just about any antenna would do!' and 'Click for the complete Advertisement'.



Preimplantation Diagnosis for Early-Onset Alzheimer Disease Caused by V717L Mutation

Yury Verlinsky, PhD

Svetlana Rechitsky, PhD

Oleg Verlinsky, MS

Christina Masciangelo, MS

Kevin Lederer, MD

Anver Kuliev, MD, PhD

ACCORDING TO THE MOST RECENT review,¹ preimplantation genetic diagnosis (PGD) has been applied to at least 50 different genetic conditions in more than 3000 clinical cycles. In addition to traditional indications, similar to those in prenatal diagnosis, PGD was performed for an increasing number of new indications, such as late-onset disorders with genetic predisposition and HLA testing combined with PGD for pre-existing single-gene disorders.^{2,3} These conditions have never been an indication for prenatal diagnosis because of potential pregnancy termination, which is highly controversial if performed for genetic predisposition alone. With the introduction of PGD, it has become possible to avoid the transfer of the embryos carrying the genes that predispose a person to common disorders, thereby establishing only potentially healthy pregnancies and overcoming important ethical issues in connection with selective abortions.

To our knowledge, this article presents the first experience of PGD for

Context Indications for preimplantation genetic diagnosis (PGD) have recently been expanded to include disorders with genetic predisposition to allow only embryos free of predisposing genes to be preselected for transfer back to patients, with no potential for pregnancy termination.

Objective To perform PGD for early-onset Alzheimer disease (AD), determined by nearly completely penetrant autosomal dominant mutation in the amyloid precursor protein (*APP*) gene.

Design Analysis undertaken in 1999-2000 of DNA for the V717L mutation (valine to leucine substitution at codon 717) in the *APP* gene in the first and second polar bodies, obtained by sequential sampling of oocytes following in vitro fertilization, to preselect and transfer back to the patient only the embryos that resulted from mutation-free oocytes.

Setting An in vitro fertilization center in Chicago, Ill.

Patients A 30-year-old AD-asymptomatic woman with a V717L mutation that was identified by predictive testing of a family with a history of early-onset AD.

Main Outcome Measures Results of mutation analysis; pregnancy outcome.

Results Four of 15 embryos tested for maternal mutation in 2 PGD cycles, originating from V717L mutation-free oocytes, were preselected for embryo transfer, yielding a clinical pregnancy and birth of a healthy child free of predisposing gene mutation according to chorionic villus sampling and testing of the neonate's blood.

Conclusion This is the first known PGD procedure for inherited early-onset AD resulting in a clinical pregnancy and birth of a child free of inherited predisposition to early-onset AD.

JAMA. 2002;287:1018-1021

www.jama.com

early-onset Alzheimer disease (AD), representing a rare autosomal dominant familial predisposition to the presenile form of dementia. Three different genes have been found to be involved in this form of AD, including presenilin 1 located on chromosome 14,⁴ presenilin 2 on chromosome 1,⁵ and amyloid precursor protein (*APP*) on chromosome 21,⁶ which is well known for its role in the formation of amyloid deposits found in the characteristic plaques of patients with AD. The

early-onset dementias associated with *APP* mutations are nearly completely penetrant and, therefore, are potential candidates for not only predictive testing but also PGD. Of the 10 *APP* mutations currently described, mutations in exons 16 and 17 have been

Author Affiliations: Reproductive Genetics Institute (Drs Verlinsky, Rechitsky, and Kuliev, Mr Verlinsky, and Ms Masciangelo) and IVF Illinois (Dr Lederer), Chicago.

Corresponding Author and Reprints: Yury Verlinsky, PhD, 2825 N Halsted St, Chicago, IL 60657 (e-mail: rgi@flash.net).

See also p 1038.

reported in the familial cases with the earliest onset. One of these mutations, with onset as early as the mid or late 30s, is due to a single G-to-C nucleotide substitution in exon 17, resulting in a valine-to-leucine amino acid change at codon 717 (V717L).⁷ This mutation was identified in 3 of 5 family members (siblings) tested, 1 of whom presented for PGD.

METHODS

The patient who presented for PGD was a 30-year-old woman with no signs of AD who carried the V717L mutation. The patient had been tested because her sister developed symptoms of AD at age 38 years and was found to be carrying this mutation.⁷ This sister is still alive, but her cognitive problems progressed to the point where she was placed in an assisted living facility. The patient's father had died at age 42 years and had a history of psychological difficulties and marked memory problems. The V717L mutation was also detected in one of her brothers, who experienced mild short-term memory problems as early as age 35 years, with a moderate decline in memory, new learning, and sequential tracking in the next 2 to 3 years. Other family members, including 1 brother and 2 sisters, were asymptomatic,⁷ although predictive testing was done only in the sisters, who appeared to be free of the APP gene mutation (FIGURE 1).

Two PGD cycles were performed, involving 2 standard in vitro fertilization cycles, coupled with micromanipulation procedures, including removal of polar body 1 (PB1) and polar body 2 (PB2) and intracytoplasmic sperm injection, for which the patient gave informed consent. The study was approved by the institutional review board of the Illinois Masonic Medical Center, Chicago. Testing for the maternal mutation was done by DNA analysis of PB1 and PB2, which were removed sequentially following maturation and fertilization of oocytes.⁹ A multiplex nested polymerase chain reaction (PCR) was performed,⁹ involving the mutation testing simultaneously with

the linked polymorphic marker, representing the short tandem repeat in intron 1 ($[GA]_n \dots [GT]_n$).¹⁰

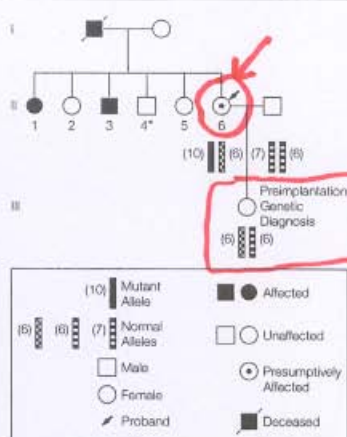
The first-round amplification cocktail for the multiplex nested PCR system contained outer primers for both the APP gene and linked marker, whereas the second-round PCR used inner primers for each gene. We designed the outer primers APP-1 (5'-GTGTTCTTTG-CAGAAGATG-3') and APP-102 (5'-CATGGAAGCACACTGATTC-3') for performing the first-round amplification and the inner primers APP-101 (5'-GTTCAAACAAGGTGCAATC-3') and APP-103 (5'-TCTTAGCAAAAAGC-TAAGCC-3') for the second round of PCR. As shown in FIGURE 2, second-round PCR produces a 115-base pair (bp) product, undigested by *MnlI* restriction enzyme, corresponding to the normal allele, and 2 restriction fragments of 72 and 43 bp, corresponding to the mutant allele. There was also an invariant fragment of 84 bp produced in both normal and mutant alleles, which was used as a control.

To perform nested PCR for specific amplification of the linked marker $(GA)_n \dots (GT)_n$ in intron 1, we designed the outer primers In1-1 (5'-CCT-TATTCAAATCCCTAC-3') and In1-2 (5'-GATTGGAGGTTAAGTTTCTG-3') for the first round and the inner primers In1-3 (5'-CAGCATCTGTCAC-TAAG-3') and In1-4 (5'-AATATT-TGTTACATTCCTC-3') for the second round of amplification. The haplotype analysis, based on the PB genotyping, demonstrated that the affected allele was linked to the 10 and the normal one to the 6 repeats.

The patient was counseled and gave consent for unaffected embryos that resulted from oocytes determined to be mutation-free, based on both mutation and short tandem repeat analysis, to be preselected for transfer back to her and those predicted to be mutant to be exposed to the confirmatory analysis using the genomic DNA from these embryos to evaluate the accuracy of the single cell-based PGD. (We did not counsel the patient about her decision to undergo the PGD testing itself.) The

Why Dax in mt?

Figure 1. Pedigree of Family With Early-Onset Alzheimer Disease (AD)



Numbers in parentheses indicate number of repeats. Preimplantation genetic diagnosis for asymptomatic carrier (II: 6) of the mutant gene linked to the 10 repeats (10) (normal gene [N] is linked to 6 repeats (6)), resulting in the birth of an unaffected child (III). Paternal genotype is also shown, with the normal alleles (N) linked to the 6 repeats (6) and 7 repeats (7). Haplotype analysis shows that the child inherited normal maternal allele (N) linked to the 6 repeats (6). The patient's sister (II: 1), brother (II: 3), and father (I) were affected by early-onset AD. No predictive testing was performed in her asymptomatic brother, indicated by an asterisk.

patient was also informed about the expected number of embryos to be transferred to achieve a pregnancy and the risks of multiple gestation, the misdiagnosis rates depending on the availability of the marker information in addition to mutation analysis, and the need for confirmation of PGD by prenatal diagnosis.

RESULTS

In the first in vitro fertilization cycle, 8 oocytes were available for testing, of which 2 were tested by both PB1 and PB2; both were affected. In the second in vitro fertilization cycle, 15 oocytes were available for testing, of which 13 were tested by both PB1 and PB2. The mutation and linked marker analysis in intron 1 revealed 6 normal and 7 affected oocytes. The results of the second cycle, resulting in the embryo transfer, are presented in Figure 2. As shown

in this figure, oocytes 4, 9, 14, and 15 were clearly normal because both mutant and normal genes were present in their PB1, with the mutant gene further extruded with the corresponding PB2, leaving only the normal gene in the resulting oocyte. In addition, oocytes 3 and 13 were also normal because their corresponding PB1s were homozygous mutant, suggesting that the resulting oocytes should have been normal, as further confirmed by the presence of the normal gene in the extruded PB2s, also in agreement with the linked markers analysis. However, because only 1 linked

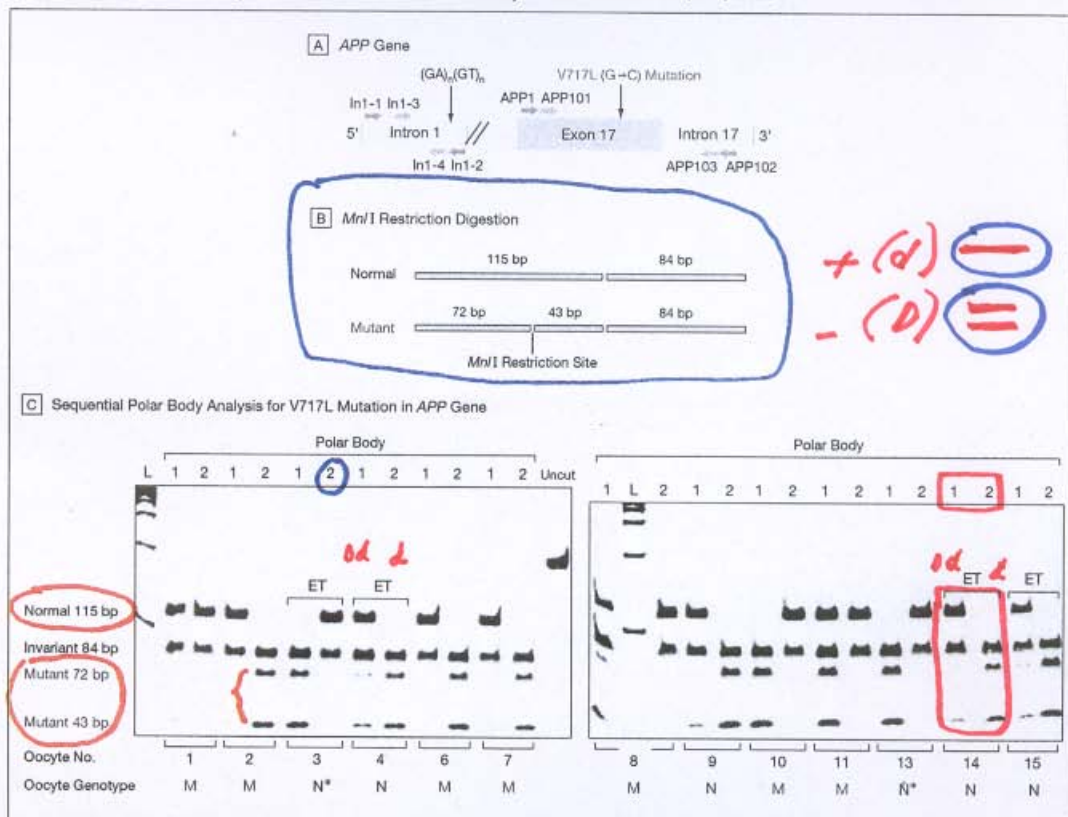
marker was available for testing, a .05 probability of allele dropout of the normal gene in the corresponding PB1 could not be excluded, as established in our previous observations.¹¹

The remaining oocytes were predicted to be mutant, based on heterozygous PB1 and normal PB2 in 4 of them (oocytes 1, 8, 10, and 11; the heterozygous status of PB1 in oocyte 10 was based on the presence of markers linked to both normal and mutant alleles, which is not shown in Figure 2) and homozygous normal PB1 and mutant PB2 in 3 (oocytes 2, 6, and 7). The follow-up

study of the embryos that resulted from these oocytes confirmed their affected status in all but 1 (oocyte 7). The latter may be explained by allele dropout of the mutant allele in the apparently heterozygous PB1, which was left undetected because of the amplification failure of the linked marker in this case.

To exclude any probability of misdiagnosis, the priority in the embryo transfer was given to 4 of the 6 normal embryos, resulting from the oocytes with heterozygous PB2 and mutant PB2. However, only 3 of these embryos developed into the cleavage stage and

Figure 2. Preimplantation Diagnosis for V717L Mutation in the Amyloid Precursor Protein (APP) Gene



ET indicates embryo transfer; L, ladder (size standard); bp, base pair; PB, polar body; and uncut, the undigested polymerase chain reaction product. Arrows indicate fully nested primer sets. A, Map of human APP gene, showing sites and location of V717L G-to-C mutation and polymorphic markers. B, Restriction map for normal and abnormal alleles. C, Polyacrylamide gel analysis of *MnlI* restriction digestion, showing 6 unaffected (N) oocytes (3, 4, 9, 13, 14, and 15) and 7 mutant (M) oocytes (1, 2, 6, 7, 8, 10, and 11). Unaffected oocytes are identified by the gel showing normal (top band) and affected (lowest 2 bands) genes in PB1. After extrusion of the affected genes in PB2, only normal genes remain. Four of these oocytes (4, 9, 14, and 15) had heterozygous PB1 and mutant PB2, and only 2 (3, 13) had homozygous mutant PB1 and normal PB2, leaving a .05 probability for misdiagnosis, noted as N*. Oocyte 10 had heterozygous PB1 detected by marker analysis (not shown).

could be transferred (4, 14, and 15), so an additional embryo (3) was preselected, originating from the oocyte with homozygous mutant PB1 and the normal PB2, since these results were also confirmed by the linked marker analysis. These 4 embryos were transferred back to the patient, yielding a singleton clinical pregnancy, confirmed to be unaffected by chorionic villus sampling and birth of a mutation-free child confirmed after birth by a blood test.

COMMENT

The results presented herein demonstrate the feasibility of PGD for early-onset AD, providing a nontraditional option for patients who wish to avoid the transmission of the mutant gene that predisposes their potential children to early-onset AD. For some patients, this may be the only reason for undertaking pregnancy, since the pregnancy may be free of an inherited predisposition

to AD from the onset. Because the disease never presents at birth or early childhood and even later may not be expressed in 100% of cases, the application of PGD for AD is still controversial. However, because there is currently no treatment for AD, which may arise despite presymptomatic diagnosis and follow-up, PGD seems to be the only relief for at-risk couples, such as the presented case and the previously reported cases of PGD for p53 tumor suppressor gene mutations.²

Therefore, prospective parents who are determined by strong genetic predisposition to be at risk for producing progeny with severe disorders should be informed about this emerging technology so they can make a choice about reproduction.^{12,13} This seems to be ethically more acceptable than suppressing information on the availability of PGD. Despite raising important ethical issues,^{14,15} the results presented

herein, together with previously described cases of PGD for late-onset disorders with genetic predisposition and HLA typing, demonstrate the extended practical implications of PGD, such as providing prospective couples at genetic risk with more reproductive options for having unaffected children.

Author Contributions: Study concept and design: Y. Verlinsky, Rechitsky, Kuliev.

Acquisition of data: Rechitsky, O. Verlinsky, Masciangelo, Lederer.

Analysis and interpretation of data: Rechitsky, Kuliev.

Drafting of the manuscript: Kuliev.

Critical revision of the manuscript for important intellectual content: Y. Verlinsky, Rechitsky, O. Verlinsky, Masciangelo, Lederer.

Obtained funding: Y. Verlinsky.

Administrative, technical, or material support: O. Verlinsky, Masciangelo, Lederer.

Study supervision: Y. Verlinsky, Rechitsky, Kuliev.

Funding/Support: Funding for this study was provided by the Reproductive Genetics Institute, Chicago, Ill.

Acknowledgment: We thank the patient for her permission to publish this information.

REFERENCES

1. International Working Group on Preimplantation Genetics. Preimplantation genetic diagnosis: experience of three thousand clinical cycles: report of the 11th Annual Meeting of the International Working Group on Preimplantation Genetics, in association with 10th International Congress of Human Genetics, Vienna, May 15, 2001. *Reprod BioMed Online*. 2001; 3:49-53.
2. Verlinsky Y, Rechitsky S, Verlinsky O, et al. Preimplantation diagnosis for p53 tumor suppressor gene mutations. *Reprod BioMed Online*. 2001;2:102-105.
3. Verlinsky Y, Rechitsky S, Schoolcraft W, Strom C, Kuliev A. Preimplantation diagnosis for Fanconi anemia combined with HLA matching. *JAMA*. 2001;285:3130-3133.
4. Sherington R, Rogaev EI, Liang Y, et al. Cloning of the gene bearing missense mutations in early onset familial Alzheimer disease. *Nature*. 1995;375:754-760.
5. Levy-Lehad E, Wasco W, Poorkaj P, et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science*. 1995;269:973-977.
6. Goate AM, Chantier-Harlin MC, Mullan M, et al. Segregation of missense mutation in the amyloid precursor protein gene with familial Alzheimer disease. *Nature*. 1991;349:704-706.
7. Murrell JR, Hake AM, Quaid KA, Farlow MR, Ghetti B. Early-onset Alzheimer disease caused by a new mutation (V717L) in the amyloid precursor protein gene. *Arch Neurol*. 2000;57:885-887.
8. Verlinsky Y, Kuliev A. *Atlas of Preimplantation Genetic Diagnosis*. New York, NY: Parthenon; 2000.
9. Rechitsky S, Strom C, Verlinsky O, et al. Accuracy of preimplantation diagnosis of single-gene disorders by polar body analysis of oocytes. *J Assist Reprod Genet*. 1999;16:192-198.
10. Tupler R, Rogaeva E, Vaula G, et al. A highly informative microsatellite repeat polymorphism in intron 1 of the human amyloid precursor protein (APP) gene. *Hum Mol Genet*. 1993;2:620-621.
11. Rechitsky S, Verlinsky O, Strom C, et al. Experience with single-cell PCR in preimplantation genetic diagnosis: how to avoid pitfalls. In: Hahn S, Holzgreve W, eds. *Fetal Cells in Maternal Blood: New Developments for a New Millennium: 11th Fetal Cell Workshop*. Basel, Switzerland: Karger; 2000:8-15.
12. Simpson JL. Celebrating preimplantation genetic diagnosis of p53 mutations in Li-Fraumeni syndrome. *Reprod BioMed Online*. 2001;3:2-3.
13. Cram D. Preimplantation genetic diagnosis for familial cancer. *Reprod BioMed Online*. 2001;3:3-4.
14. Edwards RG. A well-taken opportunity for a double blessing. *Reprod BioMed Online*. 2000;1:31-33.
15. Damewood MD. Ethical implications of a new application of preimplantation diagnosis. *JAMA*. 2001; 285:3143-3144.

EMBRYO SPLITTING CAN LEAD TO QUADRUPLETS IN PRIMATES

Fig. 1. Embryo splitting and development of non-human primates after embryo transfer. A zona-free eight-cell stage rhesus embryo, fertilized in vitro, is dissociated into eight individual blastomeres by mechanical disruption in Ca^{2+} - and Mg^{2+} -free medium. Two dissociated blastomeres are transferred into each of four empty zonae (A), thereby creating the four quadruplet embryos, each with two of the eight original cells (B). Split embryos are scored daily for development and structural normalcy, and embryos showing signs of compaction are selected for transfer 1 to 3 days after splitting. Endocrine profiles are traced daily and implantation is confirmed by ultrasound on day 31 after transfer. A miscarried pregnancy in which the fetus is absent though the placenta appears normal (C), and the quadruplet pregnancy with normal fetal development (D) that resulted in the birth of Tetra (Fig. 2) resulted from the transfer of two quadruplet embryos each to two surrogates. Bar in (A) and (B), 120 μm ; in (C) and (D), 5 cm.

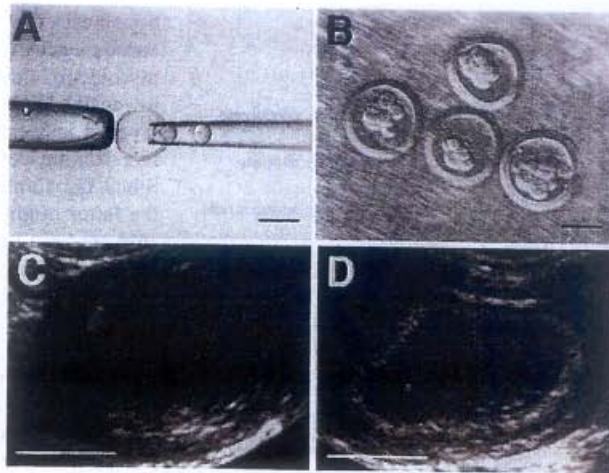


Fig. 2. Tetra, a nonhuman primate quadruplet cloned from an eight-cell embryo by splitting.

Humans? ASO/OABI Test \rightarrow good embryo \rightarrow split into 3 (2 cells each)
 implant 1 \rightarrow daughter \rightarrow 40 yo \rightarrow implant
 a second (gives birth to her daughter/sister !!
 Twin!!!)

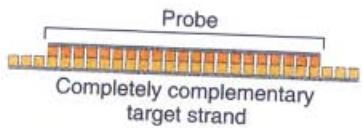
USE OF ASOs or
Allele-Specific Oligonucleotides
to Detect SNPs and
Genes

ASOs or Allele Specific Oligonucleotide Probes can be used to detect Specific Alleles/RFLPs/SNPs

This is the fastest/Simplest Approach to Fingerprinting or Monitoring disease loci - it utilizes PCR (+) Specific Annealing Conditions!

AT HIGH TEMPERATURE ONLY A PERFECT MATCH CAN ANNEAL SUCCESSFULLY! ONE MISMATCHED BASE PREVENTS HYBRID FORMATION!!!

(a) 1. 21-Base probe/target hybrid with no mismatches

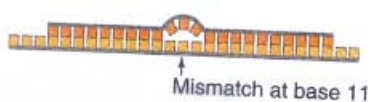


Raise temperature



Hybrid Perfect Match

2. 21-Base probe/target hybrid with middle mismatch



Raise temperature



NO Annealing Mismatch!

Figure 9.8 Short hybridization probes can distinguish single-base mismatches, longer probes cannot. (a) Researchers allow hybridization to occur between a short 21-base probe and two different target sequences. (1) A perfect match between probe and target extends across all 21 bases. When the temperature rises, this hybrid has enough hydrogen bonds to remain intact. (2) With a single-base mismatch in the middle of the probe, the effective length of the probe-target hybrid is only 10 bases. When the temperature rises, this hybrid does not have enough hydrogen bonds to remain intact, and it falls apart.

A Simpler More Inexpensive way to find SNPs/RFLPs!

USING ASOs to Detect Wild-Type (Normal) + Disease Alleles

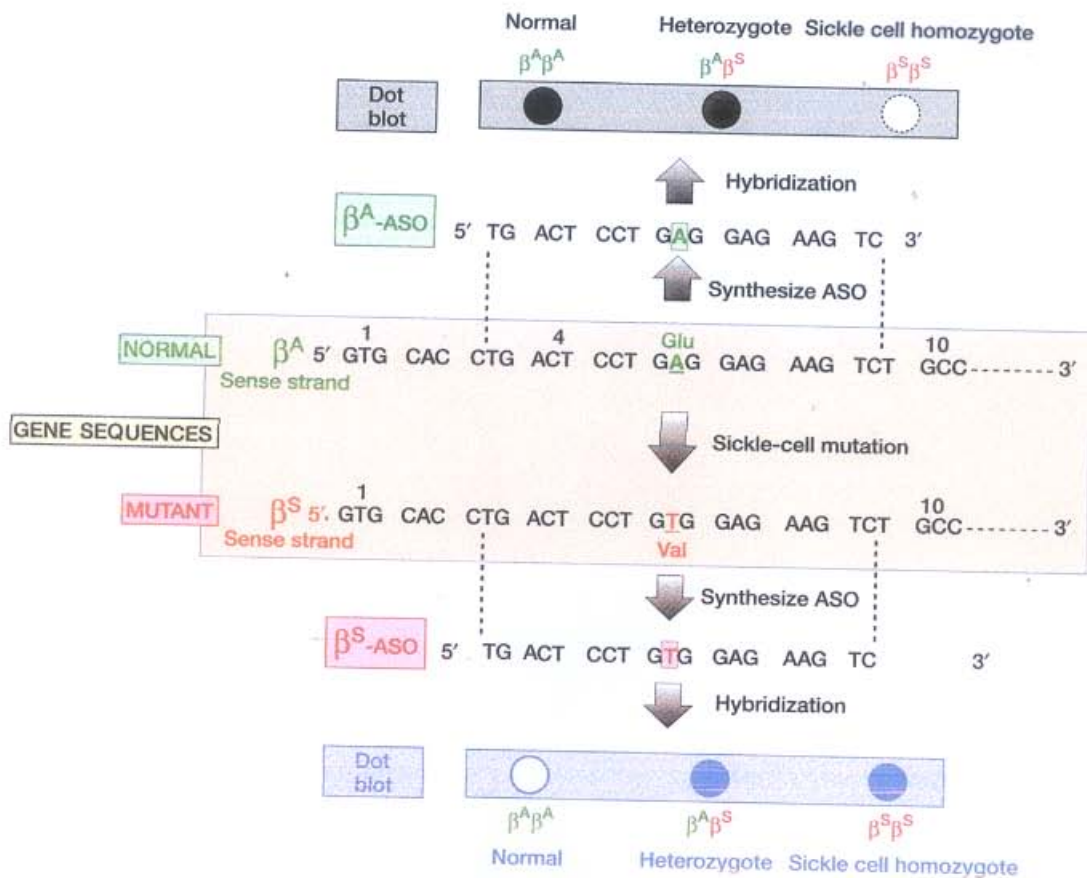
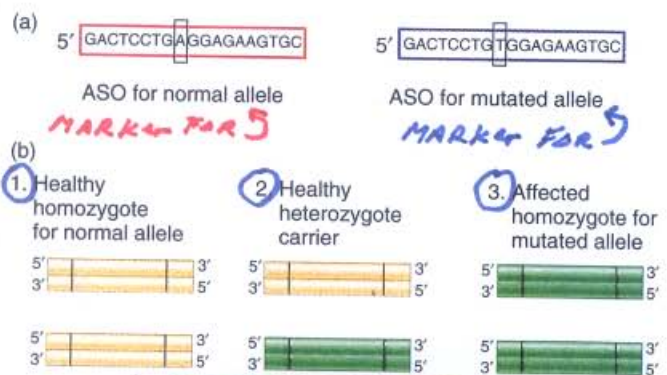


Figure 6.11: Allele-specific oligonucleotide (ASO) dot-blot hybridization can identify individuals with the sickle cell mutation.

The schematic dot blot at the top shows the result of probing with an ASO specific for the normal β -globin allele (β^A -ASO; shown immediately below). The results are positive (filled circle) for normal individuals and for heterozygotes but negative for sickle cell homozygotes (dashed, unfilled circle). The dot blot at the bottom shows the result of probing with an ASO specific for the sickle cell β -globin allele (β^S -ASO; shown immediately above), and in this case the results are positive for the sickle cell homozygotes and heterozygotes but negative for normal individuals. The β^A -ASO and β^S -ASO were designed to be 19 nucleotides long in this case chosen from codons 3 to 9 of respectively the sense β^A and β^S globin gene sequences surrounding the sickle cell mutation site. The latter is a single nucleotide substitution (A \rightarrow T) at codon 6 in the β -globin gene, resulting in a GAG (Glu) \rightarrow GTG (Val) substitution (see middle sequences).

Allele-Specific Oligonucleotides

USING ASOs to Detect sickle-cell ALLELES



PCR directly from cheek cell DNA or white blood cell DNA or Amniotic fluid, etc.

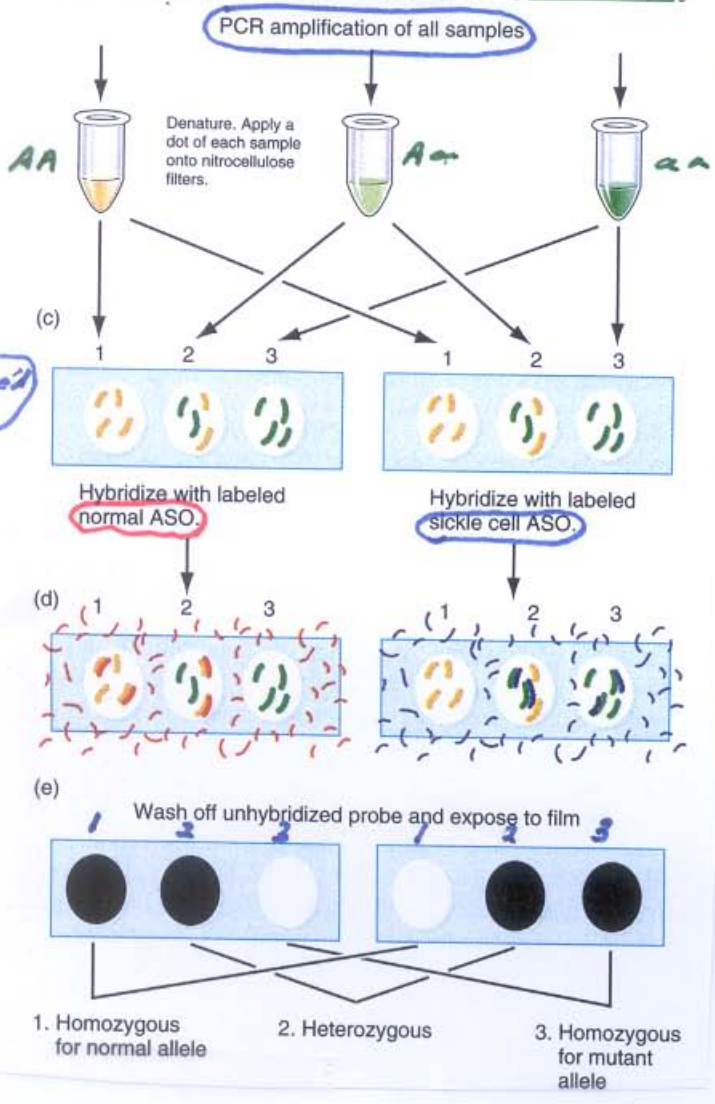


Figure 9.9 Using PCR with ASOs to determine genotype at the β -globin locus. (a) Before performing the genotyping protocol, it is necessary to synthesize two oligonucleotides that differ at only a single base; one of these oligonucleotides is complementary to the wild-type β -globin allele, the other is complementary to the sickle-cell allele. These two synthetic DNA molecules serve as the ASOs for the sickle-cell genotype assay. (b) Genomic DNA samples obtained from individual people are subjected to PCR amplification with primers complementary to nonpolymorphic sequences that flank the base that mutates to cause sickle-cell anemia. (c) The amplified sample from each individual is divided into two aliquots that are blotted directly to filter paper. (d) One aliquot from each sample is hybridized to the wild-type ASO; the other aliquot is hybridized to the sickle-cell ASO. (e) Autoradiography indicates the β -globin genotype of each individual.

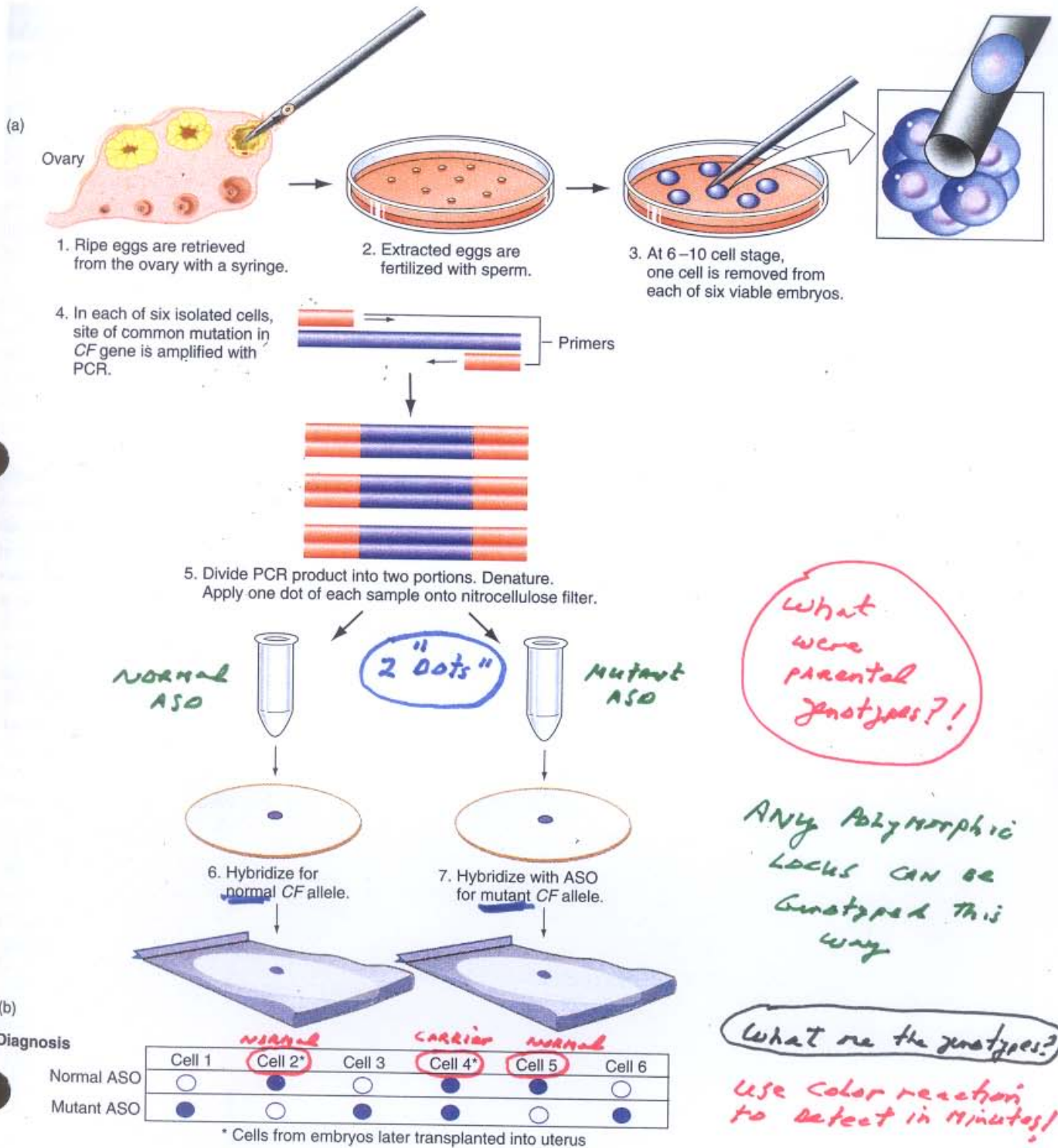
Note - Each ASO only reacts/hybridizes with a specific allele!

CAN detect with a color reaction!

FAST, simple, inexpensive!

USE FOR DISEASE ALLELES & DNA FORENSICS!!
 PATERNITY TESTS DONE THIS WAY!!!

USING ASOs + PGD to Detect Mutant Cystic Fibrosis Alleles in Post-fertilization Embryos



What were parental genotypes?!

Any polymorphic locus can be genotyped this way

What are the genotypes?

Use color reaction to detect in minutes!

Figure 9.1 Detecting the cystic fibrosis genotype of embryonic cells. (a) In vitro fertilization and preimplantation diagnosis. (b) Cell 2 is homozygous for the normal allele; cell 4 is heterozygous for the CF mutation.

Ethical Issues in Genetic Screening

Table 18.7: Requirements for a population screening program

Requirement	Examples and comments
1 A positive result must lead to some useful action	<ul style="list-style-type: none"> ▶ Preventive treatment, e.g. special diet for PKU ▶ Review and choice of reproductive options in CF carrier screening
2 The whole program must be socially and ethically acceptable	<ul style="list-style-type: none"> ▶ Subjects must opt in with informed consent ▶ Screening without counseling is unacceptable ▶ There must be no pressure to terminate affected pregnancies ▶ Screening must not be seen as discriminatory
3 The test must have high sensitivity and specificity	<ul style="list-style-type: none"> ▶ Tests with many false negatives undermine confidence in the program ▶ Tests with many false positives, even if these are subsequently filtered out by a definitive diagnostic test, can create unacceptably high levels of anxiety among normal people
4 The benefits of the program must outweigh its costs	<ul style="list-style-type: none"> ▶ It is unethical to use limited health care budgets in an inefficient way

18.6 Population screening

Population screening follows naturally from the ability to test directly for the presence of a mutation. Traditionally a distinction is drawn between screening and diagnosis. A screening test defines a high-risk group, who are then given a definitive diagnostic test. DNA tests are rather different because there are no separate screening and diagnostic tests. However, proposals to introduce any population screening test still need to satisfy the same criteria (Table 18.7), regardless of the technology used.

18.6.1 Acceptable screening programs must fit certain criteria

What would screening achieve?

The most important single function of any screening program is to produce some useful outcome. It is quite unacceptable to tell people out of the blue that they are at risk of something unpleasant, unless the knowledge enables them to do something about the risk. Proposals to screen for genes conferring susceptibility to breast cancer or heart attacks must be assessed stringently against this criterion. Predictive testing for Huntington disease might appear to break this rule – but it is offered only to people who know that they are at high risk of HD and who are suffering such agonies of uncertainty that they request a predictive test, and persist despite counseling in which all the disadvantages are pointed out.

Ideally the useful outcome is treatment, as in neonatal screening for phenylketonuria. Increased medical surveillance is a useful outcome only if it greatly improves the prognosis. One of the big risks of over-enthusiastic screening is that it can turn healthy people into ill people. A special case is screening for carrier status, where the outcome is the possibility of avoiding the birth of an affected child. People unwilling to accept prenatal diagnosis and termination of affected pregnancies would not see this as a useful outcome, and in general should not be screened, though there will be some couples who value simply knowing.

An ethical framework for screening

Ethical issues in genetic population screening have been discussed by a committee of distinguished American geneticists, clinicians, lawyers and theologians, and the reader is referred to their report for a very detailed survey (Andrews *et al.*, 1994). It is in the nature of ethical problems that they have no solutions, but certain principles emerge.

- ▶ Any program must be voluntary, with subjects taking the positive decision to opt in.
- ▶ Programs must respect the autonomy and privacy of the subject.
- ▶ People who score positive on the test must not be pressured into any particular course of action. For example, in countries with insurance-based health care systems, it would be unacceptable for insurance companies to put pressure on carrier couples to accept prenatal diagnosis and terminate affected pregnancies.
- ▶ Information should be confidential. This may seem obvious, but it can be a difficult issue – we like to think that drivers of heavy trucks or jumbo jets have been tested for all possible risks. Societies with insurance-based health care systems have particular problems about the confidentiality of genetic data, since insurance companies will argue that they are penalizing low-risk people by not loading the premiums of high-risk people.

USE OF ASOs in Forensics

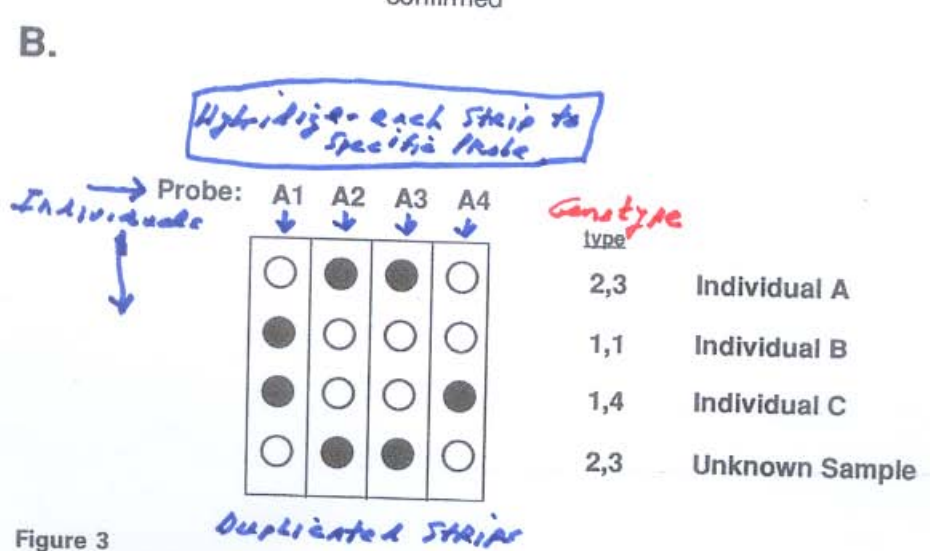
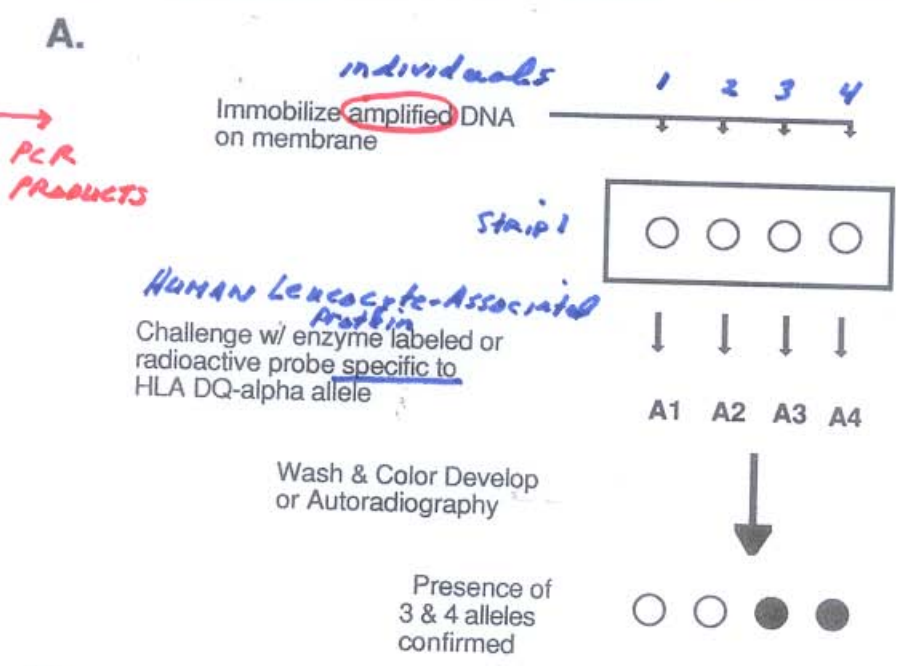
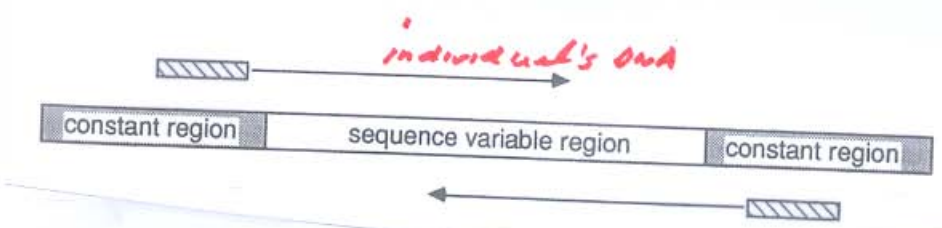


Figure 3 Format for typing amplified DQ α gene DNA. (A) Steps in the immobilization and detection of amplified DQ α alleles; (B) prototypic typing experiment. Aliquots from multiple samples are spotted in rows, as in A. Strips of membrane-containing spots from each sample are cut apart and challenged with the different probes, as in A. The developed strips are reassembled to read off the type, as shown in B.

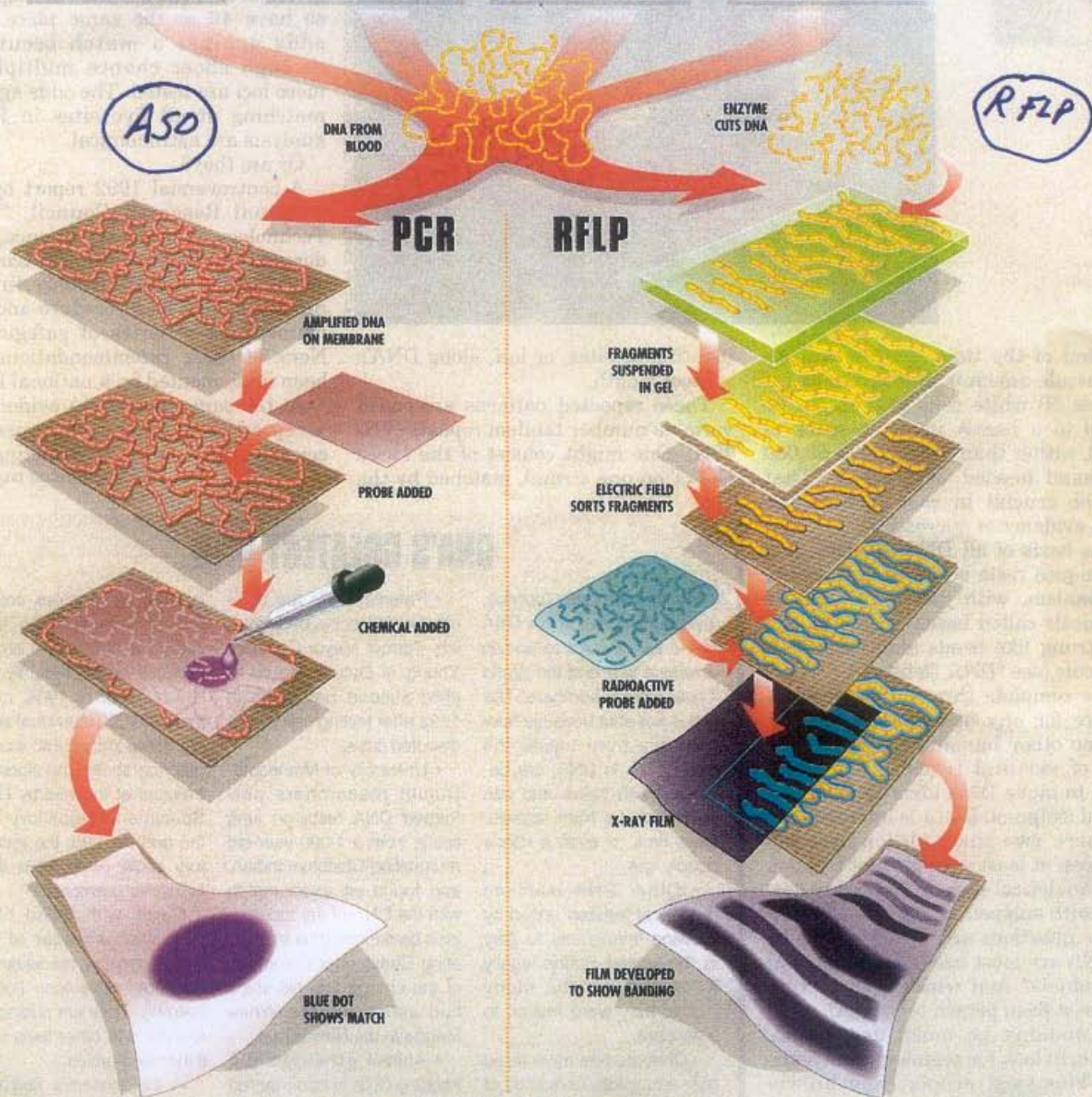
..... BUT NOT AS "SPECIFIC" AS VNTR/STR OR RFLP ANALYSIS

ASO PROBE				SAMPLE	DQ α	
2	3	4	1			
			●	POSITIVE CONTROL	1,1	} ASD control
				POSITIVE CONTROL	4,4	
	●			POSITIVE CONTROL	3,3	
●				POSITIVE CONTROL	2,2	
			●	BLOOD [S]	1,4	} Suspect Blood
			●	ONE HAIR [V]	1,4	} Victim's Hair
			●	TEN HAIRS [V]	1,4	
			●	ONE SCENE HAIR	3,3	} Crime Hair
				20 DOG HAIRS	N.A	} Negative Control
	●		●	LIQ. AUTOPSY BLOOD	1,3	} VICTIM / crime scene
	●		●	DRIED AUTOPSY BLOOD	1,3	
●		●		PANTIES SPERM DNA	4,2	} VICTIM
	●			VAG. E. CELL DNA	1,3	
●		●		VAG. SPERM DNA	4,2	
	●		●	RECTAL E. CELL DNA	1,3	

Figure 4

Case samples typed for DQ α alleles. Samples were tested in the format of Fig. 3 using horseradish peroxidase (HRP)-conjugated probes (for discussion, see text; C. Chang, in prep.).

FINGERPRINTING WORKS



being tested and transfer their radioactivity selectively to those fragments.

Finally, the membrane is placed over standard X-ray film. Radiation emitted from the P-32 gradually exposes the film and gives a precise picture of the DNA fragments.

But the process takes time. The P-32 is so weak that this approach is like sitting in your dentist's chair for two weeks to get an X-ray of

your molars. And each of the five loci must be exposed sequentially. The ten weeks of waiting for the X-ray film to be exposed accounts for most of the time it takes to complete an RFLP fingerprint.

Once the film is developed, it's inspected by the scientist conducting the test and at least one other expert. In addition, it is scanned into a computer for precise measurement and

comparison against known samples of DNA.

If lines and bars from the known and unknown DNA samples don't match, this is conclusive evidence that they came from different people.

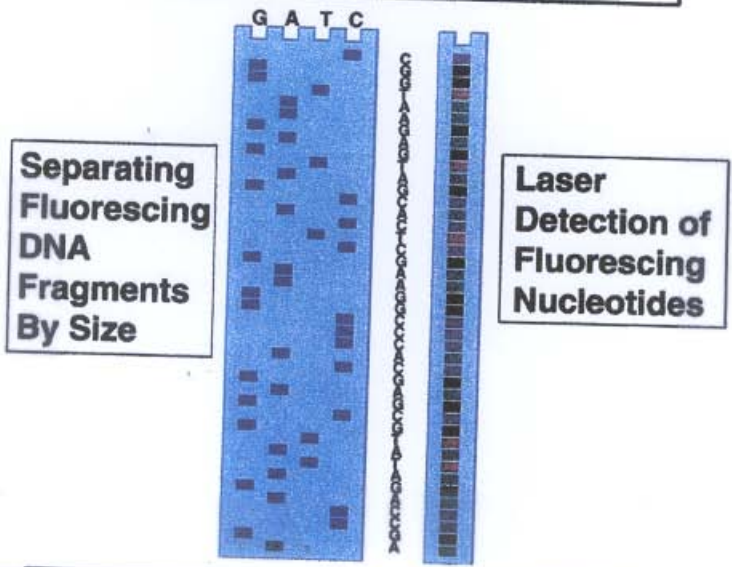
If the X-ray codes do match, some experts will argue that they almost certainly came from the same person. And other experts will challenge that conclusion.—J. S.

DIRECT DNA SEQUENCING
+
CHIPS TO DETECT
SNPs

Gene + Whole Genome Approaches

Detecting SNPs By Sequencing

Genome Sequencing Using Computers and Robotics

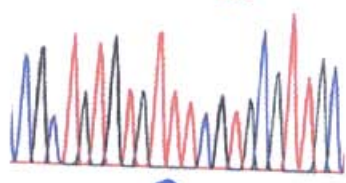


Individual 1

27-D

CGC TATG TAT T T C G T A C A T T A C
90 100

16093 T

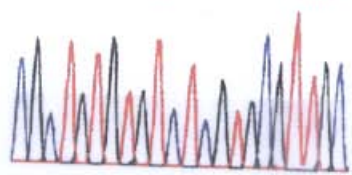


Individual 2

28-D

CGC TATG TAT C T C G T A C A T T A C
90 100

16093 C



Using Chips to Detect SNPs

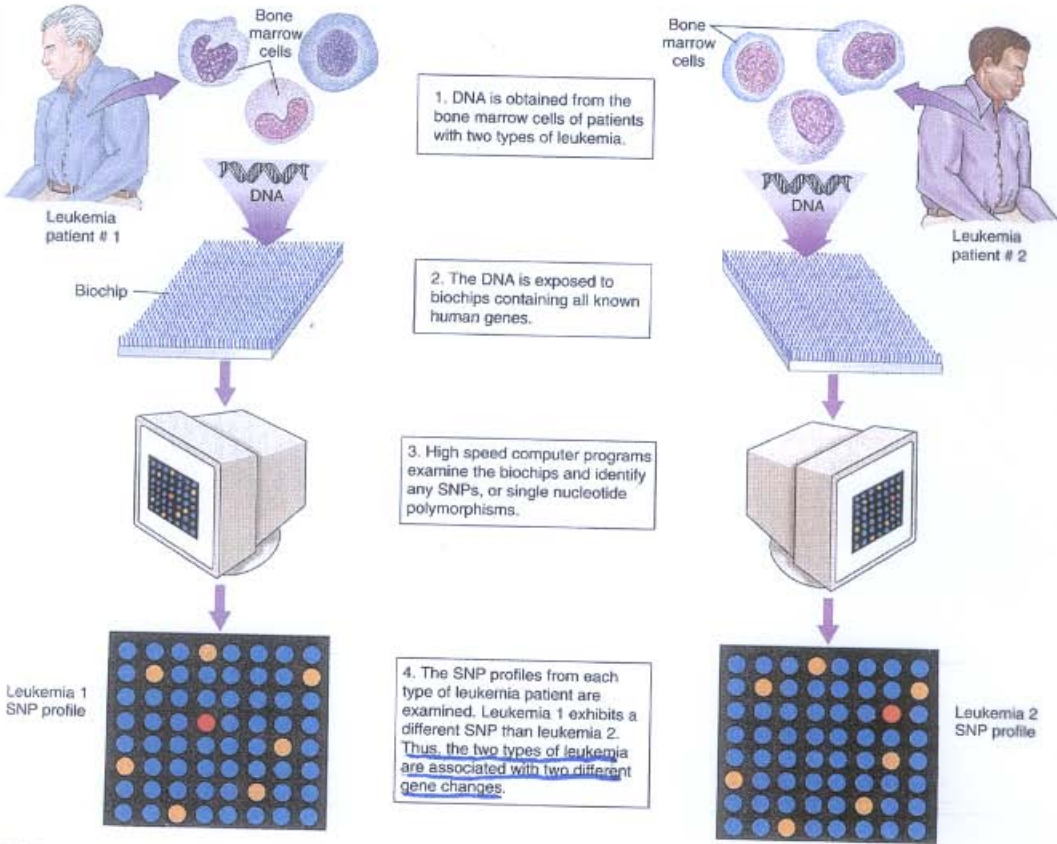


FIGURE 19.16 Biochips can help in identifying precise forms of cancer.

There are Millions of SNPs
That Differ Among
Individuals

. . . . But a small few reflect
our ancestry & "travel" in
groups on chromosomes —
are linked & may show
specific gene linkages!

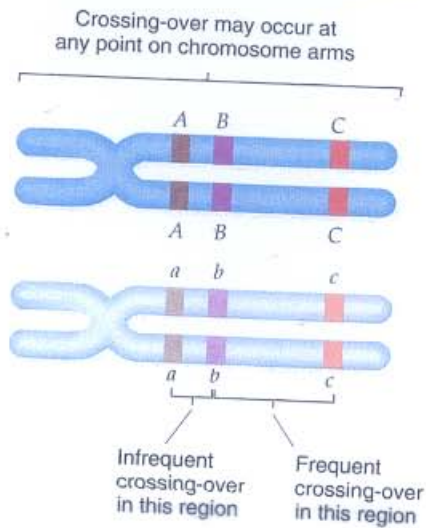
Haplotypes!

CLOSELY-LINKED SNPs ARE INHERITED AS A UNIT

Figure 5.3

The relationship between recombination and map distance.

The farther apart two genes are, the greater the number of possible sites for recombination. Thus, the probability of recombination occurring between genes A and B is much less than that between genes B and C. The percentage of recombinants can provide information about the relative genetic distance between two linked genes.

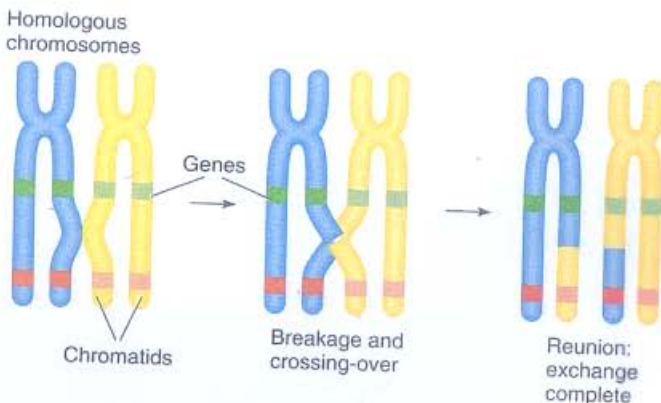


NO CROSSING OVER
 $< 5kb$

\therefore
 HAPLOTYPE
 or
 COMPLEX
 POLYMORPHIC
 LOCUS

Figure 5.2

Mechanism of crossing-over. A highly simplified diagram of a crossover between two nonsister chromatids during meiotic prophase, giving rise to recombinant (nonparental) combinations of linked genes.



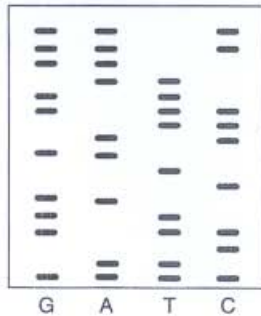
A Haplotype is a closely linked set of specific SNPs



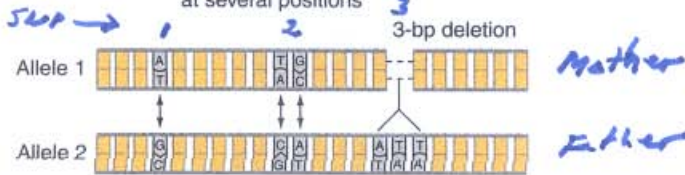
PCR amplification of HLA locus from one person who is heterozygous for two complex haplotypes

Clone from PCR products

Sequence several clones to obtain at least one sequence from each of the two alleles.



Production of two classes of clones that differ at several positions



They are ALWAYS inherited together & reflect Ancestry!

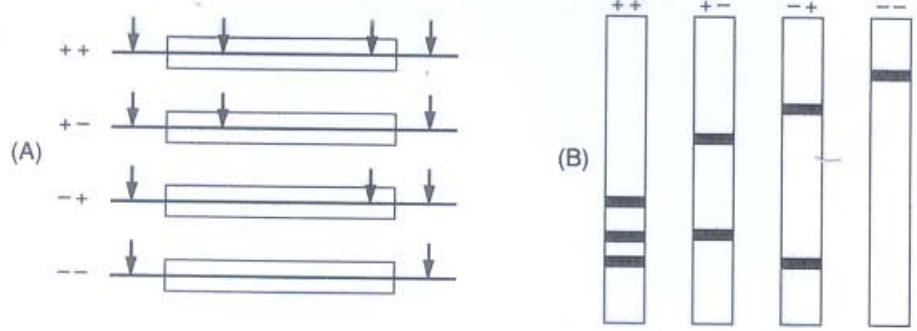
Figure 9.15 The variations associated with a complex haplotype are best defined by sequencing. Using automated protocols to sequence an entire polymorphic region is often the most rapid and accurate way to detect changes associated with polymorphic alleles at a complex locus.

A Complex Polymorphic Locus With Four Haplotypes

Figure 4-6

A complex polymorphic locus consisting of two adjacent RFLP sites. (A) The four possible haplotypes. Arrows indicate presence of a cleavage site for a restriction endonuclease. Boxed areas are the target sequences recognized by the probe. (B) Southern blots showing the relative electrophoretic mobilities of the fragments produced by restriction enzyme digestion of DNA from each haplotype. Note that all pairwise combinations of the haplotypes can be distinguished from one another; thus, these are codominant alleles.

1
2
3
4



53

Complex Haplotypes

A contraction of the phrase "haploid genotype," the term **haplotype** refers to a specific combination of linked alleles in a cluster of related genes. Immunogeneticists often use it to describe the combination of alleles of the *major histocompatibility complex (MHC)*: a large cluster of genes on human chromosome 6 that play a role in the immune response. With the resolving power to look at DNA at the level of nucleotides, "haplotype" now refers to any set of linked DNA changes along a chromosome. These changes could be in one or several genes, or in noncoding stretches. The **complex** refers to the multiple types of variation that can exist at alternative alleles, including more than one nucleotide substitution, a substitution in combination with a small deletion, duplication, or other insertion. Thus, a **complex haplotype** is a set of linked DNA variations along a chromosome, with the possibility of many differences between alternative alleles.

HAPLOTYPE PATTERNS

Person A	ATTGAT	CGGAT...CCATCGGA...CTAA
Person B	ATTGAT	AGGAT...CCAGCGGA...CTCA
Person C	ATTGAT	CGGAT...CCATCGGA...CTAA
Person D	ATTGAT	AGGAT...CCAGCGGA...CTCA
Person E	ATTGAT	CGGAT...CCATCGGA...CTAA

Building blocks. Persons B and D share a haplotype unlike the other three, characterized by three different SNPs.

Using SNP Haplotypes to identify association with Disease Genes

disease gene

Son	Mother	Grandfather	Grandmother
1	1 2	1	2 2
2	2 2	2	2 1
m	m +	+	+
1	1 1	1	1 1
2	2 1	2	1 1

∴ these SNPs associated with Disease Gene

Figure 6-4

Use of haplotypes to identify the source of a new mutation in an X-linked gene. Each column represents a hypothetical haplotype for four RFLP loci, each with two alleles (indicated by 1 or 2); and the disease locus, where + indicates the normal allele and m the mutant allele. It is assumed that the presence or absence of the mutant allele can be detected by some direct molecular assay, such as hybridization to an allele-specific oligonucleotide or PCR amplification of a portion of the gene, followed by sequencing. In either case, knowing that the mutation is present in the mother but absent in both of her parents does not tell us which of her parents was the source of the mutant gamete. Haplotype analysis, using closely linked polymorphic loci, solves that problem. In this example, it is clear that the affected boy has his grandfather's X chromosome; therefore, the mutation that he and his mother possess must have originated in his grandfather's germ cells.

The International HapMap Project

The International HapMap Consortium*

*Lists of participants and affiliations appear at the end of the paper

The goal of the International HapMap Project is to determine the common patterns of DNA sequence variation in the human genome and to make this information freely available in the public domain. An international consortium is developing a map of these patterns across the genome by determining the genotypes of one million or more sequence variants, their frequencies and the degree of association between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe. The HapMap will allow the discovery of sequence variants that affect common disease, will facilitate development of diagnostic tools, and will enhance our ability to choose targets for therapeutic intervention.

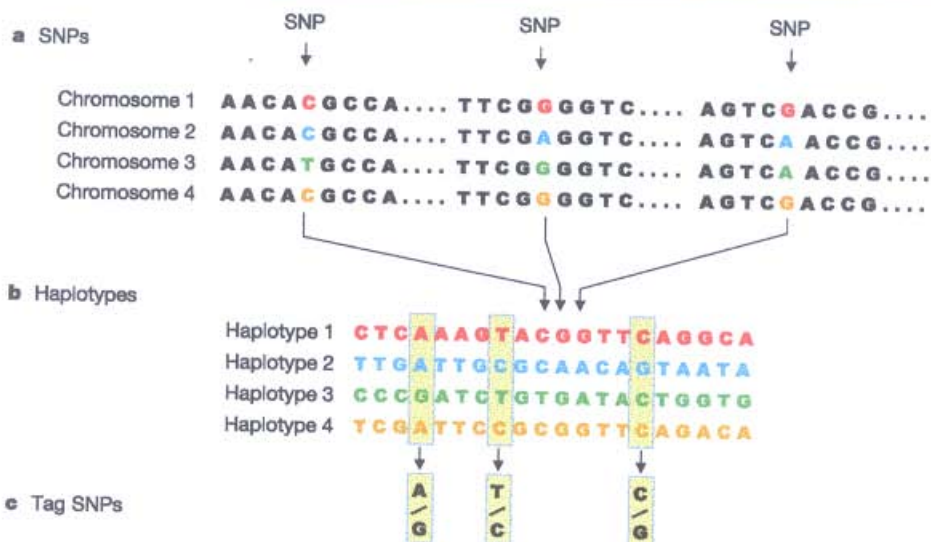


Figure 1 SNPs, haplotypes and tag SNPs. **a**, SNPs. Shown is a short stretch of DNA from four versions of the same chromosome region in different people. Most of the DNA sequence is identical in these chromosomes, but three bases are shown where variation occurs. Each SNP has two possible alleles; the first SNP in panel **a** has the alleles C and T. **b**, Haplotypes. A haplotype is made up of a particular combination of alleles at nearby SNPs. Shown here are the observed genotypes for 20 SNPs that extend across 6,000 bases of DNA. Only the variable bases are shown, including the

three SNPs that are shown in panel **a**. For this region, most of the chromosomes in a population survey turn out to have haplotypes 1–4. **c**, Tag SNPs. Genotyping just the three tag SNPs out of the 20 SNPs is sufficient to identify these four haplotypes uniquely. For instance, if a particular chromosome has the pattern A–T–C at these three tag SNPs, this pattern matches the pattern determined for haplotype 1. Note that many chromosomes carry the common haplotypes in the population.

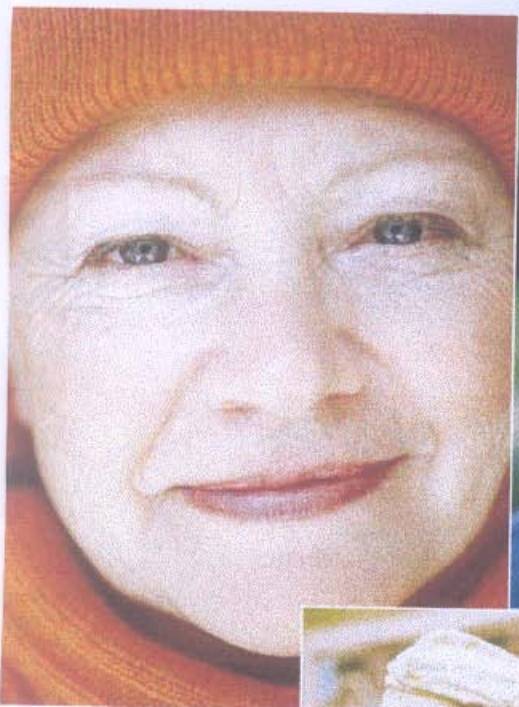
THE 0.1% THAT'S DIFFERENT!

CORRELATE WITH SEQUENCE VARIANTS AFFECTING DISEASE

6/10

International HapMap Groups

news feature



World view: the HapMap initiative will gather genetic data from African, Asian and ancestrally European populations.



- ① Northern European
- ② Western European
- ③ Yoruba/Nigerian/African
- ④ Japanese & Han Chinese/Asian

"GROUP" Genetic Diversity to Disease & Other Aspects of Biology.

⑤X

Ethical Issues Related to the HAPMAP

Box 1

Community engagement, public consultation and individual consent

As no personally identifiable information will be linked to the samples, the risk that an individual will be harmed by a breach of privacy, or by discrimination based on studies that use the HapMap, is minimal. However, because tag SNPs for future disease studies will be chosen on the basis of haplotype frequencies in the populations included in the HapMap, the data will be identified as coming from one of the four populations involved, and it will be possible to make comparisons between the populations. As a result, the use of population identifiers may create risks of discrimination or stigmatization, as might occur if a higher frequency of a disease-associated variant were to be found in a group and this information were then overgeneralized to all or most of its members⁶⁴. It is possible that there are other culturally specific risks that may not be evident to outsiders⁶⁵. To identify and address these group risks, a process of community engagement, or public consultation, was undertaken to confer with members of the populations being approached for sample donation about the implications of their participation in the project^{66,67}. The goal was to give people in the localities where donors were recruited the opportunity to have input into the informed consent and sample collection processes, and into such issues as how the populations from which the samples were collected would be named. Community engagement is not a perfect process, but it is an effort to involve potential donors in a more extended consideration of the implications of a research project before being asked to take part in it⁶⁸. Community engagement and individual informed consent were conducted under the auspices of local governments and ethics committees, taking into account local ethical standards and international ethical guidelines. As in any cross-cultural endeavour, the form and outcome of the processes varied from one population to another. A Community Advisory Group is being set up for each community to serve as a continuing liaison with the sample repository, to ensure that future uses of the samples are consistent with the uses described in the informed consent documents. A more detailed article discussing ethical, social and cultural issues relevant to the project, and describing the processes used to engage donor populations in identifying and evaluating these issues, is in preparation.

SNPs Should Be Powerful Indicators
OF Disease Susceptibility

D528-D532 *Nucleic Acids Research*, 2004, Vol. 32, Database issue
DOI: 10.1093/nar/gkh005

SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes

Bernice R. Packer*, Meredith Yeager, Brian Staats, Robert Welch, Andrew Crenshaw, Maureen Kiley, Andrew Eckert, Michael Beerman, Edward Miller, Andrew Bergen¹, Nathaniel Rothman¹, Robert Strausberg² and Stephen J. Chanock³

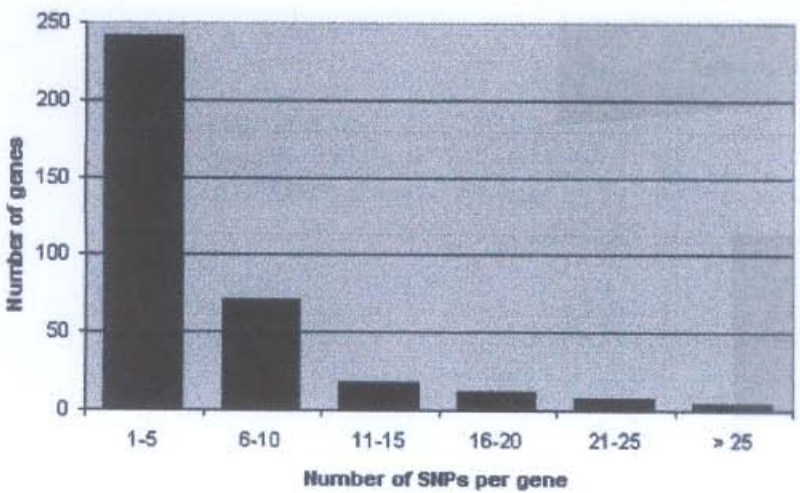


Figure 1. SNPs per gene.

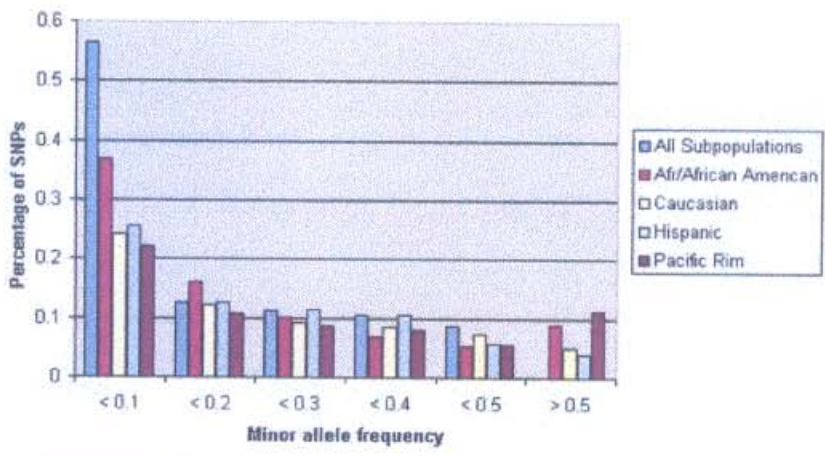


Figure 2. SNP500Cancer allele frequencies by subpopulation.

SNPs Found in IL10 Gene

dbSNP ID: **rs1800871**

SNP500Cancer ID: IL10-01
 Gene: IL10
 SNP Region: -654C>T
 Note aka -819

[dbSNP](#)
[NCBI map](#)
[Ensembl map](#)
[LocusLink](#)

Sequence of Analyzed Amplicon

```

CTTCTTCCACCCCATCTTTTAAACTTTAGACTCCAGCCACAGAAGCTTACAA
CTAAAAGAAACTCTAAGGCCAATTTAATCCAAGGTTTCATTCTATGTGCTGG
AGATGGTGTACAGTAGGGTGAGGAAAACCAAATTCACAGTTRGCAGTGGTGTGTA
CCCTTGACAGGTGATGTAA (C/T) ATCTCTGTGCCCTCAGTTTGCTCACTAT
AAAATAGAGACGGTAGGGGTCATGGTGAGCACTACCTGACTAGCATATAAGA
AGCTTTCAGCAAGTGCAGACTACTCTTACCCACTTCCCCCAAGCACAGTTGG
GGTGGGGACAGCTGAAGAGGTGAAAACATRTGCCTGAGAATCCTAATGAAA
TCGGGGTA
    
```

Frequency Data (102 anonymized subjects):



Total Completed	Genotypic			Allelic	
	CC	CT	TT	C	T
<u>102</u>	37/102 (0.363)	49/102 (0.480)	16/102 (0.157)	123/204 (0.603)	81/204 (0.397)

[View Subpopulation Frequencies](#)

Assays - these frequency results were validated on the following platforms - click to view primers, probes, and conditions:
[Sequencing](#) [Sequenom](#) [TagMan](#)

Figure 4. SNP information.

SNP Database

Cancer Genome Anatomy Project
SNP500Cancer Database

NCI > CGAP > SNP500Cancer > Search by gene

Home

Search by Gene/
Chromosome/
Pathway

Search by SNP

Links

Log In

List of SNPs on Genes matching **IL10**.

Gene Symbol: [IL10](#)
alias: CSIF TGIF IL-10 IL10A
chromosome: 1q31-q32
GenBank ID: [NM_000572](#)
[GO pathways](#)

validated SNPs:

- [rs1800871, -854C>T](#)
- [rs1800872, -527C>A](#)
- [rs1800894, -866G>A](#)
- [rs1800896, -1117A>G](#)
- [rs3021094, IVS1-192](#)
- [rs3024491, IVS1-266](#)
- [rs3024494, IVS3-112](#)
- [rs3024496, 117bp 3' of STP](#)
- [rs3024508, IVS3-171](#)
- [rs3024509, IVS3-58](#)
- [E1854 239, IVS3-126](#)
- [E1858 189, IVS1+453](#)

no variation found in SNP500Cancer population

- [rs3021093, IVS1+512](#)
- [rs3024497, 191bp 3' of STP](#)
- [rs3024510, 345bp 3' of STP](#)

Figure 3. Listing a gene's SNPs.

Frequency Data (102 anonymized subjects)						
dbSNP ID: rs1800871						
Subpopulations	Genotypic			passed <i>HWE?</i>	Allelic	
	CC	CT	TT		C	T
Total Completed	37/102 (0.363)	49/102 (0.480)	16/102 (0.157)	-	123/204 (0.603)	81/204 (0.397)
Afr/Afr American	5/24 (0.208)	14/24 (0.583)	5/24 (0.208)	passed	24/48 (0.500)	24/48 (0.500)
Caucasian	15/31 (0.484)	14/31 (0.452)	2/31 (0.065)	passed	44/62 (0.710)	18/62 (0.290)
Hispanic	13/23 (0.565)	8/23 (0.348)	2/23 (0.087)	passed	34/46 (0.739)	12/46 (0.261)
Pacific Rim	4/24 (0.167)	13/24 (0.542)	7/24 (0.292)	passed	21/48 (0.438)	27/48 (0.563)

Figure 5. Genotypic and allelic frequencies for a SNP.

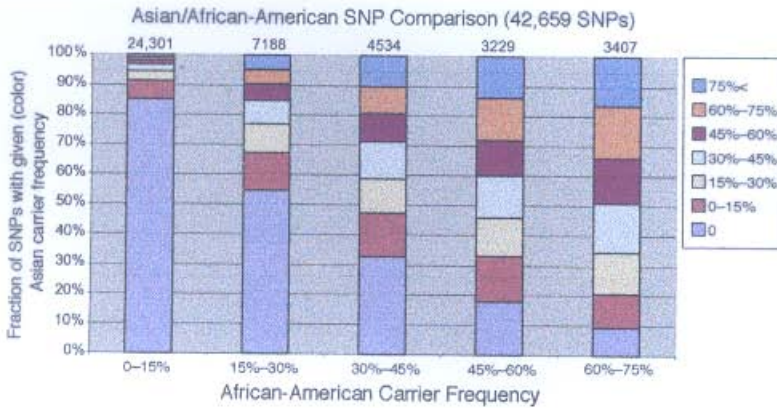
(6)

SNPs Might Be Useful for
PROVIDING CLUES AS TO
DISEASE GENES

Race and Medicine

Genetic studies of population differences, although controversial, promise clues to disease as well as new drug targets, scientists believe

24 OCTOBER 2003 VOL 302 SCIENCE www.sciencemag.org



NOTE:
Diversity
OF
ALLELES
↓
Frequency
Differences

Biodiversity. More than 42,000 SNPs (genetic variations) found in African Americans are divided into columns according to how frequently they appear in that population. Colors indicate the frequency with which these same groups of SNPs are found in East Asians. For instance, in the second column, of the 7,188 SNPs that are found in 15% to 30% of African Americans, more than half show no variation in Asians.

USING GROUP/"RACE" GENE
VARIABILITY DATA

ARE There HUMAN
"Races?"

What is the History
& the Biology?

RACE is Primarily a Sociological
Concept that has caused
MUCH HUMAN Suffering



Race is a largely non-biological concept confounded by misunderstanding and a long history of prejudice. The relationship of genomics to the concepts of race and ethnicity has to be considered within complex historical and social contexts.

Most variation in the genome is shared between all populations, but certain alleles are more frequent in some populations than in others, largely as a result of history and geography. Use of genetic data to define racial groups, or of racial categories to classify biological traits, is prone to misinterpretation. To minimize such misinterpretation, the biological and sociocultural factors that interrelate genetics with constructs of race and ethnicity need to be better understood and communicated within the next few years.

This will require research on how different individuals and cultures conceive of race, ethnicity, group identity and self-identity, and what role they believe genes or other biological factors have. It will also require a critical examination of how the scientific community understands and uses these concepts in designing research and presenting findings, and of how the media report these. Also necessary is widespread education about the biological meaning and limitations of research findings in this area (Box 6) and the formulation and adoption of public-policy options that protect against genomics-based discrimination or maltreatment (see Grand Challenge III-1).

Genome
Project
Goal

Based on a very few genes that
vary between groups far more
than majority - vast majority
of other genes



A genetic melting-pot

Marcus W. Feldman, Richard C. Lewontin and Mary-Claire King

Race as a biological concept has had a variety of meanings. In the taxonomic literature, a race is any distinguishable type within a species, such as dark-bellied and light-bellied variants of small mammals. In 1937, Theodosius Dobzhansky introduced the idea of geographical races — populations of species that differ in the frequencies of one or more genetic variants. But as no two populations have identical gene frequencies at variable (polymorphic) loci, Dobzhansky's definition of race becomes synonymous with that of population.

The classical definition of race, as applied to our species, is based on phenotypes such as skin colour, facial features and hair form that clearly differ between native inhabitants of different regions of the world. An underlying assumption is that all of these defining features (all largely genetic traits, although few of their genes have been identified) are characteristic of the genome in general. In other words, just as there are large differences between races in genes for skin colour, so there should be large genetic differences between races in general. In the previous absence of data to confirm or deny this assumption, it was not an unreasonable one to make.

But recent studies of genetic diversity indicate that the genes underlying the phenotypic differences used to assign race categories are atypical, in that they vary between races much more than genes in general. Together, the iconic features of race correlate well with continent of origin but do not reflect genome-wide differences between groups.

Discussion has arisen over the implications of these findings for the utility of racial classification in medical practice. The issue of whether race is a biologically useful or even meaningful concept when applied to humans in a medical context is controversial — holders of opposing views each claim to have evidence to support them. But there is no contradiction between these two well-substantiated bodies of data, as they actually

deal with two different questions that have become confused with one another.

The first question is: "Is it possible to find DNA sequences that differ sufficiently between populations to allow correct assignment of major geographical origin with high probability?" The answer to this question is yes, as shown by studies of genetic polymorphisms and by universal personal experience.

The second question is: "What fraction of human genetic variation, whether based on protein-coding genes or other sequences, falls within geographically separated populations, and what fraction occurs between these populations?" The answer to this question is that most genetic diversity occurs within groups, and that very little is found between them.

Why this apparent paradox? The answer is that genes that are geographically distinctive in their frequencies are not typical of the human genome in general.

It has been suggested that racial categorization has a valid role in good medical practice because many medically important genes vary between populations from different regions. But although knowing a patient's ancestry is often extremely useful in diagnosis and treatment, race is both too broad and too narrow a definition of ancestry to be biologically useful.

For any species, definitions of race can lose their discriminating power when individuals migrate to different regions and mate with their counterparts there. Among humans, large-scale migrations between continents — particularly through European colonial expansion and the commercial slave trade — has resulted in matings of individuals from different continents and the creation of new populations, especially in the Western Hemisphere and Oceania. Many people thus have ancestry from more than one major geographical region, meaning that the association of phenotype and geography breaks down.

For example, sickle-cell disease, which is often thought to be an African trait, is instead characteristic of ancient ancestry in a geographic region where malaria was endemic. Africa is one such region, but so are the Mediterranean and southern India. If sickle-cell disease is suspected, then the correct diagnostic approach is not simply to determine the patient's race, but to ask whether they have African, Mediterranean or South Indian ancestry. To use genotype effectively in making diagnostic and therapeutic decisions, it is not race that is relevant, but both intra- and trans-continental contributions to a person's ancestry.

Race and ancestry are confounded both by genetic heterogeneity within groups and by the widespread mixing of previously iso-

Race

Ancestral genetic data are far more useful for medical purposes than are racial categories, which may be correlated with disease for social or economic rather than biological reasons.

lated populations. The assignment of a racial classification to an individual hides the biological information that is needed for intelligent therapeutic and diagnostic decisions. A person classified as 'black' or 'Hispanic' by social convention could have any mixture of ancestries, as defined by continent of origin. Confusing race and ancestry could be potentially devastating for medical practice.

Other attempts to classify people into broad genetic groups based on the frequency of specific genes for, say, drug-metabolizing enzymes, are also likely to be poor predictors of medical outcome. As with racial groupings, the overall variation in the frequencies of such genes between groups is likely to be less than that within each group.

The conventional, social definition of race is useful in a medical context as it provides information about the social circumstances and lifestyle of patients. But this is a consequence of social history, so any variation is (at least in principle) transitory. By contrast, information on the likelihood that a person carries specific disease-related or treatment-response genes is grounded in their ancestry in far more complex ways. We suggest that identifying all contributions to a patient's ancestry can be useful in diagnosing and treating diseases with genetic influences. Eventually, for both diagnosis and treatment, specific genetic variants will provide concrete, useful information.

Marcus W. Feldman is in the Department of Biological Sciences, Stanford University, California 94305, USA.

Richard C. Lewontin is at the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

Mary-Claire King is in the Departments of Genome Science and Medicine, University of Washington, Seattle, Washington 98195, USA.

FURTHER READING

- Rosenberg, N. A. *et al.* *Science* 298, 2381–2385 (2002).
- González Burchard, E. *et al.* *N. Engl. J. Med.* 348, 1170–1175 (2003).
- Cooper, R. S., Kaufman, J. S. & Ward, R. *N. Engl. J. Med.* 348, 1166–1170 (2003).
- Lewontin, R. C. *Evol. Biol.* 6, 381–398 (1972).
- Barbujani, G., Magagnoli, A., Minch, E. & Cavalli-Sforza, L. L. *Proc. Natl Acad. Sci. USA* 94, 4516–4519 (1997).



Human migration, as depicted in Charles Hunt's *Landing at Madras*, blurs the boundaries of race.

605

Human pigmentation genetics: the difference is only skin deep

Richard A. Sturm,^{1*} Neil F. Box,¹ and Michele Ramsay²

Summary

There is no doubt that visual impressions of body form and color are important in the interactions within and between human communities. Remarkably, it is the levels of just one chemically inert and stable visual pigment known as melanin that is responsible for producing all shades of humankind. Major human genes involved in its formation have been identified largely using a comparative genomics approach and through the molecular analysis of the pigmentary process that occurs within the melanocyte. Three classes of genes have been examined for their contribution to normal human color variation through the production of hypopigmented phenotypes or by genetic association with skin type and hair color. The MSH cell surface receptor and the melanosomal P-protein are the two most obvious candidate genes influencing variation in pigmentation phenotype, and may do so by regulating the levels and activities of the melanogenic enzymes tyrosinase, TRP-1 and TRP-2. *BioEssays* **20**:712-721, 1998. © 1998 John Wiley & Sons, Inc.



TABLE 1. Human Pigmentation Genes

Gene symbol	Mouse homologue	Chromosome	Phenotype	Protein	Function/activity
TYR	Albino (c)	11q14-21	OCA1	Tyrosinase	Tyrosine hydroxylation; DOPA oxidase
TYRP1	Brown (b)	9p23	OCA3/ROCA	TRP-1	DHICA oxidase
TYRP2	Slaty (slt)	13q31-32	Unknown	TRP-2	Dopachrome tautomerase
P	Pink-eyed dilute (p)	15q11.2-12	OCA2, BOCA	P-protein	Melanosomal transmembrane protein
MC1R	Extension (e)	16q24.3	Red hair	MSHR	G-protein-coupled receptor

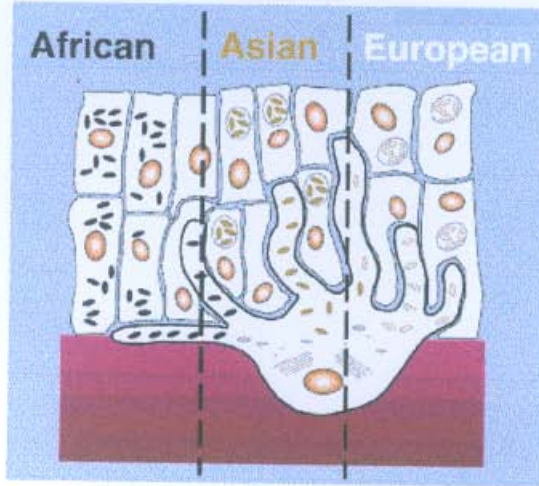
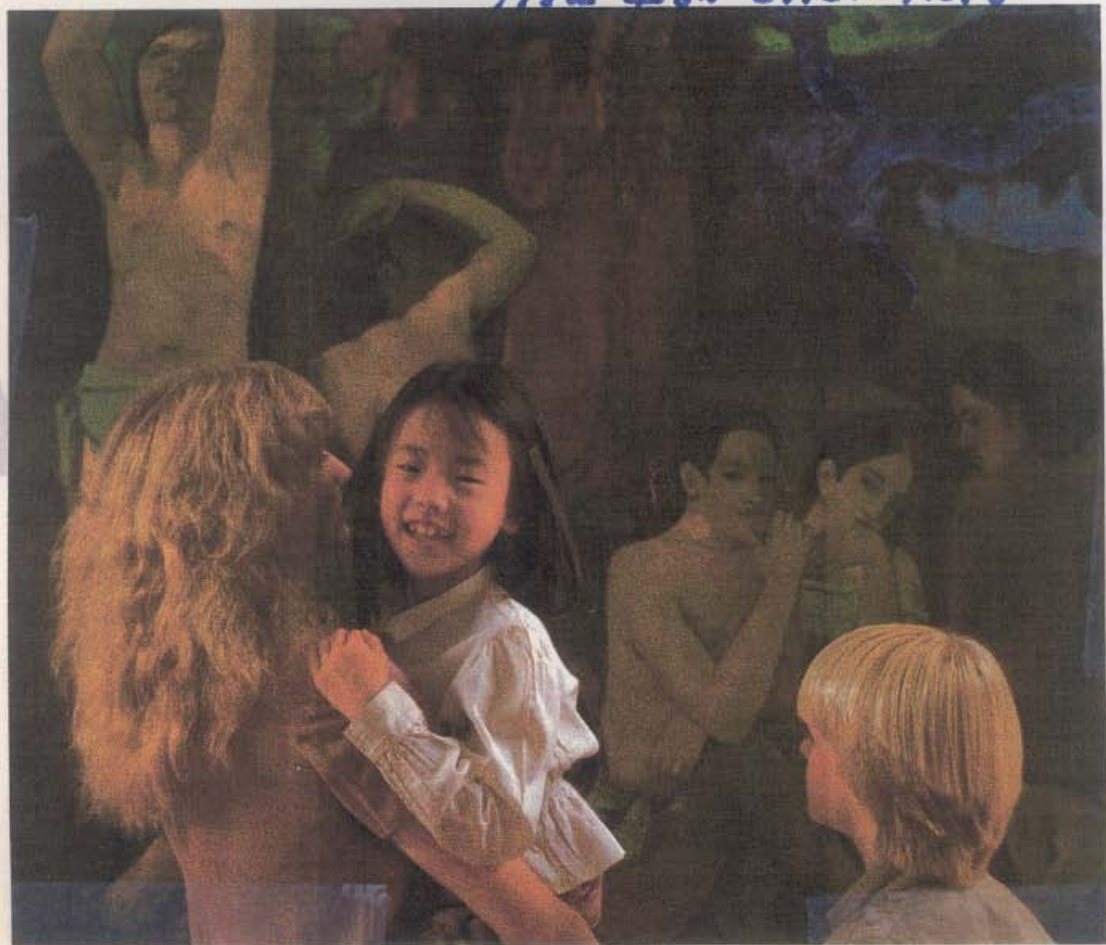


Figure 1. Variation in melanosome structure and distribution in different groups. A single skin melanocyte cell interdigitating with keratinocyte cells is partitioned into three sections. Shown within the melanocyte are the four stages of melanosome formation from budding from the Golgi apparatus, to the fully pigmented stage IV melanosomes migrating up the dendritic processes of the cell and secreted into the keratinocytes. In African populations, the melanosomes remain as singular heavily pigmented particles while in Asians and Europeans the melanosomes cluster in membrane bound organelles giving different skin complexions.

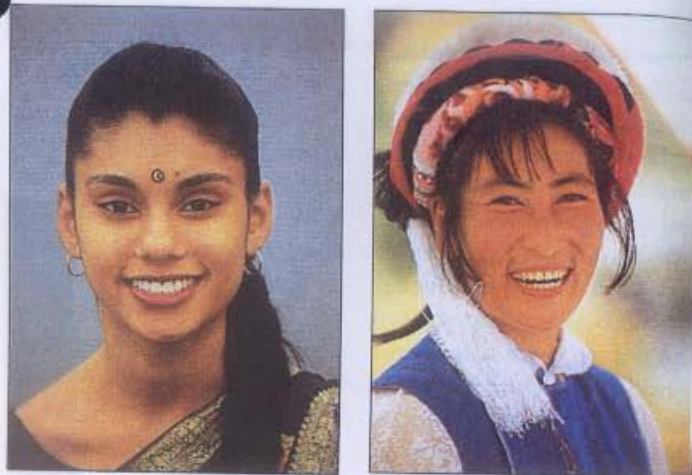
HUMAN DIVERSITY

RICHARD LEWONTIN

*Scientific American Library
1982 ISBN 07167-1469-8*



67a



A. B.

Figure 14-4 People from different geographic regions vary in such traits as height, blood type, skin color, and facial features. A. An Indian woman. B. An ethnic minority woman from China. (A, Superstock; B, Paul Greblunas/Tony Stone Images)

What Are Human "Races"?

We all know that the human species varies in appearance and in physiology geographically. People of Indian descent, for example, are recognizably different from those of Chinese descent. We can look at the shape of the nose, the eyes, and the ears, for example, and we see differences (Figure 14-4). We can look at the distribution of individual alleles and see differences. Fifty-one percent of Nigerians have type O blood, compared to only 30 percent of Japanese. Twenty percent of Russians have type B blood, while the Amerindians of Lima, Peru, have no detectable levels of type B blood at all.

Nineteenth-century anthropologists struggled to classify human groups into a few major races. Some systems identified only 12 races, while other systems listed 30 or more. One problem was that no matter how anthropologists classified humans, there always seemed to be tribes or nations that would not fit into any known group. The Basques, who live in the Pyrenees mountains between France and Spain, for example, appear European. Yet their language and culture are unlike any other in the world, and some researchers once argued that they are direct descendants of Stone Age Europeans. Similarly, the Bushmen are unique among African groups in both appearance and physiology (Figure 14-5).

A more serious problem with the grouping of humans into races is that most groups do *not* stand out from those around them; they blend. Because groups of humans inevitably mix, through migration, warfare, and trade, human "races" are never pure. Both the Japanese and British, for example, take pride

we will learn all about HUMAN VARIATION FROM GENOME SEQUENCING!

ARE THERE HUMAN RACES or HUMAN BEINGS?

DNA TESTING CAN provide information on Human origins - geographical groupings + Ancestry!

in the purity of their island races. Yet the Japanese are a grade mixture of Korean and Ainu north islanders (a people of possibly European descent). This mix shows up in the distribution of blood types from one end of Japan to the other (Figure 14-6).

The British are even more of a melting pot than the Japanese. The Bronze Age Beaker Folk mixed with the Indo-European Celts in the first thousand years B.C. In the next thousand years, the Angles, the Saxons, the Jutes, and the Picts arrived, followed by the Vikings and their descendants, the



A. B.

Figure 14-5 Classifying humans into a few discrete races has been unsuccessful. Named races frequently include markedly distinct peoples. A. A Bushman from Namibia. B. A rubber plantation foreman from the Ivory Coast. (A, M.P. Kahl/Photo Researchers, Inc.; B, Charles O. Cecil/Visuals Unlimited)

Genetic Variation = Alleles @ a given locus

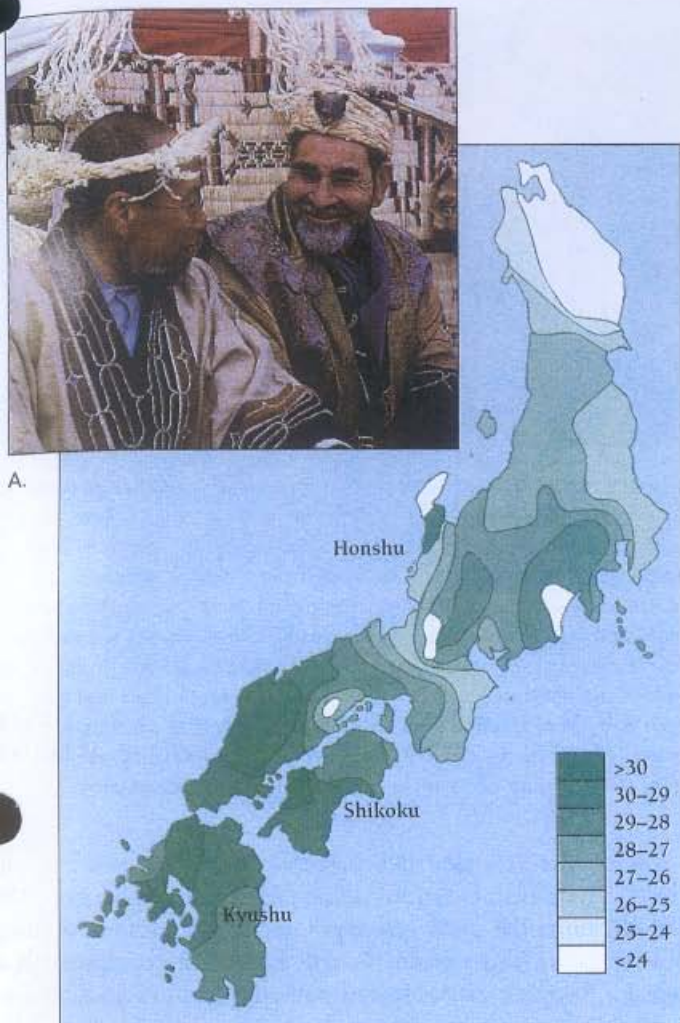


Figure 14-6 Variation in blood group A in modern Japan. A. Even the Japanese, a relatively homogeneous island race, show variation that demonstrates intermarriage with other groups. Populations closest to mainland Korea have the greatest concentration of A alleles. Populations closest to Hokkaido have the lowest concentrations of A alleles. B. In Hokkaido lives a remnant population of Ainu, sometimes called the “hairy Ainu.” The Ainu, who were once a white minority in Japan, are considered primitive by other Japanese. The Ainu had white skin, abundant body hair, and round eyes. Descendants of Causasoid peoples from northern Asia, the Ainu once lived on all four of Japan’s major islands. Today they live only in the north. Intermarriage and cultural assimilation by the Japanese have made the traditional Ainu virtually extinct. (A, Superstock; B, from *Human Diversity* by Lewontin © 1982 by Scientific American Library. Used with permission of W.H. Freeman and Company)

racess—in the biological sense—do not exist. This is because, in humans, genetic variation within populations is greater than that between nations or races. Of all human genetic variation, 85 percent is variation among the individuals within a country or a continent. Another 6 percent is variation among populations from the same continent. Only about 9 percent of all DNA reflects genetic differences among peoples (“races”) from different continents. The greatest genetic variation among humans is found in Africa. If a disaster killed everyone on Earth except for those living in Africa, the human species would still retain at least 91 percent of its genetic diversity. Less than 9 percent of all genetic diversity would be lost.

In practice, we focus on the small numbers of traits (for example, skin color and eye shape) that vary geographically—a subset of the 9 percent—and we use just those few traits to identify groups of people who differ from us more in their cultural practices than in their genetic constitutions.

Recent work, for example, shows that many differences in skin and hair color reflect variations in a single gene. That gene specifies the structure of a protein, called MC1R. MC1R affects our response to a hormone that regulates the balance of different types of skin pigments (melanins). Caucasians who tan well, for example, have the same form of MC1R that is present in Africans.

Researchers suggest that variations in skin color represent adaptations to different amounts of sunlight. A mere handful of variants are probably responsible for the most obvious traits that have historically been used to distinguish human “races.” As geneticists Mary Claire King and Kelly Owens state, “The myth of major genetic differences across “races” is . . . worth dismissing with genetic evidence.”

As we consider some peculiarities of Brazilian history and social structure, we can construct a model to explain why Color should indeed be a poor predictor of African ancestry at an individual level. Most Africans have black skin, genetically determined by a very small number of genes that were evolutionarily selected (5, 6). Thus, if we have a social race identification system based primarily on phenotype, such as occurs in Brazil, we classify individuals on the basis of the presence of certain alleles at a small number of genes that have impact on the physical appearance, while ignoring all of the rest of the genome. Assortative mating based on Color, which has been shown by demographic studies to occur in Brazil (22, 33), will produce strong associations among the individual components of Color. Indeed, we detected the presence of such positive associations at highly significant levels in the Queixadinha population (data not shown). On the other hand, we expect that any initial admixture association between Color and the PSAs will inevitably decay over time. It is easy to see how this combination of social forces could produce a population with distinct Color groups and yet with similar levels of African ancestry. Let us take as an example, the historically common Brazilian mating of a white European male with a black African slave woman: the children with more physical African features would be considered black, whereas those with more European features would be considered white, even though they would have exactly the same proportion of African and European alleles. In the next generation, the light-skinned individuals would assortatively tend to marry other whites and conversely the darker individuals would marry blacks. The long-term tendency would then be for this pattern to produce a white group and a black group, which would, nonetheless, have a similar proportion of African ancestry.

Our study makes clear the hazards of equating color or race with geographical ancestry and using interchangeably terms such as white, Caucasian, and European on one hand, and black, Negro, or African on the other, as is often done in scientific and medical literature (34, 35).

Color Marker No!

69

SIMILARITY AND DIFFERENCES in Blood Group Allele Frequencies by "Race" or Population

How DOES Genetic Variation (different alleles)
VARY between & within Populations?

Examples of extreme differentiation and close similarity in blood group allele frequencies in three racial groups

different variation
but most alleles present →

SAME VARIATION →

Gene	Alleles	Caucasoid	Negroid	Mongoloid
Duffy	Fy ^a	.0300	.9393	.0985
	Fy ^a	.4208	.0607	.9015
	Fy ^b	.5492	—	—
Rhesus	R ₀	.0186	.7395	.0409
	R ₁	.4036	.0256	.7591
	R ₂	.1670	.0427	.1951
	r	.3820	.1184	.0049
	r'	.0049	.0707	0
	others	.0239	.0021	0
P	P ₁	.5161	.8911	.1677
	P ₂	.4839	.1089	.8323
Auberger	Au ^a	.6213	.6419	
	Au	.3787	.3581	
Xg	Xg ^a	.67	.55	.54
	Xg	.33	.45	.46
Secretor	Se	.5233	.5727	
	se	.4767	.4273	

Source: R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia University Press, 1974).

- ① Most alleles in ALL "races"/populations
- ② No Homozygosity at any locus
- ③ Some Allelic Frequencies differ between "Races" & some are the same! Duffy differs & X₂ is the same. why? Adaptive value.
- ④ Auberger, Xg, & Secretor Loci show how alleles vary within populations similarly & show NO between population differences.
- ⑤ Wide range of different alleles within/between "Races"

SIMILARITY & Differences in Blood Group Gene Alleles

Variation
Between Groups

Wide
variation [

Wide
variation [

Little
variation [

26-2 TABLE Examples of Extreme Differentiation and Close Similarity in Blood Group Allelic Frequencies in Three Racial Groups

Gene	Allele	POPULATION		
		Caucasoid	Negroid	Mongoloid
Duffy	Fy	0.0300	0.9393	0.0985
	Fy ^a	0.4208	0.0000	0.9015
	Fy ^b	0.5492	0.0607	0.0000
Rhesus	R ₀	0.0186	0.7395	0.0409
	R ₁	0.4036	0.0256	0.7591
	R ₂	0.1670	0.0427	0.1951
	r	0.3820	0.1184	0.0049
	r'	0.0049	0.0707	0.0000
	Others	0.0239	0.0021	0.0000
	P			
	P ₁	0.5161	0.8911	0.1677
	P ₂	0.4839	0.1089	0.8323
Auberger	Au ^a	0.6213	0.6419	No data
	Au	0.3787	0.3581	No data
Xg	Xg ^a	0.67	0.55	0.54
	Xg	0.33	0.45	0.46
Secretor	Se	0.5233	0.5727	No data
	se	0.4767	0.4273	No data

Source: A summary is provided in L. L. Cavalli-Sforza and W. F. Bodmer, *The Genetics of Human Populations* (W. H. Freeman and Company, 1971), pp. 724-731. See L. L. Cavalli-Sforza, P. Menozzi, and A. Piazza, *The History and Geography of Human Genes* (Princeton University Press, 1994), for detailed data.

We all have the SAME Genes
but MAY differ in Frequencies
of some alleles/genes/subs

Differences in ALLELIC Frequencies in VARIOUS Populations

24-5 TABLE Frequencies of the Alleles I^A , I^B , and i at the ABO Blood Group Locus in Various Human Populations

Population	I^A	I^B	i
Eskimo	0.333	0.026	0.641
Sioux	0.035	0.010	0.955
Belgian	0.257	0.058	0.684
Japanese	0.279	0.172	0.549
Pygmy	0.227	0.219	0.554

Source: W. C. Boyd, *Genetics and the Races of Man*. D. C. Heath, 1950.

24-1 TABLE Frequencies of Genotypes for Alleles at MN Blood Group Locus in Various Human Populations

Population	GENOTYPE			ALLELE FREQUENCIES	
	M/M	M/N	N/N	$p(M)$	$q(N)$
Eskimo	0.835	0.156	0.009	0.913	0.087
Australian aborigine	0.024	0.304	0.672	0.176	0.824
Egyptian	0.278	0.489	0.233	0.523	0.477
German	0.297	0.507	0.196	0.550	0.450
Chinese	0.332	0.486	0.182	0.575	0.425
Nigerian	0.301	0.495	0.204	0.548	0.452

Source: W. C. Boyd, *Genetics and the Races of Man*. D. C. Heath, 1950.

24-2 TABLE Frequencies of Gametic Types for MNS System in Various Human Populations

Population	GAMETIC TYPE				HETEROZYGOSITY (H)	
	MS	Ms	NS	Ns	From gametes	From alleles
Ainu	0.024	0.381	0.247	0.348	0.672	0.438
Ugandan	0.134	0.357	0.071	0.438	0.658	0.412
Pakistani	0.177	0.405	0.127	0.291	0.704	0.455
English	0.247	0.283	0.080	0.290	0.700	0.469
Navaho	0.185	0.702	0.062	0.051	0.467	0.286

Source: A. E. Mourant, *The Distribution of the Human Blood Groups*. Blackwell Scientific, 1954.

Reasons FOR Allelic Variation Between Populations

① Founder Effect / Geographical Isolation

↳ Selective mating do to geography/culture or both

② Adaptive Value

↳ Have positive effect in specific environments

e.g. H^bS Sickle-Cell Globin allele
(India, Africa, Mediterranean)

Duffy

Skin Color Genes

GEOGRAPHY &/or ADAPTIVE VALUE

But do not vary across whole genome as most loci are "neutral"

NOVEL EXAMPLE OF ALLELE ADAPTIVE VALUE

- ① FyB^{ES} (Erythroid Silent) is a Major Duffy allele in African Americans & Black populations of African Heritage. Rarely in other populations!
- ② Encodes chemokine receptor protein on blood cell membrane. Parasite receptors for Malarial PLASMODIUM bind to this receptor.
- ③ FyB^{ES} is a Switch mutation allele - FyB^{ES} allele CANNOT be transcribed ∴ No FyB or PLASMODIUM receptor protein -
- ④ No binding of Malaria PLASMODIUM - ∴ Malarial Resistance
- ⑤ High Frequency in Geographical Regions with high levels of Malaria!
- ⑥ Other Duffy alleles (Chromosome 1)
@ different frequencies in other populations

There is a Large variation in
 D1S80 VNTR alleles within
 populations but little between

Table 15.2 Allele frequencies for D1S80 among U.S. population groups

Repeat number	Caucasian	Hispanic	African American	Asian
14	0	0	0	0
15	0	0.001	0	0
16	0.001	0.010	0.002	0.034
17	0.002	0.009	0.028	0.025
18	0.237	0.224	0.073	0.152
19	0.003	0.005	0.003	0.022
20	0.018	0.013	0.032	0.007
21	0.021	0.028	0.115	0.034
22	0.038	0.024	0.081	0.017
23	0.012	0.009	0.014	0.017
24	0.378	0.315	0.234	0.230
25	0.046	0.072	0.045	0.027
26	0.020	0.007	0.006	0
27	0.007	0.016	0.008	0.047
28	0.063	0.078	0.130	0.076
29	0.052	0.055	0.053	0.042
30	0.008	0.039	0.009	0.123
31	0.072	0.053	0.054	0.093
32	0.006	0.005	0.007	0.012
33	0.003	0.004	0.004	0.005
34	0.001	0.006	0.086	0.005
35	0.003	0	0.002	0.005
36	0.004	0.011	0.001	0.005
37	0.001	0.004	0	0.007
38	0	0	0	0
39	0.003	0.004	0.003	0.005
40	0	0	0	0
41	0	0.002	0.002	0.007
>41	0.001	0.006	0.007	0.002
Sample size	718	409	606	204

Source: Data from B. Budowle, et al. 1993. *Journal of Forensic Science* 40:38

Just
 as
 predicted
 from
 class
 genotype

NO ADAPTIVE VALUE!!
 Good locus for Forensics -
 Group Neutral!

74a

There is More Genetic Diversity
within Populations than *Between*
 Populations!! So much for the
 concept of racial "parity"!!!

Proportion of genetic diversity accounted
 for within and between populations and
 races

Gene	Total H_{species}	Proportion		
		<i>within any population</i> Within Populations	Within Races between Populations	Between Races
Hp	.994	.893	.051	.056
Ag	.994	.834	—	—
Lp	.639	.939	—	—
Xm	.869	.997	—	—
Ap	.989	.927	.062	.011
6PGD	.327	.875	.058	.067
PGM	.758	.942	.033	.025
Ak	.184	.848	.021	.131
Kidd	.977	.741	.211	.048
Duffy	.938	.636	.105	.259
Lewis	.994	.966	.032	.002
Kell	.189	.901	.073	.026
Lutheran	.153	.694	.214	.092
P	1.000	.949	.029	.022
MNS	1.746	.911	.041	.048
Rh	1.900	.674	.073	.253
ABO	1.241	.907	.063	.030
Mean		.854	.083	.063

*More
genetic
diversity
within any
population
than
between
populations!*

*of
same
Race*

Source: R. C. Lewontin, *Genetic Basis of Evolutionary Change* (Columbia University Press, 1974).

- ① 85% of Human Genetic Variation occurs *within* Populations + Between Individuals in that population!
- ② Remaining 15% of Human Genetic Variation Split Between different Populations of same "race" (8%) + Between different "Races" (6%).
- ③ *only 6% of Human Genetic Variation due to differences between RACES!!*

The SAME ALLELES are present
in most populations but
at different frequencies

Mapping Human History

Mary-Claire King and Arno G. Motulsky

The DNA of modern humans contains a record of the travels and encounters of our ancestors. The genotypes of people living today are the result of ancient human migrations, the continuous appearance of new mutations, selection by climate and infection for genetic alleles that conferred a survival advantage, and mating patterns determined by cultural norms. By sampling genotypes from people across the globe, geneticists have reconstructed the major features of our history: our ancient African origin, migrations out of Africa, movements and settlements throughout Eurasia and Oceania, and peopling of the Americas (1–5). As genomic technology has improved, these analyses of genotype have successively incorporated new markers: blood groups (2), protein polymorphisms (2), mitochondrial DNA sequences (1), Y chromosome haplotypes (3), and highly variable nuclear microsatellite markers (4, 5).

The most recent contribution to this literature is by Rosenberg *et al.* (6) on page 2381 of this issue. These investigators explored the genetic structure of human pop-

ulations using highly variable markers on the human autosomes of individuals from different parts of the world. The genotyped markers were microsatellite short tandem repeat sequences that do not encode any expressed genes and are generally selectively neutral. The populations studied were defined by geography, language, and culture, and participating individuals were well rooted in their populations, with several generations of ancestors known to have lived in the same locale as the participant. Genotypes from more than a thousand individuals were evaluated by a statistical method that defines clusters of people on the basis of genetic similarity at multiple loci, without using prior information about ancestry. In this method, individuals are assigned to clusters probabilistically (5, 7). Individuals may have significant probabilities of membership in more than one cluster due either to genetic similarities of groups or to ancestral intergroup matings. The world map (see the figure) illustrates variation at one microsatellite marker in 12 populations. This marker has four common alleles, each of which appears in all populations. Rare alleles are shared by fewer populations. Few alleles are unique to only one population. No allele is population specific.

Previous genetic analyses of human

history have consistently suggested that most human genetic variation is due to differences among individuals within populations rather than to differences among populations (4, 8). The Rosenberg *et al.* analysis of many more markers and many more people confirms this result: 93 to 95% of genetic variation is due to genetic differences among individuals who are members of the same population and only 3 to 5% of genetic variation is due to differences among the major population groups.

The power of the method lies in the construction of clusters on the basis of accumulated small differences in allele frequencies across many markers and many people. Statistical clustering of genotypes—composed of 4682 alleles from 377 markers in 1056 individuals from 52 populations—yields groups corresponding to major geographic regions of the world [see figure 1 in (6)]. Creation of two clusters reflects ancient human origins in Africa and rapid expansion throughout Eurasia, and migrations to the Americas from East Asia. Creation of five clusters yields groups corresponding to five major geographic regions of the world: Africa, Eurasia (Europe, the Middle East, Central and South Asia), East Asia, Oceania, and America. There is excellent agreement between membership of individuals in these clusters and their self-identified regions of origin. Similar results were obtained by the same statistical approach based on fewer populations and fewer markers [Table 2 of (5)].

The authors are in the Department of Medicine (Medical Genetics) and Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. E-mail: mcking@u.washington.edu

We learn about our origins
& diversity from the Human
Genome Project

Population substructure could be consistently identified within some geographic regions but not others. Within Africa, for example, analysis consistently yielded the same four subclusters: Mbuti Pygmies, Biaka Pygmies, San peoples, and speakers of Niger-Kordofanian languages (Bantu, Yoruba, and Mandenka populations). In contrast, within Europe, multiple analyses were not consistent. Many more individuals will need to be included to sort out European demographic history.

without discontinuities between clusters. After thousands of years—if enough markers and people are studied—allele frequency differences are collectively adequate to create clusters that correspond to the major migrations of human history.

What are the implications of the Rosenberg *et al.* findings for medicine? The current medical literature increasingly includes studies exploring population differences in disease incidence or in efficacy or adverse responses to drug treatment (5,

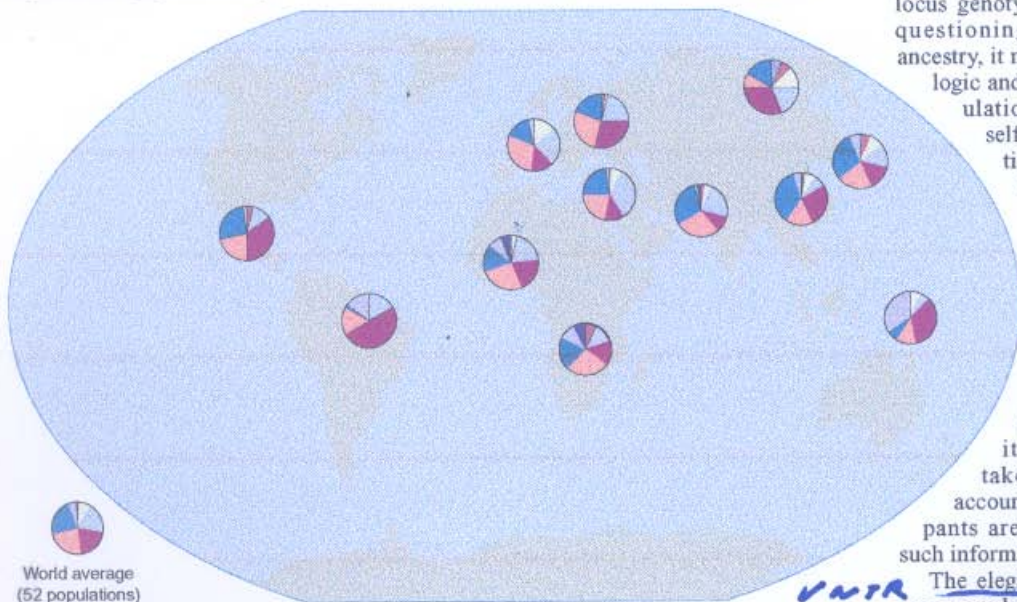
to dictate medical management for the entire group. Instead, critical alleles influencing disease risk or response to treatment are likely to be either ancient, worldwide, and relatively common in many populations, or geographically localized and individually rare (10).

Differences among populations in disease frequency and treatment outcome certainly occur but may not be genetic in origin. Given that the major population origin of groups can be defined by multilocus genotype clustering (5–7) without questioning individuals about their ancestry, it may be tempting in epidemiologic and clinical studies to omit population characterization through self-reporting. However, correlations identified by the clustering method may be falsely ascribed to genes when in fact they had nothing to do with genetics but were caused by social, economic, or discriminatory factors limited to a genetically defined population cluster (11). To evaluate medically important group differences, it is therefore necessary to take all such risk factors into account. Patients and study participants are usually the best source of such information.

The elegant statistical analysis of human population structure by Rosenberg and colleagues reflects the major human migrations out of Africa, into Europe, across Asia, into Oceania, and to the Americas. By genotyping a large sample of an individual's alleles, it is possible to identify the migrations in which his or her ancestors participated. But the link between historical genetic demography and medically important risk is complex. Disease susceptibility may be genetic but not geographically clustered, or geographically clustered but not genetic, or neither, or both.

References

1. A. C. Wilson, R. L. Cann, *Sci. Am.* 266, 68 (April 1992).
2. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
3. P. A. Underhill *et al.*, *Nature Genet.* 26, 358 (2000).
4. G. Barbujani, A. Magagni, E. Minch, L. L. Cavalli-Sforza, *Proc. Natl. Acad. Sci. U.S.A.* 94, 4516 (1997).
5. J. F. Wilson *et al.*, *Nature Genet.* 29, 265 (2001).
6. N. A. Rosenberg *et al.*, *Science* 298, 2381 (2002).
7. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* 155, 945 (2000).
8. R. C. Lewontin, *Evol. Biol.* 6, 381 (1972).
9. H. G. Xie, R. B. Kim, A. J. Wood, C. M. Stein, *Annu. Rev. Pharmacol. Toxicol.* 41, 815 (2001).
10. J. K. Pritchard, N. J. Cox, *Hum. Mol. Genet.* 11, 2417 (2002).
11. N. Risch, E. Burchard, E. Ziv, H. Tang, *Genome Biol.* 3, comment2007 (2002).



World average
(52 populations)

Humans on the move. Worldwide genetic variation at a neutral marker. Allele frequencies of one randomly chosen microsatellite marker reveal common alleles shared in all populations and the gradual and arbitrary differences in allele frequencies across geographic regions. Populations shown in this example are Yoruba and Bantu (Africa); French, Russians, Palestinians, and Pakistani Brahui (Eurasia); Han Chinese, Japanese, and Yakut (East Asia); New Guineans (Oceania); and Maya and Karitians (America). Each color on the diagrams represents one of nine alleles of GGAA29H03 (D13S1493), which range in length from 219 to 255 base pairs. By accumulating small differences in allele frequencies from hundreds of such highly variable markers and hundreds of people, statistical methods reveal genetic clusters of Africans, Eurasians, East Asians, Pacific Islanders, and Americans, corresponding to major ancient human migrations (6).

The identification of clusters corresponding to the major geographic regions may depend on the sampling of individuals from well-defined, relatively homogeneous populations. If individuals were sampled from a worldwide "grid" (or a worldwide grid weighted by population density), the clusters might be much less precisely defined. Does the correspondence of worldwide genetic clusters and major geographic regions suggest borders around genetic clusters analogous to the physical borders—oceans, mountain ranges, and deserts—separating geographic regions? No. Both the results of Rosenberg and colleagues and those of previous studies (1–5, 8) indicate that unlike separations between geographic regions, differences in allele frequencies are gradual,

9). The rationale of these studies is that alleles influencing disease susceptibility or treatment response may differ in frequency across populations. Consequently, individuals would be better served if critical genotypes were taken into account when assessing disease risk or designing treatment regimens. In the absence of knowing the identities of the critical alleles, personal ancestry as indicated by study participants is often used as an initial but potentially misleading substitute.

The Rosenberg *et al.* data suggest that with the exception of ancient highly selected loci (for example, the Duffy null blood group, which confers complete protection against vivax malaria), very few alleles will be both confined to one population and common enough in that population

CONCLUSION

① If 85% of Human Genetic Variation occurs between different people within any given population (localized)

② If only 6% of Human Genetic Variation occurs between "RACES" (Novel Alleles Specific to a "race") → e.g., $Fy0^{ES}$

③ Then Losing all "races" except one retains 94% of all Human Genetic Variation!

$$[85\% + (15\% - 7\%)] = 94\%$$

85% within population genetic variability

8% between populations of same "race"

7% between "race" genetic variability

④ ∴ Humans highly Heterozygous or Hybrids —
* if above not true — most of us would not
be here — need genetic variation to survive!

So what is a "Race"?

- ① Primarily a sociological concept — but could be a localized or inbred population that has a higher frequency of alleles at a very small number of loci. Affects few physical features.
- ② High frequency alleles in one "race" are present at lower frequency in other "races."
ALL humans have SAME genes — differ in form
Mostly within populations!
- ③ Heterozygosity (variation) high in human populations — ALL populations. None homozygous at all loci!
- ④ NO such thing as a "pure" Race — would have little variation —
- ⑤ Genes Affecting Physical Features NOT representative of genes across genome —

Geographical Ancestry is Relevant — Many "racial" groups now have multiple ancestries because of admixture & migration

∴ "Race" Classification is **Arbitrary**
& BASED on a few traits — NOT
Science BASED

CAN DEFINE BY MANY CRITERIA....

RACE BY RESISTANCE

Traditionally we divide ourselves into races by the twin criteria of geographic location and visible physical characteristics. But we could make an equally reasonable and arbitrary division by the presence or absence of a gene, such as the sickle-cell gene, that confers resistance to malaria.

By this reckoning, we'd place Yemenites, Greeks, New Guineans, Thai, and Dinkas in one "race," Norwegians and several black African peoples in another.

People who possess anti-malarial genes (alleles!) & those that do not

RACE BY DIGESTION

We could define a race by any geographically variable trait—for example, the retention in adulthood of the enzyme lactase, which allows us to digest milk. Using this as our divisive criterion, we can place northern and central Europeans with Arabians and such West African peoples as the Fulani; in a "lactase-negative race," we can group most other African blacks with east Asians, American Indians, southern Europeans, and Australian aborigines.

People who possess lactase & those that do not!

RACE BY FINGERPRINTS

Probably the most trivial division of humans we could manage would be based on fingerprint patterns. As it turns out, the prevalence of certain basic features varies predictably among peoples: in the "Loops" race we could group together most Europeans, black Africans, and east Asians. Among the "Whorls" we could place Mongolians and Australian aborigines. Finally, in an "Arches" race, we could group Khoisans and some central Europeans.

People who possess loops & those that have arches!

RACE BY GENES

One method that seems to offer a way out of arbitrariness is to classify peoples by degree of genetic distinctness. By this standard the Khoisans of southern Africa would be in a race by themselves. African blacks would form several other distinct races. All the rest of the world's peoples—Norwegians, Navajo, Greeks, Japanese, Australian aborigines, and so on—would, despite their greatly differing external appearance, belong to a single race.

People who possess one form of an allele (localized inbred population) & those that do not

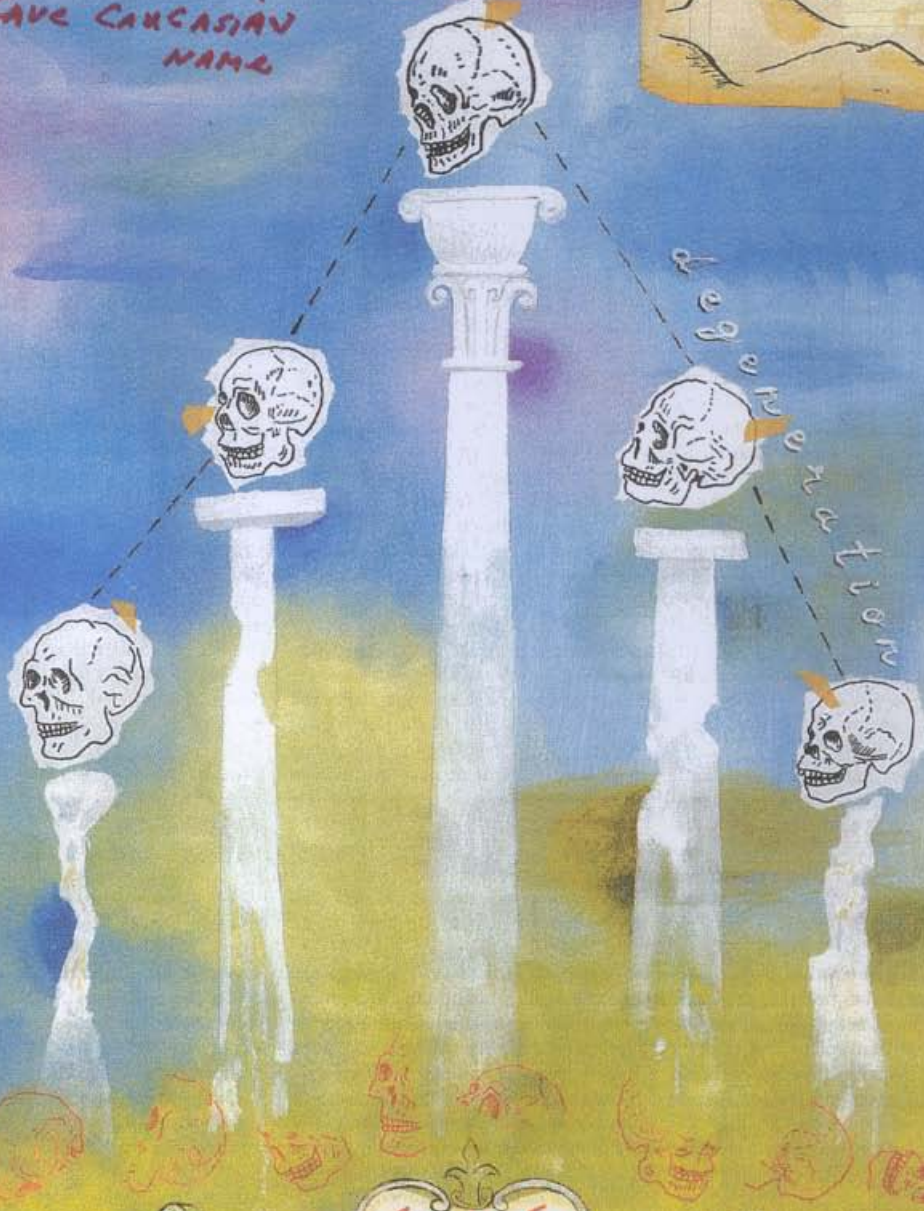
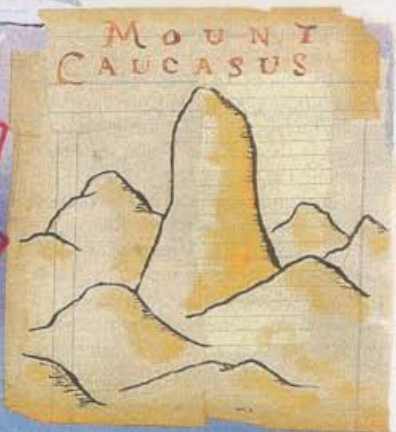
RACIAL CLASSIFICATION DIDN'T COME FROM SCIENCE BUT FROM THE BODY'S SIGNALS FOR DIFFERENTIATING ATTRACTIVE FROM UNATTRACTIVE SEX PARTNERS, AND FRIEND FROM FOE.

Physical traits

HOW DID WE GET TO WHERE WE ARE?

JOHAN FREIDRICH BLUMENBACH

INVENTOR OF MODERN
"RACIAL" CLASSIFICATION
CAVE CAUCASIAN
NAME



1795

JOHAN FREIDRICH BLUMENBACH

- ① Changed Linnaean neutral-geography-based Hermon classifications to a value-based classification—
- ② Said all "races" originated from one place/origin — in Europe around Mt. CAUCASUS — because they are the most "beautiful" race — even though he felt all races were equal in all respects & argued with people that didn't!

③ Classified five "races"

CAUCASIAN
Mongolian
Ethiopian
American
Malay

Value-based Classification & putting CAUCASIANS on top — most beautiful — had disastrous consequences!

CONCLUDE

- ① Races are Arbitrary Entities - Social Constructs that are Culturally generated.
- ② Yes - there is genetic differences between "races" or relatively inbred geographical populations that can lead to physical & other differences - due to very small # genes & not reflective of whole genome
- ③ Within population genetic variation Much greater than between population genetic variation - Many loci have same allele frequencies - some differ
- ④ Only minor differences between genomes of different people or groups of people - Unity >>> Differences!
- ⑤ We are all the same - but different.
you now know why!

Race classification is Arbitrary, unscientific, & divisive!