

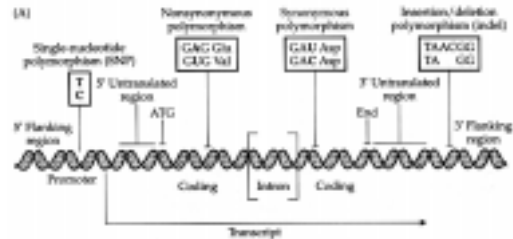
nucleic acid variation

- indirect
 - DNA/DNA hybridization
 - enzymatic degradation
 - restriction enzyme digestion (RFLP)
 - amplification polymorphism (RAPD)
 - repeat sequence variation (VNTR / STR)
- direct
 - sequencing

DNA POLYMORPHISMS

- DIRECT SEQUENCE DETERMINATION
 - CLONING
 - ORGANELLE GENOMES
 - PCR AMPLIFICATION OF TARGET GENES / SEQUENCING

GENE STRUCTURE



DNA SEQUENCES

- Sequencing procedures developed by
 - Sanger - dideoxy chain termination
 - Maxam & Gilbert - chemical termination
- Most sequencing now performed by dideoxy sequencing

DIDEOXY NUCLEOTIDES

ATCGGGCTATC

DIDEOXY NUCLEOTIDES

ATCGGGCTATC
A*.....
AT*.....
ATC*.....
ATCG*.....
ATCGG*.....
ATCGGG*.....
ATCGGGC*.....
ATCGGGCT*.....
ATCGGGCTA*

DIDEOXY NUCLEOTIDES

ATCGGGCTATC
A*.....
ATCGGGCTA*....

DIDEOXY NUCLEOTIDES

ATCGGGCTATC
AT*.....
ATCGGGCT*.....
ATCGGGCTAT*..

DIDEOXY NUCLEOTIDES

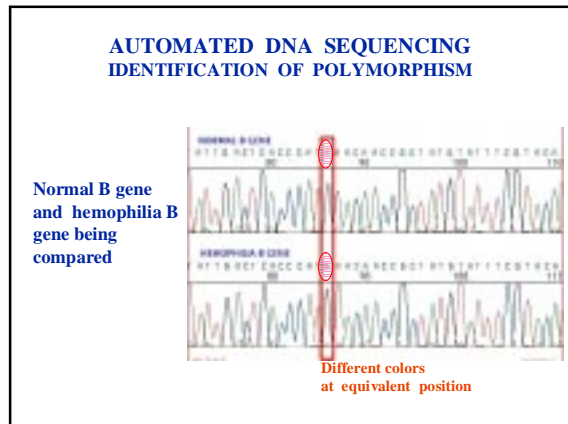
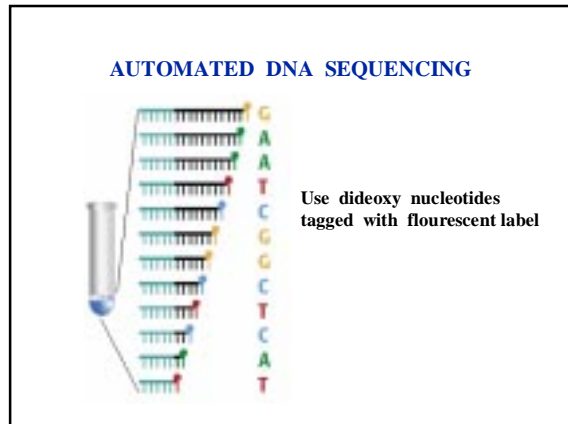
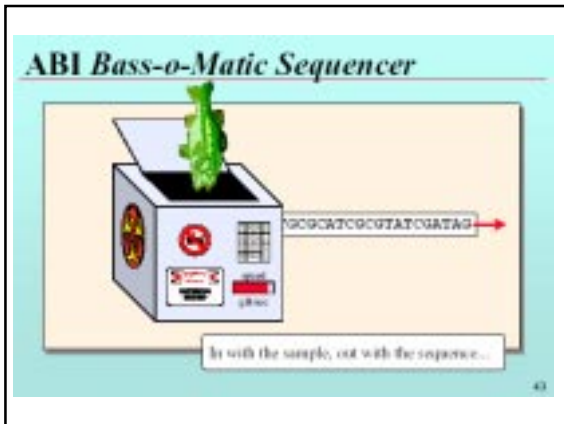
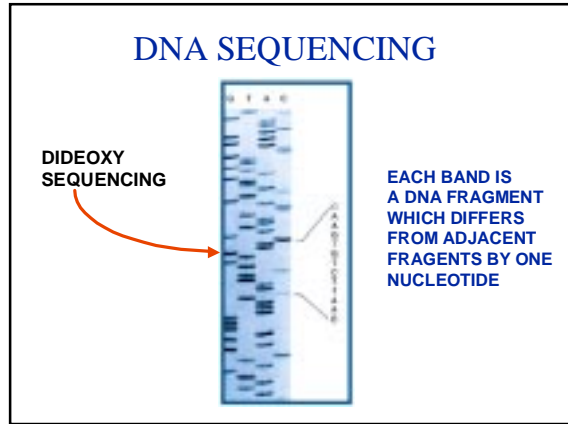
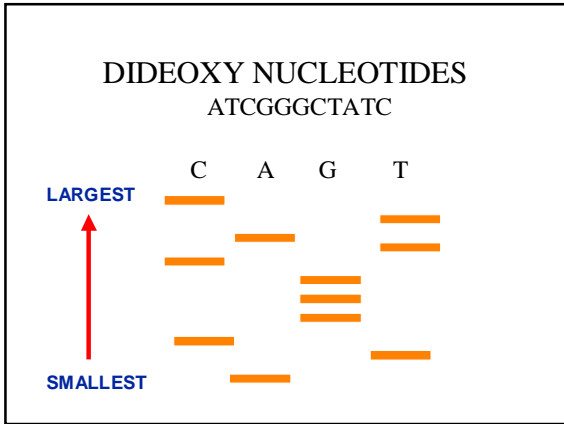
ATCGGGCTATC
ATC*.....
ATCGGGC*.....
ATCGGGCTATC*

DIDEOXY NUCLEOTIDES

ATCGGGCTATC
ATCG*.....
ATCGG*.....
ATCGGG*.....

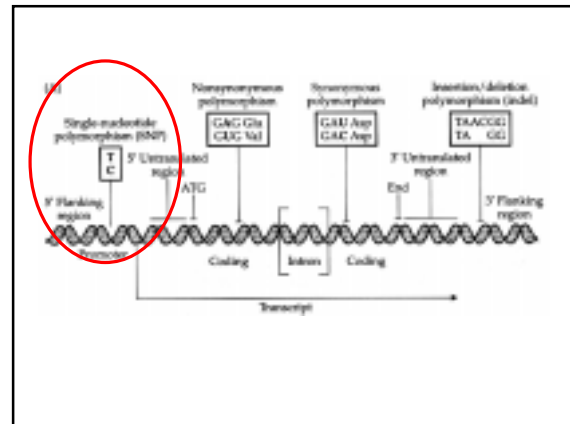
RESULT

- Set of DNA fragments which terminate on a particular nucleotide
 - Each fragment has a specific length
 - read the barcode to determine the sequence



POLYMORPHISM IN DNA SEQUENCE

- DIRECT COMPARISON OF SEQUENCES
 - SNP - Single Nucleotide Polymorphisms



Types of polymorphism

- SNP
 - Non-synonymous
 - Amino acid substitution
 - Synonymous
 - Silent

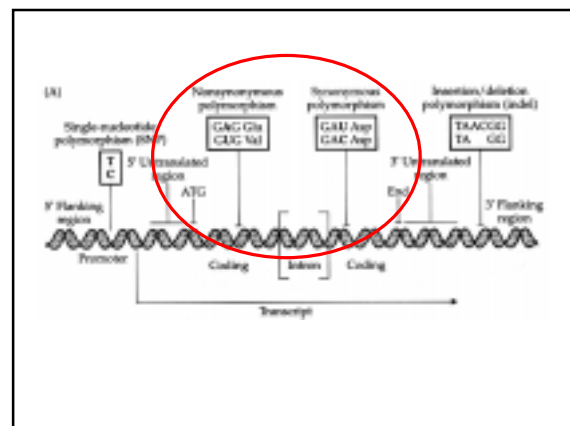
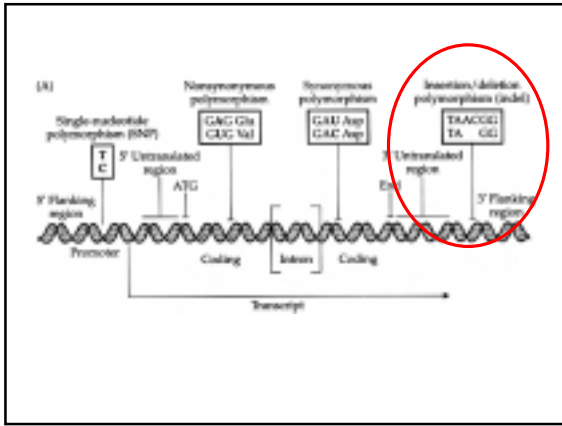


TABLE 8.2 DEGENERACY IN THE GENETIC CODE

		Second nucleotide in codon			
		T	C	A	G
T	T	TTY Phe TTR Leu	TCN Ser	TAY Tyr TAR Stop	TGY Cys TGA Stop TGG Trp
	C	CTN Leu	CCNPro	CAY His CAR Gln	CGN Arg
A	A	ATH Ile ATG Met	ACN Thr	AAY Asn AAR Lys	AGY Ser AGR Arg
	G	GTN Val	GCN Ala	GAY Asp GAR Glu	GGN Gly

Types of polymorphism

- SNP
 - Non-synonymous
 - Amino acid substitution
 - Synonymous
 - Silent
- INDEL
 - Insertion or deletion
 - Sometimes transposable element



nucleotide variation in a portion of the vertebrate 28S rRNA gene

not aligned

(A)	22 00	22 20	22 40
Mus	GTCAGCCAGGACTCTCTACCCGCTACGGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rattus	GTCAGCCAGGACTCTCTACCCGCTACGGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Homo	GTCAGCCAGGACTCTCTACCCGCTACGGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rhinorana	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Caciatua	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Xenopus	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rhizactriton	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Typhlonectes	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Lalamera	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Cyprinella	GTCAGTCAGGATTCCTCTACCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		

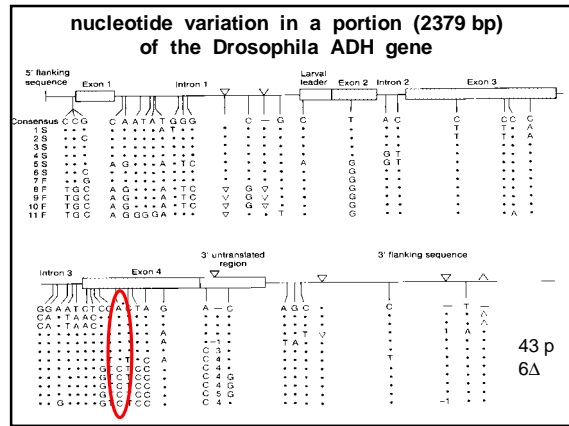
nucleotide variation in a portion of the vertebrate 28S rRNA gene

not aligned

(A)	22 00	22 20	22 40
Mus	GTCAGCCAGGACTCTCTACCCGCTACGGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rattus	GTCAGCCAGGACTCTCTACCCGCTACGGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Homo	GTCAGCCAGGACTCTCTACCCGCTACGGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rhinorana	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Caciatua	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Xenopus	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rhizactriton	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Typhlonectes	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Lalamera	GTCAGCCAGGATTCCTCTATCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Cyprinella	GTCAGTCAGGATTCCTCTACCCGCTCGCGCAAGGCTTCCTGCCCGTACCGGAGGCAAC		

aligned

(B)	22 00	22 20	22 40
Mus	GTCAG-CCAGGACTCTCTACCCGCTACGG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rattus	GTCAG-CCAGGACTCTCTACCCGCTACGG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Homo	GTCAG-CCAGGACTCTCTACCCGCTACGG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rhinorana	GTCAG-CCAGGATTCCTCTATCCGCTCGCG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Caciatua	GTCAG-CCAGGATTCCTCTATCCGCTCGCG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Xenopus	GTCAG-CCAGGATTCCTCTATCCGCTCGCG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Rhizactriton	GTCAG-CCAGGATTCCTCTATCCGCTCGCG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Typhlonectes	GTCAG-CCAGGATTCCTCTATCCGCTCGCG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Lalamera	GTCAG-CCAGGATTCCTCTATCCGCTCGCG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		
Cyprinella	GTCAGTCAGGATTCCTCTACCCGCTCGCG-CAAGGCTTCCTGCCCGTACCGGAGGCAAC		



DNA SEQUENCE POLYMORPHISM

- level of variability (Drosophila)
 - % polymorphic nucleotide sites: ~ 0.03 – 1.0%
 - heterozygosity = 0.01 - 1.0 % / nuc site average ~ 0.04%

NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com

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

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. 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.

DNA SEQUENCE POLYMORPHISM

- SNPs in the Human Genome Project
 - Map of 1.42 million SNPs
 - 1 SNP / 1.9 kbp in the genome

DNA SEQUENCE POLYMORPHISM

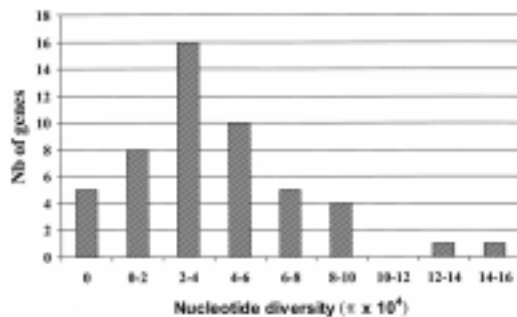
- SNPs in the Human Genome Project
 - Map of 1.42 million SNPs
 - 1 SNP / 1.9 kbp in the genome
 - Examination of a subset of coding regions
 - ~ 2 exonic SNPs per gene
 - (1 SNP / 1.08 kb) estimate may be biased
 - Total SNPs in genes = ~ 60,000

DNA SEQUENCE POLYMORPHISM

- level of variability - human genome
 - % polymorphic nucleotide sites:
 - heterozygosity (π) =
 - 7.5×10^{-4} (autosomes)
 - (1 SNP every 1,331 bp)
 - 4.5×10^{-4} (X chromosome)
 - 1.5×10^{-4} (Y chromosome)

SNP variation

- Distribution within a genome

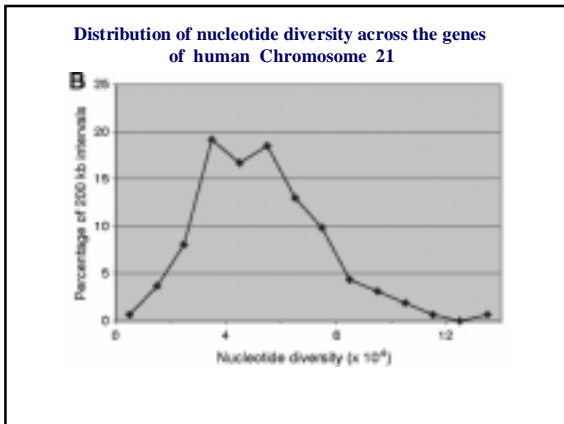
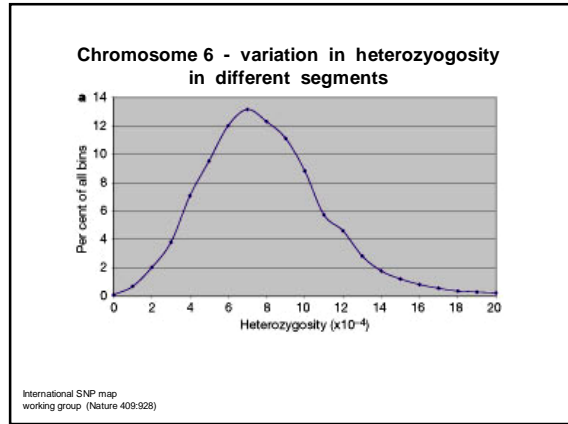
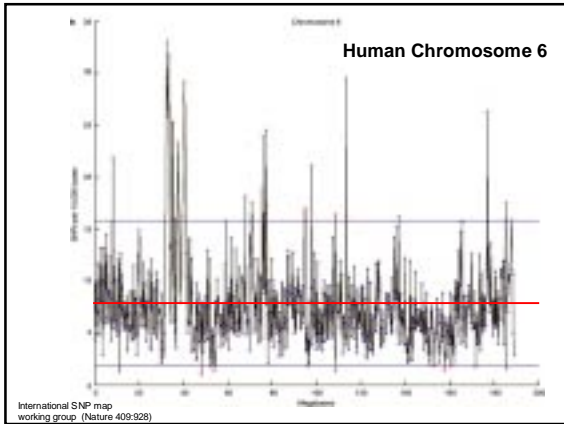


Distribution of nucleotide diversity
across 50 random genes.

Laurence Turet, et al. Human Molecular
Genetics, 2002, Vol. 11, No. 4 419-429

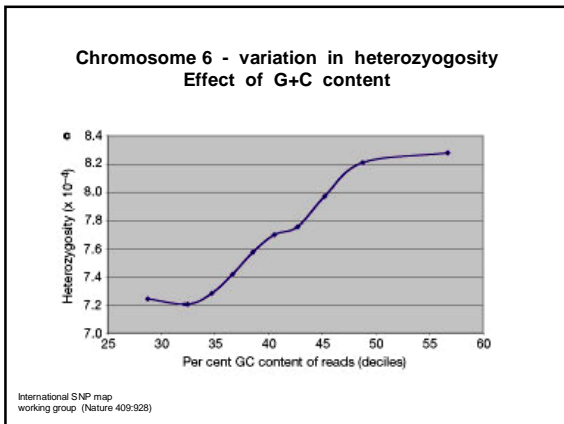
SNP variation

- Distribution within a genome
 - Distribution within a chromosome



SNP variation

- Distribution within a genome
 - Distribution within a chromosome
 - Effects of sequence composition



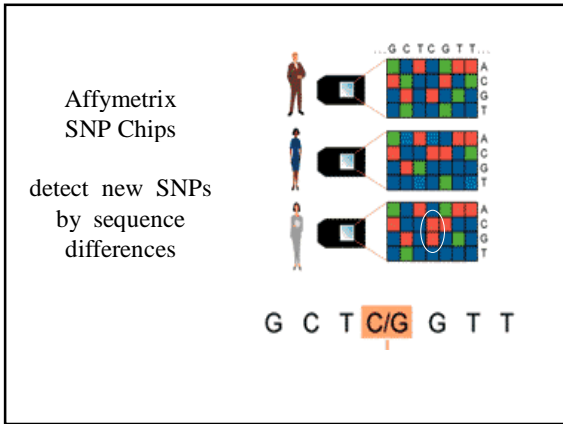
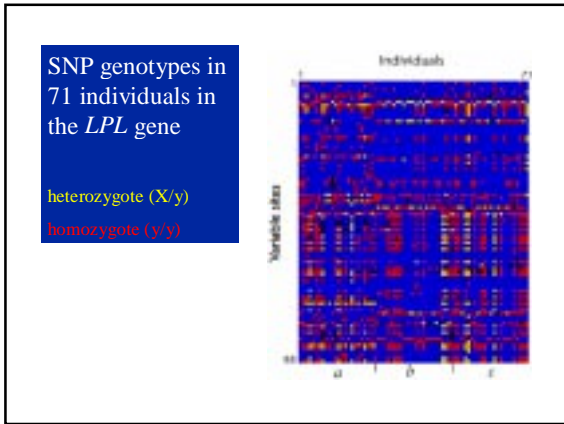
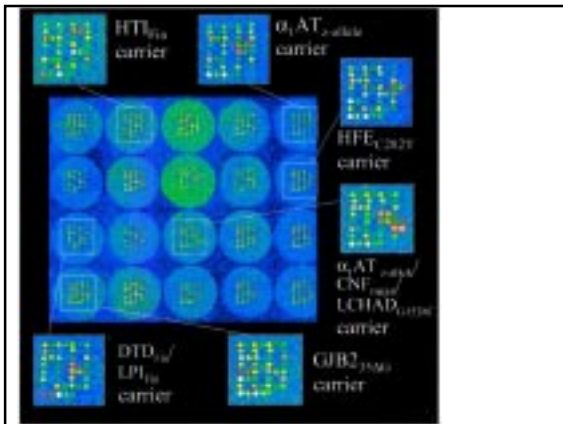
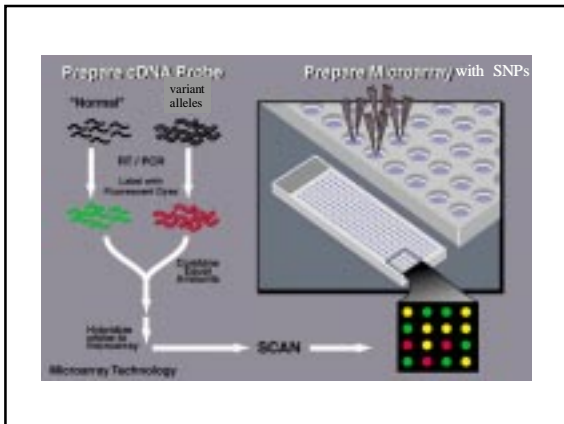
SNPs

- Future: ability to correlate variation at the genome level to specific continuous (polygenic / quantitative) traits



SNP polymorphisms studied using chip technology

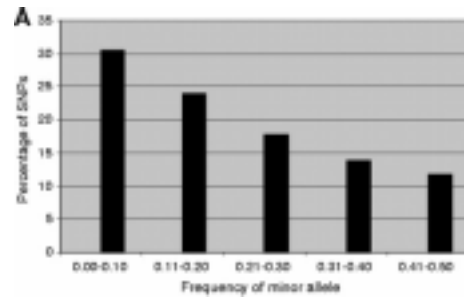
- DNA microarray technology (DNA chips)
 - Automated PCR and hybridization procedures for rapid identification of sequences and transcripts



SNP variation

- Distribution within a genome
 - Distribution within a chromosome
 - Effects of sequence composition
- How variable are SNPs?

Chromosome 21



Are SNPs in one population group also found in another population?

