Lecture 28: Polymorphisms in Human DNA Sequences

- SNPs
- SSRs

The methods of genetic analysis that you have been learning are applicable to mammals — even to humans. However, we need to combine these genetic principles with an understanding of the physical realities of the human genome. To genetics we will add genomics.

Eukaryotic Genes and Genomes

genome = DNA content of a complete haploid set of chromosomes
= DNA content of a gamete (sperm or egg)

Species	Chromosomes	cM	DNA content/ haploid(Mb)	year sequence completed	genes/ haploid
E. coli	1	N/A	5	1997	4,200
S. cerevisiae	16	4000	12	1997	5,800
C. elegans	6	300	100	1998	19,000
D. melanogaster	4	280	180	2000	14,000
M. musculus	20	1700	3000	2002 draft 2005 finished?	30,000?
H. sapiens	23	3300	3000	2001 draft 2003 finished	30,000?

Note: cM = centi Morgan = 1% recombination
Mb = megabase = 1 million base-pairs of DNA
Kb = kilobase = 1 thousand base-pairs of DNA

Let's add some columns to a table we constructed several lectures back:

Species	cM	DNA content/ haploid (Mb)	generation time	design crosses?	true breeding strains?
E. coli	N/A	5	30 min	yes	yes
S. cerevisiae	4000	12	90 min	yes	yes
C. elegans	300	100	4 d	yes	yes
D. melanogaster	280	180	2 wk	yes	yes
M. musculus	1700	3000	3 mo	yes	yes
H. sapiens	3300	3000	20 yr	no	no

You might add a column indicating the number of offspring per adult. What are the implications of this table for human genetic studies? Obviously they're difficult.

More specifically:

- Human genetics is retrospective (vs prospective). Human geneticists cannot test hypotheses prospectively. The mouse provides a prospective surrogate.
- Can't do selections
- Meager amounts of data Human geneticists typically rely upon statistical arguments as opposed to overwhelming amounts of data in drawing connections between genotype and phenotype.
- Highly dependent on DNA-based maps and DNA-based analysis

The unique advantages of human genetics:

- A large population which is self-screening to a considerable degree
- Phenotypic subtlety is not lost on the observer
- The self interest of our species

Let's consider the types and frequency of polymorphisms at the DNA level in the human genome. DNA polymorphisms are of many types, including substitutions, duplications, deletions, etc. Two types of DNA polymorphisms are of particular importance in human genetics today:

A locus is said to be polymorphic if two or more alleles are each present at a frequency of at least 1% in a population of animals.

1) SNPs = single nucleotide polymorphisms = single nucleotide substitutions

In human populations:

$\mathbf{H_{nuc}}$ = average heterozygosity per nucleotide site = 0.001

This means that, on average, at a randomly selected locus, two randomly selected human alleles (chromosomes) differ at about 1 nucleotide per 1000. This implies that your maternal genome (the haploid genome that you inherited from your mother) differs from your paternal genome at about 1 nucleotide per 1000.

Similarities and differences: This also implies that the genomes of any two individuals are 99.9% identical. Conversely, the genomes of two randomly selected individuals will differ at several million nucleotides. (Identical twins are a notable exception.)

The great majority (probably 99%) of SNPs are selectively “neutral” changes of little or no functional consequence:

- outside coding or gene regulatory regions (>97% of human genome)
- silent substitutions in coding sequences
- some amino acid substitutions do not affect protein stability or function
- disadvantageous SNPs selected against --> further underrepresentation

A small minority of SNPs are of functional consequence and are selectively advantageous or disadvantageous.

Affymetrix chip to identify SNPs

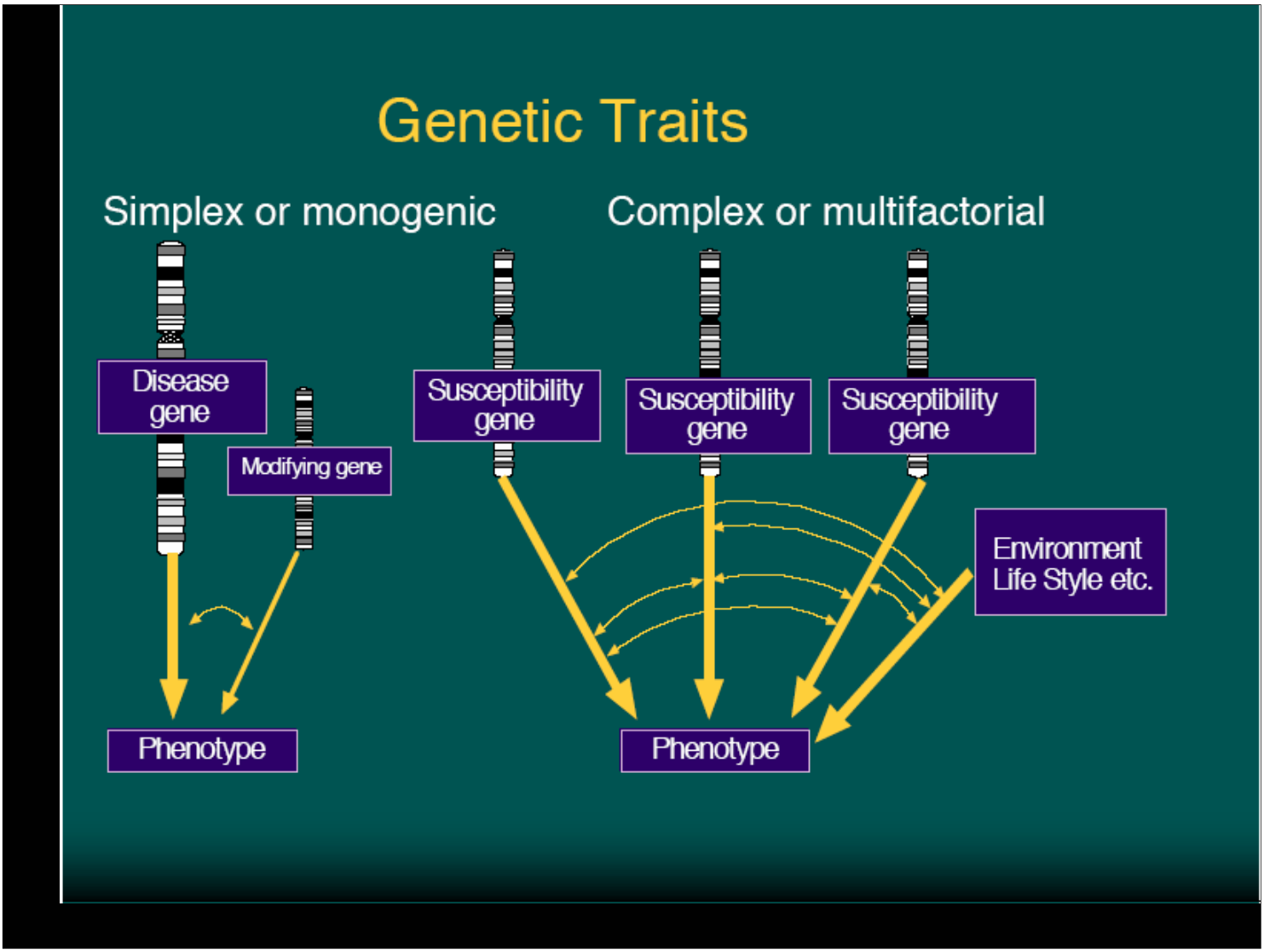
Image removed due to copyright considerations.

6000 datapoints, tabular and visual views of the data.

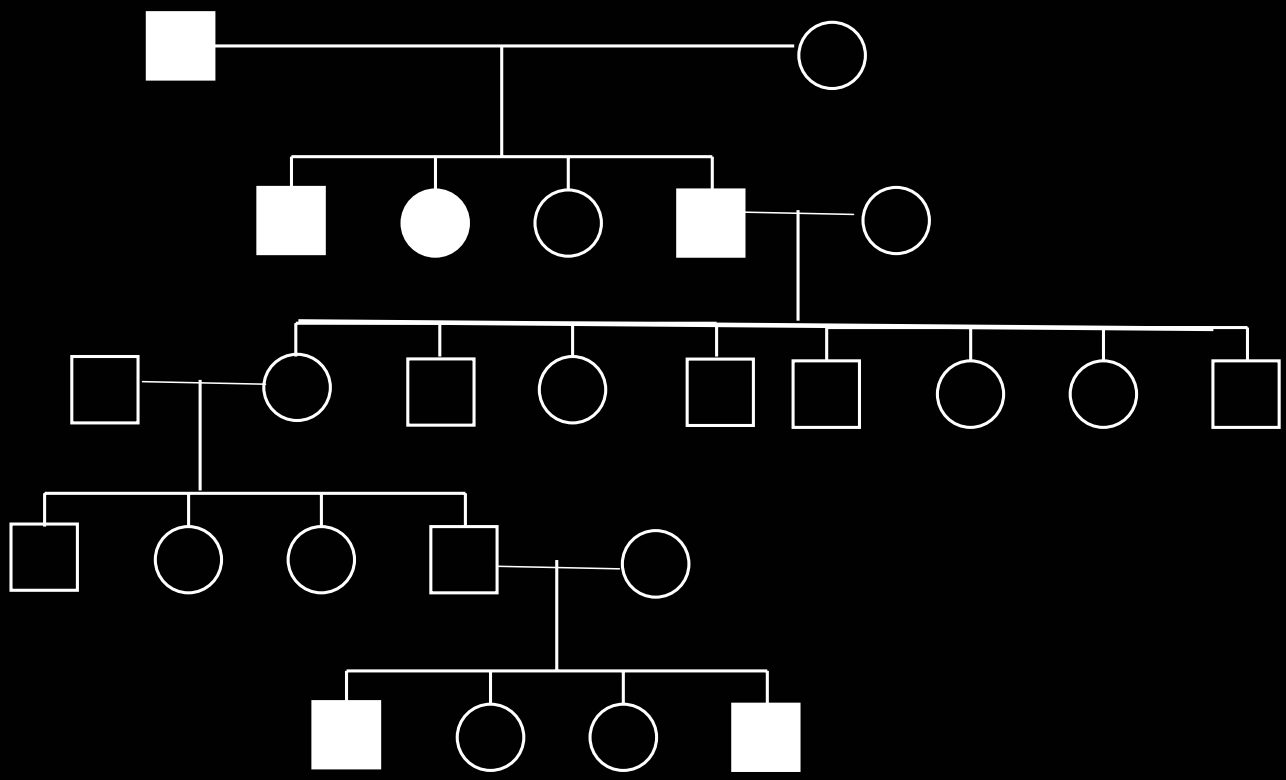
Note that only 1500 showing in image on left, a few hundred at most on right.

Following slides show…

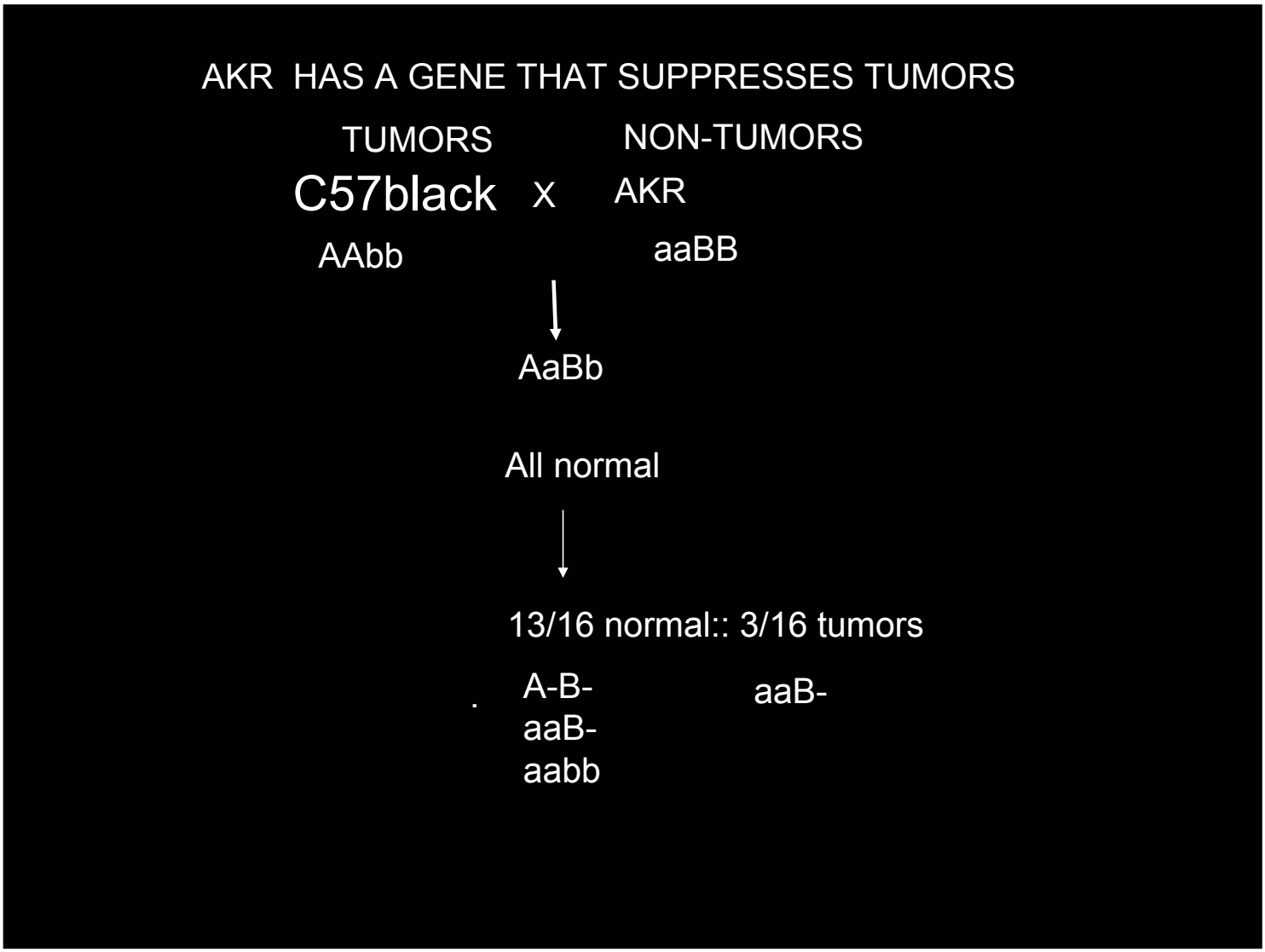
how we visualize data


Genetic Traits
Simplex or monogenic
Complex or multifactorial
Disease gene
Modifying gene
Susceptibility gene
Susceptibility gene
Susceptibility gene
Environment Life Style etc.
Phenotype
Phenotype

PEDIGREE: DOMINANT TRAIT WITH SUPPRESSOR SEGREGATING



It looks like we've been lucky. Allele A at SSR37 appears to segregate with HD. But can you be confident that the HD gene is in close proximity to the SSR37 locus, or even that it is on chromosome 4?

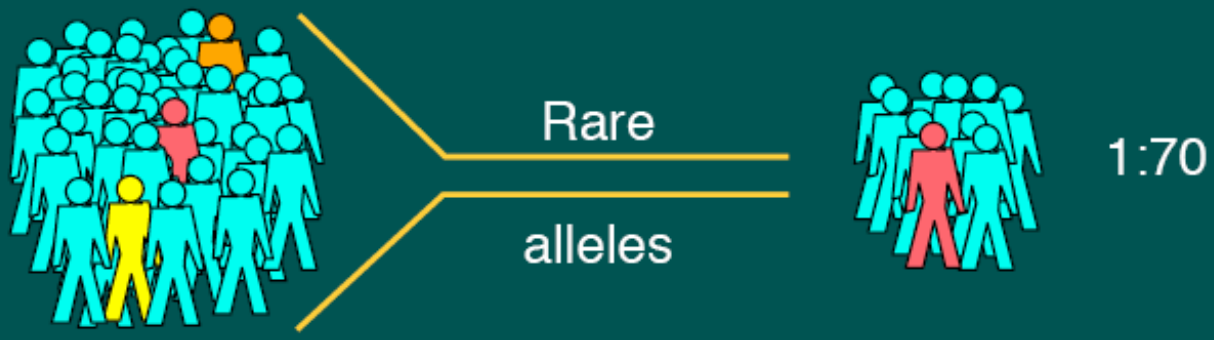


It looks like we've been lucky. Allele A at SSR37 appears to segregate with HD. But can you be confident that the HD gene is in close proximity to the SSR37 locus, or even that it is on chromosome 4?

Efforts to Simplify the Complex Genetic Background of Common Diseases

- Familial cases
- Population isolates
- Defined clinical phenotype
- Animal models

The Effect of Population Bottlenecks to Disease Alleles

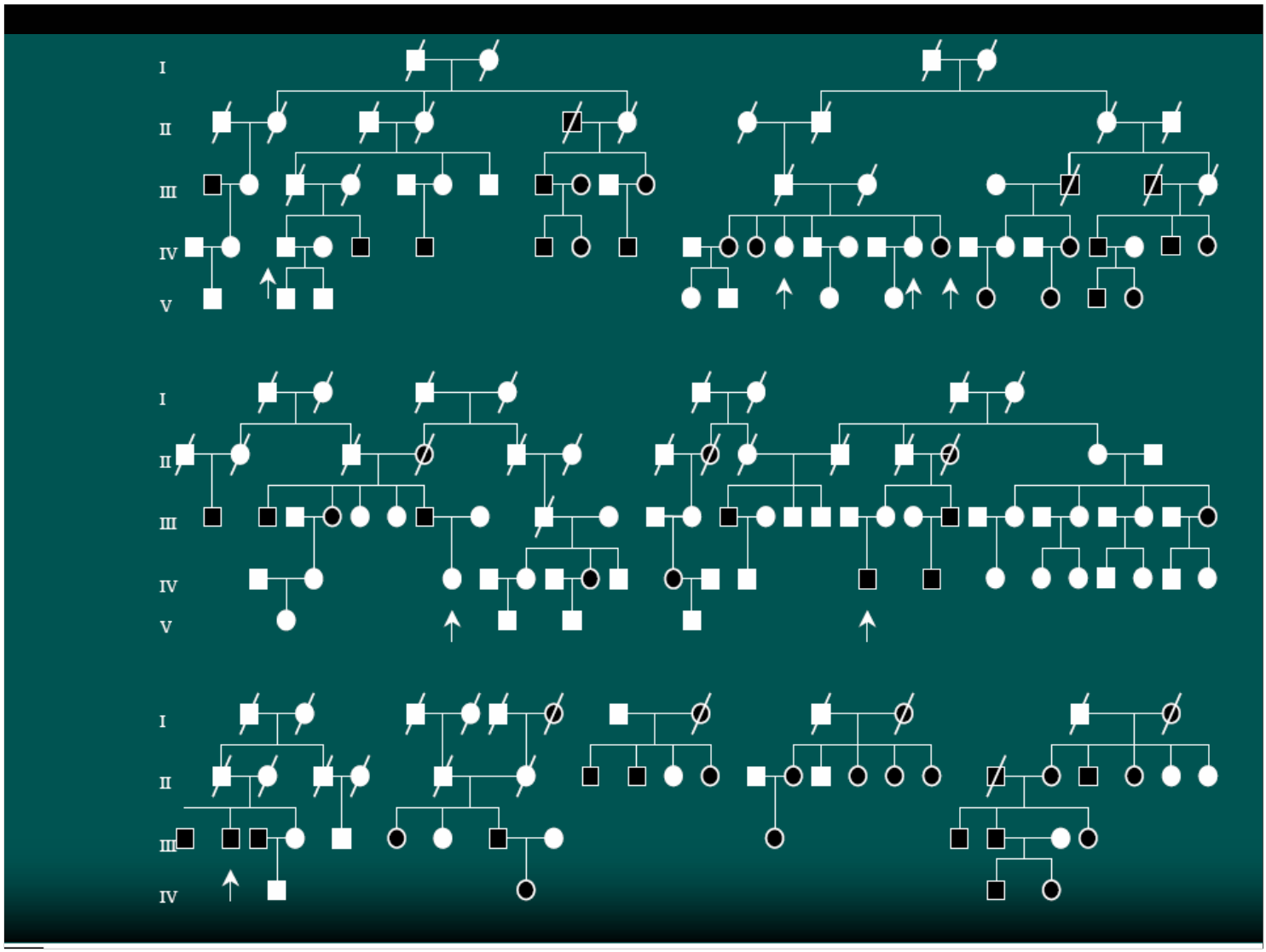


Carrier Frequencies

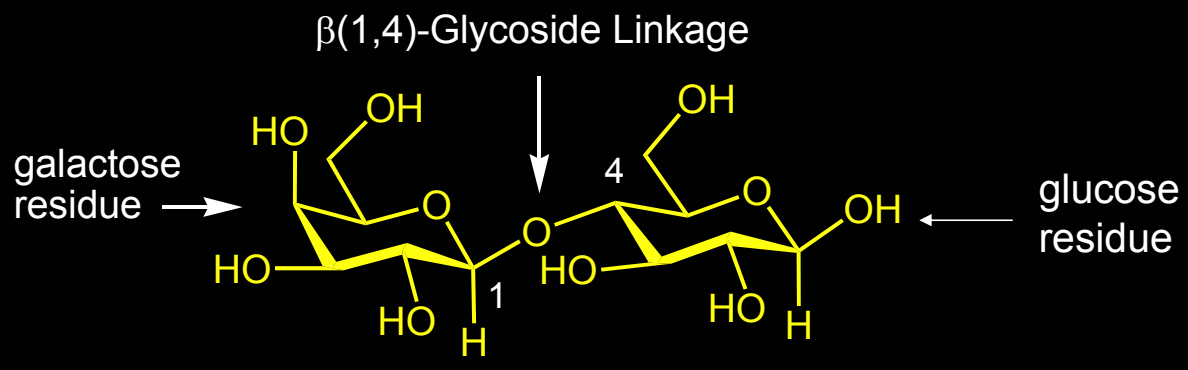
early settlement

late settlement

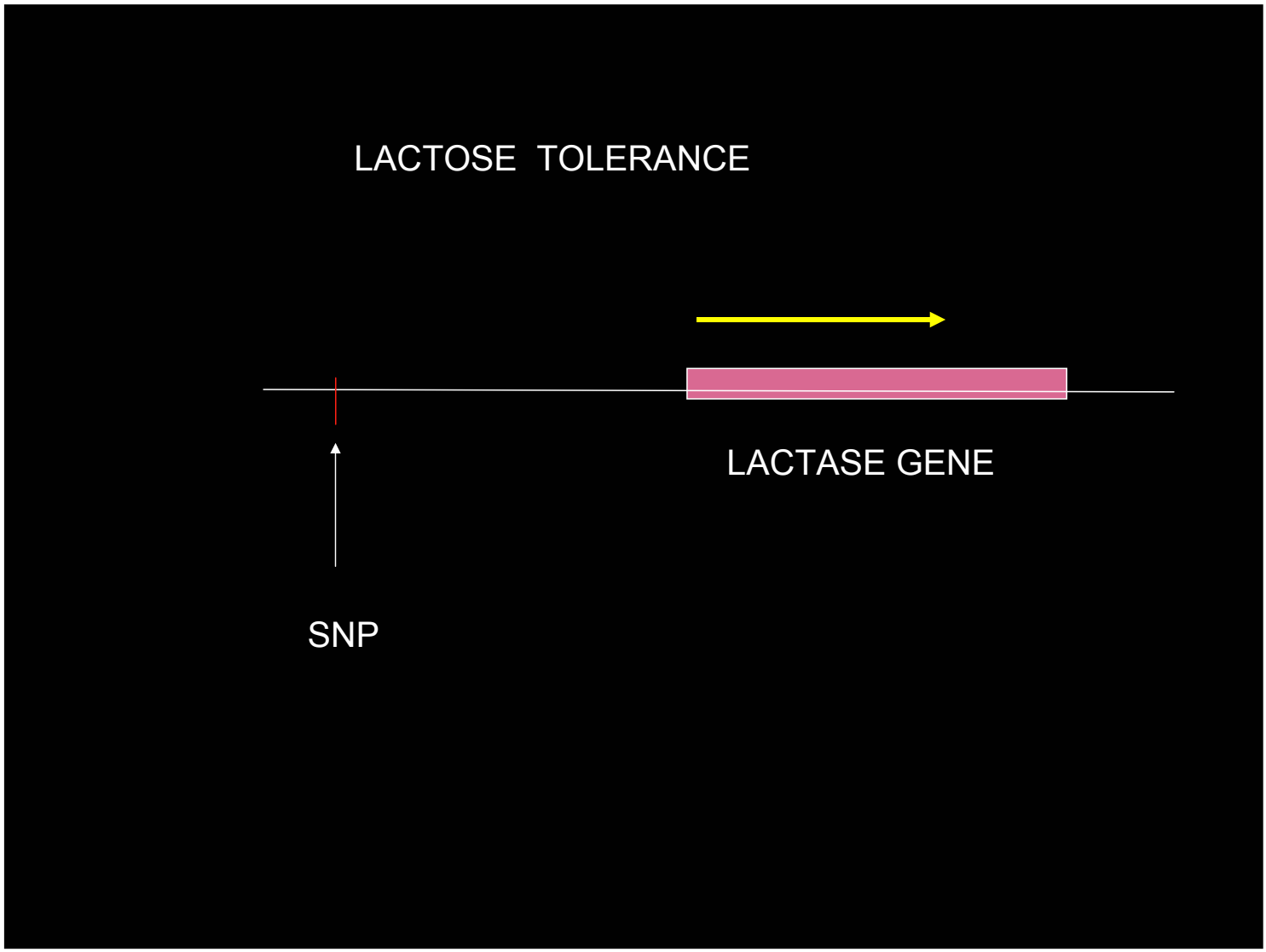
Helsinki


I
II
III
IV
V
I
II
III
IV
V
I
II
III
IV

LACTOSE

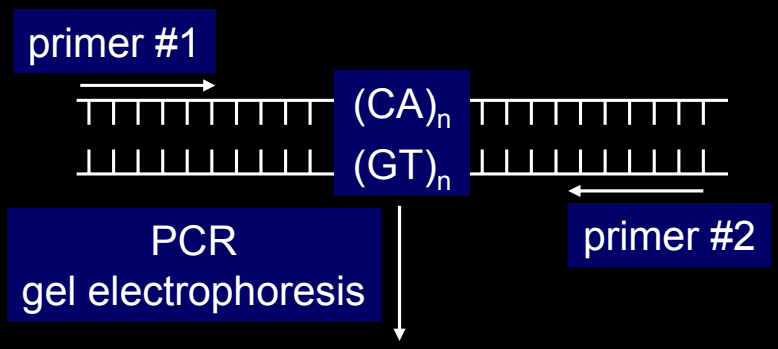


CANDIDATE GENE


LACTOSE TOLERANCE
LACTASE GENE
SNP

2) SSRs = simple sequence repeat polymorphisms = "microsatellites"

Most common type in mammalian genomes is CA repeat:



alleles	n
A	11
B	12
C	13
D	14
E	15
F	16