Concepts in Modern Genetics

Introduction

Genes, Chromosomes,

Genetics of Single Gene Disorders

M. Stoffel, MD, PhD

Concepts in Modern Genetics

HS 2017

04.12. Introduction, Genes, Chromosomes, Approaches to study human genetic disease

05.12. Genetics of polygenic, complex diseases

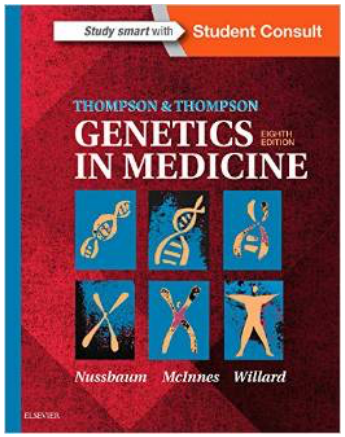
11.12. Self study (original research articles)

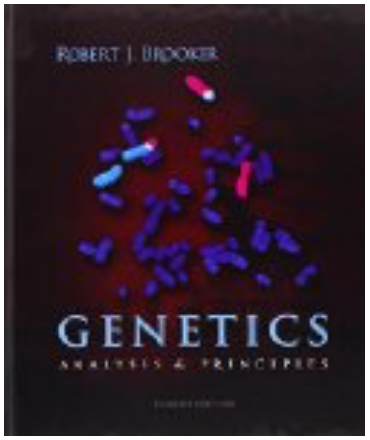
12.12. Approaches to treat genetic diseases, concepts of personalized medicine

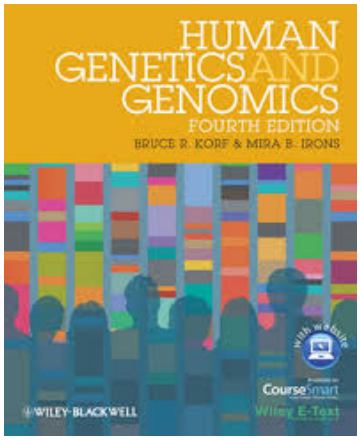
18.12. Problem solving, group study

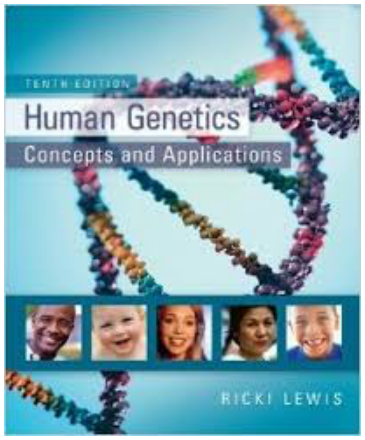
Textbooks

There is no one essential textbook!!! Find out what is good for you!





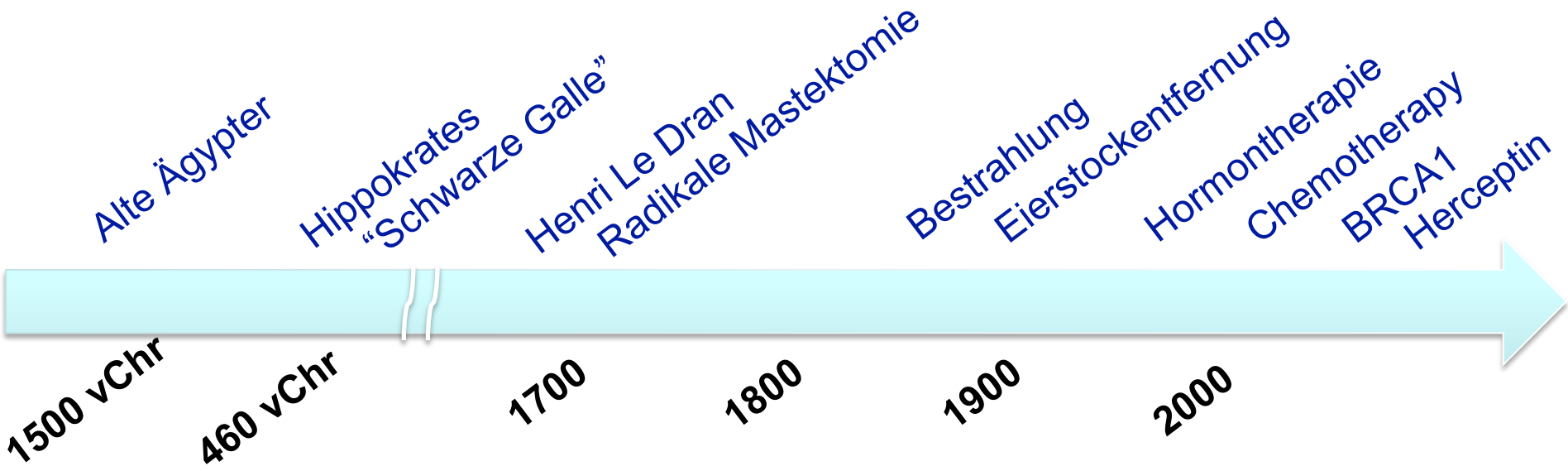


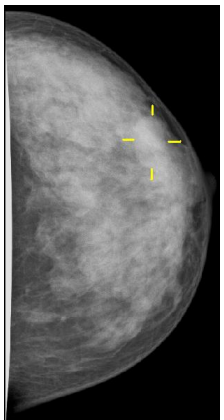


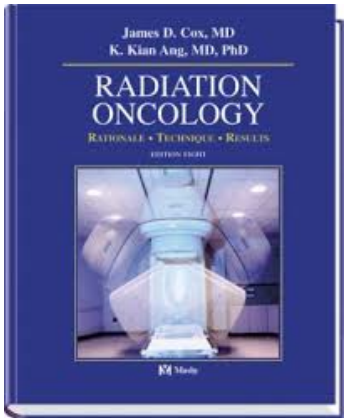
http://www.amazon.com/Thompson-Genetics-Medicine-8e/dp/1437706967/ref=zg_bs_13535_9/189-6402271-2602531

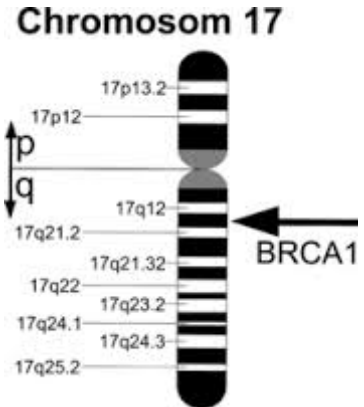
Study of Disease: A historical Perspective

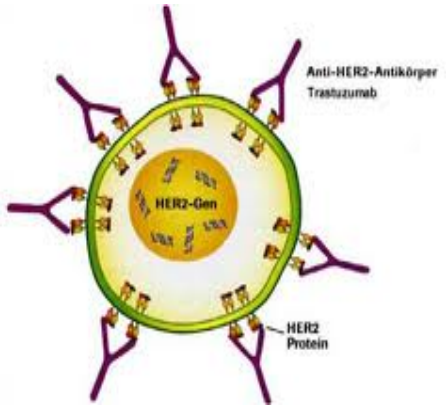
Changing pathogenesis, treatment, therapy











Disease Terminology

Precise terminology is important for communication in the medical sciences!
Make sure you understand scientific/medical terms, otherwise ask!

Health - physical, mental, and social well-being

Disease - an abnormality in body function that threatens health

Etiology - the study of the factors that cause a disease

Idiopathic - refers to a disease with an unknown cause

Signs and symptoms - the objective and subjective abnormalities associated with a disease

Pathogenesis - the pattern of a disease's development (infectious, non infectious)

Disease Terminology

Symptoms:

Greek *σύμπτωμα*, "accident, misfortune, that which befalls", is a departure from normal function or feeling which is noticed by a patient, indicating the presence of disease or abnormality.

- Subjective
- Observed by patient
- Cannot be measured directly

Disease Terminology

Symptoms:

- acute
- chronic
- relapsing
- asymptomatic (e.g. subclinical infection)
- non-specific (fatigue)

Disease Terminology

Symptoms vs sign:

A sign is noticed by other people. It is not necessarily the *nature* of the sign or symptom which defines it, but *who* observes it.

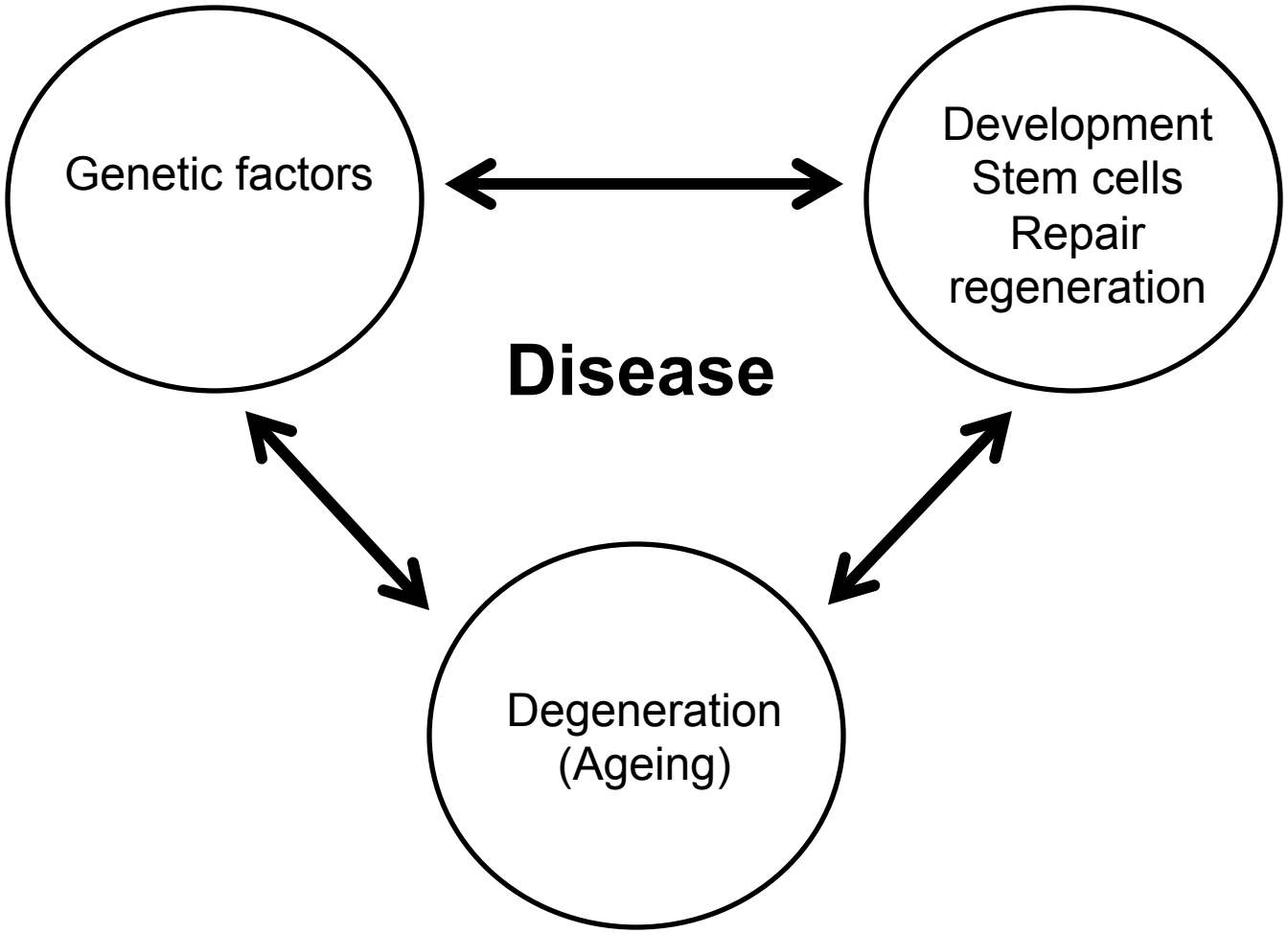
e.g. a skin rash may be noticed by either a healthcare professional as a sign, or by the patient as a symptom. When it is noticed by both, then the feature is both a sign and a symptom.

Disease Terminology

- **Syndrome**
- Syndrome is the association of several medical signs, symptoms, and or other characteristics that often occur together.

- one cause, e.g. Downs syndrome
- Multiple causes, e.g. Parkinsonian syndrome
- unknown

Disease Mechanisms



Genes, Chromosomes Genetics and Single Gene Disorders

M. Stoffel, MD, PhD

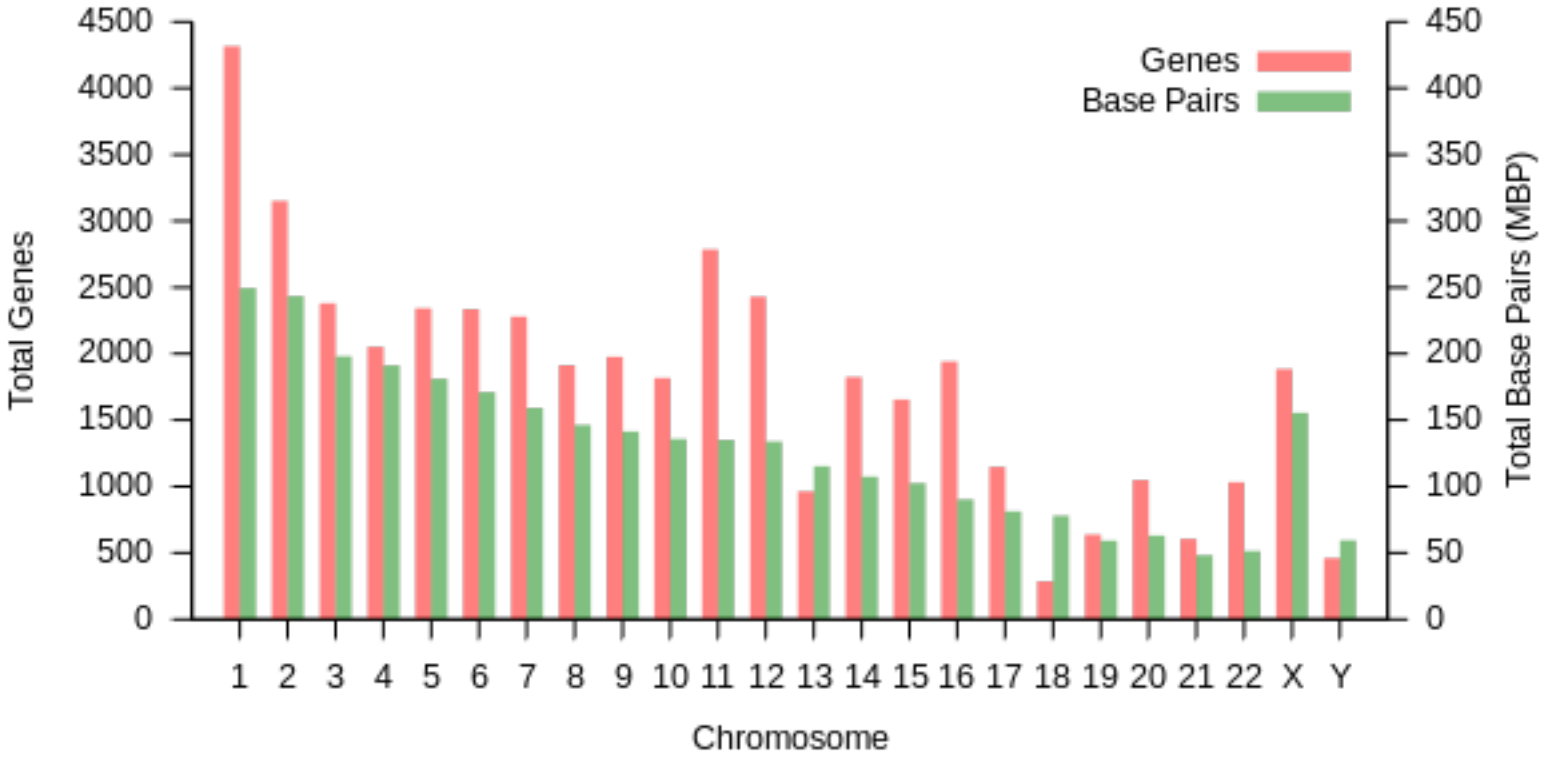
Topics

- Organization of the human genome
- Gene mapping
- Identifying the genetic basis of disease
- Mutation
  - quantitative aspects
  - qualitative aspects
- Examples of monogenic diseases

Organization of the human genome

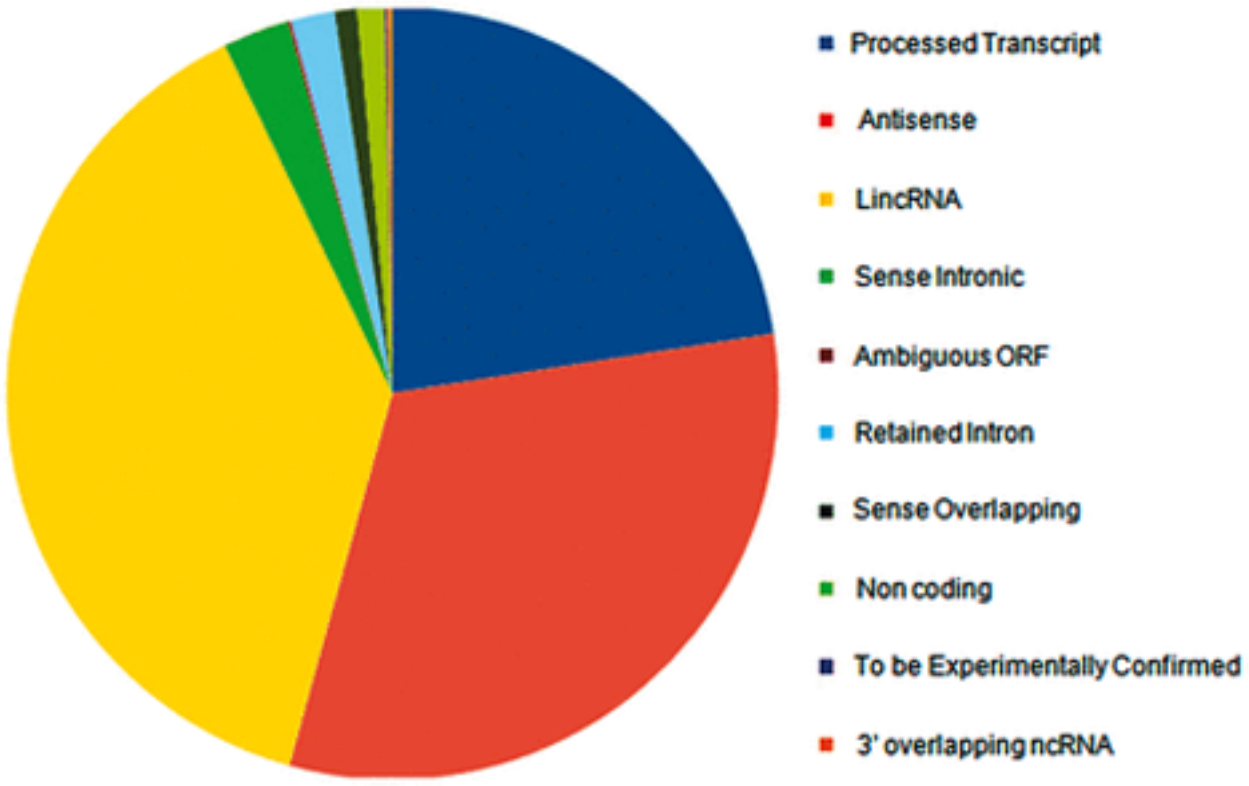
- 3.5×10^9 pairs of nucleotides in DNA
- Haploid genome, dipoid genome
- Approximately the same amount of DNA is found in cells of mice, elephants, bats, armadillos, pandas
- Humans:
  - 23 pairs of chromosomes
  - 22 Autosomes (great apes have 23 autosomes)
  - X-chromosome
  - Y chromosome
- Mitochondrial chromosome
- Karyotype: collective features of a set of chromosomes in late prophase or metaphase stages of mitosis

Estimated number of genes and chromosomes



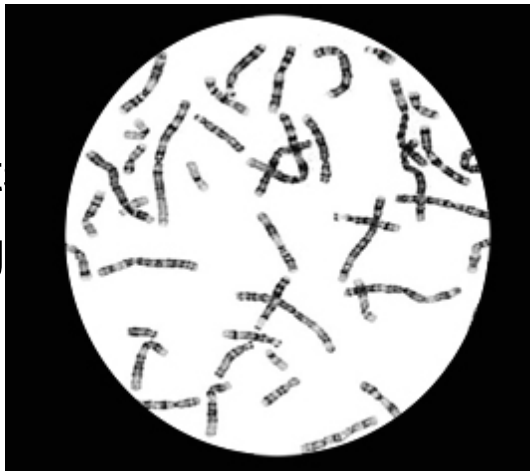
Organization of the human genome

Coding genome: ≈1.5%

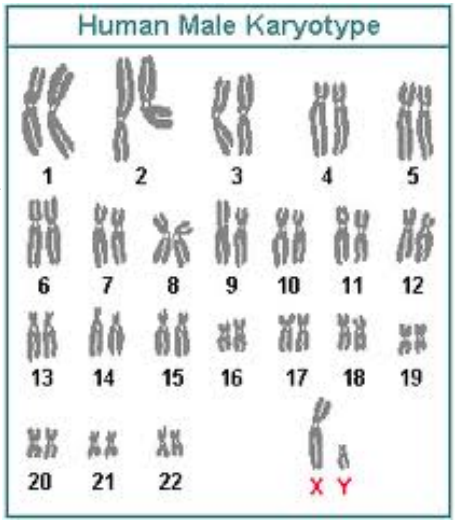


Organization of the human genome

Experimental procedure for observing human chromosomes:



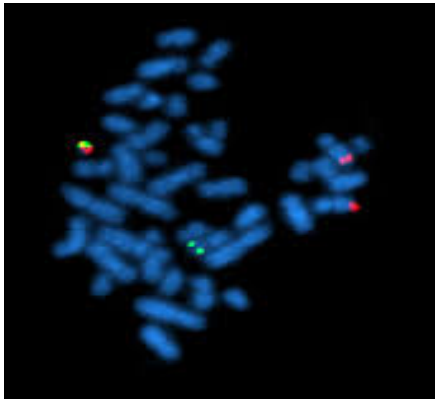
- Obtain white blood donor cells
- Incubate in culture medium with growth stimulant
- Add colchemid (microtubule-depolymerizing drug
- Sediment cells by centrifugation
- Suspend cells in hypotonic solution
- Sediment cells by centrifugation
- Suspend cells gradually in fixative
- Drop fixed cells onto a microscope slide
- Stain dried slide to produce chromosome ba
- Observe chromosomes in microscope



Organization of the human genome

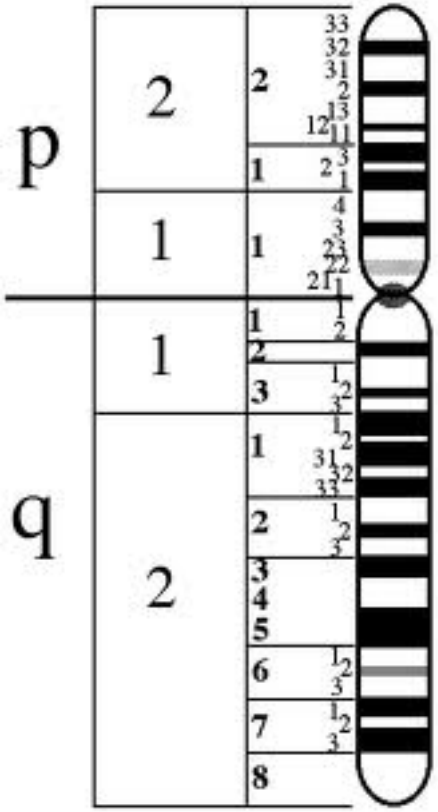
Fluorescent in situ hybridization (FISH)

Specific labeling of chromosomes using molecular probes



FISH is used to detect and localize the presence or absence of specific DNA sequences On chromosomes. FISH uses fluorescent probes. FISH uses fluorescent probes that bind to only those parts of the chromosome with which they show a high degree of sequence complementarity.
FISH can also be used to detect and localize specific RNA targets (mRNA lncRNA and miRNA) in cells, circulating tumor cells, and tissue samples.

General features of chromosomes



Chromosomes stained with dyes bound to DNA

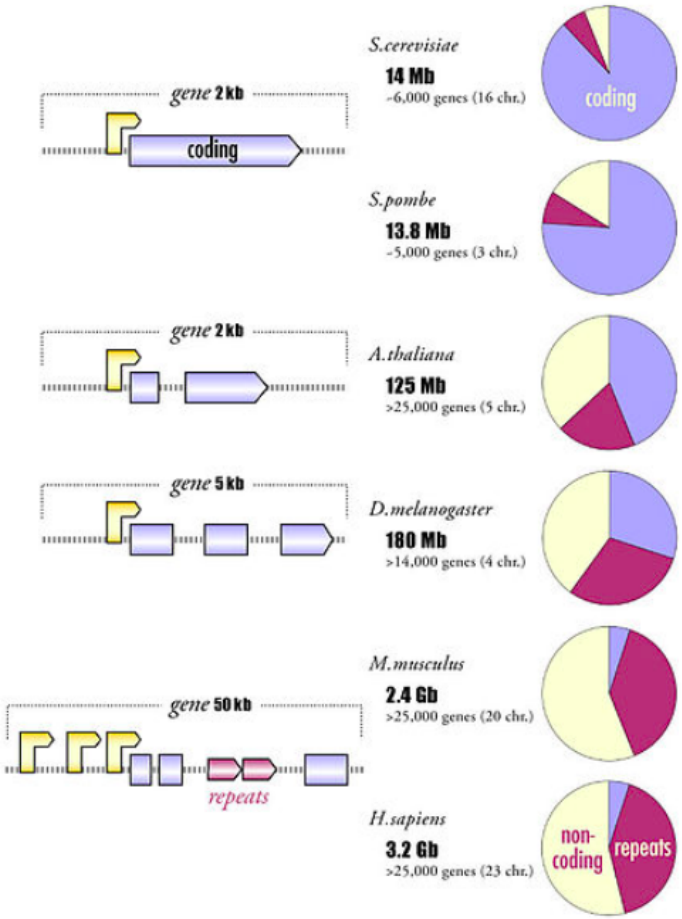
Giemsa staining:
Dark bands are relative A-T rich
Light bands are G-C rich

Giemsa banding patterns, size and centromere position are morphological features defining Chromosomes

p: indicates short arm
q: long arm

End of chromosomes: telomeres

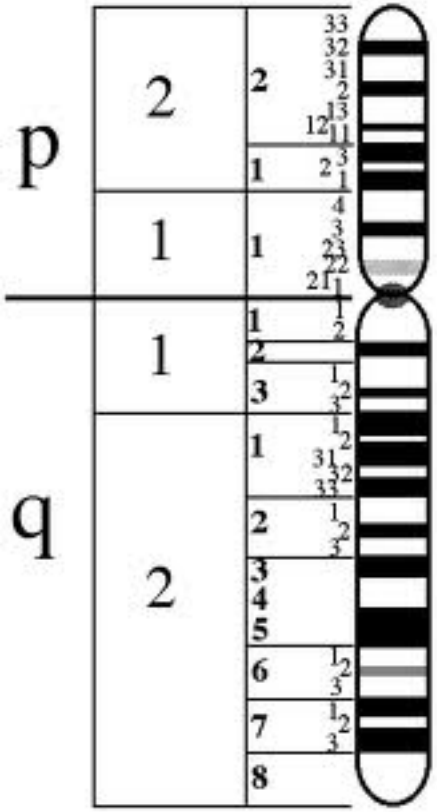
General features of chromosomes



Mammals have accumulated considerable repetitive elements and noncoding regions, which account for the majority of their DNA sequences (52% non-coding and 44% repetitive DNA). Only 1.2% of the mammalian genome thus encodes for protein function. This massive expansion of repetitive and noncoding sequences in multicellular organisms is most likely due to the incorporation of invasive elements, such as DNA transposons, retrotransposons, and other repetitive elements.

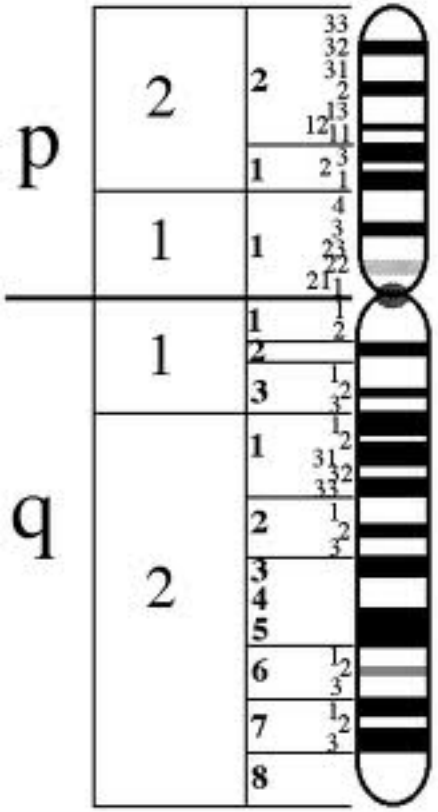
The vast expansion of the genome with non-coding and repetitive DNA in higher eukaryotes implies more extensive epigenetic silencing mechanisms.

Families of reiterated sequences



- Approximately 1/3 of the human genome consists of repeat sequences
- Satellite DNA: reiterated, tandemly arranged sequences
- Alpha family: found in centomeres, basic repeat unit of 171 bp
  - Tandem arrays up to several million bp long
  - Spindle fibres attach to the centromere via the kinetochore
- Telomeres
  - Hexameric element (TTAGGG)
  - tandem arrays up to 5,000 – 10,000 by in length

Families of reiterated sequences



Interspersed repetitive sequences

Short interspersed repetitive sequences (SINE)
Long interspersed repetitive sequences (LINE)

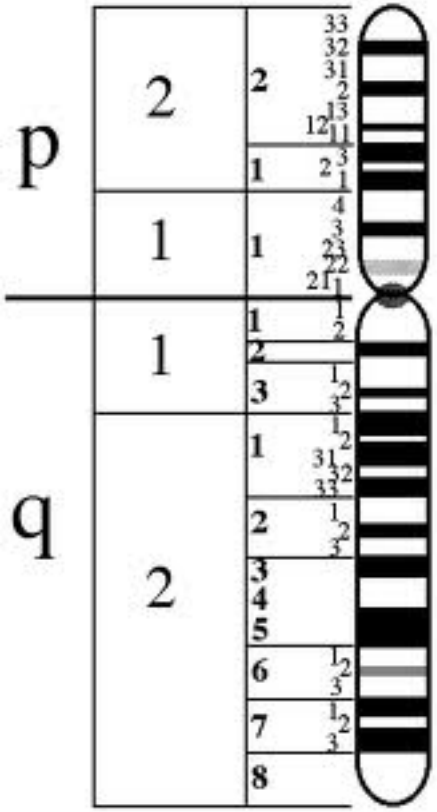
SINE Class:
Alu family: Alu elements of different kinds occur in large numbers in primate genomes (>1 mio in humans). Are the most abundant Transposable elements in the human genome. They are derived from the small cytoplasmic 7SL RNA (299 bp), a component of the signal recognition particle.

Alu sequences contain RNA pol III initiation sites

Alu insertions have been implicated in several inherited human diseases and in various forms of cancer.

gccgggcgcg gtggcgcgtg cctgtagtcc agctactcg ggaggctg**AG GCT**Gga
GGAT CGcttgAGTC CAggagttctgggctgtagt gcgctatgcc gatcgggtgt
ccgcactaag ttcggcatca atatggtgac ctcccgggag cgggggac caccaggttgcc
taaggagggg tgaaccggcc caggtcggaa acggagcagg tcaaaactcc cgtgctgatc
Agtagtggga tcgcgcctgt gaatagccac tgcactccag cctgggcaac

Families of reiterated sequences



Interspersed repetitive sequences

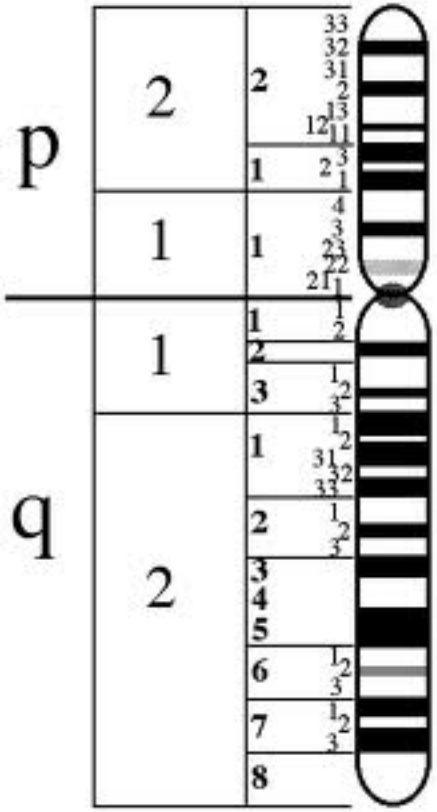
Short interspersed repetitive sequences (SINE)
Long interspersed repetitive sequences (LINE)

LINE Class:
> 500 bp long

L1 element: ≈6400 bp long, 5' end highly variable
Reiteration frequencies: 3,000 – 40,000 copies

Predominantly in dark bands

Families of reiterated sequences



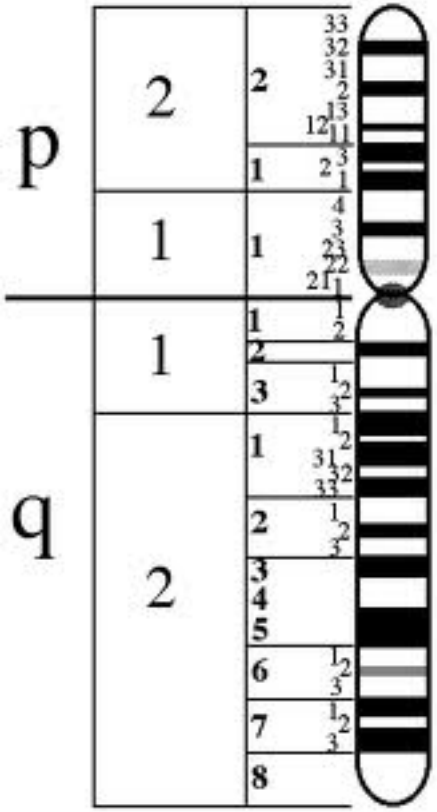
Middle repetitive sequences

Genes for 18S and 28S ribosomal RNAs

Each copy contains one 18S and one 28S sequence together with additional nucleotides and are transcribed to form large pre-rRNAs, which are then cleaved by specific nucleases.

Several hundred copies of genes for 18S and 28S ribosomal RNAs are tandemly arranges in clusters on several chromosomes.

Families of reiterated sequences



Polymorphic sequences

Single nucleotide polymorphisms
Approximately every 1000 bp
Used to be detected by restriction fragment length polymorphism (RFLP)

Sequencing is standard method of detection today

Minisatellites:
Variable number of tandem repeats (VNTR)
Usually 10-60 bp long, scattered throughout the genome
Number of repeats: 10 to 60

Microsatellites:
Di-, tri-, tetra-nucleotide repeats (CA, TG, TCC, TAA)
50,000 sequences in human genome
Detected by PCR

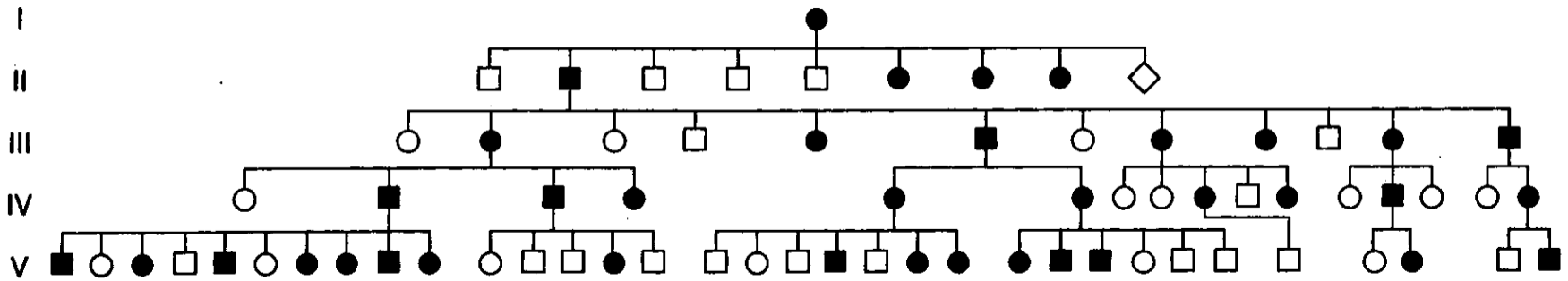
Introduction

The total load of genetic disease

Type of genetic disease	Frequency/ 1000 Population
Single gene	
- Dominant	1.8 - 9.5
- Recessive	2.2 - 2.5
- X-linked	0.5 - 2.0
Chromosomal abnormalities	6.8
Common disorders (with significant genetic component (e.g. CHD, diabetes, cancer, autoimmune)	(7 - 10)
Congenital malformations	(19 - 22)
Total (brackets indicate rough estimations)	**36.3 – 52.8**

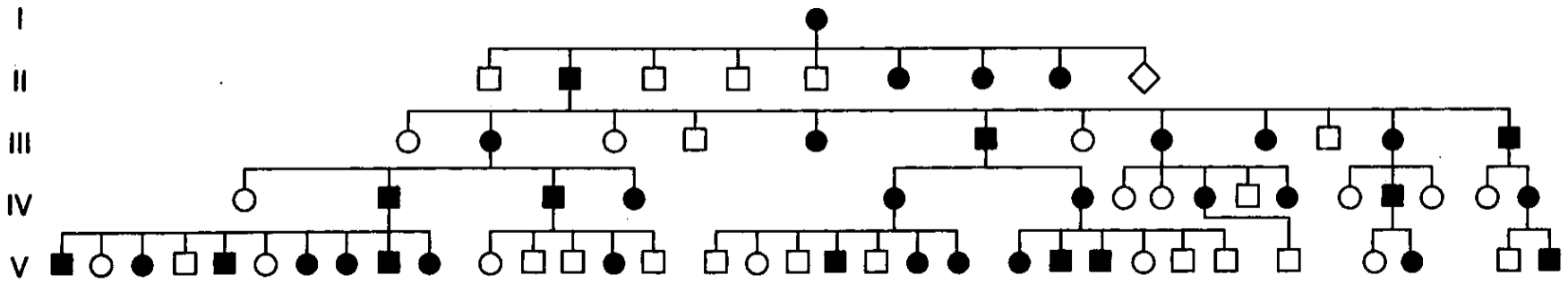
- **Specific genetic diseases are usually rare, however collectively they occur frequently!**

Identifying the Genetic basis of disease



Mode of inheritance?

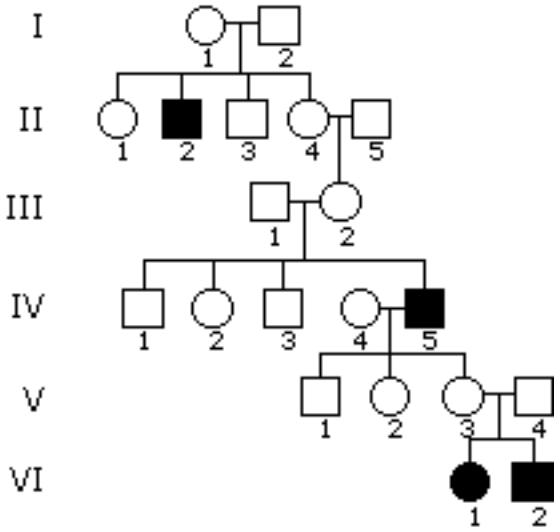
Identifying the genetic basis of disease



Autosomal dominant inheritance has the following characteristics:

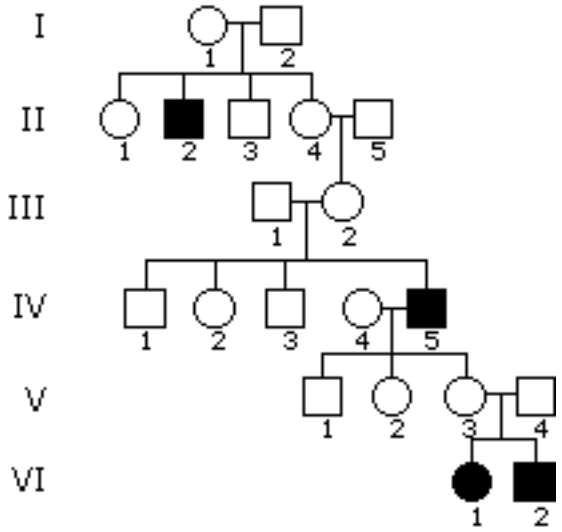
- Affected individuals have at least one affected parent
- Matings between a normal and a heterozygous affected person have a 50% chance of producing an affected offspring and a 50% chance of producing a normal offspring with each pregnancy.
- Males and females are affected in roughly equal numbers
- Both males and females transmit the phenotype

Identifying the Genetic basis of disease



Mode of inheritance?

Identifying the Genetic basis of disease



Autosomal-recessive inheritance has the following characteristics:

- Affected individuals have two normal parents
- Matings between a heterozygotes (both phenotypically normal) have a 75% chance of producing a normal offspring and a 25% chance of producing an affected offspring with each pregnancy. Thus, on a population basis, the ratio of normal to affected offspring will be 3:1.
- Matings between affected persons produce only affected children
- Males and females are affected in roughly equal numbers
- Both males and females transmit mutant alleles

Identifying the Genetic Basis of Disease

X-linked dominant inheritance has the following characteristics:

In X-linked dominant inheritance, when the mother alone is the carrier of a mutated, or defective gene associated with disease or disorder; she herself will have the disorder.

Her children will inherit the disorder as follows:

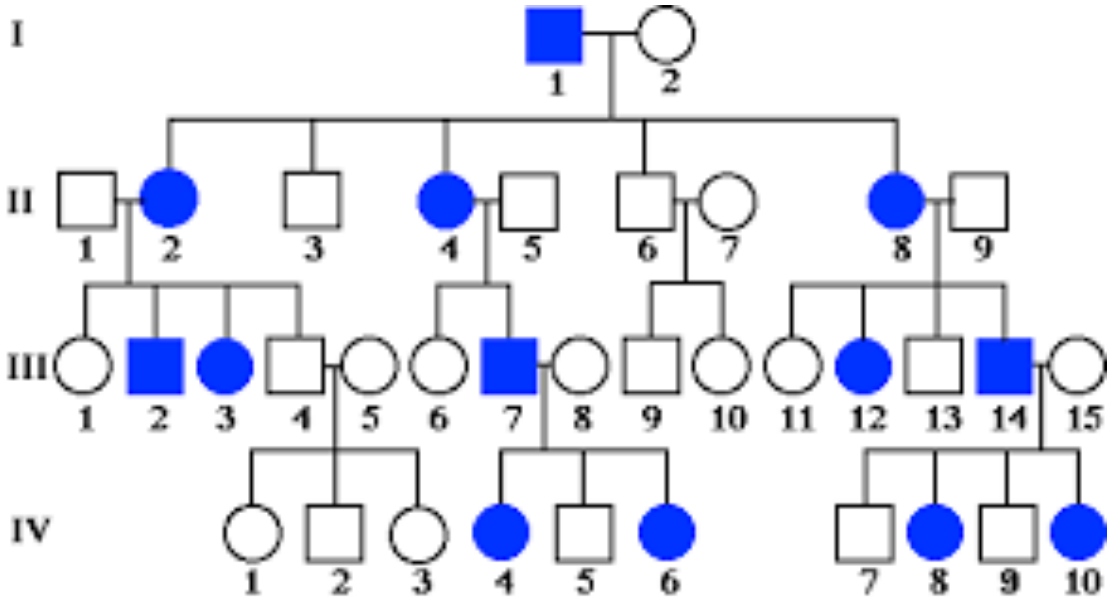
- Of her daughters and sons: 50% will have the disorder, 50% will be completely unaffected.
- Children of both genders have an even chance of receiving either of their mother's two X chromosomes, one of which contains the defective gene in question.

When the father alone is the carrier of a defective gene associated with a disease or disorder, he too will have the disorder.

- His children will inherit the disorder as follows:
- Of his daughters: 100% will have the disorder, since all of his daughters will receive one copy of his single X chromosome.
- Of his sons: none will have the disorder; sons do not receive an X chromosome from their father.

Identifying the Genetic Basis of Disease

Pedigree with X-linked dominant inheritance:

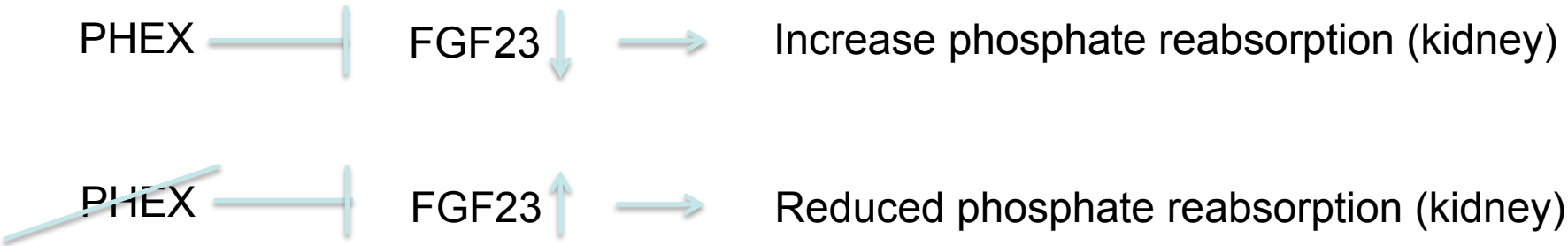


Pedigree 5. X-linked dominant inheritance.

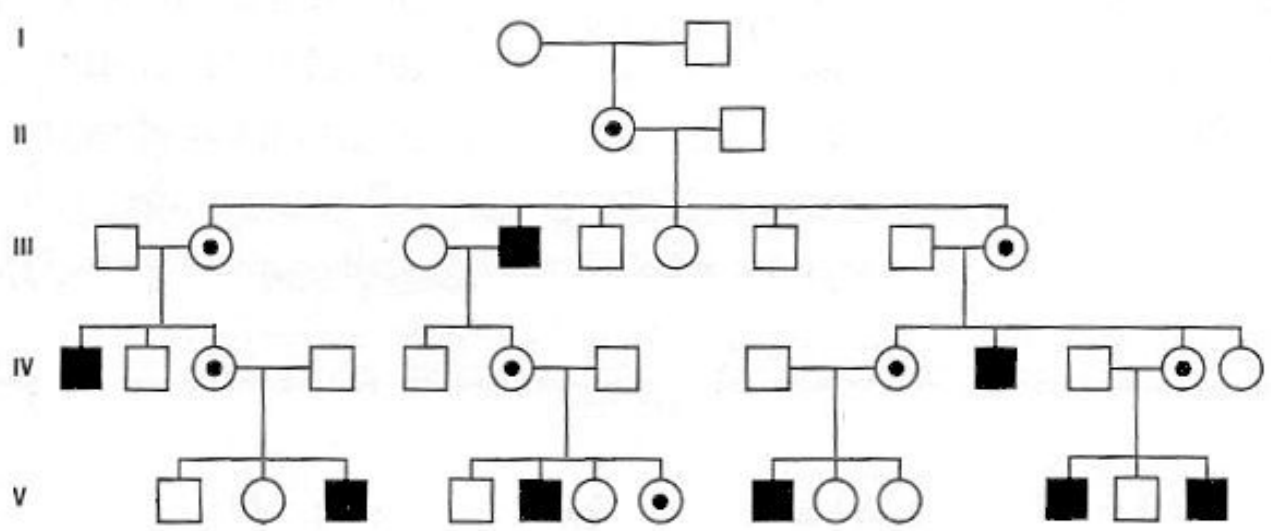
Identifying the Genetic Basis of Disease

X-linked dominant hypophosphatemic rickets

- Differs from most cases of rickets in that ingestion of vitamin D is relatively ineffective-
- Bone deformity including short stature
- XLH is associated with a mutation in the PHEX gene, located on the human X-chromosome at location Xp22.2-p22.1
- The PHEX protein regulates another protein called fibroblast growth factor (FGF) 23
- Gene mutations in PHEX prevent it from correctly regulating fibroblast growth factor 23
- The resulting overactivity of this protein reduces phosphate reabsorption by the kidneys, leading to hypophosphatemia and the related features of hereditary hypophosphatemic rickets



Identifying the Genetic Basis of Disease

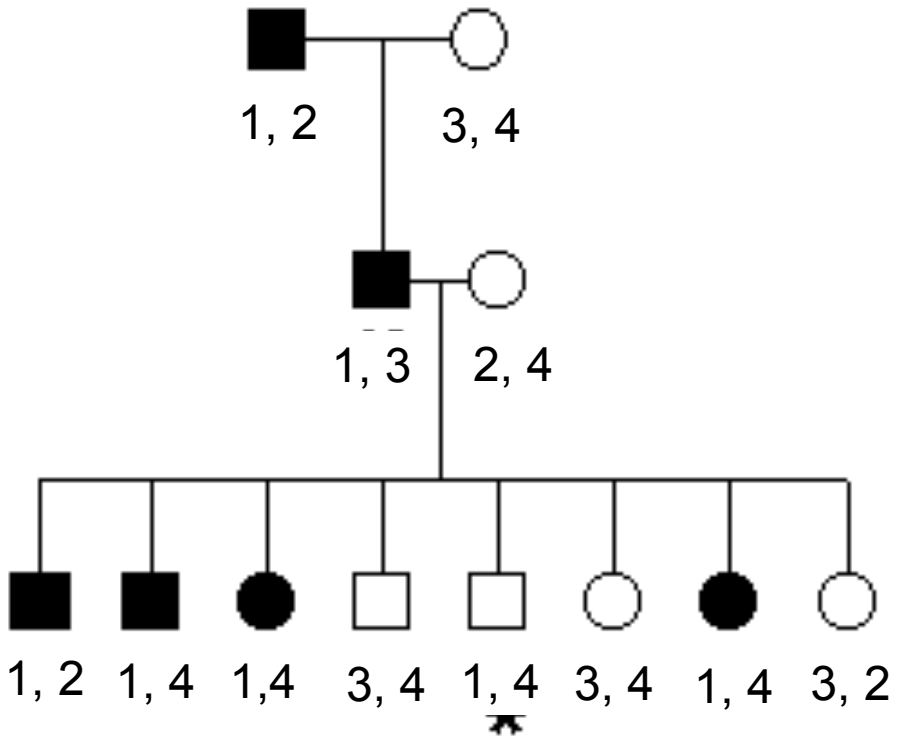


X-linked recessive inheritance

- mode of inheritance in which a mutation in a gene on the X chromosome causes the phenotype to be expressed (1) in males (who are homozygous for the mutation since they only carry one X chromosome and (2) in females who are homozygous for the gene mutation (they have a mutant copy of the gene on each of their X chromosomes.
- In humans, *generally* "men are affected and women are carriers"

Examples, Red-green color blindness, Hemophilia A, B, Duchenne, Becker's muscular dystrophy

Disease mapping using genetic linkage



Nail-patella syndrome = ● or ■

Genotypes: 1, 2, 3, 4

The **LOD score** (logarithm (base 10) of odds) is a statistical test used in linkage analysis in human, animal, and plant populations. The LOD score compares the likelihood of obtaining the test data if the two loci are indeed linked, to the likelihood of observing the same data purely by chance. Positive LOD scores favor the presence of linkage, whereas negative LOD scores indicate that linkage is less likely. Computerized LOD score analysis is a simple way to analyze complex family pedigrees in order to determine the linkage between Mendelian traits (or between a trait and a marker, or two markers).

A positional analysis experimental strategy

- Map the gene by linkage or association analysis
- Define minimal disease region
- Identify conserved sequences (coding, 5' and 3'UTRs, non-coding RNAs, promoters, enhancers)
- Identify mRNAs transcribed from each candidate gene (show that gene is expressed in affected tissues)
- Correlate abnormal mRNAs with presence of the disease (abnormal mRNA size, splicing mutations)
- Correlate DNA deletions with presence of disease
- Identify mutations in the candidate gene from affected subjects (Missense mutation, deletions, insertions, polymorphisms, functional characterization)

Identifying the Genetic Basis of Disease

Forward genetics:

- Clinical symptom – identification of biochemical step – show that protein involved in this step is mutated in affected subjects
- From protein to gene
- Example: Sickle cell disease, mutation in Hb

Reverse genetics:

- Modern gene identification strategy
- Clinical symptom – identification of allelic variant– identify gene, show that variant (mutation) renders protein function and abnormal pathway n affected subjects
- From gene to protein
- Positional cloning: strategy to identify gene of unknown function based on genomic mapping

Example of an autosomal dominant disease
Huntington Disease

- **Clinical Classification**
  - Movement/Cognitive/Psychiatric disorder
  - Mean onset age 35-55 years.
- **Prevalence**
  - Incidence >1 in 10,000.
- **Genetic Testing**

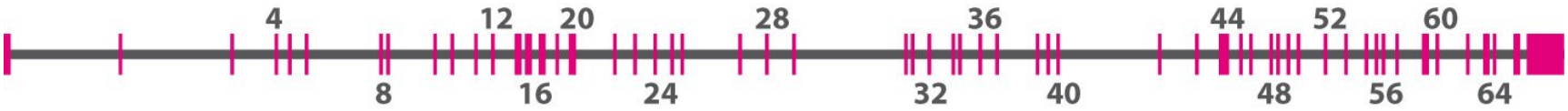
  Diagnostic

  Presymptomatic – counselling protocol.

- **Physical features:**
  - involuntary movements
  - weight loss
  - abnormal gait
  - speech & swallowing difficulties.

- **Psychiatric Manifestations:**
  - personality changes
  - depression
  - aggression
  - early onset dementia.

Example of an autosomal dominant disease
Huntington Disease



Structure of the Huntington disease gene

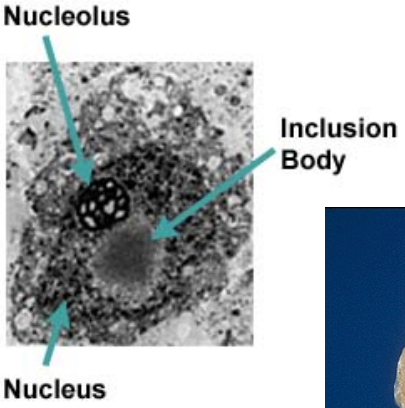
Short vertical bars represent exons.

Huntington disease - a *triplet repeat* disease

CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG ..... CAG

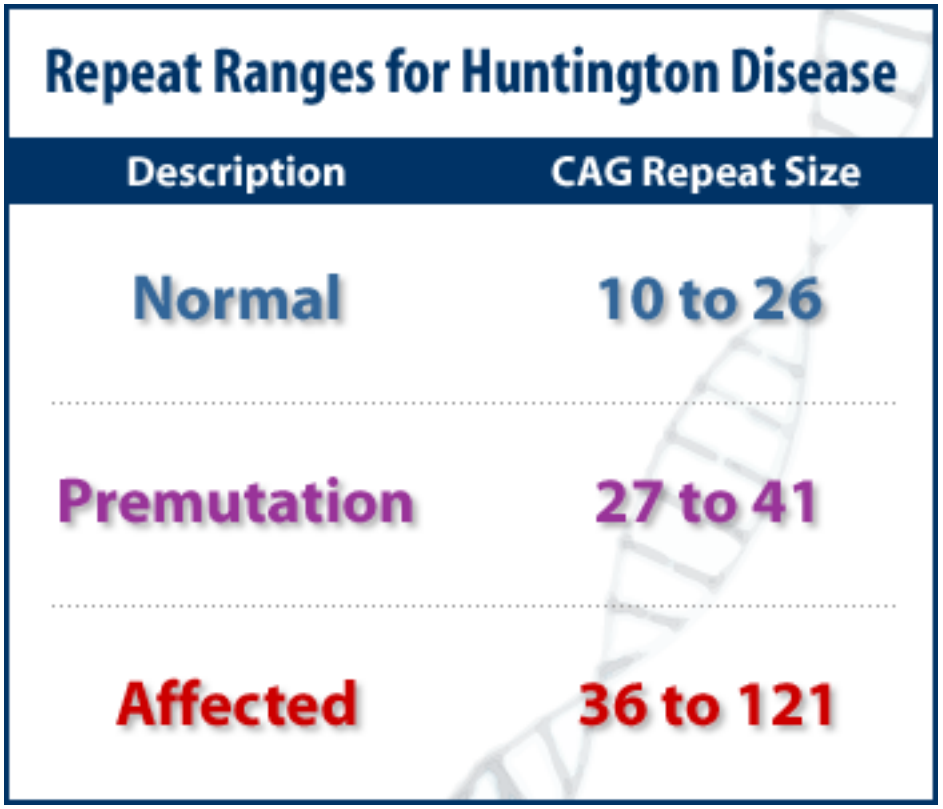
11-34 CAG triplet repeats are normal:
encodes a run of 11-34 glutamine amino acid residues in the protein.

A run of > 34 glutamine residues causes the protein to aggregate in the brain cells and cause progressive cell death.



Runs of >34 CAG repeats in the HD gene expand further (particularly during **male meiosis**) causing earlier age of onset in children of men who have the gene – **anticipation**.

Huntington disease - a *triplet repeat* disease



Repeat Ranges for Huntington Disease

Description	CAG Repeat Size
Normal	10 to 26
Premutation	27 to 41
Affected	36 to 121

Example of an autosomal dominant disease
Huntington Disease

```
  1 ttg ctg tgt gag gca gaa cct gcg ggg gca
    ggg gcg ggc tgg ttc cct ggc cag cca ttg
 61 gca gag tcc gca ggc tag ggc tgt caa tca
    tgc tgg ccg gcg tgg ccc cgc ctc cgc cgg
121 cgc ggc ccc gcc tcc gcc ggc gca cgt ctg
    gga cgc aag gcg ccg tgg ggg ctg ccg gga
181 cgg gtc caa gat gga cgg ccg ctc agg ttc
    tgc ttt tac ctg cgg ccc aga gcc cca ttc
241 att gcc ccg gtg ctg agc ggc gcc gcg agt
    cgg ccc gag gcc tcc ggg gac tgc cgt gcc
301 ggg cgg gag acc gcc atg gcg acc ctg gaa
    aag ctg atg aag gcc ttc gag tcc ctc aag
361 tcc ttc cag cag cag cag cag cag cag cag
    cag cag cag cag cag cag cag cag cag cag
421 cag cag cag caa cag ccg cca ccg ccg ccg
    ccg ccg ccg ccg cct cct cag ctt cct cag
```

CAG: Gln (Q) Glutamone

Example of an autosomal dominant disease
Huntington Disease

https://www.youtube.com/watch?v=JzAPh2v-SCQ

Huntington disease

A patient in the advanced stages of the disease showing involuntary movements of the head and face. Photos courtesy of Professor Peter Harper, Cardiff.

(b) Post mortem sections comparing normal brain (left) with brain from Huntington disease patient (right); note the loss of tissue in the Huntington disease brain. Photos courtesy of Dr David Crauford, St Mary's Hospital, Manchester.

Advantages of predictive testing for Huntington disease

- Uncertainty of gene status removed.
- If negative:
  - concerns about self and offspring reduced.
- If positive:
  - make plans for the future
  - arrange surveillance/treatment if any
  - inform children/decide whether to have children.

Disadvantages of predictive testing for Huntington disease

- **If positive:**
  - removes hope
  - introduces uncertainty (if and when)
  - known risk to offspring
  - impact on self/partner/family/friends
  - potential problems with insurance/mortgage.
- **If negative:**
  - expectations of a ‘good’ result
  - ‘survivor’ guilt.

Example of an autosomal recessive disease
Cystic fibrosis

- Affects 1 in 2,500 babies in the UK (240 babies annually)
- Recessively inherited
- Lifelong, life-limiting illness
- Affects the lungs, digestive tract and pancreas by clogging them with thick, sticky mucus
- Daily physiotherapy, dietary supplements and intensive treatment for chest infections

Cystic Fibrosis (1)

Commonest AR-inherited disease amongst Northern European Caucasians

- Incidence in UK Caucasian population 1 in 2,500 carrier risk 1 in 25
- Incidence in UK Asian population 1 in 10,000 carrier risk 1 in 50

Cystic Fibrosis (2)

- Gland secretions thicker or more viscous than normal
- Small bowel: obstruction (meconium ileus in the newborn)
- Lungs : thick bronchial mucous, recurrent chest infections, progressive lung damage, heart/lung transplantation

Cystic Fibrosis (3)

- Pancreas: failure to secrete digestive enzymes causing malabsorption, failure of growth and late development
- Men nearly always infertile – absence/atrophy of vas deferens
- Lifelong potentially fatal disease

Cystic Fibrosis Gene & Mutations

- Large gene encoding 1480 amino acids
- Over 1,100 different mutations identified in the gene
- Routinely tested for 29 commonest mutations
- If mutation/s not found in clinically affected child - send DNA to laboratory to be tested for rarer mutations

- Commonest mutation – Delta F508

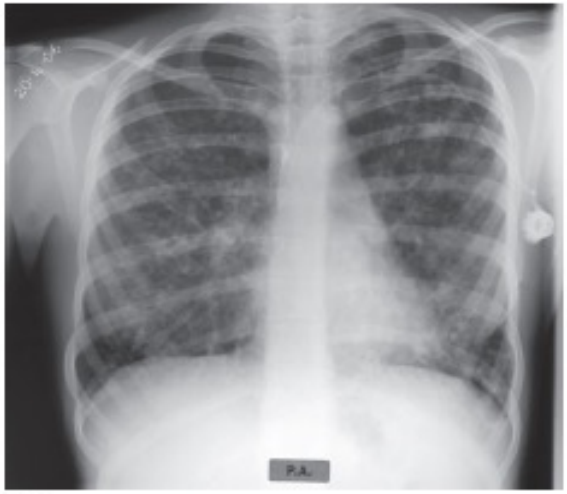
  UK Caucasians - 75%

  UK Asians - 29%

- Common mutation found in 1 in 10 UK Asians - Y569D (substitution G - T)
- Delta F508 and 28 others account for 85% mutations in the Northern European Caucasian population

Cystic Fibrosis Gene & Mutations

(a)



(b)

(c)

Cystic fibrosis

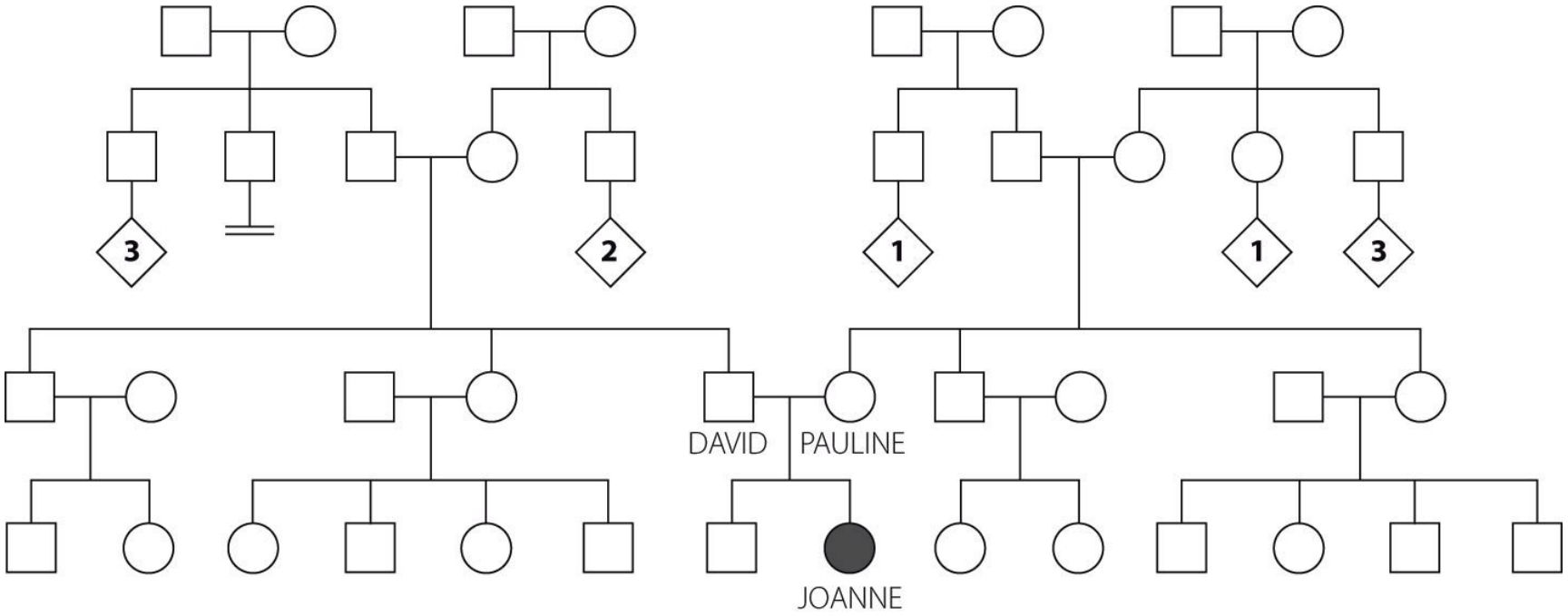
(a) The outlook for cystic fibrosis patients has improved over the years but they still need frequent hospital admissions, physiotherapy and constant medications.

(b) Chest X-ray of lungs of cystic fibrosis patient.

(c) X-ray abdominal film of newborn with meconium ileus showing multiple fluid levels.

Photos (a) and (b) courtesy of Dr Tim David, Royal Manchester Children's Hospital.

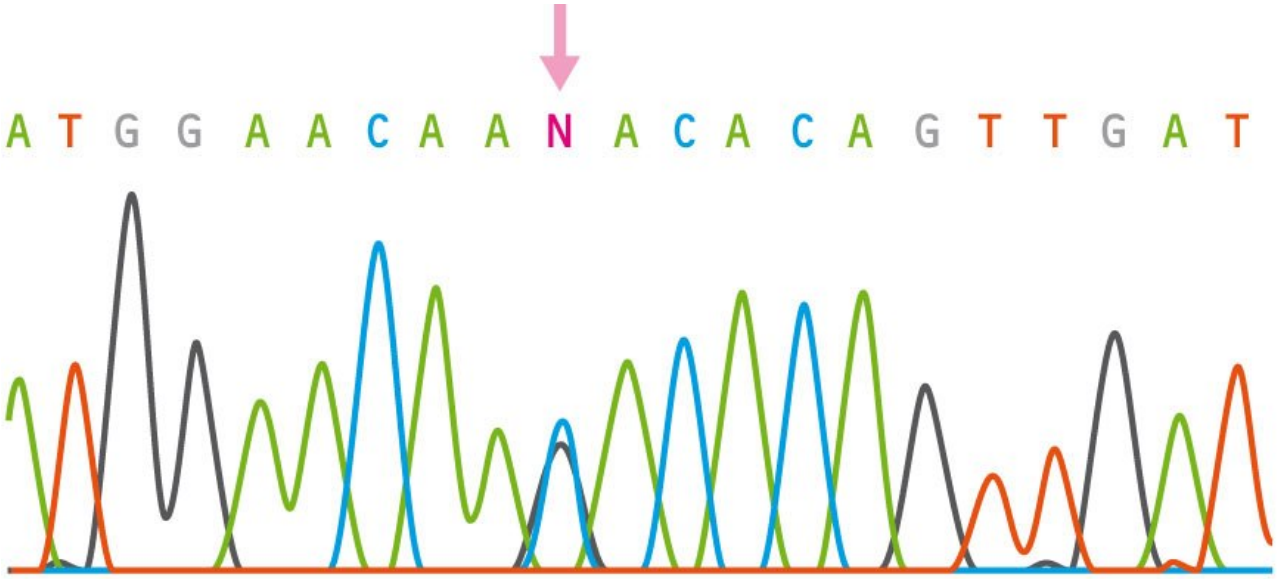
Cystic Fibrosis Gene & Mutations



Pedigree of Joanne Brown's family

Note the complete absence of any family history of cystic fibrosis. Autosomal recessive conditions commonly present as a single isolated case.

DNA sequencer trace of part of the exon 14b PCR product from Joanne Brown's *CFTR* gene



At the arrowed position G and C nucleotides are both present, showing that Joanne is heterozygous for a nucleotide substitution (remember that the products of PCR and sequencing are normally a mix of the products from the two alleles). Control samples show only the G. It is usual to sequence both strands of the DNA separately to confirm any change. In this case the sequence shown is of the reverse strand, so in the sense strand the change in Joanne is C>G.

Monogenic vs polygenic diseases

Genetic diseases

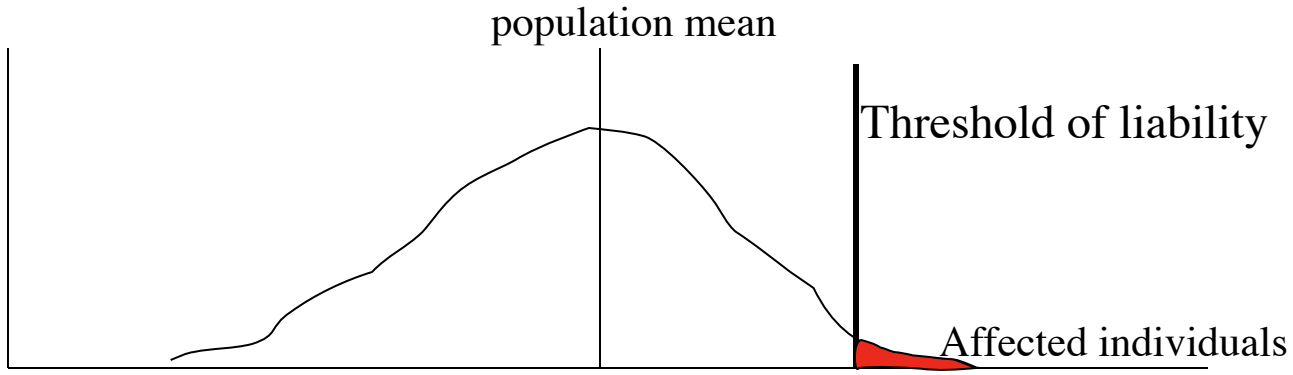
- traditionally - 3 types of diseases
  1. Genetically determined
  2. Environmentally determined
  3. Genetically and environmentally determined
- today - distinctions are blurred
- up to 20% of pediatric in-patients have genetic abnormality
- about 50% of spontaneous abort uses have chromosomal aberration
- only mutations that are not lethal are reservoir of genetic diseases

Disorders with multifactorial inheritance (polygenic)

- influence of multiple genes + environmental factors
- relatively frequent
  - Diabetes mellitus (Endocrine pathology)
  - Hypertension (see Cardiovascular ethiology)
  - Gout (Metabolic disease)
  - Schizophrenia (Psychiatric disorder)
  - Congenital heart disease - certain forms (Developmental disease)
  - Some types of cancer (ovarian, breast, colon) (Neoplasms)
- often familial occurrence - probability of disease is in 1^{st} degree relatives about 5-10%; 2^{nd} degree relatives - 0,5-1%

Key concept of complex diseases

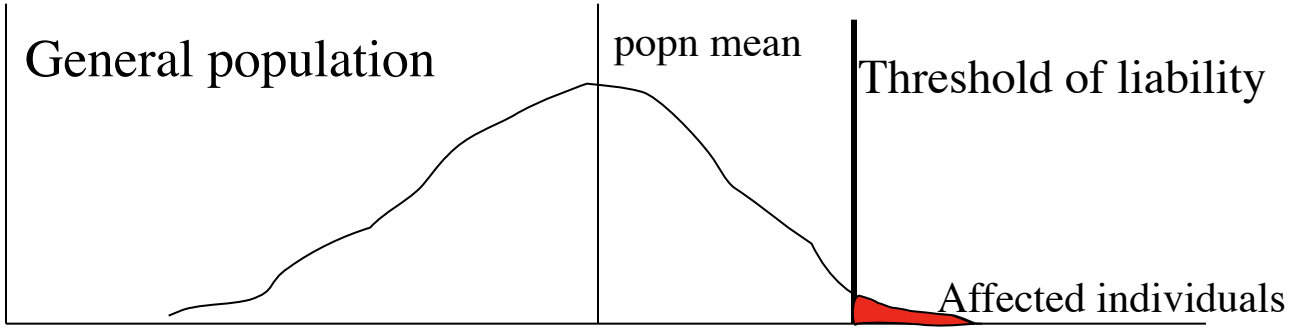
- Multiple distinct loci interact with/without other factors including the environment to result in end stage phenotype
- Expressed in population as a continuously variable susceptibility that follows Gaussian distribution
- Effectively creates a gradient of susceptibility - phenotype presents beyond a certain threshold

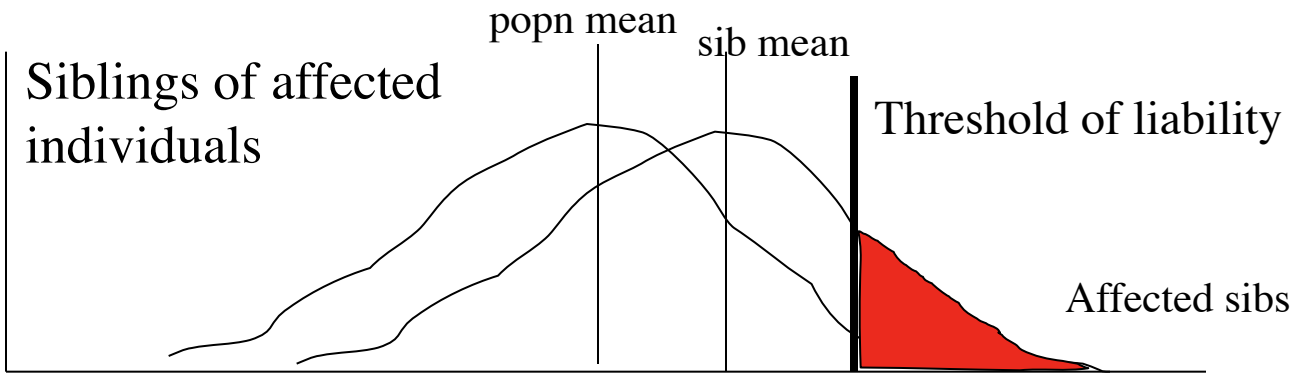


Key concepts of complex diseases

- Familial concentration of disease without specific pattern of inheritance
- Absence of clear biochemical defect resulting from single abnormal gene
- Considerable variation in severity and expression of phenotype (between and within families)
- Most affected individuals have unaffected parents
- Often sex differences
- Often ethnic/racial differences

Familial clustering





Recurrence risks

- Affected individuals have inherited combination of high susceptibility alleles.
- Relatives share these alleles
- Thus cousins, aunts, uncles, etc are also at higher risk than general population

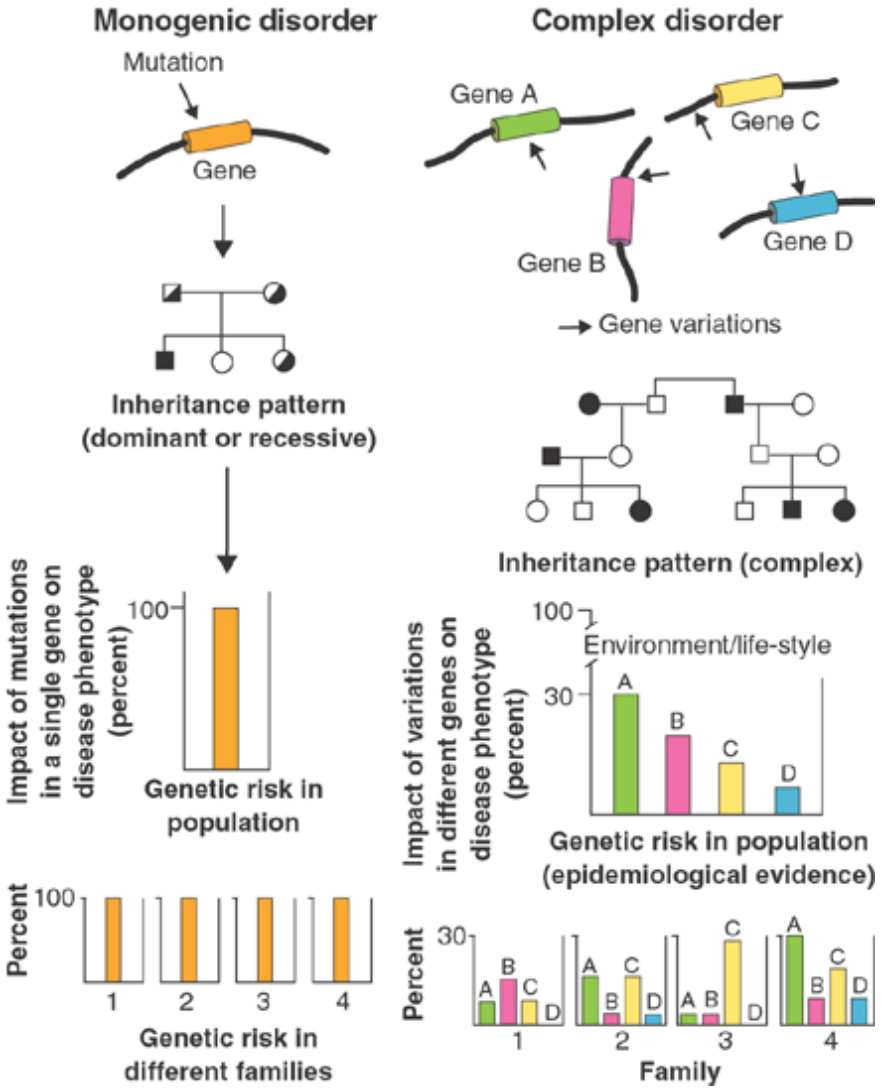
- Parents with affected child have higher than average number of high risk alleles
  - Recurrence risk is higher if >1 family member affected
  - Greater the severity of the disease the higher the recurrence risk

Multifactorial diseases

- Increasing number of loci involved
- >60 different loci identified as being involved in type 1 diabetes
  - Approximately 40% familial clustering due to the HLA loci

  (λ_S =3); other involved include INS (1.9), CTL4 (1.2)
- Each locus involved is neither sufficient nor necessary to result in the phenotype
- Also environmental factors

Monogenic and complex disease



Models to explain multifactorial disease

- 2 basic models
  - **Common disease - common variant** (restricted polymorphism model)

    Proposed that there is a small number of loci with risk alleles that are common in the population (>1%) and each exerts a considerable genetic effect *eg. APOEε4 allele in Alzheimer disease; Factor V Leiden in deep venous thrombosis*
  - **Common disease - rare variant**

    Suggested that there are a large number of loci with risk alleles that are rare in the population (<1%) and each exerts little or moderate effect

- In both models, different alleles at these loci may increase or decrease risk
  - leads to complex patterns of susceptibility

Other factors

- Environmental influence - diet, exposure to toxins, exercise etc
- Epistasis - interaction between different loci; where one particular allele at locus 1 prevents particular allele at locus 2 from manifesting its effect
- Somatic changes
- Epigenetics - methylation, imprinting, etc

All combine to produce disease state in individual

Linkage studies

- **Classic linkage studies require**
  - Large multigenerational single family (or multiple smaller families in clear homogeneous disease)
  - Defined mode of inheritance
  - Single locus responsible
  - Known penetrance
  - Genetic homogeneity

- **Clearly not the case in complex diseases**
  - however can still be used in non parametric models under different modes of inheritance, allowing for heterogeneity etc
  - thus likelihood of detecting causative locus less than in single major locus disorders
  - alleles of low or moderate genetic effect unlikely to be identified

Association studies

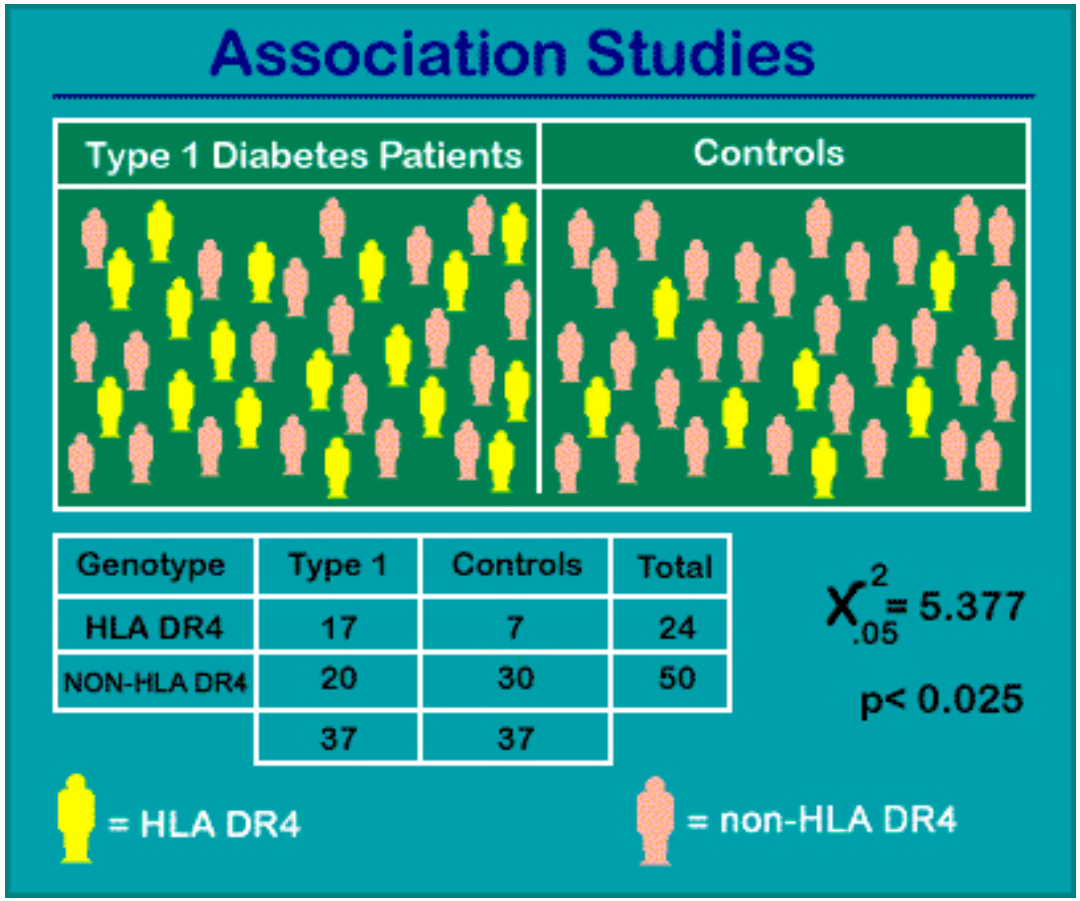
- Several methods available including
  - **Case control series**
  - Affected sib pair
  - Affected pedigree member
  - **TDT**
- Each has different advantages, disadvantages and limitations
- Complex statistical analysis

- **Can be influenced by**
  - Sample size, selection of controls
  - Population stratification, admixture
  - Epistasis, age of disease
  - Problems in multiple testing
  - Informativeness, density of markers
  - Level of risk alleles effect in disease

Association studies- population based

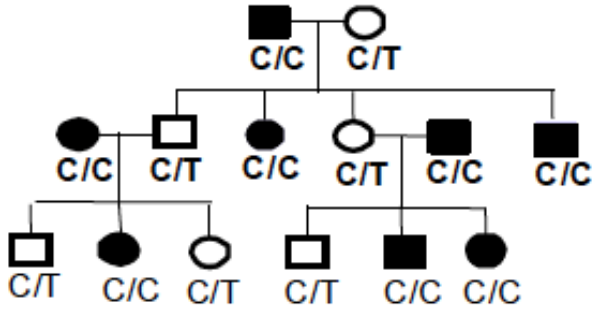
- **Case-control study**
  - Most widely applied strategy
  - Series of affected patients vs series of matched controls
  - Cases readily obtained; genotyping easy
  - Most prone to producing false positive results - usually due to incorrect control population selection. Any difference in allele frequency between groups may be due to differences between populations (independent of disease)
  - Require significant numbers to adequately power study (1000s vs 100s); especially important in study of multiple variables
- **Case-cohort study**
  - Cases and controls drawn from selected population under study to investigate broad spectrum of diseases and factors
  - Prospective; takes longer to select sufficient numbers

Association studies in diabetes type 1



Human Genetic Analysis

Families
Linkage Studies



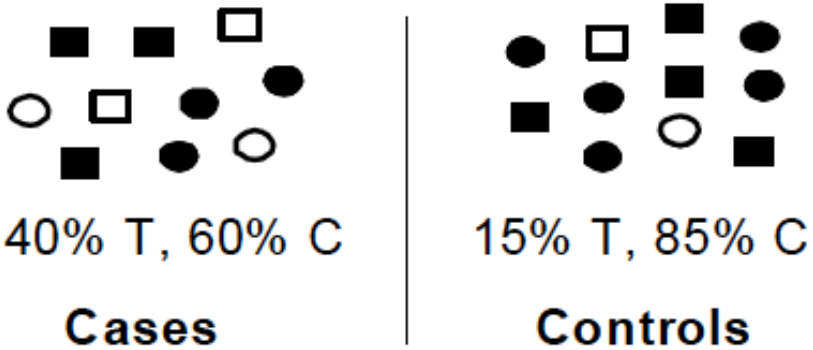
Simple Inheritance (Segregate)

Single Gene with Major Effect

Variant Rare in the Population

~600 Short Tandem Repeat Markers

Populations
Association Studies



Complex Inheritance (Aggregate)

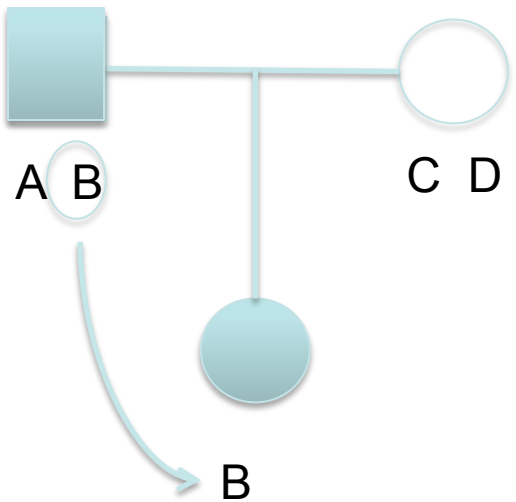
Multiple Genes with Small Contributions and Environmental Contexts

Variant(s) Common in the Population

Polymorphic Markers > 500,000 -1,000,000 Single Nucleotide Polymorphisms (SNPs)

Transmission disequilibrium test (TDT)

- **Transmission disequilibrium testing**



Allele B transmitted from heterozygous parent to affected offspring

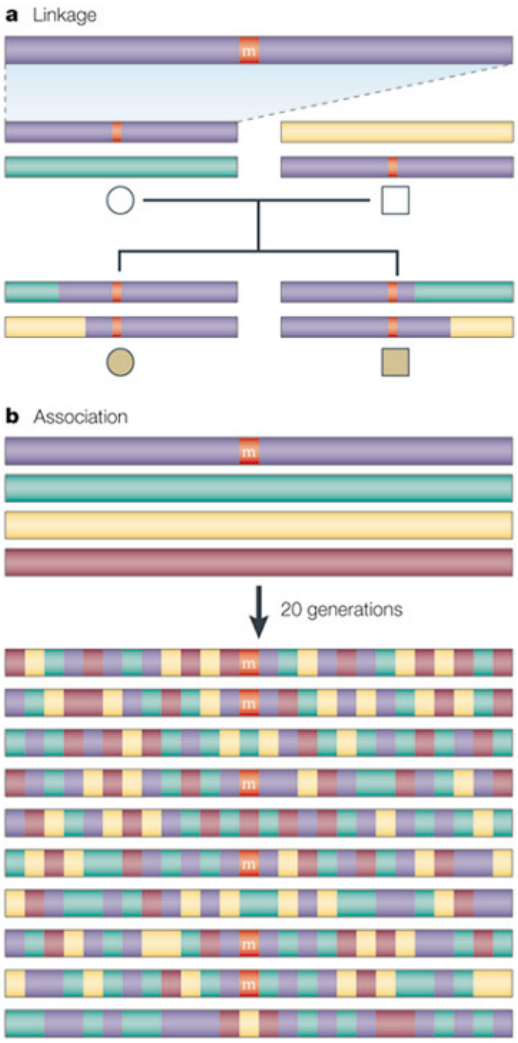
The TDT tests for distortion in transmission of alleles from a heterozygous parent to an affected offspring.

Under no association with the disease, alleles A and B have an equal chance of being transmitte from a heterozygous parent.

If however allele B increases the risk of disease this allele will be preferentially transmitted to the affected offspring.

The sampling scheme for the TDT is the family tri with DNA available from both parents and an affected offspring

Linkage vs. Association



Total sequence variation in the human genome

Population size: 6×10^{9} (diploid)

Mutation rate: 2×10^{-8} per bp per generation

Expected "hits": 240 for each bp

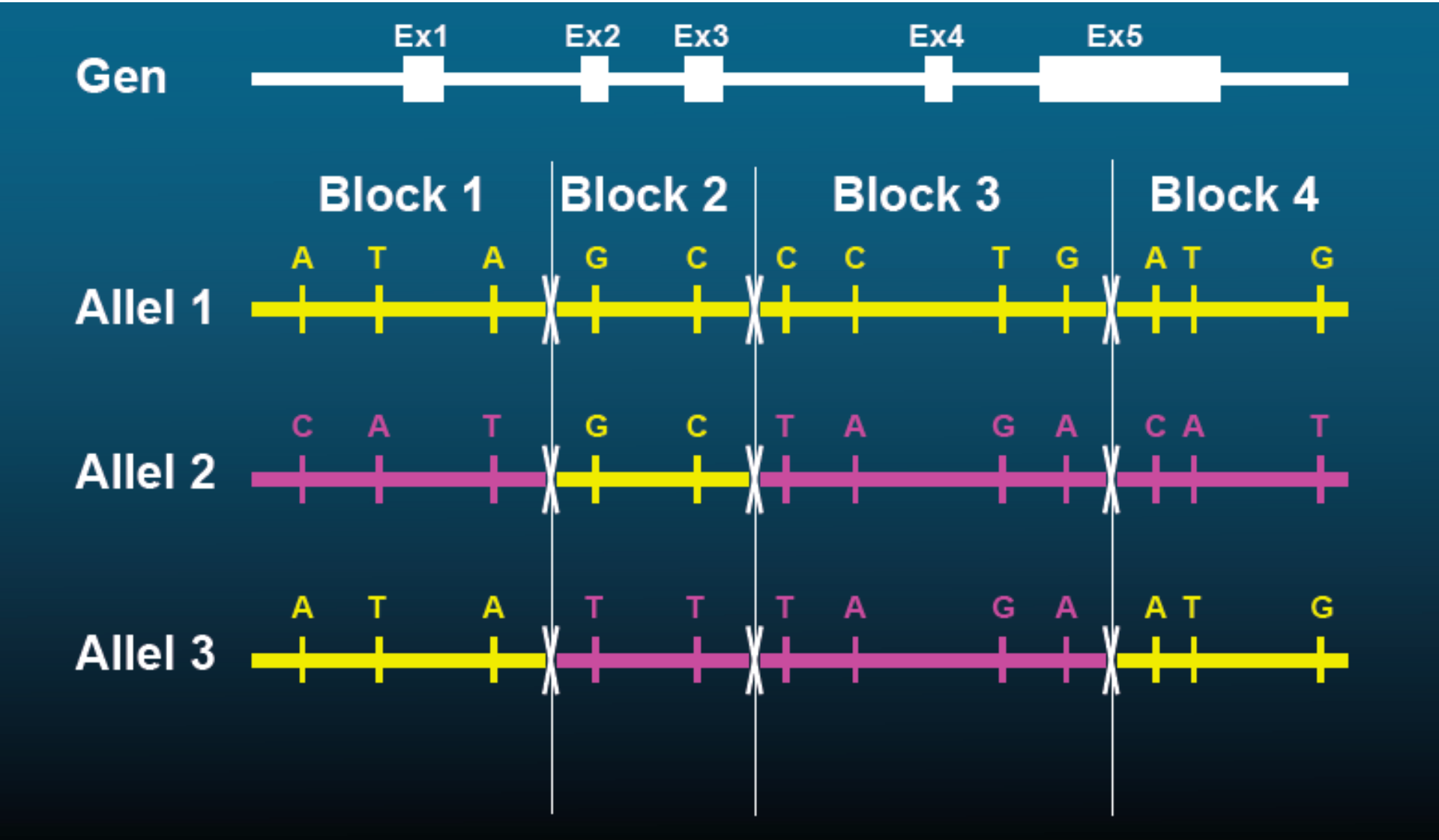
$\therefore$ Every variant compatible with life exists in the population

BUT: Most are vanishingly rare

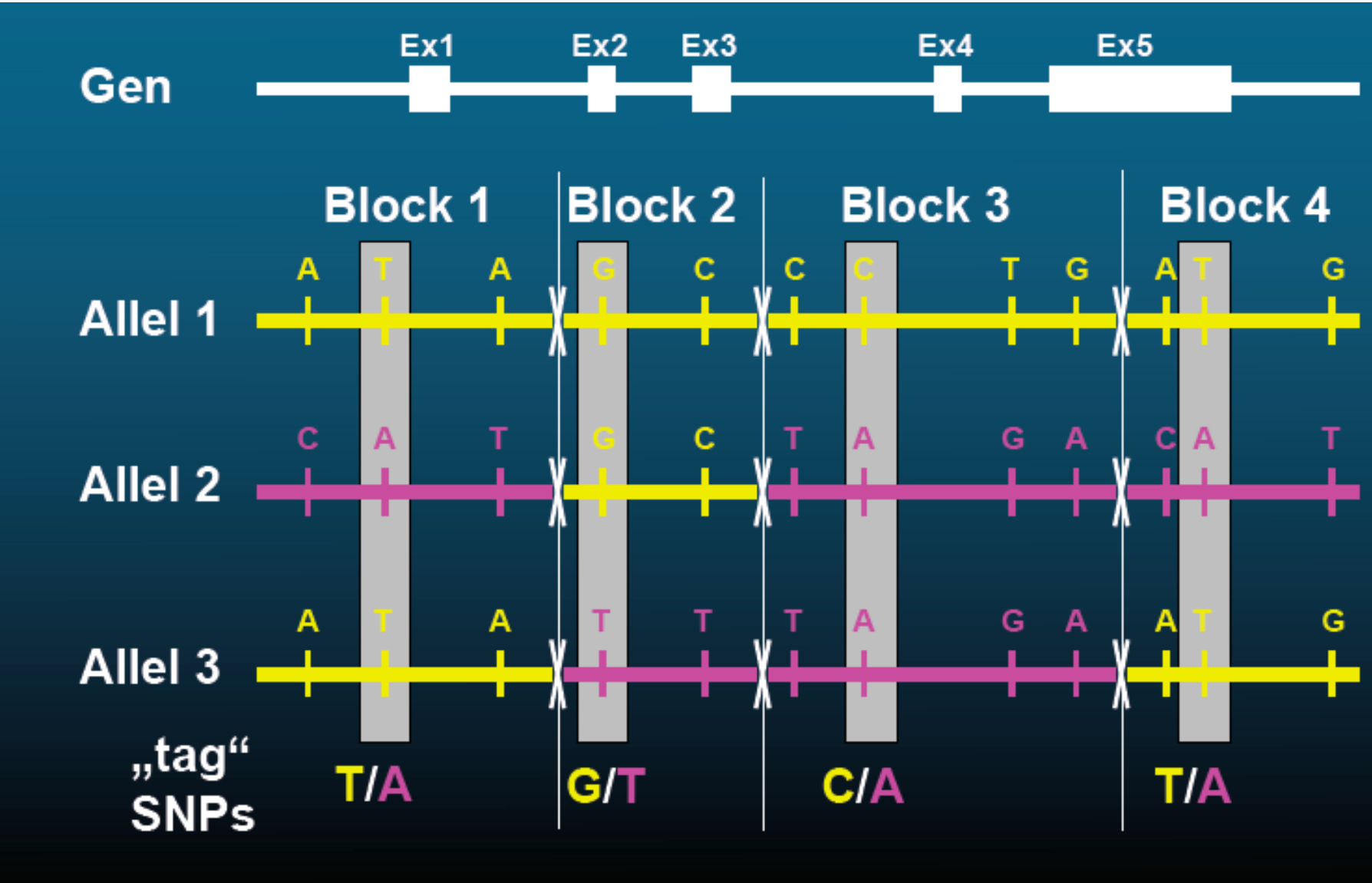
Compare 2 haploid genomes: 1 SNP per 1331 bp*

*The International SNP Map Working Group, *Nature* **409**:928 - 933 (2001)

Haplotype block structure of the human genome

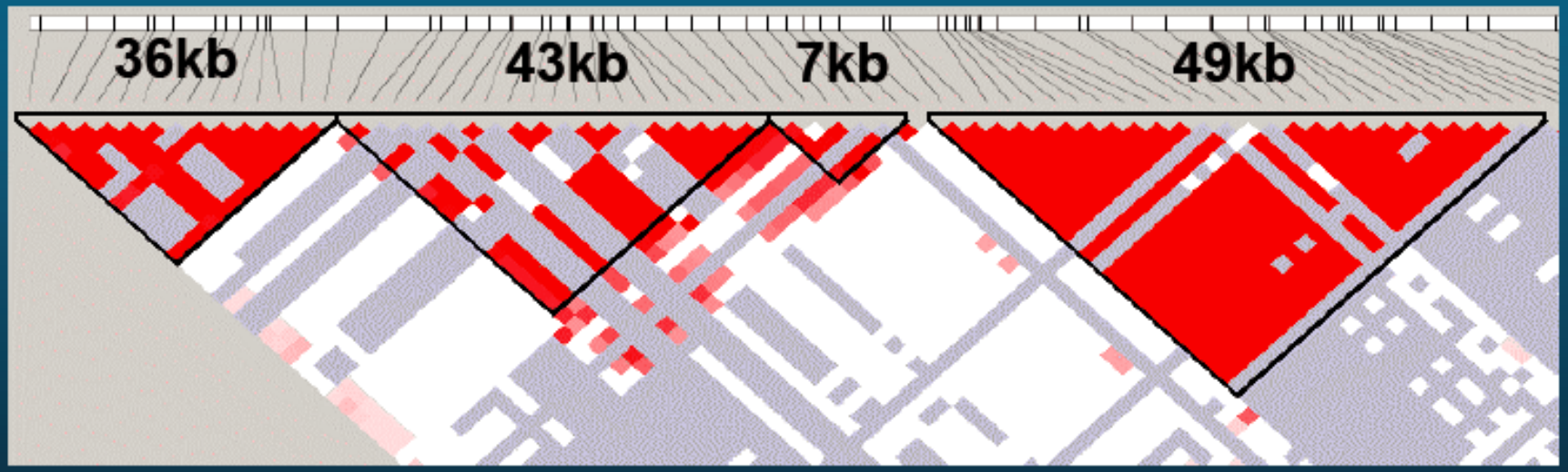


Haplotype block structure of the human genome



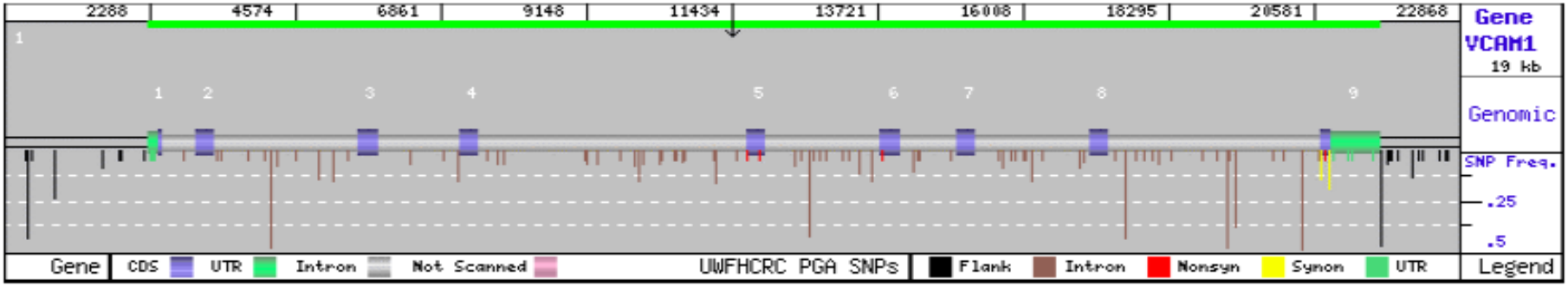
HapMap:
Genomewide identification of haplotypes

Example: Hapmap of the Igf-gene



www.hapmap.org

SNPs in the average gene



Average Gene Size -19 kb ~ Compare 2 haploid - 1 in 1,000 bp

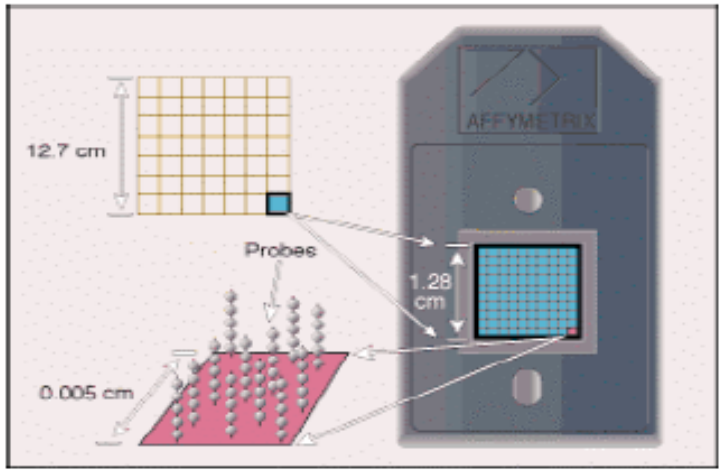
~100 SNPs (200 bp) - 15,000,000 SNPs

~ 5 coding SNPs (half change the amino acid sequence)

Crawford et al Ann Rev Genomics Hum Genet 2005;6:287-312

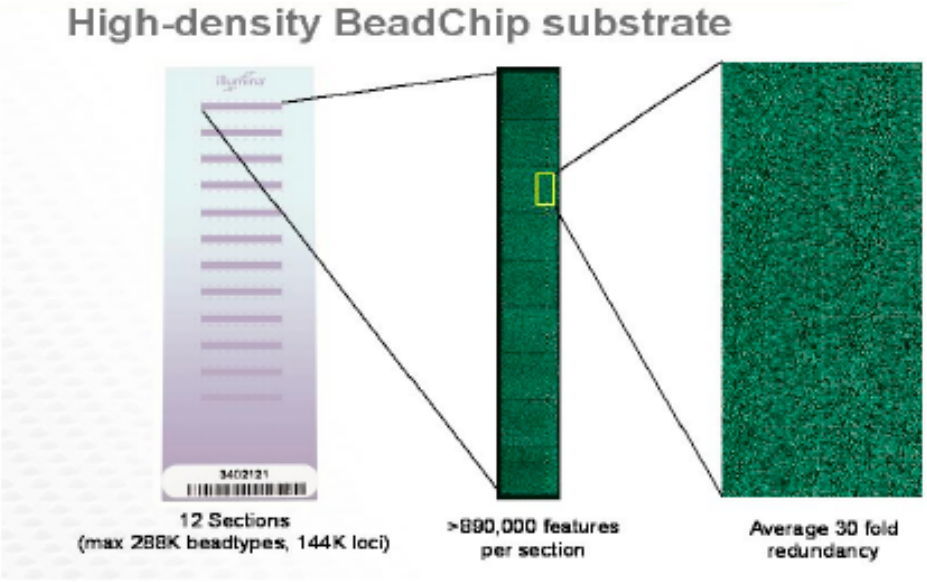
Genotyping technologies

Affymetrix



100,000 or 500,000 Quasi-Random SNPs

Illumina

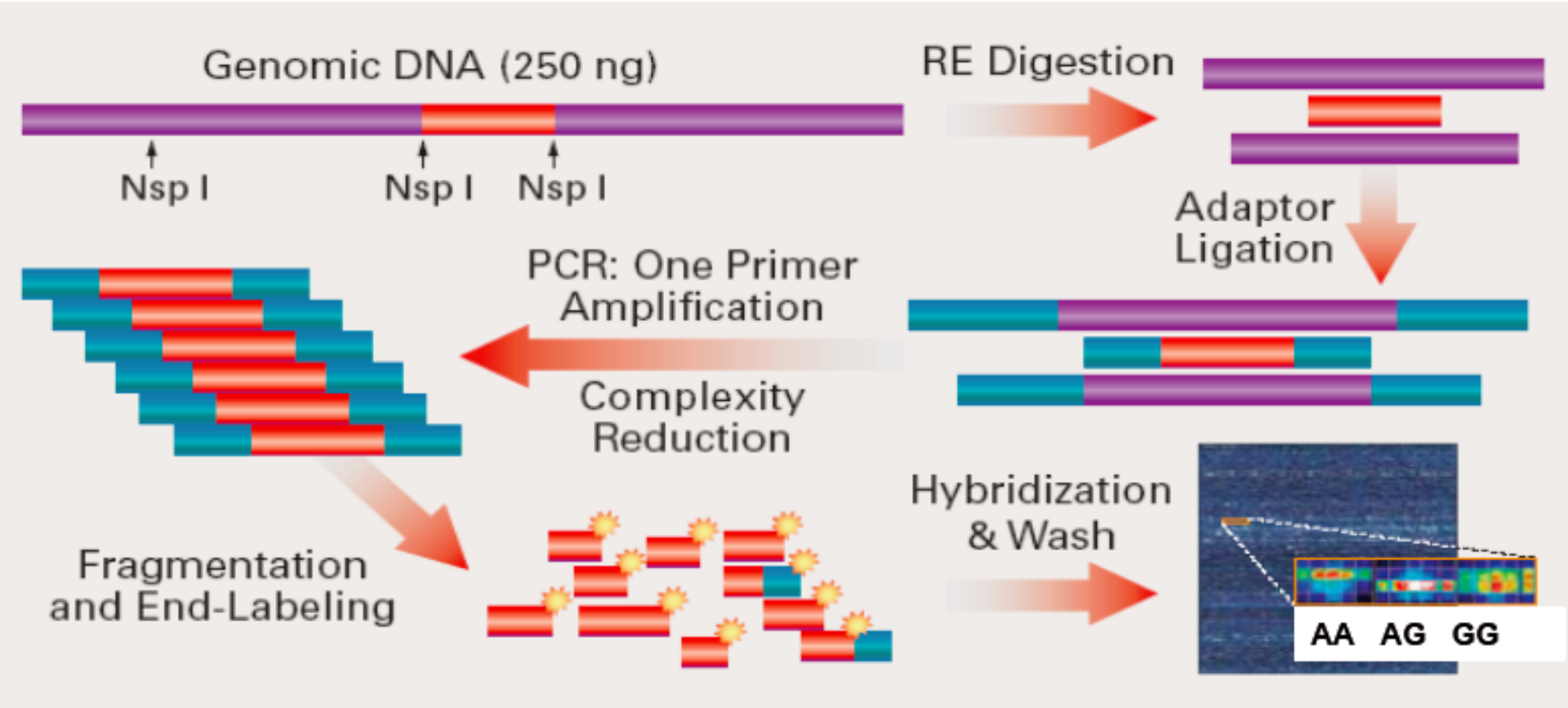


100,000, 317,000, 550,000, 650,000Y SNPs

1 Million Products are here and on the way!

A significant proportion of common SNPs can be captured

SNP Chip methodology



Factors influencing success of studies

- Control populations (stratification)
  family based controls vs matched controls

- Study population
  inbred populations reduce number of segregating loci and non allelic heterogeneity

  admixture; eg Latin American, African American; creates disequilibrium that breaks down rapidly for unlinked markers; utilised in MALD (mapping by admixture linkage disequilibrium)

- Epistasis; interaction between alleles can be accounted for by statisitical models (Markov chain Monte Carlo based methodology)

Well-known problem when case/control groups consist of two different subpopulations with different mixing proportion

- Example: comparing people's height between two places: 1. prison, and 2. nurse school
- In prison, maybe 80% are men
- In nursing school, maybe 80% are women
- Men are on average taller than women
- People in prison are taller than people in nurse school

But the cause of this difference is due to the different mixing proportions, not due to *"staying in prison makes people taller"*

Factors influencing success of studies

- Age of disease
  - old ancient diseases (restricted polymorphism model) have low range of linkage disequilibrium (~3kb). Requires high density map of markers to detect association
  - new diseases have high range of disequilibrium (10kb). Low density scans required but low power to detect
- Genetic effects of risk allele
  - Few loci exerting considerable effect
  - Power to detect reduces with increasing no. of loci
- Informativeness of markers
  - power to detect decreases with reduced heterozygosity
- Inference of linear distance
  - Distance between marker and disease not easy to predict due to non linear relationship between LD and distance below 60kb

Selected reading suggestions:

Venter G, et al., The Sequence of the Human Genome, *Science* **291**:1304-51 (2001)

Haig H. Kazazian, Jr. Mobile Elements: Drivers of Genome Evolution, *Science* **303**: 1626-32 (2004)

West M., et al., Embracing the complexity of genomic data for personalized medicine. *Genome Res.* **16**:559-66 (2006)

Antonarakis, S. E., & Beckmann, J. S. Mendelian disorders deserve more attention. *Nature Reviews Genetics* **7**, 277–282 (2006)

Badano, J. L., & Katsanis, N. Beyond Mendel: An evolving view of human genetic disease transmission. *Nature Reviews Genetics* **3**, 779–789 (2002)